

2017 NHTS Data User Guide

Draft



January 5, 2018

Submitted to:

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

<u>Chapter</u>		<u>Page</u>
1	Introduction to the 2017 NHTS	1
	1.1 Survey Sponsors and Users	3
	1.2 Overview of Survey Scope and Coverage	7
	1.3 Overview of Survey Content	8
2	Survey Procedures and Methodology Title	15
	2.1 Overview	15
	2.2 The Survey Process	17
	2.3 Telephone Data Collection.....	29
	2.4 Survey Procedures	30
	2.5 Data Editing.....	33
3	The NHTS Data Series: Changes over Time in the NPTS/NHTS	37
	3.1 Major Differences in 2017 NHTS.....	38
	3.2 Trip Distance Reporting	40
	3.3 Trip Purpose Coding.....	42
	3.4 Mean of Travel Coding.....	44

4	Survey Response Rates	47
4.1	Overview	47
4.2	Classification of All Sampled Addresses.....	48
4.3	Recruitment Survey Response Rate.....	49
4.4	Complete Households.....	52
4.5	Retrieval Survey Response Rate	52
4.6	Overall Survey Response Rates.....	54
5	Weight Calculations	56
5.1	Initial Household Weight	56
5.2	Final Household Weight.....	60
5.3	Person, Trip, and Vehicle Weights	61
5.4	Replicate Weights	62
6	Description of Data Files	64
6.1	Basic Data File Structure	64
6.2	Data File Relationship	64

Table of Contents (continued)

Chapter

Page

6.3	When is a Record on the File?	65
-----	-------------------------------------	----

6.4	Data Documentation.....	66
6.5	Data File Conventions and Special Codes	67
6.6	Derived Variables	68
6.7	Trip Purpose Variables	69
7	Using the Data.....	70
7.1	Travel Concepts	70
7.2	Trip	70
7.3	Person Trip	70
7.4	Person Miles of Travel (PMT).....	71
7.5	Vehicle Trips.....	71
7.6	Vehicle Miles of Travel (VMT)	72
7.7	Vehicle Occupancy	73
7.8	Sample Tables and Logic, Online Analysis Engine and Other Resources on the NHTS Website	73
7.9	Control Totals.....	74
7.10	Weighted Sums	75
7.11	Weighting the Data	76
7.11.1	Which Weight to Use?	77
7.12	Replicate Weights and Sampling Errors.....	78
7.13	Nonsampling Errors	79
7.14	Finding the Variables You Want.....	80
7.15	Merging Data from Multiple Files.....	81

7.16	ID Variables Not Always Sequential	83
7.17	Merging Data File Example	83

Table

1-1.	Sample sizes for national and add-on areas	4
3-1.	Differences in maximum values affect the mean	42
3-2.	Purpose codes, 2017 and 2009 NHTS	42
3-3.	Mean of travel codes, 2017 and 2009 NHTS	45
4-1.	Classification of sampled addresses	48
4-2.	Weighted recruitment response rate by study area	52

Table of Contents (continued)

<u>Table</u>		<u>Page</u>
4-3.	Weighted retrieval response rate by study area	53
4-4.	Overall survey response rate by study area	55
7-1.	Total sample sizes and weighted sums for households, persons, vehicles and trips	76
7-2.	Total sample size and weight sums for selected key travel-related data	76
7-3.	Description of 2017 NHTS weights	78

7-4.	Examples of link variables between 2017 NHTS data files	82
------	---	----

Figure

1-1.	Treemap of sample sizes for national and add-on areas.....	6
1-2.	Share of vehicle miles of travel by sector	8
1-3.	Schematic of the NHTS data	9
2-1.	Recruitment survey flowchart	19
2-2.	Retrieval survey flowchart.....	22
6-1.	Data file relationship.....	65

Exhibit

5-1.	Flowchart of NHTS weighting procedures	58
------	--	----

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. The data collected in the survey series are used by the Administration, Congress, national and local policy makers, and transportation planners to study the extent and type of daily travel in the United States. This inventory of travel behavior includes trips made by all means 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. Data from the NHTS are included in broader, bi-annual reports to Congress on the performance of the surface transportation system.

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 critical resource for understanding travel behavior in the nation. The current survey (2017 NHTS) is the eighth in the series, and allows researchers, planners, and officials at the state and federal levels to assess current travel and to develop long-term trends. The NHTS is the only 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 unique and not available from any other source since they 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 is challenging and underscores the importance of reliable data for evidence-based decisions. The NHTS provides critical 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. Each of these travel attributes can be further analyzed by the socio-demographics of the traveler and household. Importantly, the data series allows the development of trends over time to help measure changes in mobility patterns.
- **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 air quality issues, the development and use of new technologies, equity concerns, and outreach.
- **Modeling and Planning Applications:** Includes use of the NHTS in calibrating and validating state and regional travel demand and forecasting models (both four-step 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 NHTS includes a special module related to children's travel to and from school to support Safe Routes to School.

All Federally-funded surveys must be cleared by the Office of Management and Budget's (OMB) Office of Statistical Programs in accordance with the Paperwork Reduction Act. 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 what is being 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. The NHTS clearance was received in November 2015, under number 2125-0545. OMB clearance allowed for pretesting the survey, including tests of methods, materials, survey operations and other related aspects of fielding such a large, complex survey. The conduct of the full survey was from April 2016 through May of 2017, followed by quality control, weighting, and database development for public use. The survey data are available on the website at: <http://nhts.ornl.gov/>.

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

1.1 Survey Sponsors and Users

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 is charged with the overall responsibility to obtain current information on national patterns of travel, which establishes 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.

Furthermore, 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. In addition, NHTS informs policy and initiatives in other fields, including health and energy. Data on bicycling and walking have been used by the Centers for Disease Control (CDC) in its 10-year agenda, Healthy People 2020 and most recently in the Step it Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities. The Energy Information Administration (EIA) relies on NHTS-calculated vehicle miles traveled (VMT) and household vehicle data, combined with other data from the EIA and the Environmental Protection Agency (EPA) to derive vehicle fuel consumption and vehicle fuel expenditures. 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. The Institute for Highway Safety (IIHS) uses NHTS data to develop its safety ratings. Since 2009, NHTS has supported the Safe Routes to School Program by collecting specific data on children’s travel to school.

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 specific geographic 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 corridor level, interchange, and transit infrastructure and planning projects, among others. In addition, 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 larger number of samples collected through the Add-on program also provide the NHTS analysts with the ability to compare geographic areas and detailed demographic population groups with more statistical reliability. As shown in Table 1-1, the goal was to collect 129,112 samples in the 2017 NHTS, which included a national sample of approximately 26,000 households and 103,112 additional Add-on samples purchased by thirteen States or MPOs.

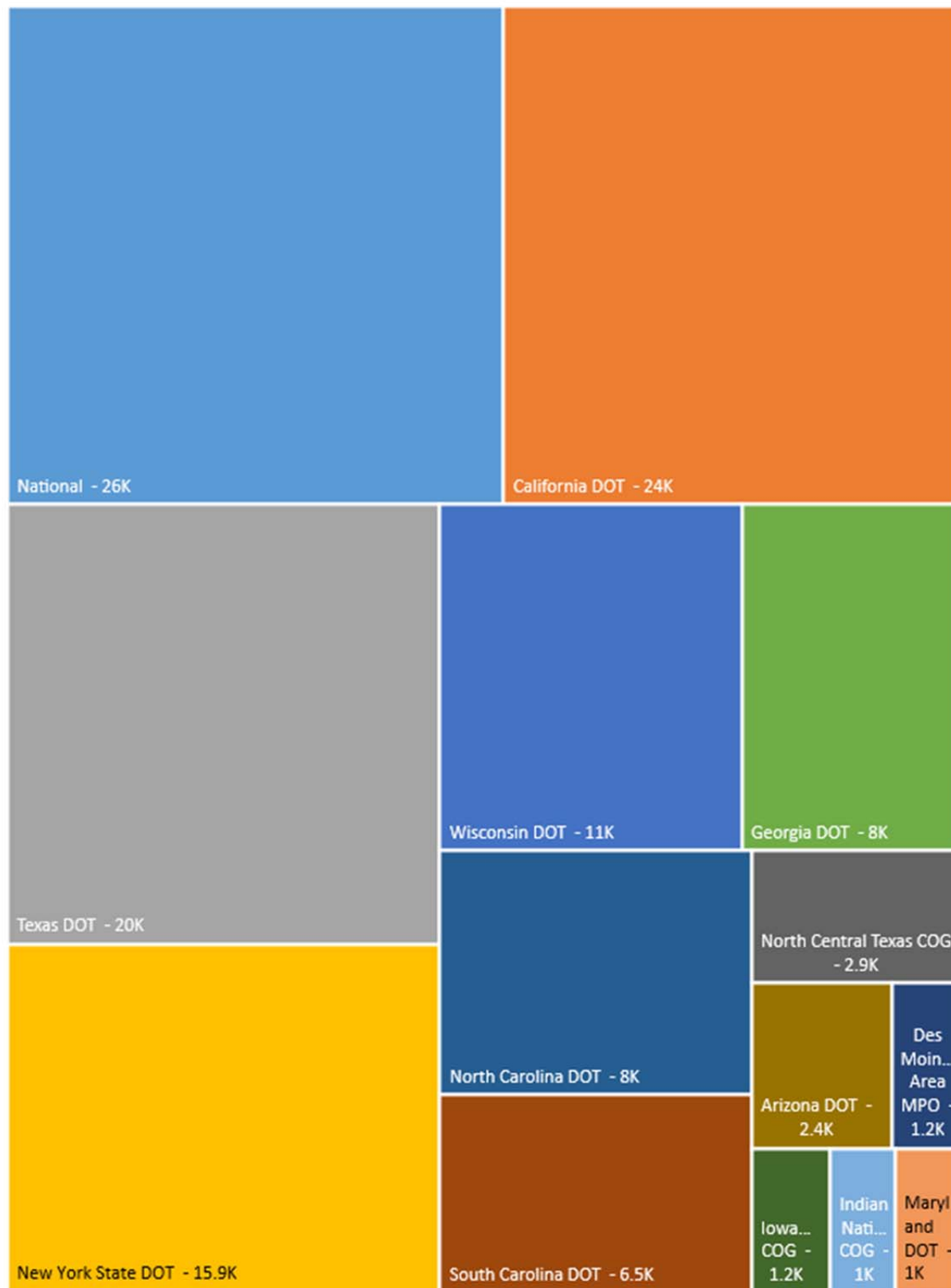
Table 1-1. Sample sizes for national and add-on areas

Study Area	Sample Size¹
National	26,000
Arizona DOT	2,444
California DOT	24,000

Des Moines Area MPO	1,200
Georgia DOT	8,000
Indian Nations Council of Governments	1,000
Iowa Northland Regional Council of Governments	1,200
Maryland DOT	1,000
New York State DOT	15,851
North Carolina DOT	8,000
South Carolina DOT	6,500
Wisconsin DOT	11,000
Texas DOT	20,000
North Central Texas Council of Governments	2,917
TOTAL	129,112

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

Figure 1-1. Treemap of sample sizes for national and add-on areas



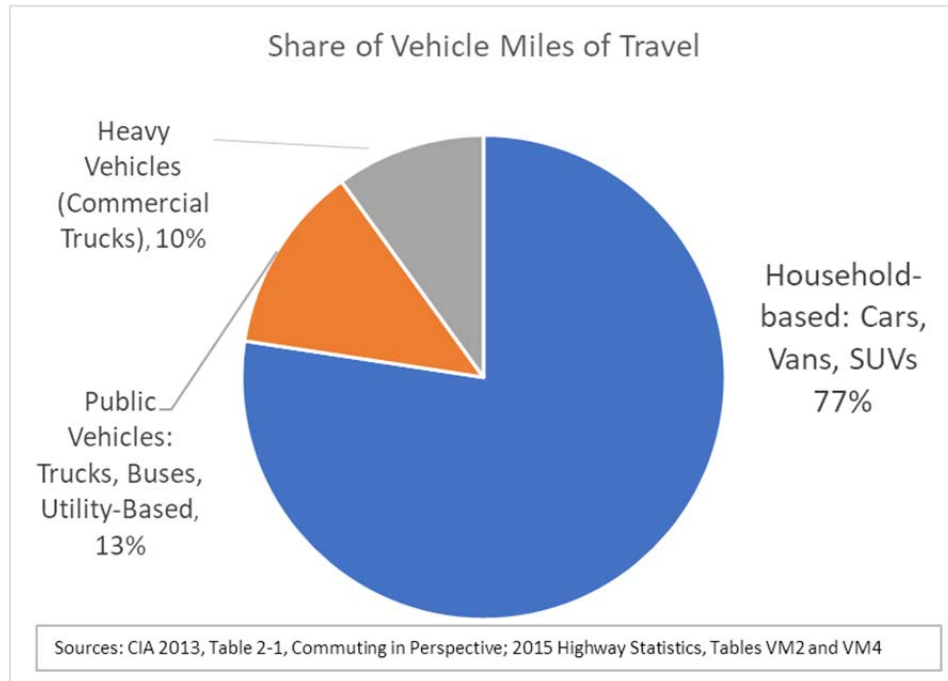
The next section describes in more detail the scope and geographic coverage of the NHTS sample.

1.2 Overview of Survey Scope and Coverage

As in previous series of the NHTS, the 2017 NHTS maintained a two-phase study, which included a household recruitment survey (phase 1) and trip level retrieval survey (phase 2). Unlike the previous series of NHTSs which used a Random Digit Dial telephone sampling method and only Computer-Assisted Telephone Interview (CATI) data collection, the 2017 NHTS used a mail out recruitment method with mail-back as the primary recruitment response mode with phone or web response options, while the travel day data retrieval survey phase offered both phone and web response options.

The NHTS is designed to collect information about travel originating from 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 accounts for over three quarters of all roadway vehicle miles in the US (see Figure 1-1).

Figure 1-2. Share of vehicle miles of travel by sector

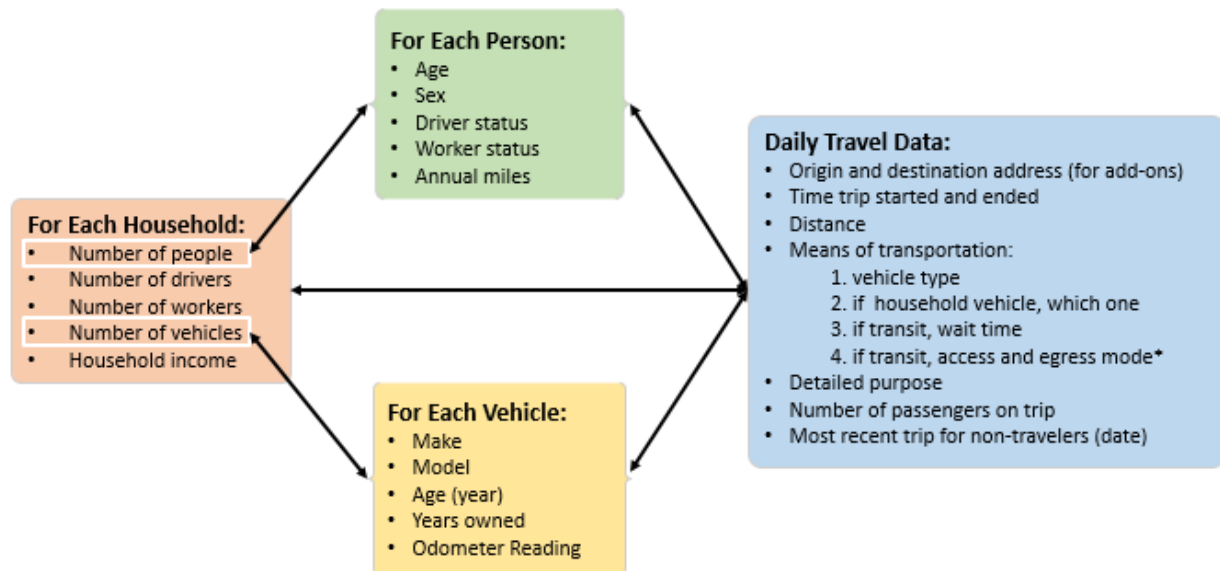


The next section describes in more detail the content of the data files collected in the NHTS.

1.3 Overview of Survey Content

The NHTS data are the only source of information on travel linked to individual personal and household characteristics, socio-economic characteristics, vehicle ownership, and vehicle attributes. Much of the core content of the NHTS has been collected since the first survey in the series conducted in 1969. The core data involves an inventory of trips taken within a 24-hour period. For each trip, respondents report their trip purpose (e.g., work, school, shopping, recreation, etc.); mode of transportation (car, bus, walking, etc.); travel time; time of day; 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 1-3. Schematic of the NHTS data¹



¹ Figure displays a subset of NHTS variables

In the NHTS content list provided below, the data items *italicized* below are considered “core” NHTS data, largely composed of the items that have been collected in all seven surveys to date.

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
- Travel log use

FOR EACH VEHICLE

- *Make/Model/Age (year)*
- Body type
- Fuel type
- *Annual miles driven*
- How long owned
- Odometer reading
- Alternative Fuel

- Primary Driver

FOR EACH PERSON

- *Age/Sex/Relation to reference person*
- *Driver status*
- *Worker status/Primary activity*
- Internet use
- 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
- School travel (children)

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

DAILY TRAVEL DATA

- Origin and Destination address (for all locations)
- *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
- Interstate Use
- Tolls Paid
- Trip Purpose
- Detailed purpose
- Travel Party Size
- Last time of travel

QUESTIONS ABOUT TRAVEL EXPERIENCES, OPINIONS

In addition to the Person, Household, Vehicle, and Daily Travel related data, other kinds of data were also collected. This included questions that related to:

- Count of number of times or days the person used different modes of travel, such as number of walk, bike, and public transportation trips during the past 7, and 30 days – including if these trips were for exercise (for walk and bike modes), and reasons for not using these modes more.
- Alternative means of transportation (such as bike, walk, public transportation, taxi, car share, ride share, etc.) – including 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.
- Commute times to work, work schedule, work locations, and others.
- The health of persons in regards to how it may affect their travel
- Technology such as internet use through PED's (personal electronic devices), smartphone app use, and internet purchases
- How many miles persons drove in the past 12 months and how long they have owned personal vehicle.

When was the Survey Conducted?

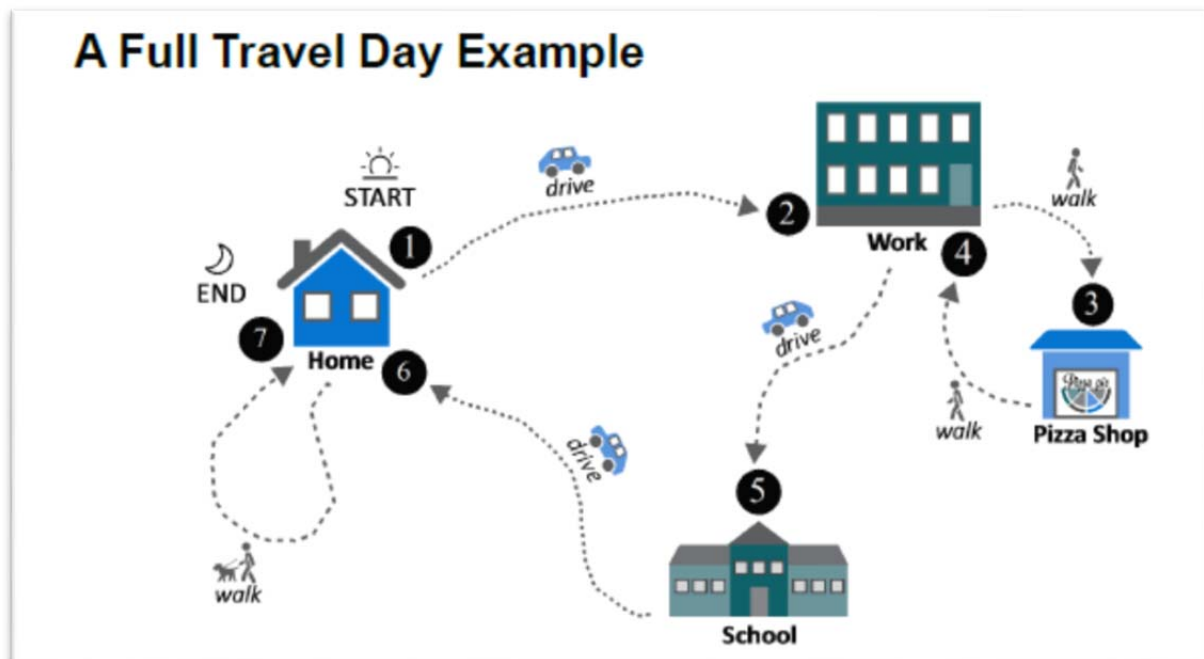
A pilot study was conducted from July to October 2015. The main survey was conducted from March 2016 through May 2017 with travel dates starting on April 19, 2016 and ending on April 25, 2017. Travel days were assigned for all seven days of the week, including holidays. The survey data were weighted 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 households that completed 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 guarantees that most household members will be at home at the start of their travel day. Trip level records that began and ended entirely outside of the United States, were not included.

Survey materials were provided to participants with instructional text on how to record their travel along with an example illustration of what a typical travel day might look like. The NHTS Field Documents contains an example of all survey materials.

The example included several routine trips as well as examples of a walk trip, as these are trips with a higher likelihood of being misreported or not reported at all. Below is an illustration of a pictorial that was included in the households travel day or retrieval package.



Survey Procedures and Methodology Title

2

2.1 Overview

An understanding of the data collection for the NHTS is essential to the proper use and interpretation of the data. In the next few sections, we 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. Survey methodology changed substantially from previous iterations, as the 2017 NHTS changed from a telephone based Random-Digit-Dialed (RDD) methodology to an Addressed Based Sample (ABS) multimode survey featuring web, paper and telephone instruments. Key aspects of the methodology are highlighted in the sections that follow.

The Data Collection Process

The NHTS was conducted as a multi-mode survey including paper questionnaire, web survey, and Computer-Assisted Telephone Interviewing (CATI).

At the sampling stage, each sampled address was randomly assigned a day of the week for their travel day assignment. After the recruitment interview was completed, each household was assigned a specific date as their "Travel Day" for which they were asked to report detailed data on travel.

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, Des Moines, Iowa, Georgia, Indian Nations Council of Governments, Iowa Northland Regional Council of Governments, Maryland, New York, North Carolina, North Central Texas Council of Governments, South Carolina, Texas, and Wisconsin. The sample file was designed and selected to yield 26,000 completed households for the national sample. Target numbers of completed households for the Add-on areas are found in Chapter 1.

Sampling Frame – The sampling frame consisted of all 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) and is updated 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 the 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).

2.2 The Survey Process

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 determined 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 there.

Recruitment Mailing – Households were invited to participate in the recruitment survey through the mail. Households received an invitation package containing an invitation letter, paper questionnaire, a business reply envelope (BRE) 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 BRE and the project website URL to provide additional information and add survey legitimacy.

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

A second recruitment package was mailed to non-responding households three weeks after the invitation package mailing. The non-response package included a non-response letter affirming that there is still time to participate, paper questionnaire, and a business reply envelope (BRE). The non-response letter urged respondents to fill out the paper questionnaire and return the complete survey using the BRE provided in the non-response package.

A second and final reminder postcard was mailed to non-responding households forty-four days after the initial invitation 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.

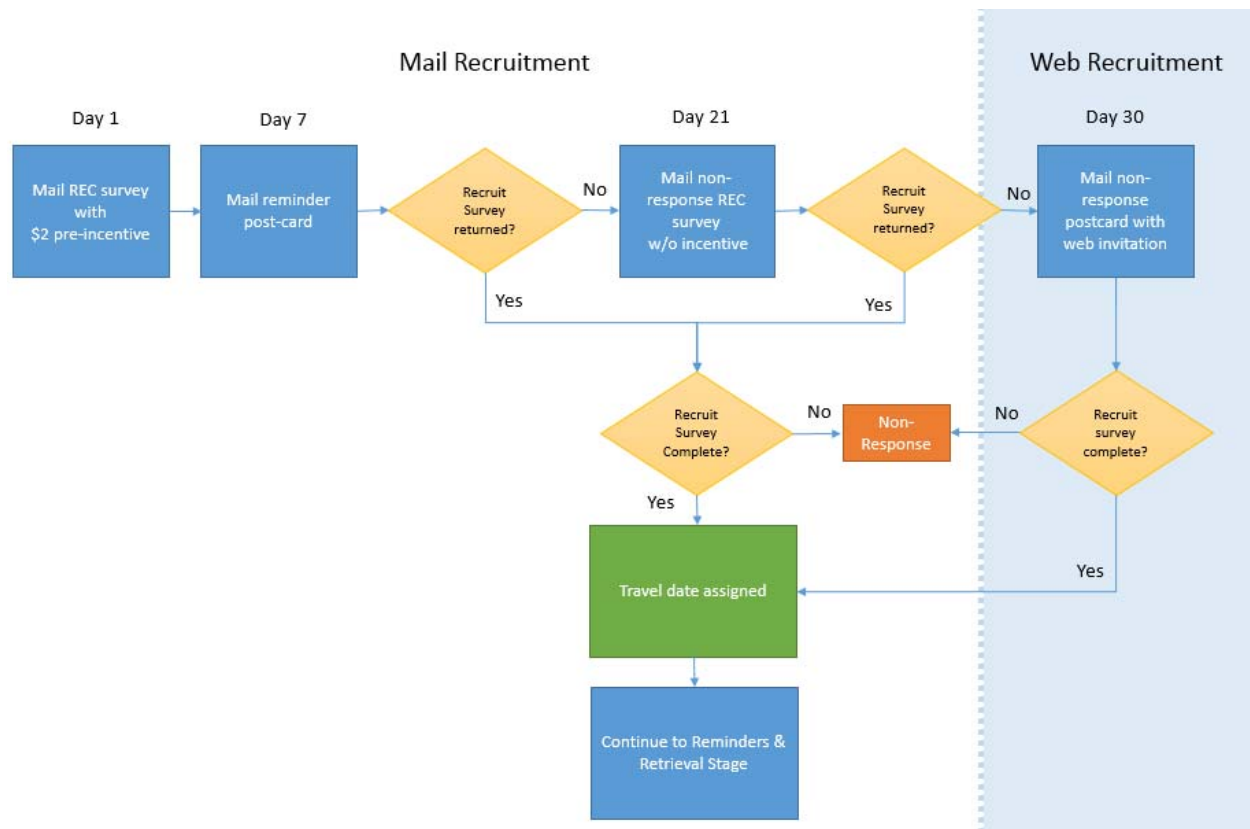
Adjustments in the mailing schedule of the non-response package and second postcard were made midway through data collection to encourage a more timely response. The mail date of the second reminder postcard was moved up from 44 days to 30 days after the invitation package mailing. In addition, the survey 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 or online.

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 contains an example of all mail materials including the recruitment package, non-responding household recruitment package, and reminder postcards.

Westat mailed invitation packages and reminder postcards to over 900,000 households. Over 239,000 (27%) households completed and returned a paper questionnaire to Westat 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.

Figure 2-1. Recruitment survey flowchart



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 this survey was to garner interest in the survey and collect contact information for the retrieval survey communication.

The primary data collection 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 completes.

Travel Day Date Assignment– 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, part of the weighting process adjusted the estimates of total persons and total households by month.

After each recruitment survey was processed, the Survey Management System (SMS) automatically assigned a travel date. The assigned travel date was communicated to each household through the travel log package mail materials. The SMS determined the household’s travel date based on the pre-selected day of the week during sample file development. The assigned travel date was at least 10 days in the future, which allowed time for the travel log package mailing to reach the household ahead of the travel date. This provided the household time to review the mailing, understand the process, and to contact Westat if they had any questions.

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). For more details, see the NHTS sampling plan.

Travel Log Mailing– The use of travel logs by respondents in household travel surveys has been shown to improve the accuracy of trip reporting. Therefore, each households that completed a household recruitment interview were mailed a travel log for each member of the household five year of age and older to record their travel. Of the over 275,000 persons who reported their travel, 66 percent reported using their travel log. The log package was mailed within a day or two following the completion of the household recruitment interview. The log package was mailed via United States Postal Service Priority Mail and 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;
- A five-dollar cash incentive to encourage participation;

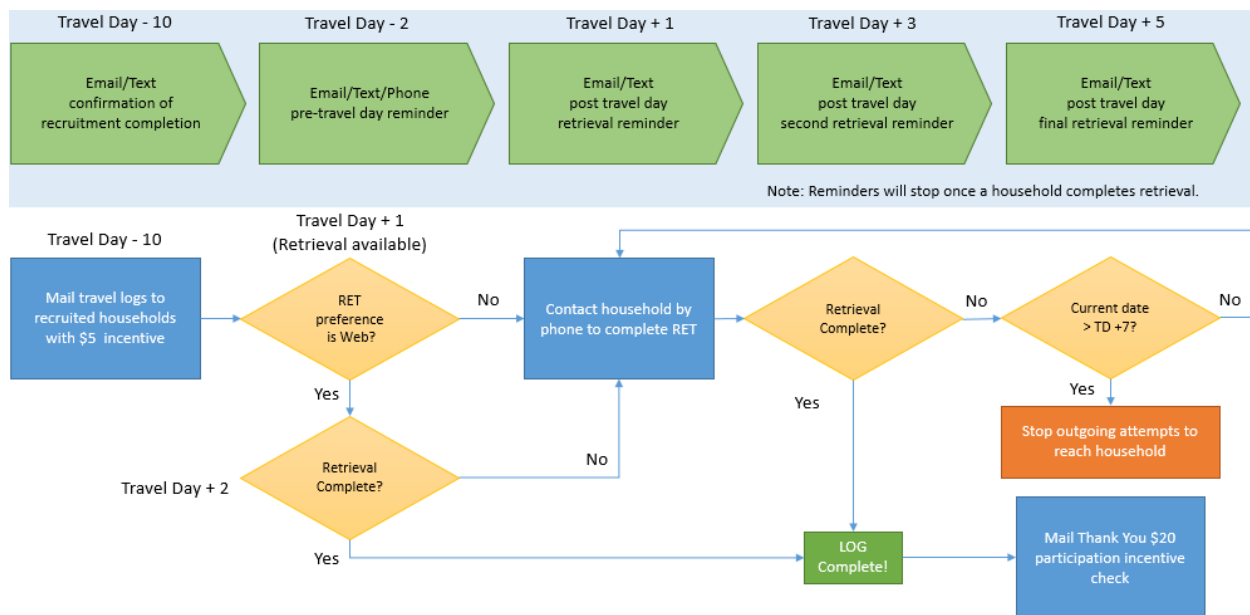
- A personalized travel log for each household member at least five years of age or older. The front side of each log provided instructions on how to record their travel and to complete the survey after their assigned travel day. The reverse side included an example of a completed log for guidance; and
- 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 contains an example of all travel log mail materials including the retrieval letter, travel log, and odometer mileage form.

Pre-Travel Day Reminders – Once a recruitment survey was processed, a confirmation email was sent 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 will receive a package inviting the household to participate in the retrieval survey.

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 log on their assigned travel date. Households were provided the survey's toll-free number if they had any questions.

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, participants were prompted to confirm and/or correct all the household information they had provided earlier at the beginning of the retrieval survey. Upon verification and/or correction of the recruitment data, additional household level information was collected including 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 is provided later in this section.

The retrieval survey could be completed online or by telephone. Sixty percent of retrieval surveys were completed online, thirty percent over the phone, and the remaining ten percent by using a combination of both modes. The online survey was available across multiple platforms including personal computer, tablet, and smartphone. Seventy-eight percent of the online retrieval surveys were completed on a personal computer, nine percent on a tablet, and thirteen percent on a smartphone.

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. Reminder messages were sent via email, text message, and Interactive Voice Response (IVR). Reminders were sent to every form of contact that was provided in the recruitment survey. 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’s toll-free number if they had any questions.

Retrieval Period – There was a seven-day period after the assigned travel date during which interviewers were permitted by the CATI system to call each household to collect a household’s travel data.. However, the web retrieval survey remained open to participants after the seven-day period.

During the recruitment survey, in addition to collecting contact information, participants indicated if they prefer to complete the second part of the 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 them the day after their assigned travel day. Retrieval calls continued for the next six additional days unless the household completed the retrieval survey or refused to participate in the survey.

If a participant provided a phone number, but reported that they preferred to complete the second part of the survey online, they were given two days to do so. If the household had not completed the survey within two days, Westat 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 they provided an email address.

Although data collectors did not make outbound calls after the seven-day period, Westat allowed participants to log in to the web survey or accepted calls to the toll-free number to collect their travel data after that period. Westat allowed participants to report their travel after the seven-day window, if the participant initiated contact.

Proxy Interview Procedures – A proxy interview is one in which someone else in the household reports for the subject. During the NHTS data collection, an adult household member always served as the proxy for any child 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 all retrieval surveys were completed entirely online with another ten percent completed over a combination of web and telephone survey. As such, enforcing traditional proxy rules for the majority of retrieval completes was not possible. Westat did encourage self-reporting and created an online mechanism 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 is 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 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 diary on their travel day, the proxy reporting household member was asked to retrieve the diary and refer to it when they served as a proxy for others in the household. 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).

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.

Westat's Instrument Design, Evaluation, and Analysis Services (IDEA) group reviewed 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.

NHTS respondent materials were also reviewed by survey design experts including an expert panel put together by FHWA with leading survey methodologists including Mick Couper, Don Dillman, Laura Erhard, Paul Lavrakas, Mike Brick, Steven Polzin, Guy Rousseau and Clyde Tucker. Westat's IDEA group also cognitively tested the respondent materials. Westat made refinements to the respondent materials based on feedback from the expert panel and the cognitive testing recommendations.

In accordance with OMB's guidelines, Westat's IDEA group cognitively tested the retrieval survey with live participants. Special attention was paid to vetting the trip-reporting module of the survey

instrument. Information gathered from observing participants navigate 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 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 address information all at once for each person. This allowed participants to familiarize themselves with the process and streamlined the geocoding aspect of the survey.

If a participant reported that another household member traveled to a place with him/her, then this place was automatically recorded on the roster for the other household member, provided that household member had not yet reported their travel. When the other household member reported their travel, (s)he merely had to confirm that they went on the trip, confirm the trip destination and start time, and add the departure time. If the currently surveyed household member confirmed that the trips were the same, the trip was retained on the roster; otherwise, the respondent would have to delete the place. If the currently surveyed household member confirmed that the trip was correct then certain trip details, such as party size, household members on trip, and household vehicle used, were copied from the previous household member. Other trip details, such as trip purpose, were still required for the currently surveyed household member to report.

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. As in 1995, 2001, and 2009, data collection on the 2017 NHTS took 14 months, rather than 12 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. Mail-back questionnaire have been found to elicit higher response rates than other modes, and the inclusion of the recruitment screener helped the 2017 NHTS achieve a response rate two to three times higher than comparable regional household travel surveys are achieving. 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. The survey weights adjust for the monthly differences in number of interviews completed. The sample of addresses was refreshed midway throughout the study field period to ensure that new addresses were included and all geographic areas were completely represented in all seasons.

NHTS Core Data – There is a group of variables considered “core” NHTS data, and it is largely composed of data collected consistently across all previous and current Nationwide Personal Transportation Survey (NPTS) and NHTS surveys. The data items considered core in the 2017 NHTS are:

- Household Level Core Data – These data were collected for each household that completed a retrieval survey and included:
 - Number of household vehicles
 - Household Size
 - Household composition
 - Age of each household member
 - Sex of each household member
 - Driver status of each household member
 - Number of vehicles
 - Race & Hispanic status of household members
 - Household location confirmation
 - Household income

- Home ownership
- Person Level Core Data – These data were collected for each household member age five and older in households that completed the retrieval survey and included:
 - Education level
 - Worker status
 - If worker: employer information
 - If worker: typical work trip
 - Driver Status
 - If driver: annual miles driven
- Vehicle Level Core Data – These data were collected for each household vehicle, owned, leased or available for regular use by household members in households that completed the retrieval survey and included:
 - Make
 - Model
 - Model year
 - Length owned
 - Annual miles driven
 - Primary driver
 - Odometer readings
- Travel Day Core Data – These data were collected for each trip each household member age five and older made on the household’s assigned travel day and included:
 - Time trip started
 - Time trip ended
 - Trip purpose

- Distance to destination
- Trip duration
- Main mode of transportation
- If public transit: access and egress modes used
- Household vehicle used
- If household vehicle: which vehicle
- If private vehicle trip: did a household member drive
- If household member drove: which household member
- If someone else on trip: any household members
- If household members on trip: which household members
- If someone else on trip: number of non-household members

2.3 Telephone Data Collection

Interviewer-Assisted Data Collection – 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 by staff at Westat’s Telephone Research Centers (TRCs) and their team of at-home interviewers (AHI) located across the nation. The TRCs are located in Frederick, MD, Merced, CA, and Rockville, MD.

Interviewer Training – Westat supervisors trained a team of 194 Westat interviewers and 32 supervisors on the 2017 NHTS. Training sessions were conducted periodically over the 14-month data collection period to accommodate data collections needs and adjust for interviewer attrition. There was a peak of 129 interviewers working on the study in any given week. Of the 194 total interviewers, 19 interviewed respondents in both English and Spanish.

All interviewers assigned to the survey completed at least 20 hours of formal project-specific training. Help Desk staff were trained for an additional 10 hours to account for the additional demands of the role such as fielding questions about the survey and guiding participants who had difficulty navigating the online survey. For interviewers with no prior interviewing experience, these hours were in addition to four hours covering general interviewing techniques such as building rapport, handling refusals, etc. and another four hours training on the use of CATI. Additionally, they received two more hours of training on Westat Remote Access (WRA), which provides instructions on how to log into the Westat systems. These hours of non-project specific training occurred prior to the interviewer's assignment to the NHTS project. For the NHTS project, 79 percent of the 194 interviewers were experienced and did not have to go through the non-specific project training. Interviewers whom the TRC supervisory staff felt were not ready for “live” interviewing at the conclusion of the formal “classroom” training received additional training time until supervisors deemed them ready for live interviewing.

Interviewer Monitoring – Interviewer monitoring is an important aspect of survey quality control. FHWA project staff members were patched-in remotely and monitored the data collection interviews. Using telephones and screen sharing software, supervisors have both audio and visual monitoring capabilities. Supervisors inconspicuously monitored on average about 10 percent of each interviewer's work over the course of the study.

2.4 Survey Procedures

Callback Procedures – Multiple call attempts spread over various times of the day and days of the week are essential to achieving a high response rate on all telephone surveys. Call attempts to complete each retrieval interview 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 – Refusal conversion attempts were an important aspect of Westat's overall response maximization effort for the NHTS. An integral component of this effort was the utilization of a

select team of refusal conversion specialists. The team was comprised of Telephone Research Center interviewer staff members who had demonstrated exceptional skills in achieving high cooperation rates. Once interviewers were familiar with the questionnaires, and common reasons for refusals were identified, Westat supervisors held special training sessions on refusal conversion techniques for the refusal conversion interviewers.

If a respondent initially refused to complete an interview over the phone, the interviewer completed a separate CATI data collection module to record any information known about the household and the respondent's reason(s) for refusing to participate. Non-hostile refusal cases were flagged and made available to specially trained interviewers to make refusal conversion calls.

Bilingual Interviewing – Data collection was conducted in either English or Spanish. Interviewing in Spanish was an important factor in gaining the cooperation of Spanish-speaking only respondents and completing interviews with them. 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. The entire CATI questionnaire, including interviewer instructions, was translated into Spanish. 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.

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. Only data for useable households are included in the four files.

For recruitments completed by paper questionnaire, a recruitment was consider 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 consider 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.

Editing the Delivery Datasets – As a final editing step, Westat data analysts compared the frequencies between the survey dataset and the delivery. 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.

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.

Thank you letter and Incentive Mailing—Upon completion of the survey, Westat thanked the participants for their participation. 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. Westat mailed respondents a \$20 completion incentive with a thank you note within two weeks of completing the retrieval survey.

2.5 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. 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. The documentation for these checks can be found in the Recruitment and Retrieval Questionnaires 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. The triggering of logic checks sometimes required modifying previously provided information. That is, the current response of the respondent being a driver on the trip may be correct but the earlier recruitment interview response that the participant was not a driver was modified to a driver.

Post-Survey Editing – In surveys with complex questionnaires and procedures, such as the NHTS, the final dataset reflects fundamental approaches taken in the data collection and editing processes. For the 2017 NHTS, two approaches may have had considerable impact on the resulting data.

First is data imputation. If the respondent did not answer a question, Westat generally did not impute the response, (i.e., determine the logical response given responses to other questions). Carefully performed imputation has its place in many statistical surveys, however Westat and FHWA determined that imputation would be limited to certain variables where a known response was required for proper weighting (e.g., race). If a variable required data imputation, an imputation/edit flag was set for the variable to indicate the values that were imputed.

Second, a conservative approach was taken regarding editing reported data. For the most part, collected survey data were not edited.

Data Collection Editing – Throughout data collection, Westat ran programs to check for data consistencies. To ensure the data were complete and of highest quality, Westat implemented a series of data editing procedures. 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, analysts used a series of Quality Control (QC) Codes to categorize the checks that failed. These households were scheduled to receive a follow-up call (Research Call) to resolve the issues. If the issues are not resolved during the research calls, QC Codes are delivered along with the household record to indicate there are still unresolved data issues remaining for one or more persons in the household. There are five categories of QC Codes, which are outlined below.

1. **Address/Location/Geocode.** These edit checks were mainly focused on Address/Geocode precision and were implemented in situations where the true address/coordinates of a

location could not be verified. For instance, if a participant reports a place's location as being on a street or a city neighborhood, and no specific address was specified.

2. **Joint Travel.** These involved discrepancies between what one person and at least one other household member reported. This happens 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 cannot 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 occurs 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 described in the Data Collection Editing section. For instance, if a participant reports very long loop trips (over six hours), or reports shopping on a loop trip.
5. **First/Last Place Misreported.** Occurs when the first and/or last place reported on the travel day does 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 afterward but the participant did not provide that information.

Imputed Variables –Several variables were imputed during the weighting process if the value for the variable was missing (-7 (refused) -8 (don't know) or -9 (not ascertained)). Those variables are provided in the weighting section.

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.

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 completed person interview);

- **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 an interviewed person made on the household's travel day (one record for each trip each person made).

The NHTS Data Series: Changes over Time in the NPTS/NHTS

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-stage methodology to collect travel information. The first stage was a telephone recruit followed by the mail-out of a travel diary for each member of the participating household to record their travel on the assigned travel day. A travel diary was found to prompt the respondents to record all 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 recruit and travel diary 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

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 stage or recruit 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 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. About 60 percent of the respondents in 2017 NHTS chose to self-report their travel day information using the web-based system, about 30 percent completed with a CATI interviewer, and the remaining ten percent were a combination of both web and telephone survey.

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 (here). The data user is responsible for applying careful thought to any trend data that they calculate, to ensure to the best of their ability that the changes tracked over time represent ‘real’ changes in travel and not an unintended effect of changes in methodology.

3.1 Major Differences in 2017 NHTS

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 NHTS². In addition, users interested in developing trends over time can obtain further details of the

² “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>

effects of the methodological changes on data estimates from the 2017 NHTS in the “Summary of Travel Trends” (link).

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 diary package materials. The household mailed back the recruit 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 travel date, send the household a diary package, and implement travel day data retrieval.
- **Reminders:** Gentle postcard reminders were mailed out at several 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 was designed to incrementally reward participation. In the 2017 NHTS design, each sampled address received a \$2 cash incentive in the recruit. The travel diary 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 choose to report their travel by telephone interview, project staff members use the same web-based data entry system so that all data are subject to the same range checks and consistency checks. These automatic edit checks reduce reporting error, reduce survey length, and maintain the flow of data reporting. This approach ensures consistency in reporting while minimizing the respondent burden, and maximizing data quality.
- **Travel Day Diary:** Provided for each eligible household member (aged five and older). The diary/memory jogger was graphically redesigned and included identification of activity at each place rather than trip purpose.
- **Weekday –weekend distribution:** Add-on participants 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.

- **Trip reporting:** Westat’s web-based retrieval system was customized for the 2017 NHTS to enable respondents to self-report their travel activity 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 provides drop down tables to allow respondents to select the correct vehicle make and model information. It also includes Google Maps type search engines that assisted in identifying specific place names and locations when destinations were entered.
- **Loop trips:** Respondents 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 changes relating to the critical elements of trip distance reporting, purpose and mode of travel are further detailed below.

3.2 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. Regional household travel surveys had successfully used a similar approach and 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 from the geocoded origin and destination. These distance estimates are provided in the trip file for all reported trips (TRPMILES). In addition, an “as-the-crow-flies” distance for each reported trip is provided in the final dataset, values of which were computed using the great-circle formula (Vincenty, 1975), for example the variable identifying the distance between home and work (GCDWORK).

This 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 (here). 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 geocoded to a location, that location was accepted as valid. These values can influence the means—or averages—that result from analysis.

An example is shown in Table 3-1 (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 geocoded addresses that were simply far away from the respondent’s home, perhaps company headquarters or the like. To develop robust and comparable averages, the user is cautioned for each analysis to check the distribution of the variable values and determine whether there are any values that could be considered outliers for that analysis.

Table 3-1. 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 take care to understand and assess how the changes in trip distance reporting may impact the specific estimates of interest obtained from the 2017 survey in comparison with previous estimates.

3.3 Trip Purpose Coding

Because the survey was primarily conducted as a self-reported travel day, the number of trip purposes was reduced so as not to overwhelm the respondent. A comparison of the trip purpose codes for 2017 and 2009 NHTS is shown in Table 3-2. As shown, the bolded 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 3-2. 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

2017 Code	2017 Category Name	2009 Code	2009 Category Name
-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
		81	Social event
		82	Get/eat meal
		83	Coffee/ice cream/snacks
14	Other general errands (post office, library)	60	Family personal business/obligations

2017 Code	2017 Category Name	2009 Code	2009 Category Name
15	Recreational Activities (visit parks, movies, bars, museums)	61	Use professional services: attorney/accountant
		63	Use personal services: grooming/haircut/nails
		64	Pet care: walk the dog/vet visits
		50	Social/recreational
		51	Go to gym/exercise/play sports
		52	Rest or relaxation/vacation
		54	Go out/hang out: entertainment/theater/sports event/go to bar
16	Exercise (go for a jog, walk, walk the dog, go to the gym)	55	Visit public place: historical site/museum/park/library
		64	Pet care: walk the dog/vet visits
		51	Go to gym/exercise/play sports
		53	Visit friends/relatives
		30	Medical/dental services
17	Visit Friends and Relatives	20	School/Religious Activity
18	Health care visit (medical, dental, therapy)	22	Go to religious activity
19	Religious or other community activities	97	Other reason
97	Something else		

3.4 Mean of Travel Coding

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

Table 3-3. Mean of travel 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
Local Transit	'11' Public or Commuter Bus	'09' Local Public Bus
		'10' Commuter Bus
		'16' Commuter Train
	'16' Subway/Elevated/Light Rail/Streetcar	'17' Subway/elevated Train
		'18' Streetcar/Trolley
	'20' Boat/Ferry/Water Taxi	'20' Ferry*
	'12' Paratransit/Dial-a-Ride	'24' Special Transit-people with disabilities
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
	'15' Amtrak/Commuter Rail	'15' Amtrak/Inter-city Train
Other	'19' Airplane	'21' Airplane
	'97' Something Else	'97' Other

Survey Response Rates

4

4.1 Overview

As discussed in Chapter 3, 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. Final response rates in regional household travel surveys tend to be lower than 10% and this low rate presented a problem for OMB clearance, and a risk to the general confidence in the survey results. Based on successes in other areas of research, Westat survey methodologists proposed a survey design where a brief paper questionnaire with a pre-incentive is used for the recruitment survey followed by a larger 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, Westat mailed sampled addresses an invitation package containing a brief recruitment survey about travel habits, opinions on transportation issues, and household demographics. The recruitment survey was available online or over the phone, if requested by the respondent.

Following the recruitment survey, Westat prepared and mailed a travel log package to households that completed the recruitment survey. The travel log package invited households to participate in the retrieval survey, introduced their assigned travel date, 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. Provided in this section is a summary of the overall response rates, as well as the rates at both stages of the survey process.

4.2 Classification of All Sampled Addresses

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. All addresses fall 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. 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** – Invitation packages returned as Postal Non-Deliverable (PND) and addresses that were identified as non-residential were marked as out-of-scope.
- **Eligibility Unknown Addresses** – All other addresses fall into the category of eligibility-unknown. These include 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

Household screener interview classification 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 Telephone Numbers in Sample	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 65.4 percent of addresses in Table 4-1 with an unknown eligibility. Westat's methodology to estimate the residency status of these cases is consistent with the American Association of Public Opinion Research (AAPOR) standards.³ A good faith effort was made to estimate the percentage of eligible addresses among those whose eligibility was not determined during the recruitment process. 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 households 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 context of response rates, the term "weighted" does **not** refer to the household, person and travel day weights used to expand the NHTS sample to estimates of the U.S. population. In the case of weighted response rates, the weight used is the inverse of the probability of selection, also called the base weight. 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.

³ Cite http://www.aapor.org/Standard_Definitions/1481.htm.

Example of Weighted Response Rate Calculation:

For example, 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, 5 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 percent⁴. 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.

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 100⁵. 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.

$$\begin{aligned} & (10 \times 90 \text{ responding HHs}) + (100 \times 5 \text{ responding HHs}) / \\ & \{ (10 \times 100 \text{ sampled HHs}) + (100 \times 10 \text{ sampled HHs}) \} \\ & = (900 + 500) / (1000 + 1000) = 1400 / 2000 \\ & = 70\% \end{aligned}$$

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.

⁴ Take the number of completes (95) divided by the number of sampled households (110) = 86.4 percent.

⁵ 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 below provides the weighted recruitment response rate for the full sample and each individual Add-on study area. 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.

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%
Des Moines, Iowa	42.4%
Georgia	28.2%
Indian Nations Oklahoma	27.4%
Iowa	37.9%
Maryland	31.1%
North Carolina	31.5%
North Central Texas	25.4%
New York	29.8%
South Carolina	31.2%
Texas	24.4%
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. The 2017 NHTS dataset contains the information collected from households where all household members age five and older completed the retrieval survey.

4.5 Retrieval Survey Response Rate

The retrieval rate is the percentage of recruited households that completed the retrieval survey. Provided in Table 4-3 below 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 above for

weighted screener 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 full sample in 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%
Des Moines, Iowa	56.9%
Georgia	50.2%
Indian Nations Oklahoma	54.5%
Iowa	58.9%
Maryland	51.8%
North Carolina	52.7%
North Central Texas	51.4%
New York	49.5%
South Carolina	48.5%
Texas	48.7%
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.

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%
Des Moines, Iowa	42.4%	56.9%	24.1%
Georgia	28.2%	50.2%	14.2%
Indian Nations Oklahoma	27.4%	54.5%	14.9%
Iowa	37.9%	58.9%	22.3%
Maryland	31.1%	51.8%	16.1%
North Carolina	31.5%	52.7%	16.6%
North Central Texas	25.4%	51.4%	13.1%
New York	29.8%	49.5%	14.8%
South Carolina	31.2%	48.5%	15.2%
Texas	24.4%	48.7%	11.9%
Wisconsin	40.3%	58.5%	23.6%

Weight Calculations

5

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 includes 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.

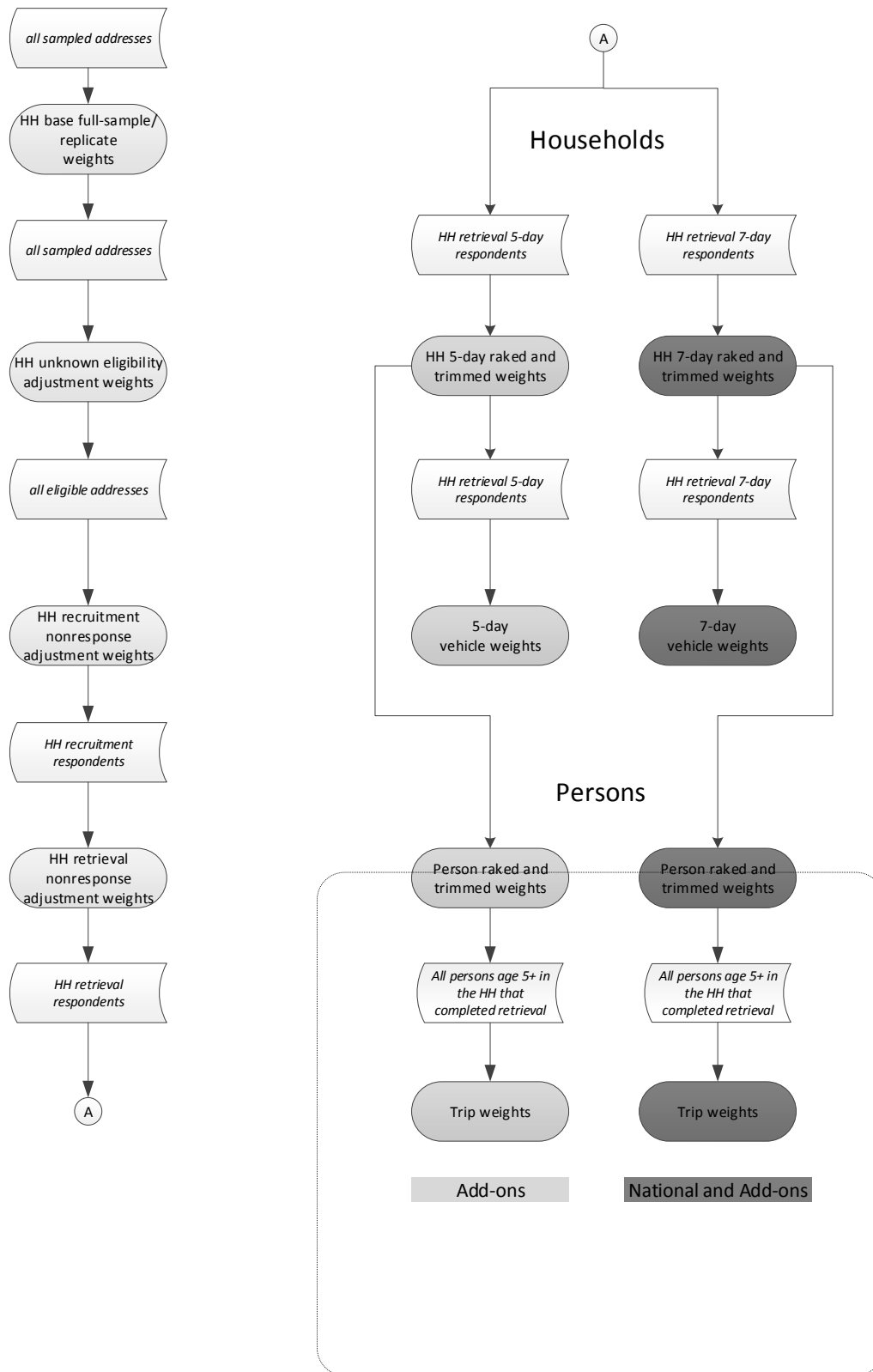
5.1 Initial Household Weight

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.

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 households from the 2015 American

Exhibit 5-1. Flowchart of NHTS weighting procedures



Community Survey (ACS). 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 differed⁶. A search algorithm was used to define a set of heterogeneous cells with as different response rates as possible. This allowed for the maximum possible adjustment of the biasing effects of nonresponse, while providing larger adjustments to households with a larger propensity to not respond (so that these households are not underrepresented because of nonresponse). The search algorithm generated a “tree” of response cells by 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

⁶ 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.

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 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, we adjusted the retrieval nonresponse adjusted weights 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, we raked to also achieve equal distribution among each of the seven days of the week and by each month of the year. The 2017 NHTS Weighting Report provides details on the eight dimensions used.

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.

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 household person weight multiplied by 365.

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.

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 data⁷. 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.

⁷ 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

6.1 Basic Data File Structure

The 2017 NHTS public use data files were developed to emulate the 2001 and 2009 NHTS programs. The four data files, **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, DBF, SAS7BDAT and SAV formats.

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

HOUSEHOLD

- Household characteristic (e.g. household size)

VEHICLE

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

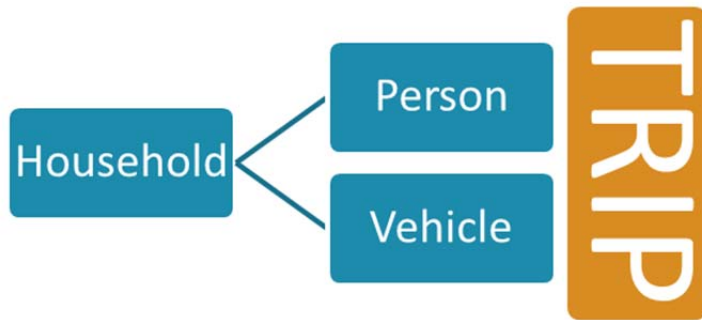
PERSON

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

TRIP

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

Figure 6-1. Data file relationship



6.3 When is a Record on the File?

HOUSEHOLD: A record in this file represents a household unit where all residents aged 5+ 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 table.

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 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 Elements Documentation, which are available in web (html) and Excel (xlsx) formats. These documents are typically referred to as "codebooks" or "dictionaries" and 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 the household, person, vehicle and trip tables. The 2017 NHTS documentation/codebook are interactive and include several pieces of information for each variable including the complete question and answer text. The Variable Documentation table below lists the metadata available in Data Elements with a description and an actual example.

Variable Documentation

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}
		-8 = I don't know
		-7 = I prefer not to answer
Value Label or Range	Answer label text for variable codes	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
		-8 = 1,728
		-7 = 36,993
Value Frequencies – Estimate	Weighted sum of each unique value group in file	01 = 74,518,546
		02 = 42,463,981
		97 = 1,187,002

Data Documentation Files Format

Name	Format	Available Metadata
Data Elements	HTML	All Available
Data Elements	XLSX	All Available except Derived
Derived Variables	PDF	Derived

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. Each data file uses the following coding schemes for consistency.

Special Values

Variable Value	Label	Extended Description
-1	[Appropriate Skip]	Selected by system when the instrument's ask-if logic is not met and no value exists
-9	[Not Ascertained]	Selected by system when the instrument's ask-if logic is met 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 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 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 edited or imputed.

6.7 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. 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 for transportation modelling purposes. 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 2017 Derived Variables document should be referenced when using any of the derived trip purposes in your analysis.

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

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 household travel survey data. However, these may also be useful to the transportation planning professional because the use of certain travel 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 study participants) 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).

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

Online Resources:

Official NHTS Website (<http://nhts.ornl.gov>)

- 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

The website 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.

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

In addition to the tools offered by the official NHTS Website, Westat has also 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 above for download instructions, interactive examples, and more.

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.

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	276,095	321,418,820
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 (Table 7-2). The user should use these as marginal checks when tabulating the NHTS data.

Table 7-2. Total sample size and weight sums for selected key travel-related data

Data	Sample Size	Weighted Sum
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)*	5,560,750	2,015,973,725,687

*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 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. Estimates of the totals are obtained 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) are used 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) 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) are used 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) are generally used 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
- 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	HHWGT1-98	WTPERFIN1-98	DAYWGT1-98

*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:

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

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 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 invaluable for checking to insure that the user's calculations show the same frequency as in the dataset. As discussed earlier the codebook or data elements documentation for the 2017 NHTS will be an online tool to facilitate the quick identification of variables.

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:

$$\text{HOUSEID} + \text{PERSONID} + \text{TDTRPNUM} = \{12345678\}\{01\}\{03\}$$

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 ID Variables Not Always Sequential

The ID variables within a file are not always sequential. There are a number of reasons for this. Examples explaining these reasons were provided in the Data Editing portion of Chapter 2. Some reasons why the numbers are not sequential are:

- Some persons and vehicles reported by the household respondent were later found did 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.

7.17 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	WKRMMH
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