2017 NHTS Data User Guide

March 2, 2018 (revised March 8, 2018)

Submitted to and revised by:
Federal Highway Administration
Office of Policy Information
1200 New Jersey Avenue, SE
Washington, DC 20590

Submitted by:
Westat
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

Table of Contents (continued)

Table of Contents (continued)

Table of Contents (continued)

Figure **Page**

Exhibit

Introduction to the 2017 NHTS 1

The National Household Travel Survey (NHTS) is the source of the nation's information about travel by US residents in all 50 States and the District of Columbia. This inventory of travel behavior includes trips made by all modes of travel (private vehicle, public transportation, pedestrian and cycling) and for all purposes (travel to work, school, recreation, and personal/family trips). Federal and state agencies use the survey results to monitor the performance and adequacy of current facilities and infrastructure, and to plan for future needs. State and regional agencies use the data to support travel demand modeling and long-range transportation planning. Data from the NHTS are included in broader, bi-annual reports to Congress on the performance of the surface transportation system. Survey data are also applied outside of transportation in the fields of public health, environmental analysis, time use studies, energy consumption, and social welfare.

The collection and analysis of national transportation data has been of critical importance for nearly half a century. Previous surveys conducted in 1969, 1977, 1983, 1990, 1995, 2001, and in 2009 served as a significant resource for understanding travel behavior in the nation. The current survey (2017 NHTS) is the eighth in the series. The NHTS is the main national source of data on how the travel behavior of the American public is changing as demographic, economic, and cultural changes are taking place in our country.

The NHTS data are collected directly from a stratified random sample of U. S. households. The diversity of information needed to support the wide range of transportation decisions in the current environment underscores the importance of reliable data for evidence-based decisions. The NHTS provides data on individual and household travel behavior trends linked to economic, demographic, and geographic factors that influence travel decisions and are used to forecast travel demand. Just a few examples of how the data collected in these surveys are used include:

- **Describing and Analyzing Current Travel and Developing Trends Over Time**: These include measures of mobility such as trip rates, miles traveled, vehicle availability, as well as the purpose of trips, the amount and nature of travel by each mode, and temporal patterns of daily travel. Data from the 2017 NHTS facilitates the analysis of each of these travel attributes by the socio-demographics of the traveler and household.

Importantly, as in past NHTS Surveys, the 2017 NHTS allows the development of trends over time to help measure changes in mobility patterns[1].

- **Energy Consumption, Environmental Concerns and Household Vehicles**: These encompass modeling air quality, energy consumption of the household vehicle fleet, relationships between gas prices and travel, the carbon footprint of travel, and the penetration and use of vehicles by type, such as hybrid and electric vehicles.

- **Travel of Specific Demographic Groups:** To analyze travel behavior of commuters, school children, millennials, the elderly, and immigrant and low-income groups to help inform policies and programs addressing mobility issues, air quality, the development and use of new technologies, equity concerns, and outreach.

- **Modeling and Planning Applications**: Includes use of the NHTS as model inputs as well as in calibrating and validating state and regional travel demand and forecasting models (both trip-based and activity-based models), the intersection of land-use and travel behavior, and linking NHTS data with American Community Survey (ACS) data.

- **Safety**: Provides travel data by age, gender, mode use, auto occupancy and time of day and gives a broader context to crash and fatality data.

The Office of Management and Budget's (OMB) Office of Statistical Programs, in accordance with the Paperwork Reduction Act, clears all federally funded surveys. The primary purpose of this clearance is to ensure a Federal agency is not burdening the public by collecting information already available from another agency or another source. Additionally, OMB checks on the reasonableness of the questions asked, the suitability of the survey for the needs it purports to fill, and the necessity for any potential sensitive information asked. The OMB clearance process is also designed to ensure that survey statistical and methodological norms are being met. In November 2015, OMB provided FHWA with clearance to conduct the 2017 NHTS under number 2125-0545. OMB clearance allowed Westat, FHWA's contractor for the 2017 NHTS, to conduct the survey. This involved developing the survey instruments, pretesting the survey, conducting cognitive and usability tests of the instruments and electronic systems, developing the sample frame, fielding the survey, cleaning and weighting the data, and developing the weighted and unweighted data files. Travel information was collected between April 2016 and May of 2017. The survey data are available on the NHTS website at: http://nhts.ornl.gov. This website includes tools to enable data access without needing to download the data files. The NHTS website also contains assistance to users in the form of the NHTS Academy, selected publications produced from the data, a compendium of all publications using NHTS data and other user support information.

[1] FHWA has tasked Westat with the development of the *2017 NHTS Summary of Travel Trends*. It is planned for release by the summer of 2018 and will analyze trends between the 2017 NHTS and past NHTS'.

The next section describes in more detail the sponsors and users of the NHTS data.

1.1 Survey Sponsors and Users

The Federal Highway Administration (FHWA), sponsored the 2017 NHTS. Additional funding was provided through thirteen Add-on agencies, detailed later in this section.

Title 23, United States Code, Section 502 authorizes the U.S. Department of Transportation (USDOT) to carry out transportation research to measure the performance of the surface transportation systems in the US, including the efficiency, energy use, air quality, congestion, and safety of the highway and intermodal transportation systems. The USDOT has overall responsibility to obtain current information on national patterns of travel, establish a database to better understand travel behavior, evaluate the use of transportation facilities, and gauge the impact of the USDOT's policies and programs.

Data from the NHTS are widely used to support research needs within the USDOT, and State and local agencies, in addition to responding to queries from Congress, the research community and the media on important issues. Within the USDOT, the Federal Highway Administration (FHWA) holds responsibility for technical and funding coordination. Other primary data users include the National Highway Traffic Safety Administration (NHTSA), Federal Transit Administration (FTA), and the Bureau of Transportation Statistics (BTS); these agencies have historically participated in project planning and financial support.

With respect to regulatory uses, NHTS data are integral to the calculation of the model year Corporate Average Fuel Economy (CAFE) standards, which are regulations issued by the National Highway Traffic Safety Administration.[2] The data are commonly used to estimate vehicle miles of travel by specific groups, such as age groups or men and women drivers.[3]

The NHTS informs other policy research, including health and energy. For example, the Centers for Disease Control (CDC) uses the data on the percent of children who walk to school (among other

[2] https://www.fhwa.dot.gov/publications/research/randt/evaluations/16082/16082.pdf

[3] http://nhts.ornl.gov/usersAndUses.pdf. Also, see for example, research such as: https://www.sciencedirect.com/science/article/abs/pii/S0001457509002966

indicators) from the NHTS as part of their Healthy People 2020.[4] The Energy Information Administration (EIA) uses the NHTS data as a source of information on household fuel consumption for transportation,[5] the Environmental Protection Agency (EPA) to provide default values for the Moves2010 model[6] and to estimate the value of various policy initiatives related to fuel consumption and greenhouse gas emissions.[7] Policy and decision makers use these statistics extensively to understand economic and environmental impacts of changing travel demand.

Advocacy groups and non-profit organizations are another set of data users. For example, the American Association of Retired People (AARP) and the American Association of Automobiles Foundation for Traffic Safety (AAAFTS), often use NHTS data in their reports and presentations to increase awareness about priority topics and to lobby Congress for action.[8]

To support planning and policy-making at the State and regional level, since 1990 the NHTS has encouraged 'Add-on' partners who wish to purchase supplemental samples of the survey in their State or Metropolitan Planning Organization (MPO) area. With the more robust sample sizes provided in the Add-on areas, States and MPOs are able to perform more in-depth analyses and to drill down to smaller geographic units, such as cities or counties.

States and MPOs use the data for a range of purposes. In most cases, these agencies use NHTS trip data to develop, calibrate, or validate State and MPO travel demand models. These models are critical to transportation planning and inform long-range planning, as well as corridor level, interchange, and transit infrastructure and planning projects, among others. In addition, some States and MPOs that are required to produce air quality reports (because of their status as a non-attainment area) use NHTS data as an input to their models.

The goal in the 2017 NHTS was to obtain completed surveys from 129,112 households, which included a national sample of 26,000 households and 103,112 additional Add-on partner samples purchased by thirteen States or MPOs. Tables 1-1 lists the nine State DOTs partners, and Table 1-2, the four regional planning agencies/MPO partners. The number of surveys actually completed in the 2017 NHTS is slightly higher than the target amount.

---

[4] See table 33-2 in https://www.cdc.gov/nchs/data/hpdata2020/HP2020MCR-C33-PA.pdf

[5] https://www.eia.gov/todayinenergy/detail.php?id=20832

[6] https://www.epa.gov/sites/production/files/2016-06/documents/fleet-activity-moves-2011.pdf

[7] For example, see: https://www.epa.gov/greenvehicles/what-if-we-kept-our-cars-parked-trips-less-one-mile

[8] https://www.fhwa.dot.gov/publications/research/randt/evaluations/16082/16082.pdf

Table 1-1. State Departments of Transportation

Add-on Partner	Target Number of Completed Households[1]
Arizona	2,444
California	24,000
Georgia	8,000
Maryland	1,000
New York State	15,851
North Carolina	8,000
South Carolina	6,500
Texas	20,000
Wisconsin	11,000

[1] These are households for which all of the household members ages five and older complete the retrieval survey.

Table 1-2. Regional Planning Agencies/Metropolitan Planning Organizations

Add-on Partner	Target Number of Completed Households[1]
Des Moines Area MPO in Iowa	1,200
Indian Nations Council of Governments in Oklahoma	1,000
Iowa Northland Regional Council of Governments in Iowa	1,200
North Central Texas Council of Governments in Texas	2,917

[1] These are households for which all of the household members ages five and older complete the retrieval survey.

Section 1.2 describes in more detail the scope and geographic coverage of the NHTS sample.

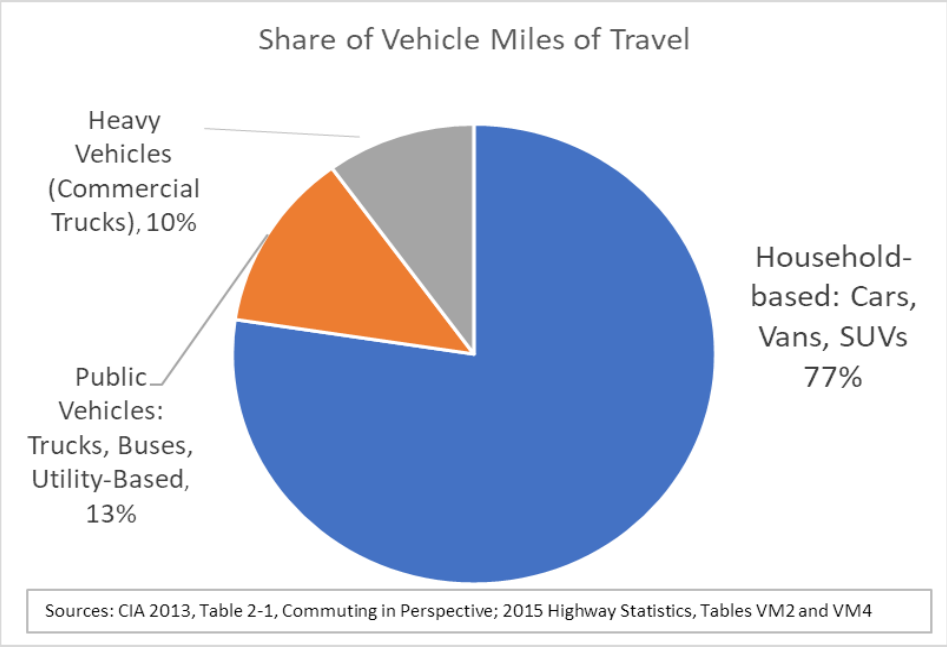
1.2 Overview of Survey Scope and Coverage

As in the previous series of the NHTS, the 2017 NHTS maintained a two-phase study, which included a household recruitment survey (phase 1) and a person level retrieval survey (phase 2). Older iterations of the NHTS refer to these phases as the short and long surveys. Unlike the previous series of NHTS' which used a Random Digit Dial (RDD) telephone sampling method and only Computer-Assisted Telephone Interview (CATI) data collection; the 2017 NHTS used Address-Based Sampling with mail-back as the primary recruitment survey response mode with phone or web as secondary response options, while the person level retrieval survey phase offered both phone and web response options.

The NHTS is designed to collect information about daily travel generated by US households—including urban and rural, multi-family or single-person, and households of people who travel a lot

or not at all. The sample does not include group housing such as dormitories, prisons, rest homes, and other housing where 10 or more unrelated persons reside. According to the most recent statistics, household-based travel in light-duty vehicles (such as cars, vans, and SUVs) accounts for over three quarters of all roadway vehicle miles in the US (see Figure 1-1).

Figure 1-1. Share of Vehicle Miles of Travel by Sector



Section 1.3 describes in more detail the content of the data files collected in the NHTS.

1.3 Overview of Survey Content

The NHTS data are the main source of information on daily travel linked to individual personal and household characteristics, socio-economic characteristics, vehicle ownership, and vehicle attributes. The daily travel data includes an inventory of all trips taken within a 24-hour period by all household members aged 5 or older. For each trip, respondents report trip purpose (e.g., work, school, shopping, recreation, etc.), mode of transportation (car, bus, walking, etc.), time of day of travel, day of the week, and vehicle occupancy. These data can be linked with the household vehicle characteristics (make, model, and year), respondent demographic characteristics (gender, age, driver and worker status, etc.), and household socio-economic characteristics (income, number of workers, and housing type/neighborhood characteristics). Figure 1-2 illustrates the relationships between the household, person, travel, and vehicle data and includes examples of the core variables collected.

Figure 1-2. Schematic of the NHTS Data

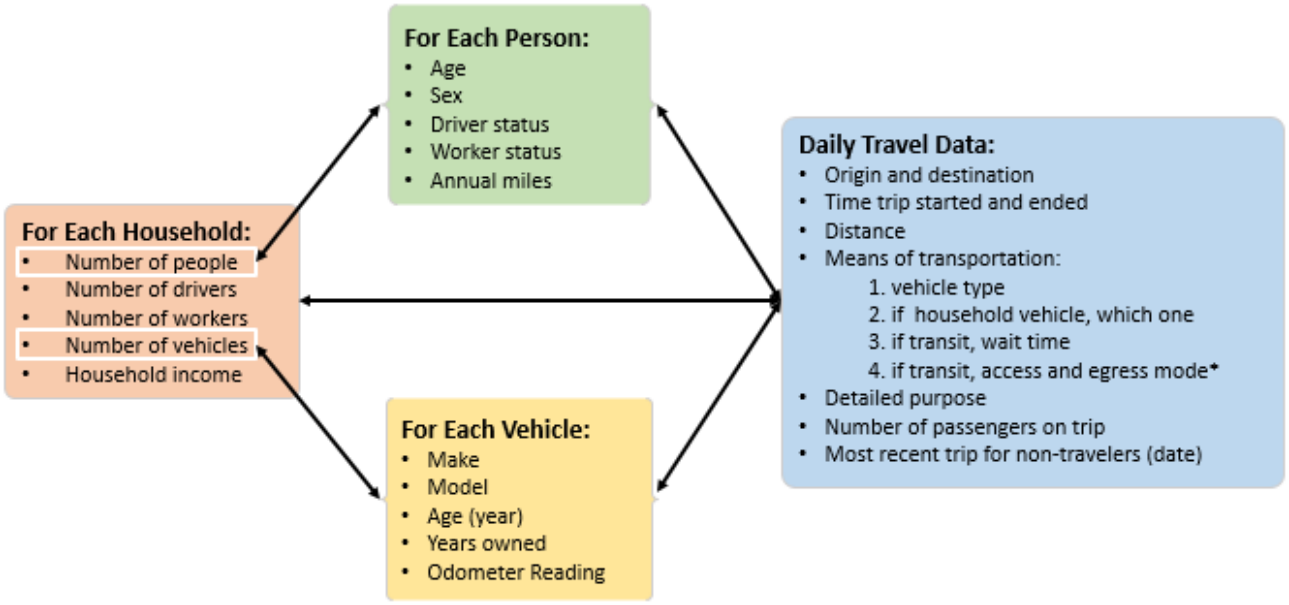


Table 1-3 lists select NHTS data available for households, vehicles, persons, workers, and travel days. We have *italicized and bolded* data considered "core" NHTS data. Core data are items collected in all seven surveys to date.

Table 1-3. 2017 NHTS Summary of Content

2017 NHTS Summary of Content

For Each Household:

- ***Number of people, drivers, workers and vehicles***
- ***Income***
- Housing type
- Owned or rented
- ***Race of reference person***
- ***Hispanic status of reference person***
- Tract and block group characteristics
- Internet Use & Delivery to households

For Each Person:

- ***Age/Sex/Relation to reference person***
- ***Driver status***
- ***Worker status/Primary activity***
- Home deliveries from Internet shopping
- Travel Disability
- Effect of disability on mobility
- ***Education level***
- Immigrant status
- Views on transportation
- ***Annual miles driven***
- Incidence of public transit use in past month
- Incidence of motorcycle use in last month
- Incidence of walk and bike trips in past week
  - Number of walk/bike trips for exercise*
- Incidence of use of ridesharing app in last month
- Incidence of use of cars haring service in last month
- Usual mode to school
- Use of travel log on travel day

For Each Worker:

- ***Full or part-time***
- More than one job
- Occupation (four categories)
- ***Workplace location***
- ***Usual mode to work***
- ***Drive alone or carpool***
- ***Usual distance to work***
- Usual time to work
- Work from home
- Usual arrival time at work
- Flexibility in work arrival time

For Each Vehicle:

- Make/Model/Age (year)
- Body type
- Fuel type*
  - If hybrid, type of hybrid*
- Annual miles driven
- How long owned
- Odometer reading
- Alternative Fuel
- Primary driver

Daily Travel Data:

- Origin and Destination address
- ***Time trip started and ended***
- ***Distance***
- ***Means of transportation:***
  - ***Vehicle type***
    - ***If household vehicle, which one***
    - If transit, wait time
    - if transit, access and egress mode
- ***Trip Purpose***
- Detailed purpose
- ***Travel Party Size***
- Last time of travel

*added in 2017

In addition to person, household, vehicle, and daily travel related data, the 2017 NHTS also collected data on respondents' opinions and travel experiences. Select topics covered are bulleted below.

- Typical travel behaviors, attitudes, and beliefs of the reference person.
- Reasons that each person does not walk or bike more often.
- Availability and use of alternative means of transportation (such as bike, walk, public transportation, taxi, car share, ride share, etc.) – including whether these modes are available to the person, the number of times used in a 30 day period (for car share and ride share), and carpool frequency in the past week.
- Work start time, if they typically work from home, commute time without traffic, and minutes spent transferring on their commute to work if they take public transportation.
- The general health and physical activity level of each person.
- Technology such as internet use through personal electronic devices (PED's), smartphone app use, and internet purchases.

When was the Survey Conducted?

A pilot study was conducted from July to October 2015. The main survey was conducted from March 2016 through May 2017 with travel dates starting on April 19, 2016 and ending on April 25, 2017. Travel days were assigned for all seven days of the week, including holidays. The survey data were weighted to a 12-month period to produce annual estimates of travel.

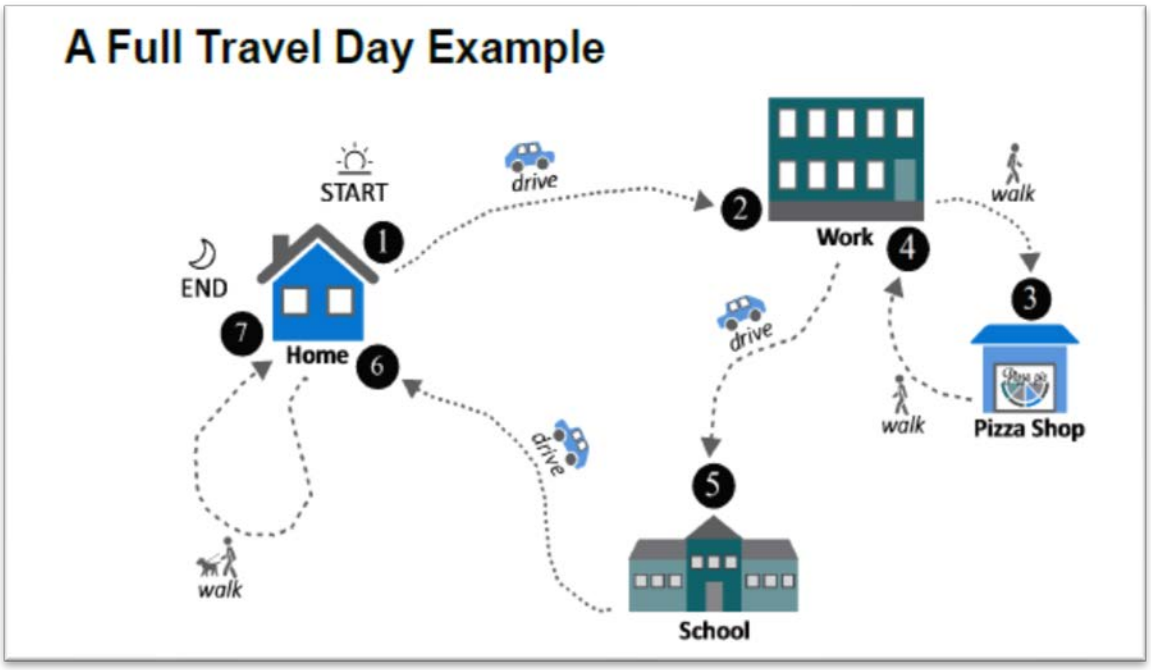
What Trips were Included?

All trips made on the assigned travel date by persons age five and older were collected from household members through the retrieval survey. The designated 24-hour travel day started at 4:00 a.m. (local time) of the assigned travel day and ended at 3:59 a.m. of the following day. On a typical day, 4 a.m. represents the time when a relatively small number of people are traveling. Starting the travel day at this time increased the likelihood that most household members would be at home at the start of their travel day.

Survey materials participants received included instructions and pictorials on how to record their travel. An example of a pictorial is Figure 1-3. It illustrates seven trips as well as examples of trips

such as walk trips that have a higher likelihood of being misreported or not reported at all. Additional survey materials can be accessed in the NHTS Field Documents.

Figure 1-3. Travel Day Graphic included on the Travel Log



2 Survey Procedures and Methodology

An understanding of the data collection for the NHTS is important for the proper use and interpretation of the data. Details in the next few sections provide a basic background on the survey process to aid in understanding a typical household's involvement in the survey.

Data collection for the 2017 NHTS, national sample and thirteen Add-ons areas, was conducted by Westat. Key aspects of the methodology are highlighted in the sections that follow.

2.1 Sample Design and Selection

This survey was designed as an address based sample (ABS) survey to yield an equal probability sample of households. The national sample was supplemented by sample from 13 Add-on areas, which included Arizona, California, Dallas-Ft. Worth TX, Des Moines IA, Georgia, Maryland, New York, North Carolina, South Carolina, Texas, Tulsa OK, Waterloo IA, and Wisconsin. The national sample file was designed and selected to yield 26,000 completed households. The distribution of the 103,112 target numbers of completed households for the Add-on areas are listed in Tables 1-1 and 1-2.

Sampling Frame – The sampling frame consisted of all residential addresses on a frame maintained by Marketing Systems Group (MSG). Their frame originates from the U.S. Postal Service (USPS) Computerized Delivery Sequence file (CDS); MSG updates it on a monthly basis. A sample of addresses was selected at two points in time: first in January 2016 before the main study data collection began and again in September 2016 about halfway through the main study data collection, using MSG's most recently updated sampling frame. This design allowed for sample size adjustments, in both the national sample and the 13 Add-on areas, for the second half of data collection in areas where response rates differed from the original assumptions.

Sample Selection – Addresses were assigned to sampling strata prior to selection, and then randomly selected from each stratum. Sample selection was implemented for the national sample and the Add-on areas at the same time, within each of the specified sampling strata.

The primary sampling strata consisted of four possible geographic designations. Additional sample stratification was incorporated into the sampling in some of the Add-on areas. The groupings were:

1) Addresses located in Metropolitan Statistical Areas (MSAs) with access to heavy rail transit and more than 1 million people,

2) Addresses located in MSAs with more than 1 million people but no access to heavy rail transit,

3) Addresses located in MSAs with fewer than 1 million people, and

4) Non-MSA addresses (i.e. rural areas or small towns).

Day of Week Assignment – The variation in travel by day of the week was balanced in the national sample by randomly assigning the travel days for one-seventh of the sample addresses to each day of the week. Several Add-on areas opted to have fewer households assigned to weekend travel. For those Add-on areas, Saturday and Sunday were each assigned travel days for one-fourteenth of the sampled addresses. The remaining six-sevenths of the households were randomly assigned evenly across weekdays (Monday – Friday). This variation in travel day assignment was offered because many states and MPOs use only weekday travel when updating their travel demand models. More information can be found in the NHTS Sample Design Plan.

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. People living in dormitory rooms, fraternity and sorority houses were also excluded from the sample.

All sampled addresses that were established by respondents as residential were eligible for the household interview. To be eligible to participate in the survey, the household must have at least one person 18 years of age or older living at the sampled address.

2.2 The Survey Process

Recruitment Mailing – Households were invited to participate in the recruitment survey through the mail. The initial recruitment package contained an invitation letter, recruitment questionnaire, a pre-paid envelope to return the completed survey, and a $2 cash pre-incentive to encourage participation. The invitation letter provided details about the survey, instructions on completing the

paper questionnaire and returning it using the pre-paid envelope and accessing the project website URL to obtain additional information and confirm the legitimacy of the survey.

A reminder postcard was mailed to households one week after the initial invitation package mailing. This first postcard reminded households to fill out the paper questionnaire and return the completed survey provided in the invitation package.

A third mailing containing a second recruitment package was mailed to non-responding households three weeks after the initial recruitment package mailing. This non-response package included a non-response letter affirming that there was still time to participate, paper questionnaire, and a pre-paid envelope. The non-response letter urged respondents to fill out the paper questionnaire and return the completed survey provided in the pre-paid envelope.

Finally, a fourth mailing which consisted of a second and final reminder postcard was mailed to non-responding households forty-four days after the initial recruitment package mailing. The second reminder postcard provided the survey website and a unique household Personal Identifiable Number (PIN) with instructions to complete the survey online.

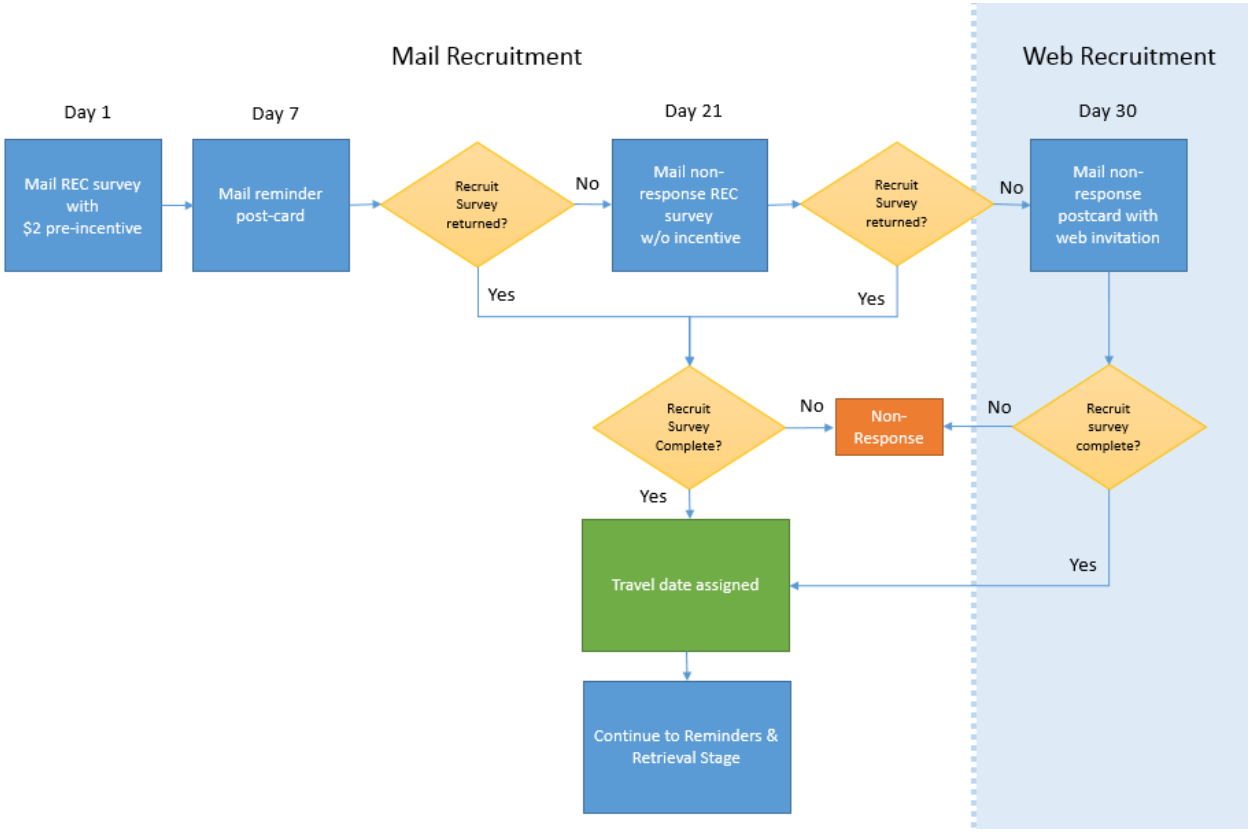
Adjustments in the mailing schedule of the non-response package and second postcard were made midway through data collection to encourage a more timely response. The mail date of the second reminder postcard was moved up from 44 days to 30 days after the initial recruitment package mailing. In addition, the survey website URL and household PIN were added to the non-response letter providing households the option of completing the recruitment survey through the paper questionnaire, online, or with a telephone interviewer using a CATI system.

All recruitment materials included the survey helpdesk toll-free telephone number for participants to call with questions about the survey or to complete the survey over the telephone. Recruitment materials also included instructions in Spanish on how to complete the survey. The NHTS Field Documents contain examples of each of the recruitment materials.

Westat mailed recruitment packages and reminder postcards to over 900,000 households. Over 239,000 (27%) households completed and returned a paper questionnaire resulting in recruitment completes. An additional 13,000 (1.4%) questionnaires were completed over the web or telephone. Approximately 7.5 percent of the recruitment packages mailed were returned to Westat as "undelivered" by the post office.

Recruitment Survey – This survey collected brief information about the household, each household member, the number of household vehicles and household contact information. The recruitment survey also included attitudinal questions about transportation and travel behavior to add survey relevance and to make it appealing to the respondent. The intent of the recruitment materials was to garner interest in the survey and collect contact information for the retrieval survey communication. Figure 2-1 summarizes the recruitment process.

Figure 2-1. Recruitment Survey Flowchart



The recruitment methodology, to push participants to respond by completing the paper questionnaire, resulted in the vast majority (95%) of recruitment completes from mail-back. Online and telephone recruitment surveys were available if requested by the participant and introduced in later recruitment mailings to combat non-response and resulted in about 5% of the recruitment completes.

Travel Day Date Assignment -- After each recruitment survey was processed, the household was automatically assigned a travel date. The assigned travel date was communicated to each household through the travel log package mail materials. The household's travel date was based on the pre-

selected day of the week randomly assigned to the household during sample file development. Travel dates were assigned at least 10 days in the future from the date the recruitment package was processed, which allowed time for the retrieval material package to reach the household ahead of the travel date. This provided the household time to review the mailing, understand the process, and to call the toll-free project helpdesk hotline if they had any questions prior to their travel date.

Travel Log or Retrieval Mailing – The use of travel logs by respondents in household travel surveys has been shown to improve the accuracy of trip reporting. Therefore, each household that completed a household recruitment interview was mailed a package that included a travel log for each member of the household five years of age and older to use in recording travel on the assigned day. Of the over 275,000 persons who reported their travel, 66 percent reported using their travel log. The package was sent via United States Postal Service Priority Mail within a day or two following the completion of the household recruitment interview. The package contained:

- A letter from the U.S. Department of Transportation thanking the household for completing the recruitment survey, inviting them to participate in the retrieval survey, and indicating their assigned travel date;[9]
- A five-dollar cash incentive to encourage participation;
- A personalized travel log for each household member at least five years of age or older. The front side of each log provided instructions on how to record their travel and to complete the survey after their assigned travel day. The reverse side included an example of a completed log for guidance; and
- A vehicle odometer mileage form to collect the make, model and year of each household vehicle, and a space for each vehicle's odometer reading and the date it was recorded.

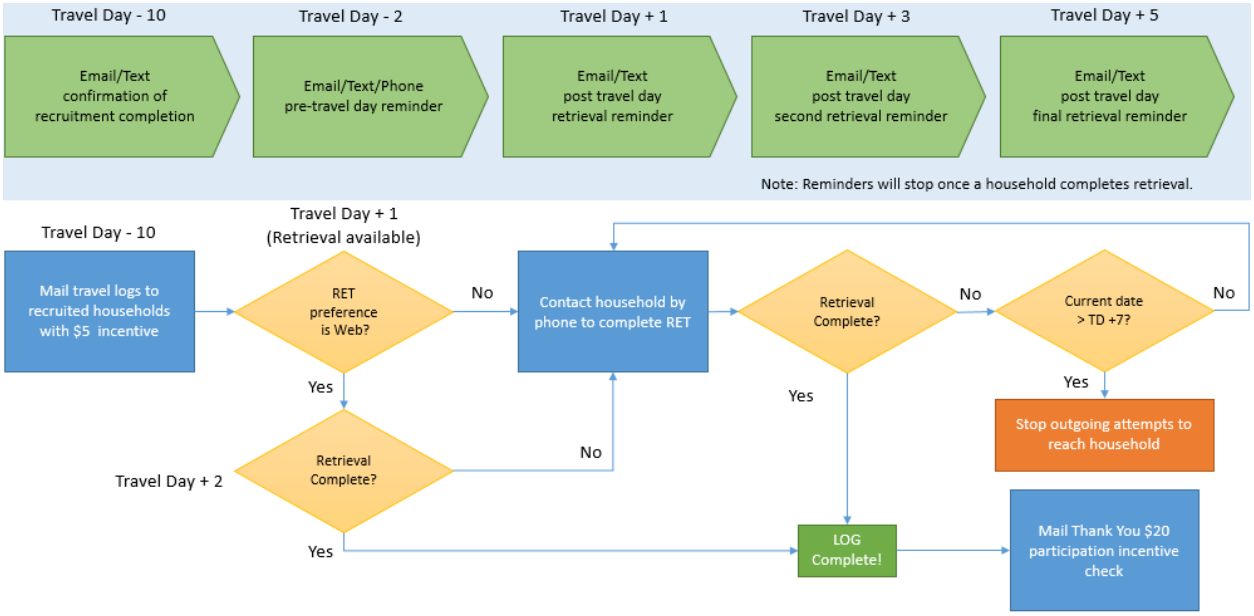
The NHTS Field Documents contain examples of all retrieval materials including the retrieval letter, travel log, and odometer mileage form.

Pre-Travel Day Reminders – Once a recruitment survey was processed, we sent a confirmation email to those that provided an email address in the recruitment survey. The email message thanked respondents for completing the recruitment survey and informed them that they would receive a package inviting the household to participate in the retrieval survey.

[9] For households sampled in an Add-on region, the name of the Add-on sponsor was added to the retrieval letter.

Each household that provided a phone number or e-mail address in the recruitment survey received a reminder the day before their assigned travel day. Reminders were available via email, text message, and Interactive Voice Response (IVR). Reminder messages were sent via every method of contact that was provided in the recruitment survey. The messages were designed to remind household members to record their travel in their logs on the assigned travel date. Households were provided the survey's toll-free number if they had any questions. Figure 2-2 summarizes the recruitment process.

Figure 2-2. Retrieval Survey Flowchart



Retrieval Survey – Because the majority of recruitment surveys were self-completed by paper questionnaire, where no data checks could be enforced, the beginning of the retrieval survey prompted participants to confirm and/or update the household information they had provided earlier. If responses were not provided in the recruitment survey, they were collected in the recruitment module of the retrieval survey. Upon completion of the recruitment module, additional household level information was collected as well as person- and vehicle-level details. Person level data were collected for each household member age five and older, including habitual location data such as school and work. Vehicle characteristics were collected for each household vehicle including the odometer reading. A list of the core data items collected in the retrieval survey was provided in Chapter 1, Section 1.3.

The retrieval survey could be completed online or by telephone. Table 2-1 provides the breakdown of completed retrieval surveys by mode of data collection.

Table 2-1. Retrieval Survey Completed Households by Mode of Data Collection

Mode of Data Collection	Number of Completed Households	Percent
Online	78,285	60.4
Desktop/laptop PC	70,231	77.6
Tablet	8,294	9.2
Smartphone	11,913	13.2
Telephone	39,259	30.3
Combination - Online and Telephone	12,152	9.4
Total Retrieval Household Completes	129,696	100.0

Post-Travel Day Reminders – Each household that provided contact information in the recruitment survey also received reminders after their assigned travel day until they completed the retrieval survey or their seven-day retrieval period elapsed. Reminders were sent via every form of contact that was provided in the recruitment survey, including email, text message, and Interactive Voice Response (IVR). The messages were designed to remind household members to log on to the survey's website or call the helpdesk to report their travel. Households were provided the survey helpdesk's toll-free number if they had any questions.

Thank you Letter and Incentive Mailing –Upon completion of the survey, we thanked participants for their participation and provided a $20 completion incentive. The thank you note also instructed participants to keep their travel logs for a couple of weeks in case any information they provided needed clarification through a research call.

2.3 Retrieval Period

There was a seven-day period, after the assigned travel date, during which interviewers initiated calls, by the CATI system, to collect household travel data. However, the web retrieval survey remained open to participants after the seven-day period. Although we did not initiate outbound calls after the seven-day period, participants could log in to the web survey or call in to the toll-free number to report their travel data. The seven-day retrieval period has been used in several surveys in the NHTS series because it is thought that daily travel recall gets too weak after 7 days. To improve recall, we closely monitored and attempted to through the use of constant reminders to reduce the number of days between the Travel Day and the day that travel was reported. We also asked respondents to refer to their completed travel log.

During the recruitment survey, in addition to providing contact information, participants indicated if they preferred to complete the retrieval survey online or by telephone. If a participant provided a telephone number and indicated a preference to complete the second part of the survey on the phone, an interviewer began calling the household the day after the assigned travel day. Retrieval calls continued for the next six additional days until the household completed the retrieval survey or refused to participate in the survey.

If a participant provided a phone number but preferred to complete the second part of the survey online, a two-day window was set prior to a CATI retrieval attempt. If the household had not completed the survey within two days, we initiated retrieval calls on day three after the travel day and continued for the next four additional days. A household that did not provide a phone number still received reminders if an email address was provided.

2.4 Data Collection Period

The 2017 NHTS survey was conducted from March 31, 2016 through May 8, 2017. The first assigned Travel Day was April 19, 2016 and the last assigned travel date was April 25, 2017.

The survey was designed to collect travel over at least a 12-month period so that seasonal variations in travel are represented. Unlike the 1995, 2001, and 2009 versions, data collection on the 2017 NHTS took 14 months to complete. The additional months of data collection in 2017 is a consequence of using a mail-back screener questionnaire as the main mode of recruitment. In Westat's recent experience, they have found that the best combination of collection modes to be mail and web. Mail-back questionnaires have been used to elicit higher response rates than other recruitment contact modes [10]. However, a drawback of the mail-back questionnaire is that the data collection ramp up and ramp down period is longer due to the amount of time that the materials are in the mail.

2.5 Proxy Interview Procedures

A proxy interview is one in which someone else in the household reports for the subject. During the NHTS data collection, adult household members were instructed to serve as the proxy for any child

[10] Han, D., Montaquila, J., and Brick, J.M. (2013). An evaluation of incentive experiments in a two-phase address-based sample mail survey. *Survey Research Methods*, 7, 207-218.

under the age of 16. The 2017 NHTS did not collect travel day details for any child under the age of five.

Regulation of proxy rules has become more complicated with the introduction of web surveys. Sixty percent of households completed their retrieval surveys entirely online. Thirty percent entirely by CATI and another ten percent used both the web and CATI to participate. As such, traditional proxy rules only applied to the 30 percent of CATI-only households. For the majority of retrieval completes this was not possible. Self-reporting was encouraged, and an online mechanism was implemented to capture the occurrence of proxy reporting. However, because there is no direct interaction between the participant and a data collector with the web survey, it is difficult to know with one hundred percent certainty which household member actually completed the web survey. In the 2017 NHTS, the proxy reporting protocol was the following:

- Before collecting person level details about each household member, participants were presented with the proxy guidelines which were:
  - Participants should report for themselves if possible,
  - For household members younger than 16 a proxy is required, and
  - For household members unavailable or unable to respond for themselves, proxy reporting is allowed.
- Participants were required to select, from a dropdown list of all household members age 16 or older, who was reporting; and
- Participants were required to select, from a dropdown list of all household members age 5 or older, who they were reporting on.

Retrieval surveys completed over the telephone followed the same proxy protocols as online, but a data collector was able to directly guide and persuade household members to self-report and accurately record when proxy reporting had occurred.

If a respondent used a travel log on the travel day, the proxy reporting household member was asked to retrieve the log and refer to it when reporting travel details for the other household member. Note that the conditions under which each interview was completed are included in the data files. There are variables indicating:

- Whether the data were reported by the subject or a proxy respondent (PROXY variable in the Person and Travel Day Files);

- If a travel log was completed by the subject or another household member (DIARYCMP variable in the Person File); and
- If a travel log was used to report a household member's travel day details (DIARYHAV variable in the Person File).

2.6 Computer Assisted Telephone Interviewing

Although the majority of respondents used mail-back for the recruitment survey and web for the retrieval survey, some respondents required interviewer facilitated data collection. Computer Assisted Telephone Interviewing (CATI) was conducted at Westat's Telephone Research Centers (TRCs) and their team of at-home interviewers located across the nation.

Multiple call attempts were made across various times of the day and days of the week to reach households to complete the retrieval survey. Call attempts began the first day after the assigned travel day if the respondent provided a telephone number in the recruitment survey and reported they preferred to complete the retrieval over the phone. If the respondent provided a telephone number but reported they preferred to complete the retrieval online, the first call attempt occurred two days after the assigned travel day. Retrieval calls continued throughout the next seven-days. A computer algorithm scheduled these calls over different days and times including weekends.

Refusal conversion attempts were made by a carefully selected a team of refusal conversion specialists. The team was comprised of 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.

2.7 Reducing Respondent Burden

Extensive efforts were made to transition the NHTS from a telephone-administered survey to a web survey. All aspects of the survey went through multiple rounds of review and testing to ensure appropriateness for a web survey and to identify ways to minimize respondent burden as much as possible. This included:

- Review of the recruitment and retrieval scripts before programming began. The focus was on ensuring all scripts were appropriate for a self-administered web survey. Special

care was taken to ensure that the structure or meaning of any core NHTS data variable was not altered.

- Review of the respondent materials by an expert panel of leading survey methodologists assembled by FHWA.
- Vetting of the trip-reporting module of the survey instrument. Information gathered from cognitive and usability testing by observing and debriefing participants who navigated the trip-reporting feature within the retrieval survey was used to help refine the tool to minimize respondent burden.

During the retrieval interview, special NHTS trip rostering procedures were applied to minimize respondent burden for household members who traveled together during the travel day. This occurred at two main points during the interview - during rostering of the places visited on the travel day ("place rostering") and during the collection of place detail on each place that was rostered.

During place rostering, each household member was asked to list all the places they visited on their assigned travel day prior to asking about the details of each individual trip. This question sequence was structured in the same way as people typically think and talk about the places where they travel. It was designed to collect the location information continuously for each person. This allowed participants to familiarize themselves with the process and streamlined the collection of locations.

If prior to a participant reporting their travel details, another household member reported that they traveled together, the place(s) was automatically populated on the participant's places roster. The participant merely had to confirm that they went on the trip, confirm the trip destination and start time, and add the departure time. If the participant confirmed that the information was accurate, the trip was retained on the roster; otherwise, the participant was requested to edit the place information. The participant confirmed certain trip details, such as party size, household members on trip, and household vehicle used, that were copied from the previous household member. The participant was required to report other trip details, such as trip purpose, as they can differ between persons on the same trip.

2.8 Spanish Surveys

Data collection for the 2017 NHTS was conducted in either English or Spanish. Each survey material received by participants included a tagline in Spanish with instructions on how to complete the survey in Spanish. The survey website and program were translated in Spanish; participants were

able to toggle between English and Spanish with the click of a button. Spanish versions of all the survey materials were also available on the Spanish version of the survey website.

In additional to the online instrument, the entire CATI questionnaire, including interviewer instructions, was translated into Spanish. Interviewing in Spanish was an important factor in gaining the cooperation of Spanish-speaking only respondents. Spanish interviews were conducted by a select set of bilingual interviewers. 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. Spanish-speaking supervisors monitored the bilingual interviewers.

All cases assigned an initial result code of "language problem" by an English-only-speaking interviewer were available only to bilingual interviewers. If the bilingual interviewer determined the respondent did not speak neither English nor Spanish, interviewers attempted to conduct the interview through the help of 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.

2.9 Data Editing

Online Edits – Most of the data corrections on the NHTS were made in real-time with the respondent through the online web and CATI instruments. Errors, both real and potential, were identified through the instrument programs such that the survey automatically proceeded to the next applicable question. If an error was identified, the participant or interviewer was prompted to either correct or confirm their response when an entry to a particular question was not a likely response. For example, if a participant reported that they were less than 14 years of age and a driver they were navigated to a follow-up question that prompted them to correct their age or driver status. The documentation for these checks can be found in the Recruitment and Retrieval Questionnaires of the NHTS Field Documents. These checks fall into three main types:

- Skip checks that moved the participant or interviewer to the appropriate next question based on a previous question response. For example, a subject that is 6 years old was not asked employment questions;
- Range checks that prompted the participant or interviewer to correct or confirm when a response entered was possibly incorrect. Responses that exceeded the allowed range were not accepted; and

- Logic checks that prompted the participant or interviewer when a value entered was within the valid range but did not pass the logic check. For example, the participant entered or 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 reported as a driver during the household recruitment interview or was not reported as being on the trip. In this scenario, a logic check was triggered, which sometimes required modifying previously provided information.

Research Follow-up Call – Data analysts processed all completed surveys through automated checks. If a case failed any automated checks, data analysts manually reviewed that specific case. If the data analyst was unable to resolve the issue, the analyst forwarded the case to a data manager for further consideration for a follow-up research call. The most common error typically related to travel day reporting, particularly shared travel. The most common issues sent to research were cases of missing information, unreasonable discrepancy, and unlikely or illogical travel behavior.

Data Collection Editing – Throughout data collection, automated programs were used to check for data consistencies. During the review of edit checks and cleaning of responses for travel day trips, slightly modifying a response was sometimes necessary to resolve incorrectly completed items.

The majority of edits performed applied to cases when reported trips did not meet the definition of a trip (e.g. a loop walk trip for exercise or for walking the dog, walk trips to access or egress from public transit), in these cases the trip was removed. For other cases, we re-contacted households to resolve issues. The most recurrent issues are categorized in the five categories below.

1. **Location.** These edit checks were mainly focused on location precision and were implemented in situations where the true location could not be verified. For instance, if a participant reports a place's location as being in a city neighborhood, and no specific address was specified.

2. **Joint Travel.** These involved discrepancies between what one household member and at least one other household member reported. This occurred when one or more of the following travel details did not match for a shared travel trip: party size, household party members reported, arrival or departure times, vehicle ID, travel mode, location address, or geocode.

3. **Time/Speed.** These applied to a trip with a high or low speed check that could not be resolved. These include trips with suspiciously high or low travel times or travel speeds relative to the trip length and travel mode.

4. **Misreported Loop Trip.** This occurred when times, travel modes, trip purpose, and geocodes are not logical in the context of the trip(s) and do not fall under the category for removal. For instance, if a participant reports very long loop trips (over six hours),

or reports shopping on a loop trip. A loop trip is a trip that started and ended at the exact same location. More information on loop trips is provided in Section 3.1.

5. **First/Last Place Misreported.** Occurred when the first and/or last place reported on the travel day did not seem like a valid location for the participant. For example, traveled to work at a bank and arrived at 8:30 am with no subsequent trips for the remainder of the assigned travel day. It is likely that the participant visited other places afterwards, but the participant did not provide that information.

Post-Survey Editing – Throughout the data collection process, significant efforts were made to understand what data points were missing or may be unreasonable or fictitious. Processes were developed for issues as they were discovered and categorized. With few exceptions, issues in the data were addressed by identifying them and flagging them at any level of the survey. Imputation or ad-hoc updates to data points were not performed on participant data but were programmed in derived or imputed variables as needed or specified by FHWA.

Editing the Delivery Datasets – As a final editing step, Westat data analysts compared the frequencies between the survey dataset and the delivery data file. Next, the analysts conducted a review and made any necessary corrections on the four delivery datasets to ensure consistency in the reporting of values across the four delivery files.

Useable Households – The four data files in the dataset contain information on only households that are "useable." A useable household in the 2017 NHTS is one in which the household recruitment survey was completed, and the retrieval survey was completed for all household members age five and older. This is different from previous studies where a useable household was one where at least fifty percent of the household members age eighteen and older completed the retrieval survey. Only data for useable households are included in the four files.

For recruitments completed by paper questionnaire, a recruitment was considered complete if a household returned the questionnaire and provided at least household size. Because all household information collected in the recruitment survey was confirmed during the retrieval survey, it was not critical to capture all the household information during the recruitment survey. The recruitment survey was an attempt to get a "foot in the door" to engage them in the survey. Further leveraging that strategy, travel log packages were mailed to all households that returned a questionnaire inviting them to participate in the retrieval survey. For recruitments completed online or over the phone, a recruitment was considered a complete if the participant responded to every question in the recruitment survey.

The retrieval survey was considered complete if all household members age five and older responded to every applicable question in the retrieval survey. This includes responding to all household level, person level, vehicle level, and trip level questions. That is, the participant reached the last question in the questionnaire and was thanked for participating in the survey.

A total of 252,304 households were recruited for the 2017 NHTS. The final dataset contains information on the 129,696 useable households.

*Derived Variables –*The data files created had several "derived variables" that were created by either renaming questionnaire variables or combining multiple variables. Survey weights and other variables not collected during the interview are appended to the final data files. The variable names and logic for creating the derived variables is available in the NTS Derived Variable Specification document.

Types of Data Files – There are four data files that are part of the NHTS dataset. The four files are the:

- **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 household member);
- **Vehicle File –** data items related to the household 's vehicles (one record for each household vehicle); and
- **Travel Day Trip File –** data items collected for each trip made by each household member age 5+ on the household's travel day (one record for each trip each person age 5+ made).

The NHTS Data Series: Changes Over Time 3

The core data in the NHTS series is the information collected about all trips made by household members over a 24-hour period. Over the NHTS survey series—from the first in 1969 to the most recent in 2017--the core data questions have not changed. That makes the survey very valuable for tracking changes in travel over time. However, the methods of obtaining the core data *have* changed, and data users interested in trends analysis need to navigate those changes intelligently. This section provides a brief overview of changes in NHTS methodology, focusing on the new protocols and methods used in the 2017 NHTS.

The earliest surveys in the series, those conducted in 1969, 1977, and 1983, were administered as face-to-face surveys using Census Bureau staff. These earlier surveys were also all conducted as retrospective (e.g. a recall of the household's travel 'yesterday'). To improve coverage and keep costs within reason, the 1990 NPTS was conducted by telephone, using a Random-Digit Dialing sample frame and Computer-Aided Telephone Interviewing (RDD-CATI).

The 1995 NPTS was the first to use a two-phase methodology to collect travel information. The phases were a telephone recruitment followed by the mail-out of a travel log for each member of the participating household to record their travel on the assigned travel day. This change was consistent with simultaneous changes being made in regional travel surveys, as research showed that the use of travel logs prompted the respondents to record more of the incidental stops and short trips that were easily forgotten in the recall method. An adjustment was made to the 1990 data weights (WTTRDADJ) to partially account for the under-reporting of these incidental trips in the 1990 survey, and to make the data more comparable across the 1990 to 1995 survey years.

The RDD sample frame with the telephone recruitment and travel log mail-out/CATI collection worked well for 1995 and 2001. By 2009, however, there was real concern about the representativeness of the RDD sample, which only included land-line telephone numbers. A growing number of U.S. households were dropping their landlines and becoming cell phone only, and these households were excluded from the RDD sample frame. To address the coverage issue with respect to cell-phone only households and households with landline phones who never use them, and obtain a more representative sample, many Federal and regional surveys began using an address-based sample frame that included all households with a postal address.

The 2017 NHTS used an address-based sample frame and a two-stage collection. Unlike the previous RDD surveys, which included only households with landline phones, the 2017 NHTS ABS sample included all households: landline-only, cell-only, both landline and cell, and no phone. The first phase (recruitment) was a mail-out/mail-back short questionnaire obtaining information about the household. The second stage was the collection of travel day information—in 2017, the travel day data was collected primarily via a web-based questionnaire. Respondents not comfortable with—or able to—respond via the web were offered a phone interview retrieval (CATI). Both the respondent self-reporting on the web and the telephone interviewer used the same web-based instrument to log the travel day information.

Additional discussion on the change in methods over the survey series and calculation of estimates of travel over time are included in the Summary of Travel Trends available on the NHTS website. The data user is reminded to exercise caution when performing trend analysis to ensure, to the best of their ability, that the changes tracked over time represent 'real' changes in travel and not an unintended effect.

3.1 Overview of Survey Methodology

Overall, the significant changes to the NHTS methodology were deemed necessary to ensure a more representative sample frame and to increase survey data quality. The details of these methods are discussed in the appropriate sections of this User's Guide, while this section presents a very brief overview. Further details can be found in the report from the Committee on New Directions for the NHTS[11]. In addition, users interested in developing trends over time can obtain further details of the effects of the methodological changes on data estimates from the 2017 NHTS in the Summary of Travel Trends (see NHTS website).

A brief overview of the methods used in the 2017 NHTS include:

- **Mail-out/mail back recruit:** Each sampled household was mailed a brief recruitment questionnaire that focused on engaging the household respondent with relevant and timely questions, gathering basic demographic information, and rostering household members, providing necessary information for assignment of travel dates, reminders, and log package materials. The household mailed back the recruitment survey to participate in the survey. When a household's completed recruitment questionnaire was received, the information was coded and processed into the sample database to assign a

[11] "Exploring New Directions for the National Household Travel Survey, Phase Two Report of Activities," TRB Subcommittee ABJ45T, October 15, 2017 http://www.trb.org/Main/Blurbs/175475.aspx

travel date, send the household a travel log package, and implement travel day data retrieval.

- **Reminders:** Postcard reminders were mailed out at two points during recruitment process. Once the household agreed to participate, multiple phone and e-mail reminders were used as needed until each household member had completed the travel day reporting.

- **Incentive structure:** The incentive plan incrementally rewarded participation. In the 2017 NHTS design, each sampled address received a $2 cash incentive in the initial recruitment mailing. The travel log package sent to each recruited household contained a $5 cash incentive. Finally, when the entire household completed the retrieval survey the household received an additional $20.

- **Travel Day data retrieval:** Primarily web-based retrieval with self-reported travel. When respondents chose to report their travel by telephone interview, CATI interviewers used the same web-based data entry system used by online respondents so that all data were subject to the same range checks and consistency checks. These automatic edit checks reduced reporting error, improved data quality, and minimized respondent burden by maintaining the flow of data reporting.

- **Trip reporting:** Westat's web-based retrieval system was customized for the 2017 NHTS to enable respondents to self-report their travel activity and answer additional demographic and attitudinal questions in a manner that reduced respondent burden while maximizing data quality. The web survey prompted them with questions and branched and skipped depending on their answer. For example, if a respondent said they were not a worker, the retrieval system skipped over the questions about usual travel to work. To assist the respondent in rostering household vehicles, the software provided drop down tables to allow respondents to select the correct vehicle make and model information.

- **Travel day log:** Provided for each eligible household member (aged five and older). The log/memory jogger was graphically designed to resemble the interface web respondents would use to report their travel details in the online survey. The log included identification of activity at each place rather than trip purpose.

Table 3-1 details the changes over time in the NPTS/NHTS series.

Table 3-1. Changes Over Time in the NPTS/NHTS Data Series

Survey year	Sample size	Sample selection	Interview method(s)	Number of contacts	Contractor	Travel day data (memory/ diary)	Travel period definition	Unique attributes	Response rate	Add-Ons
1969	15,000	235 PSU's (primary sampling units) – outgoing panels of Census' Quarterly Housing Survey	in-home interviews ; some telephone followup	one	Bureau of Census	from memory	all overnight trips ending during the 7 days before the travel day	- shortest NPTS questionnaire - asked about shopping trips to the main business district and typical school trips	Not available	none
1977	18,000	376 PSU's – outgoing panels of Census' Current Population Survey	in-home interviews ; some telephone followup	one	Bureau of Census	from memory	all trips of 75 miles or more with a return home during the 14 days preceding travel day	- NPTS and the NTS (National Travel Survey) shared a common sample - mapping of motor vehicle trips and allocation of miles to urban and rural travel - lengthy and detailed questionnaire, including stops on long-distance trips	85.3% (21,033 eligible household units, 17,949 responding)	none
1983	6,500	376 PSU's – outgoing panels of Census' Current Population Survey	in-home interviews ; some telephone followup	one	Bureau of Census	from memory	all trips of 75 miles or more with a return home during the 14 days preceding and	- more vehicle characteristics collected - designed to collect travel period and travel day trips occurring on the same day - relatively small national sample created comparability issues in the series	94.0% (6,900 eligible household units 6,438 responding)	none

Survey year	Sample size	Sample selection	Interview method(s)	Number of contacts	Contractor	Travel day data (memory/ diary)	Travel period definition	Unique attributes	Response rate	Add-Ons
							including travel day			
1990	22,317 (18,000 national and 4,300 add-ons)	- list-assisted sample of telephone numbers formed a RDD sample - sample was stratified by 34 geographic units	telephone interviews	one	Research Triangle Institute, Research Triangle Park, North Carolina	from memory	all trips of 75 miles or more with a return home during the 14 days preceding and including travel day	- first time that the survey conducted exclusively by phone - first unclustered sample design - data collected on most recent traffic accident, if within the past 5 years - first time for add-ons	73.1% (30,529 eligible and presumed eligible household units, 22,317 completing)	New York State Connecticut Indianapolis MPO
1995	42,031 (21,000 national and 21,031 add-ons)	- list-assisted sample of telephone numbers formed a Random Digit Dialing (RDD) sample - sample was stratified by 70 geographic units	telephone interviews	two – one interview at the household level, one for each person in the household	Research Triangle Institute, Research Triangle Park, North Carolina	travel diary used	all trips of 75 miles or more with a return home during the 14 days preceding and including the travel day	- first time for a 2-stage survey using a travel diary - trip rates increased considerably because of the travel diary - completed household definition changed to 50% or more of household adults - household rostering of trips (ability to confirm trips previously reported by another household member) - first time cash incentives were used - first time odometer readings were collected	37.2% (112,960 eligible and presumed eligible residential telephone numbers, 42,033 household completing)	New York State Commonwealth of Massachusetts Oklahoma City, OK Tulsa, OK

Survey year	Sample size	Sample selection	Interview method(s)	Number of contacts	Contractor	Travel day data (memory/ diary)	Travel period definition	Unique attributes	Response rate	Add-Ons
2001	26,038 national and approx. 40,000 add-ons	list-assisted sample of telephone numbers formed a Random Digit Dialing (RDD) sample	telephone interviews	two – one interview at the household level, one for each person in the household	Westat Rockville, MD (National sample and 2 add-ons); Morpace, Farmington Hills, MI (7 add-ons)	travel diary used	all trips of 50 miles or more with a return home during the 28 days preceding and including travel day	- first time the long trip survey (American Travel Survey) was combined with the daily trip survey (NPTS) - threshold for long trips was lowered to 50 miles to collect more trips in the previously ignored 50-75 mile range - if no long trips taken during travel period, most recent trip of 50 miles or more was collected - walking trips increased considerably because of multiple prompts in the questionnaire - cash incentives used in both the pre-interview letter and with the travel diary mailing - multiple data collection methods for odometer readings	41.0% (63,472 eligible and presumed eligible residential telephone numbers, 26,038 households completing)	Baltimore, MPO Des Moines MPO Hawaii Kentucky (4 counties) Lancaster PA MPO New York State Oahu (Honolulu MPO) Texas Wisconsin

Survey year	Sample size	Sample selection	Interview method(s)	Number of contacts	Contractor	Travel day data (memory/ diary)	Travel period definition	Unique attributes	Response rate	Add-Ons
2009	150,147 (25,510 national and 124,637 add-ons)	list-assisted sample of telephone numbers formed a Random Digit Dialing (RDD) sample	telephone interviews	two – one interview at the household level, one for each person in the household	Westat, Rockville, MD	travel diary used	no data collection to specifically record long-distance trips	- a total of 20 add-on programs joined the NHTS this year - geocoding of addresses was conducted online during the CATI retrieval interview - an experimental cell-phone only HH sample of 1,254 (discussed separate from the main survey)	19.8% (847,431 eligible and presumed residential telephone numbers (includes numbers with unknown eligibility but counted as residential), 150,147 households completing)	California Cedar Rapids, IA Florida Georgia Iowa Indiana North Carolina New York Omaha, NE Phoenix, AZ Piedmont, NC South Carolina South Dakota Tennessee Texas Tucson, AZ Virginia Chittenden County, VT Vermont Wisconsin

Survey year	Sample size	Sample selection	Interview method(s)	Number of contacts	Contractor	Travel day data (memory/ diary)	Travel period definition	Unique attributes	Response rate	Add-Ons
2017	129,696 (26,099 national and 103,597 add-ons)	Random sample of residential addresses selected from the US Postal Service formed an Address Based Sample (ABS)	Mail-back (recruit survey), Web and telephone for both recruit and retrieval surveys	Two – one survey at the household level, one for each person in the household	Westat, Rockville, MD	Travel log used	no data collection to specifically record long-distance trips	- a total of 13 add-on programs joined the NHTS this year - online geocoding done in real time using Google API by participants or interviewer - trip distance calculated by shortest distance along network path - trips that started and ended at the same location were collected as loop trips (Home-to-Home)	15.6%	Arizona California Dallas-Fort Worth, TX Des Moines, IA Georgia Maryland North Carolina New York South Carolina Tulsa, OK Texas Waterloo, IA Wisconsin

3.2 2017 Survey Changes

As a national source of travel behavior data over time, the NHTS is widely used for trend analysis. For that reason, it is critical for data users to understand any changes made from one survey to another. While most of the core of the NHTS remained unchanged from earlier surveys, the 2017 NHTS represents slight changes in survey methods and procedures, including:

- **Weekday –weekend distribution:** Add-on partners had the option to collect a 1/7 sample over weekend travel instead of 2/7 weekend sample. The weights adjust the travel days back to 1/7th each.
- **Loop trips:** Subjects were asked to report trips that started and ended at the same location as loop trips. For instance, if they walked the dog around the block, they would report this trip as HOME-to-HOME. The survey would then prompt the respondent to report the total trip distance in blocks or miles. This is a departure from previous surveys, where respondents were asked to report the furthest distance out and then report a return trip. The collection of loop trips may reduce the number of trips from previous iterations of the NHTS, but should not affect miles of travel.
- **Fuel Type:** Subjects were asked to report the type of fuel used by each household vehicle.
- **Hybrid:** Subjects were asked type of hybrid vehicle for vehicles that were reported to use hybrid, electronic, or alternative fuel.
- **Rideshare:** Subjects were asked how many times in the past 30 days they purchased a ride with a ridesharing app.
- **Carshare:** Subjects were asked how many times in the past 30 days they used a car sharing service.
- **Walk/Bike:** Subjects were asked additional questions about how many of their walk and bike trips were strictly for exercise and what were the reasons they do not walk or bike more often.
- **Health:** Subjects were asked to describe their general health.
- **Physical activity:** Subjects were asked to describe how physically active they are in a typical week. If they reported their activity level includes some light or vigorous physical activities, they were asked how often they performed those activities in the past week.
- **Work from home:** Subjects were asked if they usually work from home.
- **Work commute without traffic:** Subjects were asked how many minutes their work commute would take without traffic.

- **Typical travel behavior:** The main respondent was asked how often they used a variety of modes to get from place to place.
- **Attitudinal questions:** The main respondent was asked about their general attitudes in regards to their travel behavior.
- **Technology use:** The main respondent was asked how often they use a computer, smartphone, and tablet to access the internet.

For data users most familiar with the 2009 NHTS the following list provides data content that was included in 2009 and deleted in 2017:

- How many telephones in the household.
- Subjects were not asked about their travel behaviors to school (i.e., the safe routes to school question series). Mode of transportation used to get to and from school were still collected.
- Subjects were not asked if they would like to get out more often.
- Subjects were not asked if they used an interstate or turnpike on their travel day.
- Subjects were not asked if they paid any tolls on their travel day.

3.3 Trip Distance Reporting

Because of the visual and geographic aspect of trip data, the web-based tool mapped the origin and destination of each reported trip while collecting other core data such as trip purpose, mode of transportation, time of day, and other people on the same trip. This was possible because the American public is increasingly familiar with using digital maps (such as Google maps) for daily navigation.

A shortest network path distance was calculated and included for each reported trip. Google Maps was used for routing the shortest path for motorized travel on the road network. Non-motorized modes, like walk and bike, had the shortest path calculated using network routes paths. Shortest path distances were generated by the Google Maps API. These distance estimates are provided in the trip file for all reported trips (TRPMILES).

Using Google Maps API automated shortest path distance is a major break from the self-reported distances from previous surveys, and impacts the estimates of person miles traveled (PMT), vehicle

miles traveled (VMT), and mean trip lengths, such as commute distances. The analysis of the difference between self-reported trip distance and network-coded distance is detailed in the Summary of Travel Trends (see NHTS website). **Any user analysis which includes comparison of distance estimates between the survey years should be cognizant of this change in methods and how it impacts those comparisons.**

In addition, because the web-based travel day retrieval was primarily self-reported data, there are some very large values for some of the variables describing distances. For instance, if a respondent reported a legitimate work address, that location was accepted as valid. These values can influence the mean trip distances—or averages—that result from analysis.

An example is shown in 3-2 (unweighted distribution). The example variable—GCDWORK—represents the 'as the crow flies' measure of the distance between the respondent's home and workplace. In the 2017 NHTS the maximum value is much higher than the previous two surveys. In reviewing these reported locations, it was determined that they were legitimate locations that were simply far away from the respondent's home, perhaps the respondent's company headquarters or the like rather than a work location nearer the respondent's home. To develop robust and comparable averages, the user is cautioned to check the distribution of the variable values and determine whether there are any values that could be considered outliers for that analysis.

Table 3-2. Differences in Maximum Values Affect the Mean

GCDWORK Quantiles (Definition 5)			
Level	2001	2009	2017
100% Max	1,796.00	3,898.78	9,744.49
99%	54.00	110.95	177.73
95%	27.94	32.44	34.97
90%	20.00	22.71	23.98
75% Q3	11.12	13.00	13.49
50% Median	5.63	6.63	6.63
25% Q1	2.40	2.95	2.82
10%	1.00	1.25	1.12
5%	0.82	0.70	0.59
1%	0.24	0.20	0.04
0% Min	0.00	0.00	0.01
Mean Values	9.67	15.43	18.56

As with the changes in methods and protocols, data users must use caution to understand and assess how the changes in trip distance reporting may affect the specific estimates of interest obtained from the 2017 Survey in comparison with previous estimates.

3.4 Trip Purpose Coding

Because the survey was primarily conducted as a self-reported travel day, FHWA reduced the number of trip purposes displayed so as not to overwhelm the respondent. A comparison of the trip purpose codes for 2017 and 2009 NHTS is in Table 3-3. As shown, the bolded and italicized categories in 2017 represent purpose codes that do not have a 2009 equivalent. On the other hand, one purpose code in 2009 may logically be included in two different categories in 2017: "pet care, including walk the dog." Users are advised to carefully consider how to combine these purposes for trends analysis.

Table 3-3. Purpose Codes, 2017 and 2009 NHTS

2017 Code	2017 Category Name	2009 Code	2009 Category Name
-7	Refused	-7	Refused
-8	Don't know	-8	Don't know
-9	Not ascertained	-9	Not ascertained
1	Reg. Home Activities	1	Home
2	***Work from Home (Paid)***		
3	Work	11	Go to work
		12	Return to work
4	Work related/Trip	13	Attend business meeting/trip
		14	Other work related
5	Volunteer activities (Not Paid)	65	Attend meeting: PTA/home owner's association/local government
6	Drop-off/pickup someone	71	Pick up someone
		72	Take and wait
		73	Drop someone off
7	***Change type of Transportation***		
8	Attend school as a student	21	Go to school as student
9	Attend child care	24	OS - Day care
10	***Attend adult care***		
11	Buy Goods (groceries, clothes, appliances, gas)	40	Shopping/errands
		41	Buy goods: groceries/clothing/hardware store
		43	Buy gas
12	Buy services (dry cleaners, banking, service a car, pet care)	42	Buy services: video rentals/dry cleaner/post office/car service/bank
13	Buy meals (Go out for a meal, snack, carry-out)	80	Meals
		82	Get/eat meal
		83	Coffee/ice cream/snacks
14	Other general errands (post office, library)	60	Family personal business/obligations
		61	Use professional services: attorney/accountant
		63	Use personal services: grooming/haircut/nails
		64	Pet care: walk the dog/vet visits
15	Recreational Activities (visit parks, movies, bars, museums)	50	Social/recreational
		52	Rest or relaxation/vacation
		54	Go out/hang out: entertainment/theater/sports event/go to bar
		55	Visit public place: historical site/museum/park/library
		81	Social event
16	Exercise (go for a jog, walk, walk the dog, go to the gym)	64	Pet care: walk the dog/vet visits

2017 Code	2017 Category Name	2009 Code	2009 Category Name
		51	Go to gym/exercise/play sports
17	Visit Friends and Relatives	53	Visit friends/relatives
18	Health care visit (medical, dental, therapy)	30	Medical/dental services
19	Religious or other community activities	20	School/Religious Activity
		22	Go to religious activity
97	Something else	97	Other reason

3.5 Means of Transportation Coding

The means of travel categories were also changed in the 2017 NHTS. Table 3-4 shows the categories for 2017 and 2009. The two categories in bold and italics font represent shifts in coding that may be important to data users.

Table 3-4. Means of Transportation Codes, 2017 and 2009 NHTS

Summary Mode of Travel	Mode of Travel 2017 NHTS	Mode of Travel 2009 NHTS
Private Vehicle	'03' Car	'01' Car
	'04' SUV	'02' Van
	'05' Van (Minivan)	'03' SUV
	'06' Pick Up Truck	'04' Pick-up Truck
	'08' Motorcycle/Moped	'07' Motorcycle
Other Vehicle	'18' Rental Car (Inc. Zipcar and Car2Go)	'05' Other Truck
	'09' RV (motorhome, ATV, Snowmobile)	'06' RV
	'07' Golf cart/Segway	'08' Light Electric Vehicle
	'17' Taxi/Limo (including Uber/Lyft)	'19' Taxicab
Walk	'01' Walk	'23' Walk
Bike	'02' Bicycle	'22' Bicycle
School Bus	'10' School bus	'11' School bus
Public Transit	'11' Public or Commuter Bus	'09' Local Public Bus
		'10' Commuter Bus
	'15' Amtrak/Commuter Rail	'15' Amtrak/Inter-city Train
		'16' Commuter Train
	'16' Subway/Elevated/Light Rail/Streetcar	'17' Subway/elevated Train
		'18' Streetcar/Trolley
Other Transit	'13' Private/Charter/Tour/Shuttle Bus	'12' Charter/Tour
		'14' Shuttle Bus
	'14' City-to-City Bus (Greyhound, Megabus)	'13' City-to-City Bus
	'20' Boat/Ferry/Water Taxi	'20' Ferry*

Summary Mode of Travel	Mode of Travel 2017 NHTS	Mode of Travel 2009 NHTS
	'12' Paratransit/Dial-a-Ride	'24' Special Transit-people with disabilities
Other	'19' Airplane	'21' Airplane
	'97' Something Else	'97' Other

4 Survey Response Rates

4.1 Overview

As discussed in the previous chapters, the 2017 NHTS transitioned from a Random-Digit Dialing sample to an Address-Based Sample (ABS). The 2017 NHTS was redesigned to address the general trend of declining response rates in surveys, and to adjust to the dwindling share of households using landline telephones. As the survey climate has changed, response rates have dwindled across the board on all types of surveys. Because the NHTS is conducted by a Federal agency it has to obtain OMB clearance. Response rates are a key decision factor for OMB as lower rates make the survey results less representative of the population being surveyed. Understanding this declining response rate environment and the need to propose a design that would maximize response rates, Westat proposed a survey design where a brief paper questionnaire with a pre-incentive was used for the recruitment survey followed by a longer retrieval survey to collect household details and travel behavior.

The 2017 NHTS data collection occurred between March 2016 and May 2017, with assigned travel dates from April 19, 2016 through April 25, 2017. As described in Chapter 2, Survey Procedures and Methodology, there were two stages of data collection for each sampled address. First, sampled addresses were mailed an invitation package containing a brief recruitment survey about travel habits, opinions on transportation issues, and household demographics. Following the recruitment survey, a travel log package was prepared and mailed to households that completed the recruitment survey. The travel log package invited households to participate in the retrieval survey, introduced their assigned travel date, included a small incentive, and provided individualized travel logs for each household member age five and older to record their travel on the assigned day. The retrieval survey collected travel details as well as additional household, person, and vehicle level information.

Sections 4.2 and 4.3 provide a breakdown the 929,077 sampled cases by their survey outcome and present the 2017 NHTS overall response rate, as well as the rates at both stages of the survey process.

4.2 Classification of All Sampled Addresses

Using the American Association of Public Opinion (AAPOR) guidelines, the first step in the calculation of response rates is to classify all addresses in the sample into residential (in-scope) and non-residential (out-of-scope) categories. We began this process by placing all sampled addresses into three main groups:

- **In-Scope Addresses –** Addresses that completed the recruitment survey are considered residential. A paper recruitment survey was considered complete if it was returned with at least a response to the household size question, as this information was required to send the household the appropriate number of travel logs. A web or telephone recruitment survey was considered complete if the participant responded to all the questions in the recruitment survey.
- **Out-of-Scope Addresses –** Recruitment Survey packages returned as Postal Non-Deliverable (PND) or returned indicating the address was non-residential were marked as out-of-scope.
- **Eligibility Unknown Addresses –** All other addresses fell into the category of eligibility-unknown. These included non-responding households, refusals, and paper questionnaires that were returned blank.

Table 4-1 presents the final distribution of all sampled addresses in the 2017 NHTS sample. A total of 929,077 addresses were mailed an invitation package resulting in 252,304 completed recruitment surveys.

Table 4-1. Classification of Sampled Addresses

Recruitment Survey Result	Number	Percent
In-Scope - Total	252,304	27.16
Completed Recruitment Survey	252,304	100.00
Out-of-Scope - Total	68,698	7.39
Postal Non-Deliverable	68,645	99.92
Non-Residential	53	0.01
Eligibility Unknown - Total	608,075	65.45
Non-Response	604,105	7.96
Refused	386	0.06
Blank	3,584	0.59
Total Sampled Addresses	929,077	100.00

4.3 Recruitment Survey Response Rate

The overall survey response rate is a product of the response rate for the recruitment survey and the response rate for the retrieval survey. In this section, we calculate the recruitment survey response rates.

In order to calculate the recruitment survey response rate, it is necessary to estimate the residency status of the 65.45 percent of addresses in Table 4-1 with an unknown eligibility, using the American Association of Public Opinion Research (AAPOR) standards[12]. The calculation of the eligibility rate for the unknown eligibility addresses is a function of the sum of the base weights for the known eligibility status addresses, the sum of the base weights for unknown eligibility addresses, and the most recent estimate of the total number of occupied housing units from the 2015 American Community Survey (ACS). The calculation results in the estimate of the proportion of addresses with unknown eligibility considered eligible, and therefore treated as nonresponding eligible households for purposes of calculating the recruitment response rate.

Discussed below is the calculation of the weighted response rate. Note that in the response rate computations, the weights used are the base weights (inverse of the probability of selection for each sampled address). The weighted response rate (weighted using the base weight) is the weighted total of recruited households divided by the estimated aggregation of total households.

We use weighted rather than unweighted response rates so that disproportionate sampling across a region does not artificially inflate the response rate.

Example of Weighted Response Rate Calculation: Consider one sample area comprised of two regions, A and B, with each region having 1,000 households. In region A, 100 households are sampled (or 1 in 10) and in region B only 10 households are sampled (or 1 in 100). In region A, say for example that 90 of the 100 sampled households (90%) are respondents, and in region B, five of the 10 sampled households (50%) are respondents. A simple unweighted response or completion rate for these sample areas, combined, would be calculated as 86.4 percent[13]. A total of 90 percent of region A's sample responded and 50 percent of region B's sample responded, but because the number of sampled households was much larger in region A, the unweighted mean is **86.4 percent**.

[12] http://www.aapor.org/Standard_Definitions/1481.htm.

[13] Take the number of completes (95) divided by the number of sampled households (110) = 86.4 percent.

For weighted response rates, the base weight is applied. In this example, the base weight for region A is 10 and base weight for region B is 100[14]. By weighting the households by these base weights, we effectively expand region B to its correct population size, which is 50 percent of all households in the sample area, which includes regions A and B combined. Remember each region has 1,000 households.

The weighted response rate is calculated as **70 percent** as shown below.

(10*90 responding HHs)+(100*5 responding HHs)/
{(10*100 sampled HHs)+(100*10 sampled HHs)}
= (900+500)/(1000+1000)=1400 / 2000
=70%

Each region has 50 percent of the households in the population so each gets 50 percent of the weighted response rate.

The weighted response rate represents an unbiased estimate of the expected response rate when conducting a census of the entire population (no sampling). It is the true percentage of responsive households. The unweighted response rate on the other hand will tend to skew towards the regions that were sampled at higher rates relative to the rest of the region.

Table 4-2 provides the weighted recruitment response rate for the full sample and each individual Add-on study area. The table shows that while the overall weighted recruitment response rate is **30.4 percent**, large rate variations were observed across study areas, with the lowest at 24.4 percent and the highest at 42.4 percent.

[14] The base weights are the reciprocal of the probability of selection (e.g., 1 in 10 sampled has a base weight of 10).

Table 4-2. Weighted Recruitment Response Rate by Study Area

Study Area	Weighted Recruitment Response Rate
Full Sample	30.4%
National	31.3%
Arizona	31.0%
California	27.0%
Dallas-Ft. Worth, Texas	25.4%
Des Moines, Iowa	42.4%
Georgia	28.2%
Maryland	31.1%
North Carolina	31.5%
New York State	29.8%
South Carolina	31.2%
Texas	24.4%
Tulsa, Oklahoma	27.4%
Waterloo, Iowa	37.9%
Wisconsin	40.3%

4.4 Complete Households

The 2017 NHTS defined a complete household as one in which the retrieval survey was completed for all household members age five and older. That is, all household members invited to participate; household members under five years of age were not eligible.

4.5 Retrieval Survey Response Rate

The Retrieval survey response rate is the percentage of recruited households that completed the Retrieval survey. Provided in Table 4-3 are the weighted retrieval rates for the full sample and each individual Add-on study area. We use weighted rather than unweighted rates for the same reasons discussed earlier for weighted recruitment survey response rates. In this case, the weight is the product of the base weight and a recruitment nonresponse adjustment.

The weighted retrieval response rate for the 2017 NHTS was **51.4 percent**. Similar to the recruitment response rates, the retrieval rates varied across study area. The lowest rate observed was 47.9 percent and the highest was 58.9 percent

Table 4-3. Weighted Retrieval Response Rate by Study Area

Study Area	Weighted Retrieval Rate
Full Sample	51.4%
National	52.1%
Arizona	54.0%
California	47.9%
Dallas-Ft. Worth, Texas	51.4%
Des Moines, Iowa	56.9%
Georgia	50.2%
Maryland	51.8%
North Carolina	52.7%
New York State	49.5%
South Carolina	48.5%
Texas	48.7%
Tulsa, Oklahoma	54.5%
Waterloo, Iowa	58.9%
Wisconsin	58.5%

4.6 Overall Survey Response Rates

This section presents the overall weighted response rates for the full NHTS sample and for each individual study area. This response rate is the product of the weighted recruitment response rate as reported in Table 4-2 and the weighted retrieval response rate as reported in Table 4-3.

The overall 2017 NHTS response rate was **15.6 percent**. Response rates varied greatly across regions with the lowest at 11.9 percent and the highest at 24.1 percent. The range of observed responses rates at the retrieval stage was half that of the recruitment stage, making the variability of the overall rates mostly attributed to the recruitment rates.

Table 4-4. Overall Survey Response Rate by Study Area

Study Area	Weighted Recruitment Response Rate	Weighted Retrieval Rate	Overall Response Rate
Full Sample	30.4%	51.4%	15.6%
National	31.3%	52.1%	16.3%
Arizona	31.0%	54.0%	16.8%
California	27.0%	47.9%	12.9%
Dallas-Ft. Worth, Texas	25.4%	51.4%	13.1%
Des Moines, Iowa	42.4%	56.9%	24.1%
Georgia	28.2%	50.2%	14.2%
Maryland	31.1%	51.8%	16.1%
North Carolina	31.5%	52.7%	16.6%
New York State	29.8%	49.5%	14.8%
South Carolina	31.2%	48.5%	15.2%
Tulsa, Oklahoma	27.4%	54.5%	14.9%
Texas	24.4%	48.7%	11.9%
Waterloo, Iowa	37.9%	58.9%	22.3%
Wisconsin	40.3%	58.5%	23.6%

5 Weight Calculations

The public use dataset contains household and person weights for all households in which retrieval surveys were completed for all eligible household members (i.e., ages five and older). The NHTS 2017 database consists of completed surveys from all eligible household members from 129,696 households.

Weights are needed to produce valid population-level estimates. Several stages of nonresponse adjustment and post-stratification are conducted during the computation of weights to reduce sampling error and bias. Different weights are required for households, persons, trips and vehicles. A discussion of which weight to use for different types of estimates is contained in Chapter 7. Replicate weights are also available for the NHTS, allowing the user to compute unbiased variance estimates.

The steps to develop Household, Person, Trip, and Vehicle weights include preparation of initial and final household and person weights, each of which engages steps to compute base weights, to adjust for nonresponse at the household level, to post stratify both households and persons to externally sourced control totals, and to trim excessively large (or small) weights. The entire process is illustrated in Exhibit 5-1.

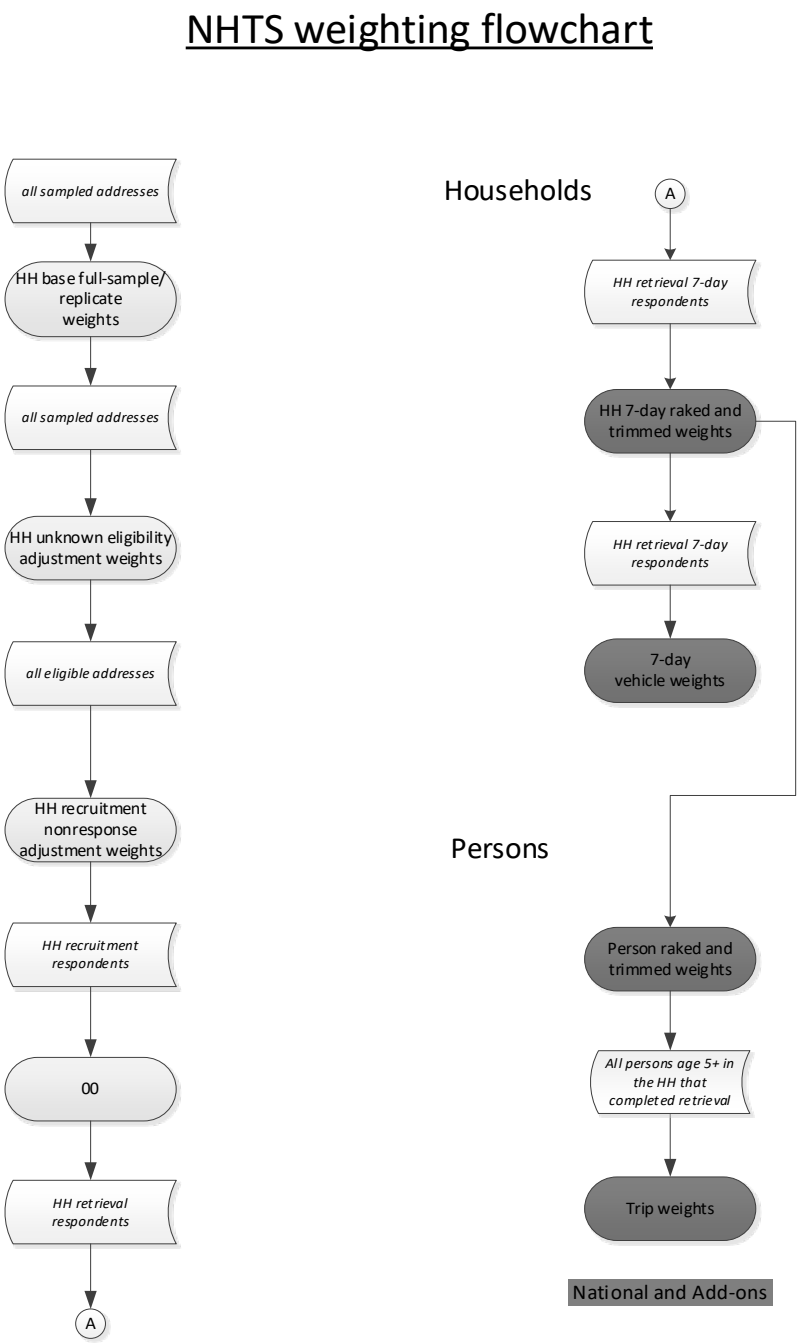
5.1 Initial Household Weight

The base weight is the reciprocal of the known probability of selection of an address from the address sampling frame. This probability of selection differs by sampling stratum due to the unique requirements of the national sample and each Add-on sample.

The first adjustment to the base weight is for recruitment nonresponse, addressing those households that did not complete the household recruitment survey. A special calculation is needed prior to the recruitment nonresponse adjustment because when no response at all is received from a sampled address, it is not possible to determine whether that address is residential (eligible). The eligibility rate for these addresses is calculated as a function of the sum of the base weights for addresses for which eligibility is known, the sum of the base weights for addresses for which eligibility is not known, and the most recent estimate of the total number of occupied housing units from the 2015

American Community Survey (ACS). Exhibit 5-1 provides an overview of the 2017 NHTS weighting process.

Exhibit 5-1. Flowchart of NHTS Weighting Procedures



The calculation results in an estimate of the proportion of addresses with unknown eligibility that should be considered eligible, and therefore treated as nonresponding households in the subsequent household weighting adjustments.

For the recruitment nonresponse adjustment, cells (collections of sampled addresses categorized by various characteristics) were formed with a separate nonresponse adjustment factor computed for each cell. Characteristics of the addresses, such as whether or not there is a telephone number associated with the address, and characteristics at the Census tract level from the ACS were examined to determine where response rates differed[15]. A search algorithm was used to define a set of cells with response rates that were as different from each other as possible. This allowed for the maximum possible adjustment for differential nonresponse (which may introduce bias), while providing larger adjustments to households with a higher propensity not to respond (so that these households are not underrepresented because of nonresponse). The search algorithm generated a "tree" of nonresponse adjustment cells using a sequence of next-best binary splits. Each split in the sequence checks all the cells on the tree at that point, and finds the best existing cell to split in terms of getting two new cells with response rates as different as possible (using the frame and Census characteristics available to define splits). Cells with too-small sample sizes or too-extreme nonresponse adjustments were ignored. The 2017 NHTS Weighting Report provides the full list of variables used to define the cells.

The weight for each cell, consisting of the product of the base weight, the adjustment for household unknown eligibility, and the household recruitment nonresponse adjustment, defines the "initial household weight."

5.2 Final Household Weight

To determine the final household weight, a retrieval nonresponse adjustment was applied to the initial household weight. The respondents in this case were households that completed the retrieval survey, and the nonrespondents were those recruitment respondent households who did not complete the retrieval survey. Information on characteristics collected in the household recruitment survey, as well as Census tract-level information available from the ACS, was available for determining nonresponse adjustment cells at this level. As in the household recruitment nonresponse adjustment, a search algorithm was used to determine the best definition of cells (the

[15] Other examples of characteristics used in recruitment nonresponse adjustment include dwelling type (single or multi-family), Census tract level median income, Census tract level median home value, and the proportion of households above or below the median for categories of race and ethnicity at the Census tract level.

most heterogeneous cells in terms of the criterion of response). The 2017 NHTS Weighting Report lists the variables used in the retrieval nonresponse adjustment.

Next, the retrieval nonresponse weights were adjusted to be consistent with independent controls based on various demographic categories, in a process called raking. Raking reduces the variances for any characteristics that are correlated to these demographic characteristics (as the independent controls have much lower variability). The source for these controls was the 2015 ACS; the controls consisted of attributes such as geography, race, ethnicity and number of household vehicles. The 2017 NHTS Weighting Report provides details on the raking process and the variables and control totals for each domain. First, weights were adjusted to assure agreement on the first raking dimension, or marginal distribution of one variable, such as ethnicity. Next, weights were adjusted for the second raking dimension, then for the third dimension, and so forth. The process was repeated again assuring agreement with each of the raking dimensions. The process continued to be repeated, with iterative controlling to each variable, until simultaneously contiguous agreement for each variable was achieved. In addition to convergence to ACS data, raking was also conducted to achieve equal distribution among each of the seven days of the week and by each month of the year.

The processes described in this section resulted in the final household weight, which was used as the starting weight for the additional weights defined below.

5.3 Person, Trip, and Vehicle Weights

The starting point for person weights is the final household weight. Each person in the household receives the household weight as their initial weight, as there is no subsampling of persons within households.

Person Weights – Person weights were calculated by controlling survey estimates to person-level control totals from the 2015 ACS estimates. The 2017 NHTS Weighting Report provides the control totals and the average adjustment factors for each of the six categories used in this weighting step, which included MSA/heavy rail original sampling strata, race, ethnicity, sex by age category, travel month, and travel day of week.

Trip Weights –Trip weights are simple functions of the person weights, modified only for the purpose of producing annual estimates of the number of trips. The trip weight is simply equal to the final person weight multiplied by 365 to produce annualized data.

A record of each trip (travel data) was collected from each household member, age five and older, and describes individual trips made by each person in the household, thus such data represent a person-level attribute.

Vehicle Weights – Each vehicle in the household receives the final household weight, as every vehicle in the household represents a characteristic of that household and others like it.

Travel behavior varies by season and day-of-the week. There was some variation in the number of completed interviews by month. To adjust for this variability and make the monthly distribution of completes equal, the estimates of total persons and total households were adjusted by month in the last step of weighting for both households and persons.

5.4 Replicate Weights

Provided with the final weights at each level (household, person, trip, vehicle) is a set of 98 replicate weights. The 98 replicate weights were generated using a Jackknife procedure. These replicate weights are provided to support calculation of estimated variance associated with any estimate from the weighted data[16]. These replicate variances will include all components of variance, including the increase in variance from the nonresponse adjustments, and the decrease in variance from the post-stratification to external control totals, as the replicate weights carefully incorporate all of these adjustments. The replicate weights are not on the downloadable 2017 NHTS Public Use dataset; they are available on the NHTS website (http://nhts.ornl.gov) as a separate file.

Software such as SUDAAN, WESVAR, and the most recent version of SAS allow for easy computation of replicate variance estimates.

[16] For each replicate, a mutually exclusive and exhaustive 1/98th part of the sample is deleted, and the remaining 97/98 sample reweighted. Any estimate can be computed using each replicate weight in place of the main weight. The sum of the squared differences between these replicate estimates and the full-sample weight estimate provides an unbiased estimator of variance.

Description of Data Files 6

6.1 Basic Data File Structure

The 2017 NHTS public use data files were developed to emulate the previous surveys in the NPTS/ NHTS series. The four data files, displayed in Table 6-1, HOUSEHOLD, PERSON, VEHICLE and TRIP, are hierarchically structured for intuitive merging using each file's identifier (primary key) variable. The data files are available in CSV, SAS7BDAT and SAV formats.

Table 6-1. 2017 NHTS Data Files

File	Record Level Description	ID Variables
HOUSEHOLD	One record per household unit	HOUSEID
VEHICLE	One record per household vehicle, if vehicle is present.	HOUSEID VEHID
PERSON	One record per household person	HOUSEID PERSONID
TRIP	One record per household person's travel day trip, if at least one trip made.	HOUSEID PERSONID TDTRPNUM

6.2 Data File Relationship

Section 6.2 describes how the four data files relate to each other. We display this pictorially in Figure 6-1.

HOUSEHOLD

- Household characteristic (e.g. household size)

VEHICLE

- Household vehicle characteristics (e.g. vehicle manufacturer)

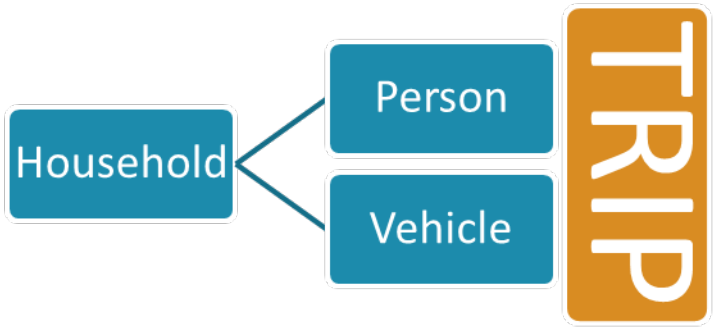
PERSON

- Household person characteristics (e.g. person age)

TRIP

- Household person trip characteristics (e.g. trip travel time)

Figure 6-1. 2017 NHTS Data File Relationship



6.3 When is a Record on the File?

Section 6.3 explains the records on each of the four data files. As mentioned earlier, all household members had to participate in the retrieval survey for the household to be useable.

HOUSEHOLD: A record in this file represents a household unit where all residents aged five and older completed all portions of the survey.

VEHICLE: A record in this file represents a household vehicle. Participants listed any vehicle that a household member owned, leased or had available for regular use, including motorcycles, mopeds and RVs. Households that reported having zero vehicles will not be present in this file.

PERSON: A record in this file represents an individual household member. Households rostered all persons living in the household. The person number assigned to each household member corresponds to the order that each were rostered by the household member who initially responded to the recruitment survey.

TRIP: A record in this file represents a trip that was reported by a household member age five and older on the household's travel date. Participants were asked to report all locations they went to from 4 am to 3:59 am on their assigned travel date, regardless of how long they were there, including trips that may have started and ended at the same location, like walk or bicycle trips for exercise. Households or persons that reported having zero travel day trips will not be present in this file.

6.4 Data Documentation

In addition to this User's Guide, data users should reference the NHTS Data Dictionary and Codebook, which are available on the NHTS website. These documents provide key metadata information on the files, variables, and variable values. In addition, sample sizes and weighted sums for every variable response is available for each file. The 2017 NHTS documentation include several pieces of information for each variable including the complete question and answer text. Table 6-2 provides an example of the metadata available in the Data Dictionary using the variable HOMEOWN.

Table 6-2. Variable Documentation for a Sample Variable, HOMEOWN

Variable Metadata	Description	Example
Name	Variable or column name in file	HOMEOWN
Table Level	The table file source of the variable	HOUSEHOLD
Data Type	Numeric or Text	TEXT
Label	Brief summarization of variable	Home Ownership
Question Text	The instrument question text	Do you own or rent your home?
Ordinal Position	Numeric column position in file	11
Derived	Computed status and logic	{Not Applicable}
Value Label or Range	Answer label text for variable codes	-8 = I don't know -7 = I prefer not to answer 01 = Own 02 = Rent 97 = Some other arrangement
Value Frequencies – Sample Size	Sum of each unique value group in file	-8 = 3 -7 = 32 01 = 98459 02 = 30268 97 = 934
Value Frequencies – Estimate	Weighted sum of each unique value group in file	-8 = 1,728 -7 = 36,993 01 = 74,518,546 02 = 42,463,981 97 = 1,187,002

6.5 Data File Conventions and Special Codes

There are a number of data file conventions and special codes that may be useful to data users. Table 6-3 displays these codes.

Table 6-3. Special Values

Variable Value	Label	Extended Description
-1	[Appropriate Skip]	Selected by system when the question was appropriate skipped and no value exists
-9	[Not Ascertained]	Selected by system when the question was skipped inappropriately and no value exists
-7	I prefer not to answer	Selected by participant (available when no answer given)
-77	I prefer not to answer	Selected by participant (always available)
-8	I don't know	Selected by participant (available when no answer given)
-88	I don't know	Selected by participant (always available)

The data files also use specific formatting conventions. These include:

- **Numeric or Character Data Type** - By default, variables are set to a character/text format unless the variable is truly numeric, for which one can perform meaningful calculations on the values,
- **"0" Prefix Variable Values** - Character format variable values in the one through nine range are prefixed with a leading zero: "01," "02," "03," "04," "05," "06," "07," "08," "09,"
- **"17" Suffix Variable Names** - Variables that are similar in design to previous NHTS variables of the same name but have at least one notable difference in the 2017 version, and
- **Yes/No (True/False) Question Values** - Variable values are "01" (Yes or True) and "02" (No or False).

6.6 Repeated Variables

It is a convention of the NHTS public files to have survey variables attached across multiple table file levels. For example, the trip file, which contains information collected at the daily trip level, will have many household level variables attached to it. This is convenient for data users not interested in or capable of merging data files using computer software in their analysis. Data users should be

aware that the person, vehicle, and trip file contain variables attached from a higher level file in the NHTS file hierarchy. The file structure and hierarchy are discussed earlier in this chapter. Table 6-4 lists the variables that repeat across the four NHTS data files.

Table 6-4. Variables Used In All Four NHTS Data Files

		NHTS Data File			
Variable Name	Variable Description	Household	Person	Vehicle	Trip
CDIVMSAR	Census Division, MSA status, & presence of a rail system	X	X	X	X
CENSUS_D	Census Division	X	X	X	X
CENSUS_R	Census Region	X	X	X	X
DRIVER	Driver status of subject		X		X
DRVRCNT	Number of household drivers	X	X	X	X
EDUC	Highest grade completed		X		X
HBHTNRNT	Percent renter-occupied - Block group	X	X	X	X
HBPPOPDN	Population per sq mile - Block group	X	X	X	X
HBRESDN	Housing units per sq mile - Block group	X	X	X	X
HH_CBSA	CBSA FIPS code for HH address	X	X	X	X
HH_HISP	Hispanic status of HH respondent	X	X	X	X
HH_RACE	Race of HH respondent	X	X	X	X
HHFAMINC	Total household family income	X	X	X	X
HHRESP	Household respondent ID number	X	X		X
HHSIZE	Number of household members	X	X	X	X
HHSTATE	State where the household is located	X	X	X	X
HHSTFIPS	State FIPS for HH address	X	X	X	X
HHVEHCNT	Number of household vehicles	X	X	X	X
HOMEOWN	Housing unit owned or rented	X	X	X	X
HOUSEID	Eight digit household ID number	X	X	X	X
HTEEMPDN	Workers per square mile living in Tract	X	X	X	X
HTHTNRNT	Percent renter-occupied - Tract level	X	X	X	X
HTPPOPDN	Population per sq mile - Tract level	X	X	X	X
HTRESDN	Housing units per sq mile - Tract level	X	X	X	X
LIF_CYC	Household life cycle	X	X	X	X
MSACAT	MSA category of household	X	X	X	X
MSASIZE	Population size of household MSA	X	X	X	X
NUMADLT	Number of adults in household	X	X	X	X
PERSONID	Person ID number		X	X	X
PRMACT	Primary activity last week		X		X
PROXY	Trip info from respondent or proxy		X		X

6.7 Derived Variables

A number of derived or computed variables were developed to increase the dataset's usefulness when executing widely used queries. These variables are documented in the Derived Variables document, available on the NHTS website, with detailed logic for reproducing the variable or understanding its derivation. Reasons for developing a derived variable include:

- Renaming a questionnaire variable to match names used during previous NHTS programs or new names provided by FHWA,
- Calculating a variable from one or more variables in the questionnaires to provide summary variables to aid data users,
- Obtaining the variable from external sources to provide additional descriptors, and
- Creating flag variables to identify data records that had been imputed.

6.8 Trip Purpose Variables

Trip purpose, the participant's coded response for why they made the trip, is one of the most common variables used in travel behavior analysis. As listed earlier, the 2017 NHTS has several variables for describing each trip's purpose, divided into two coding schemes, referred to here as "one-way" and "round-trip" purposes. These are displayed in Table 6-4. The one-way scheme, also described as a "from-to" scheme, has been used for collecting trip purpose since the 1995 NPTS.

- A one-way scheme provides enough data for trip-chaining or round-trip analysis and more advanced coding and processing.
- A round-trip purpose scheme can be derived from the surveyed trip purpose and is necessary for comparing NHTS trip purpose to earlier NPTS. The roundtrip scheme is often used because it allows a focus on the purpose of the travel, for example, a trip from work to home would be categorized as "work' in the roundtrip scheme but as ' home' in the whyto scheme. When using the whyto scheme, trips to return home, will not provide a descriptive reason for the travel. The roundtrip scheme allows the miles from work to home to be associated with a trip to work.

Table 6-5. 2017 NHTS Trip Purpose Variables

Variable Name	Derived Status	Description
WHYTO	No	Asked-of respondent, for every place, "What was your main activity at…"
WHYFROM	Yes	In converting places to trips, WHYFROM is derived from WHYTO so that origin and destination purposes are related at the trip level
WHYTRP90	Yes	Aggregation of WHYTO for relating trip purposes across NHTS programs back to 1990
WHYTRP1S	Yes	Aggregation of WHYTO for relating trip purposes across NHTS programs back to 2001
TRIPPURP	Yes	Aggregation of WHYTO to five main purposes to be consistent with 2001 and 2009

Using the Data 7

7.1 Travel Concepts

The following travel concepts are central to using the NHTS Trip data and are provided primarily for data users who are not familiar with NHTS data. However, these may also be useful to all data users because the use of certain terms and concepts often vary by individual survey.

7.2 Trip

Definition: On a given travel day, participants were asked to report all locations they went to from 4 am to 3:59 am the next day, regardless of how long they were there, including cases where they started and ended at the same location, like walk or bicycle trips for exercise. A Trip represents that start and end movement from location to location by any mode of transportation.

Example: A person starts their day at home, goes for a jog around the neighborhood and returns back home. They then go to work and return home later where they end their travel day. This example represents a travel day with three trips (the loop trip for exercise, the trip from home to work, and the trip from work to home).

7.3 Person Trip

Definition: A trip by one person using any mode of transportation. This is the most basic and universal measure of personal travel. Each record in the Trip file in the NHTS dataset represents one person trip.

Example: Two household members travelling together in one car are counted as two person trips. Three household members walking to the store together are counted as three person trips.

7.4 Person Miles of Travel (PMT)

Definition: The number of miles traveled by each person on a trip. The purpose is to account for all miles traveled by all people using any mode.

Example: Two people travelling together take a six-mile subway trip to the airport. That trip results in 12 person miles of travel (the sum of all the miles traveled by all the people who traveled). A ten-mile van trip with a driver and three passengers (assuming all passengers were household members) results in 40 person miles of travel (4 people each traveling 10 miles).

7.5 Vehicle Trips

Definition: A trip by a single privately operated vehicle (POV) regardless of the number of persons in the vehicle. In order to compute vehicle trips, trips records must be filtered for trips where a household member was the driver (DRVR_FLG = 01) of a POV (TRPTRANS = (03, 04, 05, 06, 08, 09, 18)). The conditions are:

- TRPTRANS = (03, 04, 05, 06, 08, 09, 18)
- DRVR_FLG = 01

Example: Two people travelling together in a car are counted as one vehicle trip. Four people going to a restaurant in a van is considered one vehicle trip.

7.6 Vehicle Miles of Travel (VMT)

Definition: Vehicle miles of travel is the total movement in miles of one privately operated vehicle (POV), regardless of the number of people in the vehicle. When computing VMT, trip records must follow the same filtering criteria as vehicle trips. See the conditions above.

Example: When one person drives a car 12 miles to work, 12 vehicle miles of travel are generated (number of vehicles times the number of miles traveled). If two people travel three miles by pickup, they generated three vehicle miles of travel.

7.7 Vehicle Occupancy

Definition: For NHTS data, vehicle occupancy is computed as person miles of travel per vehicle mile (referred to as the travel method). Note that the other commonly used definition of vehicle occupancy is persons per vehicle trip (referred to as the trip method).

Example: Three people travelling together on an eight-mile trip to the park would result in a vehicle occupancy of three.

7.8 Sample Tables and Logic, Online Analysis Engine and Other Resources on the NHTS Website

The NHTS website (http://nhts.ornl.gov) should be the data user's first stop to determine which of the many NHTS resources can assist the user in finding or creating the data needed. Many users seek data that is already available in the Frequently Asked for Tables. The standard tables are an invaluable resource to those starting to tabulate the NHTS data. To use this resource, click on the Analysis Tools tab, then on the Frequently Asked for Tables tab.

The Online Analysis Engine allows users to create tables without having to download the files and develop their own table statement. It produces properly weighted user-specified tables in either Excel or HTML format. The website also provides for user support by going to the "Contact Us" section.

The NHTS Website offers:

- Analysis tool with user-defined table creation capability
- A component for exploratory analysis of the data
- A number of standard NHTS tables
- Collection of papers and articles analyzing the NHTS data
- Repository of dataset users and uses as well as user support comments

7.8.1 Additional Resources

In addition to the tools offered by the official NHTS Website, Westat has developed a package in R (an open source software environment for statistical computing) to support reading, organizing, analyzing, and reporting using the NHTS dataset. View the Github page in the link below for download instructions, interactive examples, and more.

R Package – summarize NHTS (https://github.com/Westat-Transportation/summarizeNHTS)

- Manages the downloading, organizing, and loading of NHTS datasets
- Computes weighted aggregates and standard errors

- Produces interactive HTML visualizations
- Supports custom derived variable creation

7.9 Control Totals

Control totals are known values, external to the survey itself, which are used to adjust the survey weights for non-response and non-coverage. Control totals were used to adjust the 2017 NHTS weights for:

- The number of U.S. households, and
- The number of persons in these households.

The control categories chosen for the 2017 NHTS and the weighting procedure are described in Chapter 5 of this User's Guide. The full complement of control numbers for the 2017 NHTS data set is contained in the 2017 NHTS Weighting Report available on the NHTS website.

7.10 Weighted Sums

Weighted sums are simply the calculated sums of the survey weights. These values are helpful to users in verifying the correctness of data tabulations. The 2017 NHTS total sample sizes and weight sums for the four data files are shown in Table 7-1.

Table 7-1. Total Sample Sizes and Weighted Sums for Households, Persons, Vehicles and Trips

Data File	Sample Size	Weighted Sum
1. Household	129,696	118,208,251
2. Person	264,234	301,599,169
3. Vehicle	256,115	222,578,947
*Vehicles (VEHTYPE 01-07)	254,954	221,692,415
4. Trip (annualized)	923,572	371,151,971,524

*Note: Vehicles whose type is not 01-07 are excluded to ensure that only motorized vehicles that can be licensed for highway use are included. This condition is also required for deriving the variables HHVEHCNT and VMT_MILE.

Other travel indicators are sample sizes and weighted sums. These are displayed in Table 7-2. The user should use these as marginal checks when tabulating the NHTS data.

Table 7-2. Total Unweighted and Weight Sums for Selected Key Travel-Related Data

Data	Unweighted	Weighted
Workers	128,288	156,988,243
Drivers	217,452	223,277,172
Person Trips	923,572	371,151,971,524
Person Miles of Travel (PMT)*	10,567,913	3,970,286,733,838
Vehicle Trips	611,342	220,429,661,377
Vehicle Miles of Travel (VMT)*	5,826,298	2,105,881,711,626

*Based on calculated trip distance

7.11 Weighting the Data

Chapter 5 describes how the weights were calculated for the 2017 Survey. The weights reflect the selection probabilities and adjustments to account for eligibility, nonresponse and undercoverage. To obtain estimates that are minimally biased, weights must be used. Note that the 2017 NHTS sample was designed to oversample households in 13 add-on states and or metropolitan planning areas. The weighting process adjusted that oversampling to provide correctly balanced estimates by geographic area. Because the weighting also involved adjustments for demographic factors, such as household size, race and ethnicity, tabulations without weights may be significantly different than weighted estimates and may be subject to large biases. Users can obtain estimates of the totals by multiplying each data value by the appropriate weight and summing the results.

7.11.1 Which Weight to Use?

There are several different weights, and it is important that the appropriate weight be used for a particular estimate. There are household weights, person weights and trip weights.

- **Household weights** (WTHHFIN). Use these when tabulating an estimate at the household level (e.g., number of households by household vehicle ownership and distribution of households by number of household drivers).
- **Vehicle weights** (WTHHFIN). These are the same as the household weight since the vehicle is considered a household attribute. Use the household weight for items such as vehicles by vehicle type or by vehicle age.
- **Trip weights** (WTTRDFIN). Use these for estimates involving numbers of trips or miles of travel, for example, number of vehicle trips by trip purpose. Only trips in privately operated vehicles (POV) that are reported by the driver should be counted in estimating vehicle trips. For example, if a person reports being a passenger in a vehicle driven by another household member, that trip would not be counted.

- **Person weights** (WTPERFIN). Use these for person-level estimates of non-household and non-travel day items of interest, for example workers by gender, drivers by annual miles estimated, etc.

As an example, to estimate the number of daily trips per household by Census region, for each region, calculate:

- The weighted count of households = Sum of the household weights, and
- The weighted count of trips = Sum of the trip weights.

The estimate of daily trips per household for that region is then simply its weighted trip count divided by its weighted household count. The previous calculation needs to be performed over each replicate weight to be able to calculate standard error, which is explained further below

7.12 Replicate Weights and Sampling Errors

Table 7-3 provides the variable names for the weights and the replicate weights and this section discusses how they are used to estimate sampling errors.

Table 7-3. Description of 2017 NHTS Weights

	Household	Person	Trip
Weight	WTHHFIN	WTPERFIN	WTTRDFIN
Replicates	WTHHFIN1-98	WTPERFIN1-98	WTTRDFIN1-98

*Note: to calculate vehicle estimates, use the Household Weight (WTHHFIN).

Because we did not invite every person and household in the U.S. to participate in the 2017 NHTS, the sample estimate could differ from the result that would have been obtained if we had conducted a census under the exact same circumstances. Calculating sampling errors provides the basis for measurement of the variability in the estimated statistics, and allows analysts to make probability statements about how large the difference may be between an estimated sample statistic and what would have been obtained for that statistic had a census been conducted.

The replicate weights may be used to calculate standard errors. The idea in replicate variance estimation is that sample estimates are made for a number of subsamples of the fully conducted survey. One then looks at the difference between each replicate sample estimate and the full sample estimate and squares the difference. Finally, one sums up the squared differences across all the replicates, with an appropriate multiplicative factor.

The replicate weights were calculated using the Jackknife method. Standard error estimates can also be easily calculated using the following formula:

$$SE = \sqrt{\sum_{i=1}^{98} (\frac{6}{7})[REP_i - x]^2}$$

$$x = Final\ Weighted\ Estimate$$
$$REP = Replicate\ Weighted\ Estimate$$
$$i = Replicate\ Number$$

Where x is the full sample estimate (calculated by using the full sample weights) and REP(i) is the estimate calculated by using the replicate weights and the summation over the index i is from 1 to 98. For example, suppose one is interested in an estimate of persons. The weight WTPERFIN is used to calculate the overall estimate x. The weight WTPERFIN1 is used to calculate the estimate REP(1), the weight WTPERFIN2 is used to calculate the estimate REP(2), etc.

As an example of the use of standard errors, the weighted survey estimate of total transit trips is 9,444,506,727 with an estimated standard error of 212,640,677. This standard error estimate allows one to conclude with 95 percent confidence probability that the interval 9,022,528,224 to 9,866,485,231 (the weighted survey estimate of total transit trips plus or minus 1.96 times the estimated standard error) contains the estimated number of total public transit trips that would have been obtained if a census of households were conducted using the same procedures.

7.13 Nonsampling Errors

There are many sources of error in addition to error occurring because only a sample was selected. Some examples of nonsampling errors include:

- A respondent misunderstands a question and answers it incorrectly,
- A respondent does not recall a trip or remembers details of the trip incorrectly,
- An interviewer does not correctly record what the respondent says, and
- A person does not answer a specific question.

Undercoverage may also be a source of error. In a national address-based sample, such as that used for the NHTS, undercoverage can occur when respondents reside in very newly constructed homes whose addresses are not yet available on the sampling frame, when households have simplified addresses (e.g., John Doe, Anytown, MD 12345), or when the household respondent, either accidentally or purposely, does not report all the people living in the household.

Note that nonsampling errors can sometimes be much larger than sampling error. Furthermore, for this survey, accurate estimates of sampling error are possible but, as in most surveys, it is impossible to estimate nonsampling error.

7.14 Finding the Variables You Want

The 2017 NHTS datasets are large, complex and contain numerous survey and non-survey variables (e.g., metadata or derived variables). The following documentation is available on the NHTS website to assist users in locating NHTS variables:

Codebook. The Codebook is the most commonly-used and comprehensive source of information of the data files. There is a separate Codebook section for each of the four data files – Household, Person, Vehicle and Trip. The Codebook provides the variable name, a description of the variable and its characteristics (length, character or numeric) the questionnaire item used, the possible responses to each variable and the unweighted and weighted frequency of each possible response. Thus, the Codebook is useful for checking to insure that the user's calculations show the same frequency as in the dataset.

Data Dictionary. The Data Dictionary documents the variables contained in each of the four 2017 NHTS data files in a single alphabetical listing by variable name. Because many variables are in more than one file, the Data Dictionary list contains four columns indicating which data files contain the variable.

7.15 Merging Data from Multiple Files

Despite the effort to include key variables on multiple files (see Chapter 6), an analyst may need to use information from separate files. For example, to study the daily trip patterns of different types of privately operated vehicles (POVs), one needs to use the variable VEHTYPE (vehicle type) from

the Vehicle file and link it to trip characteristics maintained in the Trip file. In these types of circumstances, one needs to merge together two or more of the four files.

File merging can be complicated and confusing, and a mistake can lead to invalid results. However, understanding the structure and relationship of the four files can significantly clarify the process.

ID Numbers – Each unit (i.e., households, persons) in the survey has a unique identification number (ID). Specifically, each household is identified by a unique eight digit household ID (HOUSEID). Within each household, household members are identified by a two-digit person number (PERSONID) and, similarly, household vehicles are identified by a two digit vehicle number (VEHID). Finally, trips made by an individual are numbered by a trip number (TDTRPNUM) for a travel day trip.

With this numbering system, the number that identifies a unit within a household (e.g., the household's vehicles and household members) needs to be used in conjunction with the household ID to uniquely identify that unit. For example, if a household has a HOUSEID of 12345678, its first member has a PERSONID of 01, and its second member has a PERSONID of 02, then the first household member is uniquely identified by an ID of 1234567801 and the second member 1234567802.

Similarly, the number that identifies a trip taken by an individual needs to be used in conjunction with the person's unique ID (i.e., HOUSEID and PERSONID) to uniquely identify that trip.

Continuing the above example, assume that the first household member took three travel day trips on the assigned travel day. Thus, TDTRPNUM for the first trip is 01, the second trip 02 and the third trip 03. An ID of 123456780101 will uniquely identify the first trip taken by the first household member of Household 12345678. Likewise, an ID of 123456780102 and an ID of 123456780103 will uniquely identify the second and the third trips taken by the same person, respectively. The third trip ID is represented as:

HOUSEID + PERSONID + TDTRPNUM = {12345678}{01}{03}

Table 7-4 shows the most common data linking of any two data files. The linking ID must be common to both the "from" and "to" files. For example, in linking Person file data with Trip file data, the variable TDTRPNUM would not be used because it is only on the Trip file, not on the Person file.

Table 7-4. Examples of Link Variables between 2017 NHTS Data Files

File 1 (From)	File 2 (To)	Linking ID Variables
Household file	Person file	HOUSEID + PERSONID
Household file	Vehicle file	HOUSEID
Household file	Trip file	HOUSEID
Person file	Vehicle file	HOUSEID + PERSONID
Person file	Trip file	HOUSEID + PERSONID
Vehicle file	Trip file	HOUSEID + VEHID

7.16 Merging Data File Example

Below is an example of a scenario in which one would need to merge multiple data files before performing analysis. In this example, the user wants to analyze the impact of occasional telecommuting on the number of daily trips. WKRMHM, the variable indicating occasional telecommuting, is located in the Person file and the trip information is in the Trip file. Consequently, the two files need to be merged on common identifiers.

The variables HOUSEID and PERSONID combined enable one to use the Person file to identify those who occasionally telecommute and those who do not. Using the combined identification number for HOUSEID and PERSONID, one can identify trips taken by that person in the Trip file. In this case, HOUSEID and PERSONID combined is the common identification needed to merge the Trip and Person files.

The illustration below shows how the two files are "linked" by common household and person identifiers. After the merge, each record in the resulting table should correspond with a unique trip, like the Trip file. This is because the Trip file contains information that is more "granular" than the Person file. Thus, the variable, WKRMHM, is repeated for each person trip, as shown below.

Trip file

HOUSEID	PERSONID	TDTRPNUM
10000000	01	01
10000000	01	02
10000000	02	01
10000000	02	02
20000000	01	01
20000000	01	02
20000000	01	03

Person file

HOUSEID	PERSONID	WKRMHM
10000000	01	02
10000000	02	01
20000000	01	02

Combined Trip and Person file

HOUSEID	PERSONID	TDTRPNUM	WKRMHM
10000000	01	01	02
10000000	01	02	02
10000000	02	01	01
10000000	02	02	01
20000000	01	01	02
20000000	01	02	02
20000000	01	03	02

Here is another example. Let us say your goal is to analyze the impact of occasional telecommuting on the number of daily trips. To do this you would need to merge the trip and person data files. Instructions on how to accomplish this follow.

Step-by-step Instructions:

1. Read the person table into your software, selecting WKRMHM and the person table identifiers, HOUSEID and PERSONID.

2. Read the trip table into your software, selecting the trip table identifiers, HOUSEID, PERSONID, and TDTRPNUM.

3. Instruct the software to join the person table and trip table on common identifiers, HOUSEID and PERSONID, so that every trip record is matched to its corresponding person record.

7.17 ID Variables Not Always Sequential

The ID variables within a file are not always sequential. Some reasons why the numbers are not sequential include:

- Some persons and vehicles reported by the household respondent were later found to not belong with the household and were deleted from the data set,
- Some trip segments reported as separate trips were combined during editing, and
- Some trip segments reported as a single trip were split into two.

In each of the examples above, the person, vehicle and trip IDs were not renumbered to be sequential.