

# **SUPPORTING STATEMENT**

## **Next Generation National Household Travel Survey**

This is a request for an Office of Management and Budget (OMB)-approved clearance for the reinstatement of a periodic information collection entitled “Next Generation National Household Travel Survey” (NextGen NHTS)

### **PART B**

#### **STATISTICAL METHODS**

The NextGen NHTS will be conducted using two independent yet parallel surveys. The main survey will be a probability-based sample of households selected from an ABS frame representative of the entire United States (ABS study). The second data collection effort will survey U.S. households selected from an online probability-based panel which was initially recruited from the same ABS frame. This second effort will be referred to as the Panel Frame Sample Study, or PFS study.

Data collected from the ABS study will serve as the official statistical information, while data and information gathered from the PFS effort will be used to study the validity and feasibility of using the PFS method for future NHTS administrations.

#### **1. Sampling Methodology**

For both the ABS and the PFS work, a nationally representative sample of 7,500 households for each study will be completed to provide statistically valid representation (SVR). These sample sizes will allow the resulting survey estimates from both samples to carry error margins no larger than  $\pm 5\%$  with at least 95% confidence. This will allow the results of the ABS study to estimate FHWA’s required characterization and reporting personal travel behaviors based on SVR samples for key subdomains, including the Nation as well as urban and rural geographies.

The ABS study will complete all surveys with households over the entire 365 days of the year to address all of the previously mentioned analytical requirements. The needed sample of addresses will be selected on a monthly basis to avoid the exclusion of new dwellings. Travel days will be assigned to respondents by asking them to provide travel information for the day prior to completing the household roster.

For the PFS study, the sample will also provide SVR for the Nation as well as urban/rural reporting subdomains. This will allow the generation of comparable metrics for a full assessment of the ABS versus the PFS by an independent third party. The PFS study will also complete all households for the same 365-day period, with samples selected on a monthly basis and with travel days assigned for the day prior to completion of the household roster.

While some fluctuations will be inevitable due to seasonality and other varying yield rates, the application of the regimented sample allocation is expected to secure about 20.55 completed surveys each day for each study (41 total across both the ABS and PFS studies). The sample sizes will be sufficient to roll up the data to various levels of temporal aggregations, such as weeks, months, quarters, and year.

### 1.1 Sample Source

The sample for the ABS study will be randomly selected from the U.S. Postal Service delivery sequence file (DSF).

For non-response follow-up and assessment purposes, each address within the DSF will be geocoded to a specific latitude and longitude and then appended with a long list of ancillary data from commercial databases and the Census Bureau. Residential telephone numbers will also be appended to the DSF to allow for telephone follow-ups with non-responding households.

The panel members in the PFS sample will be randomly selected from within the existing online probability-based panel sampling frame.

### 1.2 Precision of Survey Estimates

The final national (ABS sample) weighted data will allow users to characterize and report personal travel behavior for key domains, including the following:

- National (the entire Nation, including all the 50 States and the District of Columbia combined).
- Urban areas (national combined).
- Rural areas (national combined).

Precision of the survey estimates is a direct function of sample size, universe size, point estimate of interest, and required level of confidence. For example, when estimating population proportions, the needed sample size can be calculated by:

$$n = \frac{Np(1 - p)}{(N - 1)\frac{\epsilon^2}{z^2} + p(1 - p)}$$

In the above formulation,  $N$  is the universe size (often assumed to be infinitely large),  $\epsilon$  is the error bound, and  $z$  is the percentile of the standard normal distribution.

The sampling variances will increase due to design effect, which is approximated by the following formula, in which  $W_i$  represents the final weight of the  $i^{\text{th}}$  respondent:

$$Deff(\delta) = 1 + CV_w^2 = 1 + \frac{\sum_i \frac{(W_i - \bar{W})^2}{n - 1}}{\bar{W}^2}$$

The final PFS sample will be weighted using similar formulas.

## 2. Data Collection Procedures

### 2.1 Survey Communication Protocol

For the ABS study, households will receive an invitation to complete the survey through the mail. The PFS study households will receive an invitation consistent with the online panel protocols. In both surveys, the primary household respondent will complete a short roster to collect key household information (e.g., enumeration of household members and household vehicles). Then, all travel information about a specific day from every household member 5 years of age and older will be collected using the online travel diary or equivalent paper form.

For households choosing to complete the survey online, the primary respondent will complete the household roster and then complete his or her diary as well as serve as a proxy responder for all children ages 5–15 years old in the household. Household members 16 and older will be invited to complete their own online diaries. If they fail to do so in a reasonable amount of time after multiple reminders, the primary household member may be asked to serve as a proxy for non-responding teens and adults in the household. Households electing to complete the survey by mail will be provided equivalent paper forms, with similar proxy-reporting instructions.

Reminders to complete the survey will be sent via mail and email to non-responders and to partial completes. A partial complete is any household that has completed a portion of the recruitment or a portion of a travel diary and/or if some of the household’s travel diaries remain incomplete. Non-respondents and households with partial completes will be sent two reminder mailings/emails to prompt response.

### 2.2 Weighting Methodology

To adjust for nonresponse and other biases associated with data collection, all survey data will be weighted in order to produce reliable estimates of population parameters.

For the ABS study, the weighting process will follow a four-step methodology similar to what was used for the 2017 NHTS. Weights will be generated at the household and person levels. As with the 2017 NHTS, the household weights will apply to the household and household vehicle files, while the person weights will be used to adjust the person and trip files.

1. In the first step, design weights will be computed to reflect selection probabilities. With  $\pi_{hi}$  representing the inclusion probability for the  $i^{\text{th}}$  member in stratum  $h$ , the design weight will be computed by the following equation:

$$DW_{hi} = \frac{1}{\pi_{hi}}, \begin{cases} h=1, \dots, H \\ i=1, \dots, n_i \end{cases}$$

2. In the second step, design weights will be adjusted to correct for observable nonresponse patterns based on sampling frame information available for both respondents and nonrespondents. With  $n_k^i$  representing the responding subset of  $n_k$  members in nonresponse adjustment cell,  $k$ , this weight for the  $r^{\text{th}}$  respondent will be computed by the following equation:

$$NW_{kr} = DW_{kr} \times \frac{\sum_{l=1}^{n_k} DW_{kl}}{\sum_{l=1}^{n_k^c} DW_{kl}}$$

3. In the third step, design weights will be adjusted for nonresponse in cells deemed to be homogeneous. With  $n_k^c$  representing the responding subset of  $n_k$  members in nonresponse adjustment cell,  $k$ , this weight for the  $r^{\text{th}}$  respondent will be computed by the following equation:

$$NW_{kr} = DW_{kr} \times \frac{\sum_{l=1}^{n_k} DW_{kl}}{\sum_{l=1}^{n_k^c} DW_{kl}}$$

4. Finally, nonresponse adjusted weights will be poststratified (raked) to a set of population totals to true up their weighted distributions to their corresponding reported benchmarks. While this adjustment will be carried out iteratively along several distributions, symbolically, the final weight for the  $r^{\text{th}}$  respondent in the poststratification cell,  $p$ , will be computed by the following equation:

$$FW_{pr} = NW_{pr} \times \frac{T_p}{\sum_{r=1}^{n_p} NW_{pr}}$$

Since weighting tends to increase variance of estimates, use of standard variance calculation formulae with weighted data can result in misleading inferences and tests of significance. With weighted data, two general approaches for variance estimation can be distinguished. One method relies on different forms of replication, while a more common approach is Taylor Series linearization. While the 2017 NHTS used balanced repeated replication (BRR) for variance estimation, FHWA will test the differences between BRR and linearization. This change in the approach to variance estimation is consistent with current industry practice.

With respect to trimming the resultant weights, the contractor will use an empirical approach that considers the observed distribution of survey weights. The following approach will be used to determine such values:

- The interquartile range (IQR) of the weights, which is the difference between the third ( $Q_3$ ) and first ( $Q_1$ ) quartiles (75th and 25th percentiles) of the weight distribution, will be computed as follows:

$$IQR = Q_3 - Q_1$$

- The upper ( $U_1$ ) and lower ( $L_1$ ) limits for identifying the so-called mild outliers will be computed as follows:

$$L_1 = Q_1 - 1.5 \times IQR$$
$$U_1 = Q_3 + 1.5 \times IQR$$

- The upper ( $U_3$ ) and lower ( $L_3$ ) limits for identifying extreme outliers will be computed as follows:

$$L_3 = Q_1 - 3 \times IQR$$
$$U_3 = Q_3 + 3 \times IQR$$

When nonresponse is moderate, the above trimming thresholds often correspond to the 1st and 99th percentiles of the weight distribution. However, under excessive nonresponse, more aggressive trimming thresholds might be necessary to improve the weighting efficiency. The unequal weighting effect (UWE) (sometimes referred to as design effect – see above for design effect formula) will also be taken into account.

The above is the factor by which the effective size of a sample is reduced, reflecting the actual number of independent observations from which inferences are made. Under an ideal situation when respondents are self-weighting, this factor would be equal to 1.0. With nonresponse and other coverage issues, however, typical UWE for general population surveys can exceed 2.0 or even higher values. In deciding the final trimming thresholds, the goal will be to keep the final UWE to no more than 2.0.

### **2.3 Nonresponse Bias Analysis**

Both the ABS and PFS methods will include a nonresponse follow-up analysis that includes nonresponse adjustments of design weights. The development of the design weights will rely on a series of comparisons between responding and nonresponding households to identify key differentiators between these two groups. These comparisons will be facilitated by a long list of ancillary data that will be appended to the DSF from public and commercial sources. Findings from these investigations will guide the weighting adjustment processes to reduce biases that can result due to differential nonresponse. Moreover, analysis weights can also receive additional calibration adjustments based on reliable travel statistics that might be available from relevant sources, such as the ACS data, FHWA HPMS data, and other related data sources.

### **2.4 Imputation of Missing Data**

All surveys are subject to missing values, which can result from item nonresponse or when observed values fail edit checks and are then set to missing. For the NextGen NHTS, missing data will be imputed using the weighted sequential hot-deck method of SUDAAN, with the respondent survey data serving as donors to provide surrogate values for records with missing values. The basic principle of this methodology involves construction of homogeneous imputation classes, which are generally defined by cross-classification of key covariates, and then replacing missing values sequentially from a single pass through the survey data within each imputation class.

### 3. Maximizing Response Rate

The NextGen NHTS redesign aims to address response rate challenges through a comprehensive review of all survey elements and possible sources of error (i.e., total survey error). This includes a reduction in survey length by reducing the number of questions asked, having a focused effort on improving the web-interface for the online program, improving the shared trip reporting aspect of the program, and refining survey materials and processes. As a result of this effort:

- Communications around the NextGen NHTS will convey the importance, relevance, and societal value associated with completing the survey.
- The online survey will be optimized for completion via smartphones and other mobile devices. In addition, the online survey will meet Section 508 compliance standards using the rules specified in section 1194.22.
- Incentives will be used to engage respondents and entice completion as follows:
  - ABS study respondents will receive \$5 per completed trip reporting (up to a maximum of \$25 per household).
  - PFS study respondents will be awarded the equivalent of \$2 for completing the household roster and \$5–\$10 equivalents when all householders report their trips (based on household size).
- All NextGen NHTS survey materials will be available in both English and Spanish to help maximize participation by Hispanic households.

### 4. Material Review and Testing

All field materials have been tested to assess their utility and validity as survey instruments. These tests are described in the following subsections.

#### 4.1 Expert Review

The first step in developing a revised questionnaire and field materials for the NextGen NHTS was to conduct a review of existing materials to identify potential improvements. In the review process, a questionnaire design expert conducted a thorough appraisal of the questionnaire, providing guidance on refinement and further development activities factoring in the new and emerging modes of transit. The design expert judged features of the instrument, such as ease and efficiency of completion, avoidance of errors, ability to correct mistakes, and how the content is presented (overall appearance and layout of the instrument).

#### 4.2 Cognitive Testing

Cognitive testing is routinely used for survey questionnaire development<sup>1,2,3</sup> and to determine question comprehension (e.g., what do respondents think the question is asking? What do specific words and phrases in the question mean to them?); information retrieval (e.g., what do respondents need to recall to answer the question? How do they do this?); decision processes

<sup>1</sup>Forsyth, B. and Lessler, J. (2011). *Cognitive Laboratory Methods: A Taxonomy*. 10.1002/9781118150382.ch20.

<sup>2</sup>DeMaio, T.J. and Rothgeb, J.M. (1996). "Cognitive interviewing techniques: In the lab and in the field." In N. Schwarz and S. Sudman (Eds.), *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, pp. 177–195. Jossey-Bass, San Francisco, CA.

<sup>3</sup>Groves, R., Fowler, F., Couper, M., Lepkowski, J., & Singer, E., and Tourangeau, R. (2004). *Survey Methodology*, First Edition, p. 561, Wiley & Sons, Hoboken, NJ.

(e.g., how do respondents choose their answers?); and usability (e.g., can respondents complete the questionnaire easily and as they were intended to?).

The NextGen NHTS cognitive testing process focused on confirming core data survey question wording made sense in the online survey environment, new survey questions on emerging research topics were appropriately worded, and the reporting of trips were aided by programming features. Questions included the following:

- Are respondents knowledgeable of all travel made by all members of the household? Can they accurately report household travel behavior by proxy or directly?
- Are all trips—including loop trips and short trips—understood, recalled, and reported adequately, directly, and by proxy?
- How frequently do households have online purchases delivered to their home location?

A cognitive testing approach was executed to improve the survey and reduce measurement error. Cognitive testing of the NextGen NHTS protocol consisted of conducting nine individual online interviews with participants who had demographic and other key characteristics similar to potential survey respondents. Interviewers engaged participants in one-on-one, open-ended cognitive interviews in which they probed a participant's approach and thinking when responding to questions within the context of the full survey administration. Cognitive interviews covered both the household roster and trip reporting for both self and proxy reporting.

The results of these tests were used to modify survey questions and response options to limit misunderstandings and inconsistencies in responses. The modified survey instrument was used to create the final set of survey questions.

## **5. Independent Analysis of Results**

Data collected from the ABS study will serve as the official statistical information, while data and information gathered from the PFS effort will be used to study the validity and feasibility of using the PFS method for future NHTS administrations. The independent evaluation of results will include the following:

**5.1 Survey Design Questions** (focusing on Core Data Elements): how do the distributions of core data elements compare across the 2 studies and with 2017 NHTS?

- How many households participated in each study? How many persons, vehicles, and trips per household were reported?
- How does reported travel behavior differ across the two studies with respect to travel metrics such as person and vehicle trip rates, PMT, VMT, trip durations, trip purposes, mode split, and time of day of travel. (overall, urban/rural, and for specific household types)
- In reviewing the validation performed by the survey contractor, summarize how the resultant travel metrics for each study compare to outside data sources (such as HPMS, NTD, HMPS, VMT, etc.).
- How do the distributions of core data elements compare across studies and with respect to 2017 NHS responses?
- What is the statistical valid representation (SVR) at the national, urban, and rural geographies for core data elements (each study and 2017 NHTS)?

- What are the key differences in results between the two studies?

### **5.2 Response Rate Evaluation**

- Using the appropriate AAPOR calculators, what is the cumulative response rate for each study? How do these compare to the 2017 NHTS?
- What were the attrition rates for each study, overall and for each survey stage?

### **5.3 Sample evaluation** (focusing on national, urban, and rural samples) – assessment of specific types of bias and implications of identified bias on the overall representativeness of results.

- Coverage bias – by design and based on fielding results, what evidence of coverage bias exists in the data for each study? (what populations are not included in the sampling frame?)
- Non-response bias - How does the distribution of key respondent characteristics differ across the two studies and as compared to both decennial and ACS Census data (based on availability at the time of the evaluation)?
- Item non-response bias - How well did each survey perform in terms of item non-response on the core data elements? How does this compare to the 2017 NHTS?
- Attrition bias - How does the presence and composition of 100% completed households compare across the study? What type of and how many households reported partial data (per the study definition) and what type of bias does this introduce?

### **5.4 Weighting evaluation**

- Overall, how did the weighting processes differ between the two studies? How does this compare to 2017 NHTS?
- What variables were used in the weighting process in each study? How well did the weights correct the biases in the data?
- Does application of the weights force key metrics to trend in an unexpected or unintended pattern?
- Was trimming of weights required/conducted in each study? To what extent was trimming of weights performed? How does this compare to 2017 NHTS?

### **5.5 Overall**

- How do the study results compare with respect to providing a representative sample?
- What are the pros and cons of conducting the NextGen NHTS using probability panels (including representativeness, completeness of responses, cost, and quality of estimates produced)?
- How transparent and reproducible was the documentation for each study? Are there any details unclear and/or missing?
- How well do the results support the continuity and trend analysis of the NHTS?

## **6. Contact Information**

Daniel Jenkins, PE  
National Travel Behavior Data Program Manager  
Federal Highway Administration  
[daniel.jenkins@dot.gov](mailto:daniel.jenkins@dot.gov)





## **APPENDICES**

1. Federal Register 60-Day Notice
2. Federal Register 30-Day Notice
3. Federal Register Second 30-Day Notice
4. Comments Submitted to Docket for all Notices
5. ABS Study Materials
6. PFS Study Materials
7. NextGen NHTS Questionnaire
8. Transportation Research Circular No. E-C238, August 8–9, 2018