

CHAPTER 3. SURVEY PROCEDURES AND METHODOLOGY

3-A. OVERVIEW

3-A.1. HOUSEHOLDS ELIGIBLE FOR THE NHTS

The NHTS collected travel data from the civilian, non-institutionalized population of the United States. People living in medical institutions, prisons and in barracks on military bases were excluded from the sample. However, telephone numbers in dormitory rooms, fraternity and sorority houses were included so long as no more than 10 people shared the same telephone number.

The focus of this User's Guide is the 69,817 useable households in the full sample. This includes 26,038 households in the national sample, 28,899 households in the Westat add-on sample, and 14,880 households in the Morpace add-on sample. The first public release of data for these households was made in January 2003. The January 2003 dataset contained the daily trips for the national sample households. The final NHTS dataset, provided to DOT in January 2004, contained 69,817 households. The dataset did not include travel period data. BTS has assumed responsibility for publishing this dataset.

All telephone numbers in households in the sample that were found to be residential were eligible for the household interview. The household interview had to be completed by a household member who was at least 18 years old. The exception to this rule was emancipated households in which the age of every household member was less than 18 years.

Interviewers could directly interview household members who were 16 years and older. Proxy interviews were requested for all younger household members. However, if asked by an adult household member, an interviewer could directly interview a household member who was 14 or 15 years old.

3-A.2. HOW THE DATA WERE COLLECTED

The NHTS was conducted as a telephone survey, using Computer-Assisted Telephone Interviewing (CATI) technology. The sample Westat used for the national survey and New York and Wisconsin add-ons was a list-assisted random digit dialing (RDD) telephone number sample.

Westat randomly pre-assigned each telephone number in the sample a day of the week. During the household interview, each household was assigned a specific date as their "Travel Day" and a four-week "Travel Period" for which detailed data on travel were collected.

Some households (those that could be associated with an address through their sampled telephone number) were first contacted by an advance letter containing a pre-survey monetary incentive to participate, followed about a week later by a telephone interview. The remaining households were first contacted by telephone. After the first telephone interview, referred to as the household interview, travel diaries and an additional monetary incentive were mailed to the household so that each household member could record their travel on the assigned travel day. A reminder call was made to each household on the day before their assigned travel day.

Household members were contacted by telephone during a six-day window beginning with the day following the travel day to complete a person interview and provide their travel.

Odometer readings for each household vehicle in completed households in the national, New York and Wisconsin add-on samples were also collected by contacts at two points in time.

3-A.3. WHEN THE DATA ARE COLLECTED

The 2001 NHTS interviews for the national sample and New York and Wisconsin add-ons were conducted from March 19, 2001 through May 9, 2002. Interviews for households in the seven Morpace add-ons were conducted between May 31, 2001 and July 5, 2002.

The survey must be conducted over at least a 12-month period so that seasonal variations in travel are represented. As in 1995, the 2001 NHTS took 14 months, rather than 12 to complete. This was because interviewers were trained in waves and it took a few months to train all the interviewers needed for the study. The weighting adjusts for the monthly differences in number of interviews completed.

Travel day dates were assigned to all seven days of the week, including holidays. The assigned travel period was the four-week period ending with the assigned travel day. The intent was to represent travel across an entire year.

3-A.4. GEOGRAPHIC COVERAGE

Interviews were conducted with households in all 50 States and the District of Columbia. Westat drew a new sample of telephone numbers every quarter to ensure that new exchanges and telephone numbers were included and all geographic areas were completely represented in all seasons.

3-B. SAMPLE DESIGN AND SELECTION

3-B.1. OVERVIEW

This survey was designed as a list-assisted random digit dialing survey, to yield an equal probability sample of households with telephones. The national sample was increased in several add-on areas: New York State, Wisconsin, Texas, Kentucky, Hawaii, Lancaster Pennsylvania, Baltimore Maryland, Des Moines Ohio and Oahu Hawaii. The supplemental sample in these areas was not included in the January 2003 dataset but was included in the January 2004 dataset.

The target sample size was 25,000 completed households for the national sample, 10,884 completed households for the New York add-on and 16,000 completed households for the Wisconsin add-on.

3-B.2. SAMPLING FRAME

Westat's sampling frame consisted of all telephone numbers in 100-banks of numbers in which there was at least one listed residential number. A 100-bank is a set of 100 telephone numbers with the same first eight digits, that is, the same area code, exchange, and the next two digits. Each quarter, a new sampling frame was constructed and sample was selected for use until a new sample was drawn. Sampling frames were constructed as of December 2000, March 2001, June 2001, September 2001 and December 2001.

3-B.3. SAMPLE SELECTION

Telephone numbers were sorted according to several variables and a systematic sample was then selected from the sorted list. For the national sample, all telephone numbers in the frame of 100-banks had an equal probability of selection.

For the national sample, the sort of telephone numbers was first by the nine Census Divisions and second by metropolitan area/non-metropolitan area. For metropolitan areas, the initial sort was by population of the metropolitan statistical area (MSA) and primary metropolitan statistical area (PMSA) (largest to smallest). Within an MSA/PMSA, telephone exchanges were ordered by those serving the county (or counties) containing the central city, followed by those serving the remaining non-central city county (or counties). Within each county, exchanges were ordered numerically - lowest to highest.

For non-metro areas, the initial sort was by state within a Census Division, with a serpentine ordering¹ from north to south and from east to west. Within state, non-metro counties were similarly ordered in a serpentine fashion, north to south and east to west. Finally, within county, exchanges were ordered numerically from lowest to highest.

¹ Serpentine Ordering: The listing begins in the most Northeast state in a given Census Division, followed by the state just south and still at the eastern edge of the Division. After the far Southeast state in the division, the listing proceeds to the state just west of the most Southeast state. The sort continues with the next state to the north. The listing continues in this fashion until all states in the Division have been included.

3-C. DATA COLLECTION PROCEDURES

3-C.1. OVERVIEW

Data collection for the national sample of the 2001 NHTS, as well as the New York state and Wisconsin add-on samples, was conducted by staff at six Westat Telephone Research Centers (TRCs). The centers used were located in Frederick, MD, Sarasota, FL, Sacramento and Merced, CA, Greeley, CO and Chambersburg, PA. Westat is a social science research firm headquartered in Rockville, Maryland. Data collection for the seven Morpace add-ons was conducted from Morpace's telephone center in Sterling Heights, Michigan.

3-C.2. INTERVIEWER TRAINING

A staff of approximately 345 Westat interviewers and 58 supervisors were trained on the 2001 NHTS. These interviewers were trained during 16 separate training sessions conducted periodically over the 14-month data collection period. The peak number of interviewers working on the study in any week was 186. This includes interviewers who worked on the national, New York, and Wisconsin survey samples. Of the 345 interviewers, 38 interviewed in both English and Spanish.

All Westat interviewers assigned to the survey participated in training sessions and completed at least 24 hours of formal project-specific training. For interviewers with no prior interviewing experience, these hours were in addition to four hours spent in training on general interviewing skills and another four-hour-plus training on the use of the CATI system. These eight hours of non-project specific training occurred prior to the interviewer's assignment to the NHTS project. For the NHTS project, 60 percent of the 345 interviewers were experienced and did not have to go through the non-project specific training. Interviewers whom the TRC supervisory staff felt were not ready for "live" interviewing at the conclusion of the formal "classroom" training received additional training time.

3-C.3. INTERVIEWER MONITORING

Interviewer monitoring is an important aspect of survey quality control, and Westat, Morpace, and DOT staff devoted considerable time and attention to it. Using extension telephones and personal computer displays linked to the interviewer's computers, supervisors silently monitored about 10 percent of each interviewer's work over the course of the study. DOT and non-TRC Westat staff monitored interviewers in all six TRC's from monitoring rooms located at Westat in Rockville, MD. Staff from DOT also monitored interviews in-person from Morpace's telephone monitoring center in Sterling Heights, Michigan on two occasions and remotely by connecting to Morpace's CATI.

3-C.4. CALLBACK PROCEDURES

Effective calling patterns are essential to achieving a high response rate on all telephone surveys. Westat made at least seven attempts to establish contact to screen a household and a minimum of eight attempts to establish subsequent contact to complete each person-level interview with each household member. A computer algorithm scheduled these calls over different days and included day, evening and weekend calls.

3-C.5. REFUSAL CONVERSION

Refusal conversion was an important aspect of Westat's overall response maximization effort for the NHTS. An integral component of this effort was the utilization of a select team of refusal conversion specialists. The team was comprised of TRC interviewer staff members who had demonstrated exceptional skills in achieving high cooperation rates. Once interviewers were familiar with the questionnaires, and common reasons for refusals were identified, Westat supervisors held special training sessions on refusal conversion techniques for the refusal conversion interviewers.

Whenever a respondent initially refused to complete an interview, the interviewer completed a separate CATI data collection module to record any information known about the household and the respondent's reason(s) for refusing to participate.

A project supervisor reviewed each case and non-hostile refusals were returned to interviewers specially trained in refusal conversion for additional calls to the household.

3-C.6. BILINGUAL INTERVIEWING

Interviewing on the NHTS was conducted in both English and Spanish. Interviewing in Spanish was an important factor in gaining the cooperation of Hispanic respondents and completing interviews with them. Westat bilingual interviewers completed the full survey interviewer training in English, and conducted interviews in English until they were thoroughly familiar with the questionnaires and CATI system. They also attended additional training on the Spanish CATI instruments. Westat translated the entire CATI questionnaire, as well as all instructions to interviewers and clarifying comments into Spanish. Spanish-speaking supervisors monitored the bilingual interviewers.

All cases assigned an initial result code of "language problem" by an English-speaking-only interviewer were available only to bilingual interviewers. If the bilingual interviewer determined that the respondent spoke neither English nor Spanish, attempts were made to conduct the interview using an English-speaking household member as a proxy. If these attempts were not successful, a final code of "language problem" was assigned to the case. Only 610 households (1 percent), out of the estimated 63,472 residential households in the national sample, could not be interviewed because of a language barrier.

3-C.7. CONFIDENTIALITY PROCEDURES

All data on the national survey were collected with an assurance that all information that could identify a specific respondent would remain confidential. All Westat and DOT personnel, including interviewers and professional staff, signed an affidavit stating that they would maintain the confidentiality of all survey data.

3-C.8. ADVANCE LETTER AND CASH INCENTIVE TO HOUSEHOLDS

As discussed in Sections 2-B.1 and 3-A.2, a subset of households in the sample were mailed an advance letter. These were households for whom Westat and Morpace were able to obtain mailing addresses. The mailing of these advance letters was timed so that the household received the letter shortly before the first telephone call to the household. This was accomplished by releasing the sample to the interviewers in small groups. A mailing occurred prior to each release of the sample.

The advance letter mailing included the letter from the Secretary of Transportation, a five-dollar cash incentive for the national and New York samples and a two-dollar cash incentive for the Wisconsin sample, and a brochure introducing the survey. The letters sent to respondents in the seven Morpace add-ons were signed by a dignitary for the add-on region. Appendix N, NHTS Field Documents, contains a copy of the advance letter and the brochure. The primary purpose of the mailing was to improve cooperation rates by informing prospective respondents that this was a legitimate survey, not a marketing or fundraising call.

Westat was able to obtain addresses for 86 percent of the residential numbers in the national sample. About 5 percent of the advance letters were returned as "undelivered" by the post office. Therefore, approximately 81 percent of residential households received the advance letter.

3-C.9. TRAVEL DAY AND TRAVEL PERIOD ASSIGNMENT

During the administration of the household interview, Westat's CATI program automatically assigned a travel day and travel period to each household. The interviewer identified the travel day to the household respondent during the interview. The travel period assigned was the four-week period ending with the travel day. Though the system assigned the travel period, the interviewer did not provide the household with the travel period during the household interview. However, the travel period dates were included in the mailout materials, along with a map showing a 50-mile radius from the household.

Travel characteristics are known to vary by season of the year and day-of-the week. There was some variation in number of completed interviews by month. For example, for the national sample April had more sample because interviewing was done in both April 2001 and April 2002. Because of national events such as September 11, 2001 and the anthrax scare there were also some months in which there were relatively fewer completed interviews. To control for this, part of the weighting process adjusted the estimates of total persons and total households to be equal for each calendar month. See Chapter 5 and Appendix H, for further details.

The variation in travel by day of the week for the Westat sample was balanced by assigning the travel days for one-seventh of the sample telephone numbers to each day of the week. When the calls to a sample phone number resulted in a completed household interview, the CATI system determined the household's travel date on the selected day of the week 10 to 14 days in the future, which allowed time for dairy mailings to reach the household. Morpace also followed a procedure to balance travel days by day of the week.

3-C.10. THE TRAVEL DIARY MAILING

The use of travel diaries on household travel surveys has been shown to improve the accuracy of trip reporting. Therefore, all household members in households who completed a household interview were sent diaries for their travel day. Of the 160,578 persons who completed person interviews in useable households in the national and add-on samples, 71.1 percent or 114,353 reported having filled out the travel diary. Westat mailed the diary package within a day or two following the completion of the household interview. It was sent via Priority Mail and contained:

- A letter from the U.S. DOT thanking the household for completing the household interview and agreeing to participate in the survey;
- A brochure describing the survey;
- A travel day diary and a two-dollar cash incentive that were included in individual envelopes addressed to each household member and placed in the Priority Mail package. The back of each diary provided guidance on completing the diary and included an example of a completed diary;

- An eye-catching bright yellow reminder card identifying the household's travel day;
- A colored map displaying the household 's home address at it's center with a circle around the home demarcating a distance of 50 miles from the home; and
- An odometer mileage form identifying the make, model and year of each household vehicle, with spaces to enter the odometer readings and the dates they were taken.

3-C.11. REMINDER CALL

Each household received a reminder call on the day before their assigned travel day. The call was designed to find out if the household had received its diary package, answer questions, and remind household members to record their travel in the diary the following day. Interviewers attempted to speak with the household respondent but spoke with any household member 16 and older if the household respondent was not available. If the interviewer reached an answering machine, the reminder to collect travel information was left on the answering machine. Households were asked to call the study's toll-free number if there were any questions.

3-C.12. CALL-BACK PERIOD

There was a six-day period during which interviewers were permitted by the CATI system to call each household member to collect their travel. This limit was established by US DOT because of memory problems beyond six days. Phone calls to collect the diary information from the household usually began the day after the travel day, and continued for the next five days. Though no outgoing calls were made after the close of the six-day period, respondents could call into the toll-free number to provide their information. Such information was recorded up to 10 days after the travel day provided the subject had completed a diary. For the national survey, New York and Wisconsin add-ons 0.6 percent or 773 persons provided their travel information after the end of the six-day window. For the remaining seven add-ons 1,883 or 5.5 percent of persons provided travel information after the six-day window. Overall, 78 percent of the

160,758 person interviews in the 2001 NHTS were completed within three days following the household's travel day.

3-C.13. PROXY INTERVIEW PROCEDURES

A proxy interview is one in which someone else in the household reports for the subject. In the NHTS data collection, an adult household member always served as the proxy for a child under age 14. Proxies were also requested for persons age 14 and 15 years. However, if an adult household member requested that the interviewer speak directly with these teenagers, the interview was conducted with the subject. Proxies were not initially requested for household members 16 years and older.

An issue with proxy interviews is under what circumstances to allow proxies for adult household members, defined here as 16 and older. In the 2001 NHTS, proxies were allowed for these subjects if:

- the subject was not capable of being interviewed because of an impairment or a language barrier;
- the interviewer was told that this subject would not be available for the entire six-day recall period;
- the interviewer was told that this subject would never participate, and the proxy was knowledgeable about the subject's travel on the assigned travel day; or
- the interviewers attempted to reach the subject for the first three days of the six-day callback period, and were not successful.

If the respondent filled out a travel diary for travel day, the proxy household member was asked to find the diary and use it when they served as a proxy for the respondent. Note that the conditions under which each interview was completed are a part of the data files. There are variables for:

- whether the interview was with the subject or a proxy respondent (variable PROXY on the Person, Travel Day and Travel Period Files), and

- if a travel diary was completed by the subject or another household member (variable DIARYCMP on the Person File).

3-C.14. REDUCING RESPONDENT BURDEN

During the person interview, special NHTS trip rostering procedures were applied to reduce respondent burden for household members who traveled together during the travel day or travel period. Burden was reduced at two main points during the interview - during trip rostering and during the collection of trip detail on each trip that was rostered.

During trip rostering, each household member was asked to list all trips taken prior to asking about the detail on each individual trip. If the household member currently being interviewed reported that another household member went on a trip with him/her, then this trip was automatically also recorded on the roster for the other household member, provided that household member had not yet been interviewed. When the interviewer talked with this other household member, (s)he merely had to confirm that the household member went on the trip. The household member had to agree with the trip destination and start and end times for the trips to be considered identical. If the household member agreed that the trips were the same, the trip was retained on the roster, otherwise, it was deleted. For travel day, if the household member agreed that the previous household member was correct and the trip was retained on the roster, then this subsequent household member, if not the driver on the trip, was not required to provide any detail on the trip. Trip detail was copied from the previous household member. The driver on a particular travel day trip was always required to report the trip details. For travel period, trip detail was not obtained from a subsequent household member if that household member indicated that during the entire duration of the trip they were with the household member who had already provided the trip detail.

3-D. DATA EDITING

3-D.1. ONLINE EDITS

Westat conducted most of the editing on the NHTS online while the interviewer had the respondent on the telephone. This editing was accomplished by programming the edits into the CATI software so that the interviewer automatically asked the appropriate next question and was prompted when a response entered to a particular question was not a likely response. All the online edits used in the NHTS are documented in the household and person questionnaires that are included as Appendix J to this report. These online edits fell into three main groups:

- Skip edits that moved the interviewer to the appropriate next question based on responses provided to earlier questions. For example, a subject that was 6 years old was not asked employment questions;
- Range edits that prompted the interviewer when a response entered was possibly incorrect. For example, a response that indicated that a particular vehicle was driven 45,000 miles in the past 12 months. For some variables, both hard and soft ranges were programmed into the CATI. In the example above, the soft range was 2,000 to 30,000. The hard range was 0 to 200,000. Therefore, in this example, the interviewer was prompted to re-ask the question because the response was not within the soft range. If the subject provided a response of 45,000 a second time, the response was accepted. Responses that exceeded the hard range were recorded in "comments." If needed, ranges were modified post-data collection to accommodate values that exceeded the hard range; and
- Logic edits that prompted the interviewer when a value entered was within the valid range for a variable but did not pass a logic check. For example, the subject informed the interviewer during the person interview that a particular household member was the driver on a particular trip. However, that household member was not recorded as a driver during the household interview or was not reported as being on the trip. In both these scenarios a logic edit would be triggered. The triggering of logic edits sometimes required modifying previously provided information. That is, the current response that the respondent was the driver on the trip may be correct but the earlier response that the subject was not on the trip may need to be modified.

3-D.2. UPDATING CATI DURING DATA COLLECTION

This editing, which was ongoing throughout data collection, occurred after a household or person interview was completed. It involved editing information recorded for a particular item during the CATI interview with information provided by the interviewer but recorded some place else. This information had to be updated later because it had not been possible to code it into the appropriate variable during the interview. Instead the information was:

- Recorded online in "comments" in CATI if the interviewer had not left the case and if the information being provided was not very extensive;
- On a "problem sheet" if the interviewer had left the case or the information being provided was extensive;
- On a form designed to collect specific information that was modified, forgotten or provided after the interviewer could enter the information into CATI; or
- Recorded in an "other specify" category.

Examples of the types of information Westat interviewers entered in "comments" include:

- The response provided by the respondent was out of range and would not be accepted by the CATI software. When this happened, the response was reviewed and if likely, after approval by DOT the range for the specific variable was broadened. If the response was unlikely, it was coded -9 (not ascertained);
- A subsequent response modified an earlier response and the respondent agreed that the earlier response needed to be changed. For example, the household respondent enumerated three household vehicles. But later, when the interviewer asked about the primary driver of each vehicle, they found that one of the vehicles was not licensed and was not in working condition. Information on this vehicle was deleted after the interview; and
- The interviewer entered an incorrect response and left the variable before the response could be modified. For example, the note from the interviewer said that she entered person A as a female when in fact person A should be male.

Examples of information recorded in "problem sheets" include:

- The interviewer completed the household interview and found out at the end of the interview that the respondent was an adult but was not a household member (e.g., (s)he usually lived elsewhere and was just visiting). In such a scenario, the person was deleted from the household;
- The respondent initially refused to provide his/her home address and provided just a mailing address. However, after the interview was completed, (s)he was more comfortable with the study and provided a home address; and
- The interviewer entered data in the wrong proxy case (e.g., the interviewer thought the mother was responding for daughter A when in fact the information was being provided for daughter B). Since the interviewer was midway through the interview, the interview was not interrupted. The error was recorded on a problem sheet and the cases were switched later.

Examples of specific forms used by interviewers included:

- A "missed trip" form. Often respondents informed the interviewer they had forgotten to mention a trip after the interviewer had left the trip roster. In such cases, the missed trip information was recorded on a form and added to the CATI file later; and
- An "odometer" form. If the household did not provide their first odometer readings during the person interview, they were contacted later and the information was recorded on a specific form (see Appendix N, NHTS Field Documents).

Sometimes an interviewer may have felt that the response categories for a particular question did not accurately describe the response provided by a respondent or the categories provided were too numerous and the interviewer felt the need to code the response quickly to keep the interview moving. In such cases the interviewer checked the "other specify" response category and recorded the response in open-ended text. On completion of the interview, these responses were reviewed and were appropriately coded into an existing category, a new category was added, or the response was left in "other specify" as an open-ended response.

3-D.3. APPROACH TO POST INTERVIEW EDITING

In surveys with complex questionnaires and procedures, such as the NHTS, the final dataset reflects fundamental approaches taken in the data collection and editing processes. For the 2001 NHTS, two approaches may have had considerable impact on the resulting data.

The first is the reluctance to impute data. If the respondent did not answer an item, its value was generally not imputed, (i.e., determine what the logical response would be given the response to other items). Carefully performed imputation has its place in many statistical surveys, however Westat and DOT determined that imputation would be limited in the NHTS data. If data was imputed, an imputation/edit flag was set for the variable to indicate the values that were imputed.

Second, a conservative approach was taken regarding changing reported data. If it was determined that what was reported could not have happened, the unlikely response was set to a "not ascertained" (-9) code. The exception to this rule was if the same information could be obtained from another household member or from elsewhere in the subject household member's interview. For example, household member A reported a "start time" for a trip that failed an edit. But, household member B went on the same trip and his reported time did not fail the edit. In this example, the start time that was reported by household member A was modified to reflect the time that was reported by household member B. This was only done when trip data failed an edit. In general, differences in data reported by household members on the same trip were allowed to remain. For example, Person A reports a trip starting at 8:30 am when Person B says the same trip started at 8:35 am.

3-D.4. POST DATA COLLECTION EDITING

On completion of all data collection on the 2001 NHTS, Westat programmed and ran edits designed to check for data consistency. A list of the edits that were run is included in Appendix P, Data Editing. When the value of a variable failed an edit, it was manually reviewed. If the value was highly unlikely, it was revised to "not ascertained" (a code of -9). No flags were set when a response was set to "not ascertained".

However, during the cleaning of responses for travel day trips, it was sometimes necessary to slightly modify a response or impute a missing response. In such cases edit and imputation flags were set to indicate the variables whose data was modified. The situations in which these flags were set are described below.

3-D.4.A. EDIT FLAGS

There are edit flags for each of the following travel day variables: STRTTIME, ENDDTIME, TRPDIST, TRVL_MIN, TRPTRANS and WHYTRPO1. A flag indicates that the value for the variable has been adjusted. The goal of editing these variables was to decrease the number of trips that had identical or overlapping trip times.

All cases where two or more travel day trips for the same household member had identical start and end times were examined. In most cases these trips were duplicates, i.e. the trip detail on both trips indicated that the exact same trip had been reported twice. Duplicates could have been generated as a result of the trip rostering procedures that were in place to reduce respondent burden. Once it was determined that the trips were duplicates the trip with the least amount of trip detail was deleted from the trip roster.

In addition to cases with duplicate trips, the editing process found travel day trips with times that overlapped with other trips or were completely embedded in other trips. In some cases information was available on another household member's record that could help clarify conflicting information. Whenever possible this information was used to make adjustments to the record with embedded trips.

In some cases the most logical edit was to "split the travel time", these were mostly walk or bike trips. For example, if a walking trip from home was reported from 8:00AM to 9:00AM, and was followed by a walking trip to home from 8:55AM to 9:00AM, then the end time of the first trip was adjusted to 8:30AM and the start time of the second trip was adjusted to 8:31AM. In some of these cases trip distance and travel time had to be adjusted as well.

Other cases involved embedded trips that were on the way to a destination. These were often reported after the interviewer had finished gathering trip information,

thus were recorded on a "Missed Trip" form and manually entered during data editing at a later stage. The most common types that were reported incorrectly involved trips to the gas station, and picking up or dropping off someone on the way to a destination. For example, if a trip to home (recorded in CATI during the interview) was reported from 5:00PM to 5:30PM, and a trip to the gas station was reported from 5:00PM to 5:10PM (recorded on a missed trip form as it was recalled later), then we assumed that the respondent stopped at the gas station on the way home. Therefore, the original trip in CATI from 5:00PM to 5:30PM was modified to a trip from the gas station to home. The start time of the trip was changed to 5:10PM. Unfortunately, when there was an embedded trip reported there was no information on dwell time. That is, the person arrived at the gas station at 5:10PM and left for home from the gas station also at 5:10PM.

For travel period the editing process also identified duplicate, non-travel period and embedded trips. Duplicate trips were those where a subject had multiple identical trips on the trip roster. In such cases the duplicate trip with the least trip detail was deleted. The next category of trips edited involved trips that did not end during the 28-day travel period. Such trips were considered non-travel period trips as they did not meet the definition of a travel period trip for the household and were deleted. The final category of travel period trips that were edited involved embedded trips where the start and end dates of one or more trips were embedded in another trip. In such cases, the great circle distance was used to identify the trip that was the farthest destination from home and the embedded trips were edited to overnight stops on this main trip. A flag EDITSTP was set to indicate that the travel period trip was edited to an overnight stop.

3-D.4.B. IMPUTATION FLAGS

There are imputation flags for each of the following variables: STRTTIME, ENDTIME, TRVL_MIN, AGE, TRIPDIST, OTHRPHON, HHR_RACE, SEX, HOMEOWN, HOMETYPE and if a whole trip was imputed. A flag for these variables indicates that the variable has been imputed.

One of the goals of this imputation was to decrease the number of travel day trips with missing start and end time values. If both start time and end time were missing the trip was left in the roster in the original order reported by the respondent. If

start time was missing but end time was known or vice versa, and trip distance and mode were reported, then we were able to estimate the travel time of the trip and subsequently impute the missing start or end time. The following rules were used to estimate travel time based on mode and distance:

- If the mode is local transit bus, school bus, subway, trolley (TRPTRANS=10,12,17,18), and the trip distance is ≤ 15 miles, then we used an average estimated speed of 10 mph,
- If the mode is local transit bus, school bus, subway, trolley (TRPTRANS=10,12,17,18), and the trip distance is > 15 miles, then we used an average estimated speed of 20 mph,
- If the mode is car, van, SUV, pickup truck, other truck, recreational vehicle, motorcycle, commuter bus, charter bus, city to city bus, Amtrak, commuter train, taxi cab, limo, shuttle (TRPTRANS=1, 2, 3, 4, 5, 6, 7, 11, 13, 14, 15, 16, 22, 23, 24), and the trip distance is ≤ 15 miles, then we used an average estimated speed of 25 mph,
- If the mode is car, van, SUV, pickup truck, other truck, recreational vehicle, motorcycle, commuter bus, charter bus, city to city bus, Amtrak, commuter train, taxi cab, limo, shuttle (TRPTRANS=1, 2, 3, 4, 5, 6, 7, 11, 13, 14, 15, 16, 22, 23, 24), and the trip distance is > 15 miles, then we used an average estimated speed of 50 mph,
- If the mode is bicycle (TRPTRANS=25), for any trip distance, the average estimated speed used was 10 mph,
- If the mode is walk (TRPTRANS=26), for any trip distance, the average estimated speed used was 3 mph, and
- If the mode is anything else (airplanes, ships, ferry's boats, other) the data was left as is as the variation was too great to estimate the trip time duration.

Once travel time was estimated, the imputed start or end time was calculated by subtracting the estimated travel time from the end time. The imputed end time was calculated by adding the estimated travel time to the start time. In all cases where a start or end time was imputed, the trip fit into the travel day roster without causing any overlapping trips.

In addition to missed trips reported for the subject on missed trip forms, trips not reported for the subject during the person interview were also imputed. These trips were imputed when a subsequent household member reported that a household member who had completed a person interview earlier had accompanied them on a trip. Since the earlier household member had already completed the interview, it was assumed that (s)he had forgotten to report the trip. The missing trip from the subsequent household member was copied to the travel day record for the household member who had completed the interviewer earlier. An imputation flag to set to indicate trips that were added.

The other variables such as AGE, SEX, OTHRPHON, etc. were imputed during the weighting process if the value for the variable was missing (-7. -8 or -9).

3-D.5. IMPUTING DATES FOR RECURRING TRAVEL PERIOD TRIPS

After a person reported a long distance trip, they were asked if they had taken this trip more than once during the 28-day reporting period, and if so, how many times they took it. Trips taken more than once during the travel period were defined as recurring trips.

We imputed for the missing departure date (IMPTLEDT), the date the trip ended (IMPTREDT), and the number of times (IMPTNTIM) the recurring trip was made during the travel period. In addition, for some reported first trips, to reduce the number of trips that had identical or overlapping trip times, we edited the start date, end date, number of times or the flag variable showing whether the trip is a recurring trip (EDITRECU). A flag for these variables indicates that the variable has been imputed or edited.

To reduce respondent burden trip detail was collected for just the first trip in the recurring trip series. The travel period file shows each recurring trip as a separate trip. To do this a procedure was developed to impute dates for each such recurring trip. The variables IMPTREDT and IMPTLEDT indicate whether the date was imputed.

The imputation procedure evaluated several factors prior to assigning a date. When possible, the date was assigned to the same day of the week. For example, if a

trip was reported for the first Tuesday in the 28-day period and there was one additional such trip taken, it would be assigned to the third Tuesday of the 28-day period. When there were more trips reported than could be assigned to the same day of the week, then recurring trips were assigned to another “weekday” if the reported trip was on a “weekday” and to another “weekend” day if the reported trip was on a “weekend”. Depending on the purpose of the trip, “weekend” was defined as Saturday and Sunday, or as Friday, Saturday and Sunday. Other factors evaluated included whether the subject had any other trips overlapping with the imputed trips. For example, if the dates imputed were from March 1 to March 3, 2002 but the subject had another travel period trip on March 2, then another date was imputed. The procedure also looked at the other household members that were on the trip and ensured that the date assigned did not conflict with other trips these household members may have taken. Finally, recurring trips were assigned the same date as the travel day if the subject’s travel day trip information indicated that s(he) took a trip of 50 miles or more from home if certain conditions were met. If the recurring trip started and ended the same day, then it was assigned the same date only if the travel day trip information indicated that the trip started and ended the same day. Similarly, if the recurring trip involved being away overnight, then it was assigned the same date only if the travel day trip information indicated that a trip started before the travel day.

3-D.6. TRIP REPORTING

The travel day trip roster for each household member who completed a person interview provides a listing of all trips taken on the travel day. However, to reduce respondent burden, not all household members were asked to provide trip detail for each trip taken. For example, trip detail was not asked on proxy interviews if the same trip was self-reported earlier by another household member. Post data collection, trip details recorded during the interview with the self-reported household member were copied to the record for the proxy household member who reported being on the same trip.

A similar procedure was implemented for travel period. That is, after travel period trip detail was recorded for a particular trip, subsequent household members who reported being on the same trip were asked if they were with the previous household

member at all times on the trip. If they were, then trip detail was not collected from subsequent household members.

3-D-7. DATA MOVED FROM CATI TO DATA FILES SPECIFIED BY DOT

The CATI data set was converted into a SAS data set and separated into public use data files based on the specifications provided in the Codebook included in Appendix B. The data files created had several "derived variables" that were created by either renaming CATI variables or combining multiple variables. The specifications for creation of the derived variables are included in this report as Appendix D, Derived Variables. During this step, the survey weights and other variables not collected during the survey were also appended to the data files.

The version of the User's Guide for the Public Use data set released in January 2003 did not include data collected on long-distance travel. The Version 2 release to DOT in the Summer of 2003 contained both travel day and travel period data and therefore included the four files below in addition to the travel period trip file. The final January 2004 delivery included the following four files. BTS will be responsible for the release of the travel period trip file.

- **Household file** - data collected once for the household (one record per household),
- **Person file** - data items collected once for each interviewed household member (one record for each completed person interview),
- **Vehicle file** - data items related to the household 's vehicles (one record for each household vehicle),
- **Travel day trip file** - data items collected for each trip an interviewed person made on the household 's travel day (one record for each travel day trip each person made), and

3-D.8. USEABLE HOUSEHOLDS

The data files contain information on only households that are "useable." A useable household in the 2001 NHTS is one in which the household interview was

completed, and person interviews were completed with at least 50 percent of the adult (age 18+) household members. Though all completed household and person interviews in the CATI database were edited, only information for useable households has been provided in the four files.

A household interview was considered complete if the:

- Interviewer asked every applicable question in the household questionnaire and set an appointment to call the household back to collect diary information;
- The household respondent provided the complete household roster information for the household; and
- The household respondent provided an address for mailing the travel diaries to the household.

The person interview was considered complete if the interviewer administered every applicable question to the subject during the person interview. That is, the interviewer got to the last question in the questionnaire and was able to thank the subject for participating in the survey.

Table 3-1 shows the number of household and person interviews completed during the 2001 NHTS. This public use dataset contains information on the 69,817 useable households in the national sample. Although the definition of a useable household required only 50 percent of adults to complete a person interview, Table 3-1 shows that for the January 2004 release (Version 3) sample for 60,520 or 86.7 percent of useable households, person interviews were completed with all adult household members in the household.

Table 3-1. Number of Completed Household and Person Interviews: Overall, in Useable, 100% Households, and Non-Useable Households

Survey Completion Level	Completed Household Interviews	Completed Person Interviews
Overall Number of Completed Interviews	106,598	163,856
Useable Households	69,817	160,758
Number of Completed Interview in 100% Households (All Adult Household Members completed a Person Interview)	60,520	144,884
Non Useable Households	36,781	3,098

3-D.9. HUNDRED PERCENT HOUSEHOLDS

For some applications, particularly those involving planning models, the data user may want to access only those households where all adults in the household were interviewed. These are the 60,520 households shown in Table 3-1 above. If a user wants to limit analysis to these 100% households, a separate weighting factor can be used to expand the 100% households to annual, national estimates. This weighting factor is EXPFLLHH on the household and vehicle files, EXPFLLPR on the person file and EXPFLLTD on the travel day trip file.

3-D.10. EDITING THE DELIVERY DATASETS

As a final editing step, frequencies for useable households on both the CATI dataset and the delivery datasets were compared. Next, edits were run on the four delivery data sets to ensure consistency in the reporting of values across the four delivery files.

3-E. SURVEY METHOD AND PROCEDURE CHANGES

3-E.1. 2001 NHTS CHANGES

The 2001 NHTS represents a significant change in survey methods and procedures from earlier national travel surveys. In Section 2-D, Survey Content Changes in 2001, we presented changes to the questionnaire content in 2001. Therefore, in the items that follow we focus on modifications to survey methods and procedures introduced during the 2001 NHTS.

1. The 2001 NHTS saw the combination of two travel surveys, the National Personal Transportation Survey (daily travel) and the American Travel Survey (long-distance travel). Each household that completed a household interview was assigned both a travel day and travel period. Detailed travel information was collected on both daily and long-distance travel.
2. The advance mailing to households for whom addresses were available included a five-dollar cash incentive and a brochure.
3. All household members were eligible for a person interview, not just household members who were over four years old.
4. Proxy rules were modified. Only subjects 16 years and older were asked to respond for themselves. Proxies were requested for all others. However, if asked by an adult household member, interviewers could directly speak with household members that were 14 and 15 years old.
5. More detailed address information was collected. The off-line geocoding operation used multiple databases and detailed manual searches to determine the latitude and longitude of a location when the address information failed an automated batch geocode search.
6. The second odometer reading for household vehicles was collected using multiple modes of data collection. In addition to mail out/mail back and interviewer initiated telephone data collection, modes also included the use of the Internet, facsimile machines and a toll-free 800 number.
7. The process for calculation of weights was more detailed. Steps combined in the past were now separated. For example, non-response adjustments and benchmarking to key variables were performed as separate steps.

8. During the editing process, certain travel day trips reported as single trips were split creating two trips from the original one trip reported by the household member. These usually involved trips that originated from home and were the last trip taken on the travel day without returning home. If such a trip had walk or bike as the transportation mode or the trip destination and had a purpose that was go to the gym/exercise/play sports, go to public place, walk the dog/vet visits, or pick up or drop off someone, the trip was split. In these cases, the survey procedure was to split the trip into an outgoing and a return portion to make them more parallel to the handling of travel day trips by other modes of transportation.
9. During the editing process, at the request of DOT, select trips reported by a household member (e.g., Person B) that were not reported by a previously interviewed household member (e.g., Person A) were added to Person A's trip data if Person B indicated that Person A also traveled on the trip.

Exhibit 3-1 summarizes key changes mentioned in Section 2-D and in this section. The reason for change has also been provided to indicate the probable impact the change may have on the 2001 survey.

Exhibit 3-1. Changes in the 2001 NHTS Survey Methodology and Content and Their Probable Impacts

TOPIC	FROM	TO	PROBABLE IMPACTS
What is collected?	Two separate surveys - the NPTS and the ATS	Combined survey that collects both travel day and travel period information	Enables analysis of relationship between daily and long-distance travel characteristics of each person
Which household members are eligible?	Household members age 5 and older	All household members	More complete trip reporting
When proxy needed?	Proxy for household members 5 to 13 years	Proxy for household members under 16 years	Increase in number of interviews by proxy Obtain parental approval when speaking with 14 and 15 year olds

Exhibit 3-1. Changes in the 2001 NHTS Survey Methodology and Content and Their Probable Impacts (contd.)

TOPIC	FROM	TO	PROBABLE IMPACTS
Respondent Contact	Advance letter	Advance letter with a \$5 cash incentive and a brochure	Improved response Legitimizes the survey with respondents
Use of a diary for long trips	The ATS used a diary to record long-distance trips	No travel period diary included	Lower respondent burden and reduce the possibility of confusion due to the mailing of both a travel day and travel period diary
Travel day trip definition	Any stop from one address to the next is a separate trip	Basically the same - stops only to change a mode of transportation excluded	May improve reporting of trips by public transportation as subjects were specifically reminded about these trips No change mode trips were recorded except where public transportation was involved.
Walk and bike trips on travel day	No specific mention of walk and bike trips	Specific reminder to include walk, bike rides and trips that started and ended in the same place	Will increase the reporting of walk and bike trips
Travel period length and travel period trip definition	The NPTS included trips of 75 miles or more and used a 2-week recall period. The ATS included trips of 100 miles or more taken over a full year (4 interviews)	The travel period was a four-week period. Trips of 50 miles or more from home were defined as long-distance	Four-week travel period and shorter criterion distance provides information on a larger sample of long-distance trips than NPTS and better recall of trips than ATS (if not recorded in ATS diary), but a smaller sample of trips and greater difficulty estimating annual long-distance trip rates than ATS. The 4-week travel period may have increased the potential for telescoping (i.e., bringing trips into the travel period)
Travel day trip purpose	There were 17 trip purpose categories	There are 36 trip purpose categories	The new categories more accurately capture responses

Exhibit 3-1. Changes in the 2001 NHTS Survey Methodology and Content and Their Probable Impacts (contd.)

TOPIC	FROM	TO	PROBABLE IMPACTS
Most recent long-distance trip	Not collected	Collected	Facilitate the imputation of trips for persons with no reported long-distance trips in travel period
Odometer readings	Readings collected by contacting the respondent by phone or by mail	Data collection modes also included the Internet, fax, and a toll-free 800 number	Improved response
Geocoding	Limited use of manual geocoding	Extensive use of manual geocoding	Higher geocoding success rates and more accurate geocoding
Splitting walk and bike trips at the end of travel day	Not conducted	Conducted	Walk and bike trip rates may be higher than on past NPTSS
Adding trips not reported by household members interviewed earlier	Not conducted	Conducted	More complete trip reporting
Weighting	Raking to control totals ²	Several stages of separate nonresponse adjustment and trimming as well as raking. Changes to cells used for raking	Presently unknown. An evaluation is to be conducted.

² In raking, one adjusts estimates to agree to one set of controls (e. g., ethnicity), then adjusts estimates to a second set of controls (e. g., region), etc. This process is then repeated until all estimates are simultaneously close to the full set of controls.

3-F. PROCEDURES TO ENSURE CONFIDENTIALITY

The following measures were taken to produce the public use dataset that accompanies this User's Guide to ensure respondent confidentiality:

- All direct identifiers, such as telephone numbers, zip codes, county codes, names of individuals, and addresses were removed from the dataset;
- Metropolitan Statistical Areas (MSAs) of less than 1 million population, states with less than 2 million population, and states for non-MSA households in states that have a total non-MSA population less than 1/2 million are not specifically identified on the dataset;
- The specific dates when travel day and travel period trips were made were removed from the file; and
- Data values for certain other variables were coded into intervals or suppressed, and some data distributions were capped. For example, detailed year/make/model information for antique and classic autos could compromise respondent confidentiality if fully revealed. In the public use dataset, rare make and model codes were not displayed.

Note: Identifying the MSA or CMSA of a household in an area with a population of one million or more requires the use of two variables: HHC_MSA and MSASIZE.