

2017 Local Area Transportation Characteristics for Households

Methodology

BACKGROUND

The National Household Transportation Survey (NHTS), a survey of the U.S. Department of Transportation's Federal Highway Administration (FHWA), assesses the mobility of the American public. The NHTS gathers data on daily personal travel, including information on household and demographic characteristics, employment status, vehicle ownership, trips taken, modal choice, and other related transportation data pertinent to U.S. households. This survey is a continuation of the Nationwide Personal Transportation Survey (NPTS), which was conducted in 1969, 1977, 1983, 1990 and 1995. The NHTS was conducted in 2001, 2009, and 2017. The 2017 NHTS collected travel data from a national sample of civilian, non-institutionalized population of the United States - 26,099 households in the national sample - and separate samples from 13 add-on areas, which together provided data on 129,696 households. Data collection in 2017 differed from that in earlier years, because it used address-based sampling rather than random digit-dialing (as was the case in the 1990, 1995, 2001, and 2009 surveys) to obtain survey responses. The change to address-based sampling enabled the inclusion of cellphone-only households – a group excluded in previous years. The 2017 NHTS also utilized both online data and phone data collection instead of phone only as in previous years. Address-based sampling also allows for better geographic controls to ensure the geographic distribution of the sample since phone numbers are no longer tied to a geographical area like they once were.

The NHTS is an excellent source of travel information for large geographic areas in the U.S., but its limited sample sizes for small areas makes it a less suitable source for small geographic areas. To address this issue, the Bureau of Transportation Statistics (BTS) developed a model that allows for small area estimation using the NHTS data along with American Community Survey (ACS) data from the Census Bureau. The model divides the NHTS data into six geographic areas, classifies these areas as urban/suburban/rural, and then estimates average weekday household: person miles traveled, person trips, vehicle miles traveled, and vehicle trips for each geographic area. The division of the data into geographic areas follows from research by Henson and Goulias (2001), which shows that people in different geographic areas travel differently even if they share the same socio-demographic characteristics. The BTS model then transfers the estimates to individual Census tracts using the household and demographic data from the ACS for each Census tract. The resulting Census tract estimates provide beneficial indicators to local governments and other customers who may not have the budget and/or time for conducting their own local survey. Additionally, the use of a standard set of questions across all geographies in the NHTS enables comparison across geographies that otherwise would be captured in separate local surveys with potentially different methodologies.

The BTS model used to produce Census tract level estimates from the 2017 NHTS replicates the BTS model used with the 2009 NHTS with minor modifications meant to improve the quality of the results. The changes made include the following:

- Elimination of variables from the model measuring the same concept. Specifically, BTS removed the count of household members from the model, because the categorical life-cycle variables (e.g., whether the household is 1 person with 0 persons less than 65, etc.) already capture the size of the household.
- Use of categorical variables to capture variation in household vehicle ownership and number of workers better than the previously used statistical average (e.g., whether there are 0, 1, or 2 plus workers in the household rather than average calculated by dividing the total number of workers in a Census tract by the total number of households)
- Identification of Census tracts with less reliable estimates. Specifically, BTS flagged Census tracts where the margin of error from the 2012-2016 American Community Survey 5-year estimates is larger than the estimate for socio-economic and demographic characteristics used in the model.

Full details on the model used with the 2009 NHTS can be found in the 2009 Local Area Transportation Characteristics by Household methodology along with detail on previous studies and related transferability research.¹ The following describes the methodology used to produce Census tract estimates from the 2017 NHTS.

METHODOLOGY

The BTS model divides the NHTS data into six geographic areas (based on Census region/division boundaries) and classifies these areas as urban/suburban/rural. The division of the data into geographic areas follows from research by Henson and Goulias (2001), which shows that people in different geographic areas travel differently even if they share the same socio-demographic characteristics.

Census Region/Division Groups

The 2017 NHTS sample of 129,696 is large enough for division of the data into six geographic regions (based on Census region/division boundaries) and then into three urban groups, for a total of 18 separate categories. The model produces estimates for each category separately. The geographic disaggregation enables more homogenous groupings of the households for the model. The selected groupings follow Census divisions except where the NHTS sample size is too small for estimation at the Census division level. In those instances, the model uses the Census region instead. The geographies selected are as follows:

1. Northeast Region
2. Midwest Region

¹ Local Area Transportation Characteristics by Household data and details can be found at: <https://www.bts.gov/statistical-products/surveys/local-area-transportation-characteristics-households-latch-survey>

3. South Atlantic Division
4. East South Central Division and West South Central Division
5. Mountain Division
6. Pacific Division.

Figure 5 shows the States in each of the above Census regions/divisions.

Development of Urbanicity Index

BTS classifies Census tracts as urban/suburban/rural using a method similar to the one used by Nielsen Claritas, Inc. to create the urban-rural continuum included in the 2009 NHTS.² The classification uses information on the population density of a Census tract (converted to a centile score) and on whether the Census tract is in an urban area or urban region/division. Because the 2017 NHTS uses the 2010 Census boundaries, the classification uses the 2010 Census tract and urban boundaries.³ Data on population come from the 2012-2016 American Community Survey 5-year estimates – the dataset with tract-level population data closest in year to the data collection year of the 2017 NHTS. Table 1 shows the assignment of Census tracts to the following categories: urban, suburban, and rural.

Table 1. Urbanicity Index: Count of 2010 Census Tracts by Category

Category	Census Tracts	With population density centile	Number of Tracts
Urban	In UAs	60 to 100	28,356
	In UCs	30 to 100	2,729
	Total (urban)		31,085
Suburban	In UAs	Greater than 0 and less than 60	18,173
	In UCs	Greater than 0 and less than 30	598
	Total (suburban)		18,771
Rural	Not in an UA or UC	N/A	22,563
	Total (rural)		22,563
No population (but land area)			320
Total			72,739

NOTES:

There are 317 Census tracts defined in the 2017 Tigerline Census Tract files with no land area. There are 320 tracts with land area but with no population in the 2017 Tigerline Census Tract files.

Urbanized Areas (UAs) are areas with 50,000 or more people; Urban Region/divisions (UCs) are areas with at least 2,500 and less than 50,000 people. Census Tracts in UAs are those with their centroid in an UA; Census Tracts in UCs are those with their centroid in an UC.

² For more information about the variables created by Nielsen Claritas, Inc. and included in the 2009 NHTS, see “Tract and Block Group Variables” in the *2009 NHTS Derived Variables Description*, available at: <https://nhts.ornl.gov/2009/pub/DerivedAddedVariables2009.pdf>

³ Census Tract boundaries obtained and urban boundaries obtained from the 2016 Census TIGER/Line Shapefiles.

Density centile was calculated by sorting Census Tracts with a population greater than 0 by their population density in ascending order and then assigning a score from 0 to 100 to each Census Tract according to this order. The Census tract with the smallest population density was assigned a score of 0 and the Census tract with the largest population density was assigned a score of 100.

Table 2 shows the number of 2017 NHTS households in each Census region/division and urban group. In the 2017 NHTS, the Census region/division information of the household did not match, in some cases, with the state identifier for the household. For example, the NHTS identified the household as belonging to the Pacific Division but listed a state other than one in the Pacific Division. Communication with the NHTS data collection team reveals this resulted from sample protocol treatment of households that moved during the data collection period for the NHTS. Westat – the firm that collected and processed the NHTS data – kept the weighting and Census data linked to the original sampled location, not the actual one reported on the travel day, and updated all other variables (including the state identifier) to the home location reported on the travel data collection day. Some households sampled in the NHTS actually moved; others appear to be at their vacation “home” on the travel data collection day; and some appear to be on a trip and list their temporary residence as their “home”. Households’ travel behavior likely changes at their vacation home or on a trip. For this reason, the BTS model excludes NHTS households where the Census division/region did not match the state identifier, i.e., households possibly on vacation or other trip on the travel data collection day.

Table 2. Breakdown of NHTS Sample by Census Region/Division and Urban Group

Census Region/Division	Urban	Suburban	Rural	Total
Northeast Region	5,871	5,865	8,625	20,361
Midwest Region	5,838	5,699	8,404	19,941
South Atlantic Division	4,449	11,293	12,982	28,724
East South Central & West South Central Divisions	11,780	7,748	7,881	27,409
Mountain Division	1,806	1,113	2,204	5,123
Pacific Division	15,998	4,630	6,993	27,621
Total	45,742	36,348	47,089	129,179

NOTE: Excludes households in Manhattan, NY and where the 2017 NHTS Census region/division information of household did not match with state identifier.

Mean and Confidence Intervals of Travel Variables

The objective of dividing the NHTS households into the 18 groups is to improve the accuracy of the estimates from the model. One way to assess the groupings is to look at the differences in means and confidence intervals for each travel variable (person miles traveled, person trips, vehicle miles traveled and vehicle trips). Figure 1 through 4 shows the means and confidence intervals for household travel estimated from the 2017 NHTS. The figures show considerable variation, both across geographical divisions and between urban groups. In particular, the Mountain Division shows the most uncertainty in mean person miles traveled, person trips, vehicle miles traveled, and vehicle trips. This results partly

from the smaller NHTS sample size in the Mountain Division. The other regions/divisions have NHTS samples sizes 2.5 or more than the Mountain Division. The Mountain Division sample consists of the national plus add-on NHTS sample. Unlike other regions/divisions, the NHTS add-on sample size is small (table 3).

Figure 1. NHTS Mean Person Miles Traveled and 90% Confidence Intervals

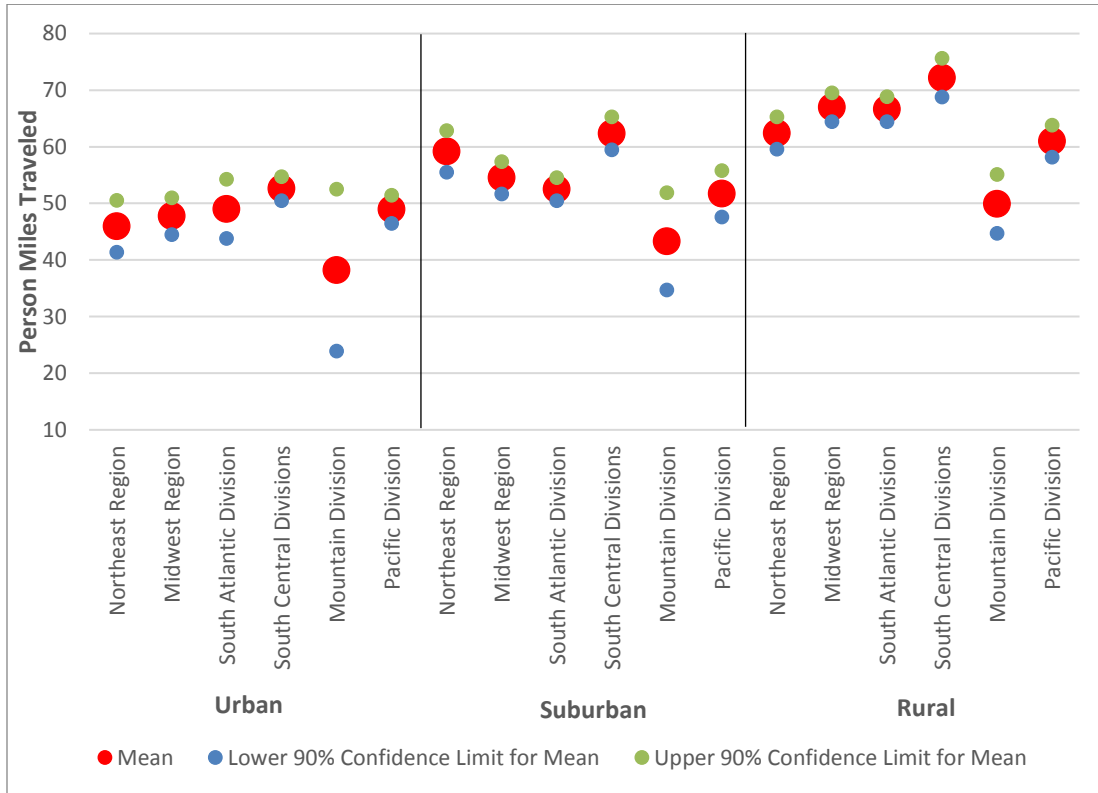


Figure 2. NHTS Mean Person Trips and 90% Confidence Intervals

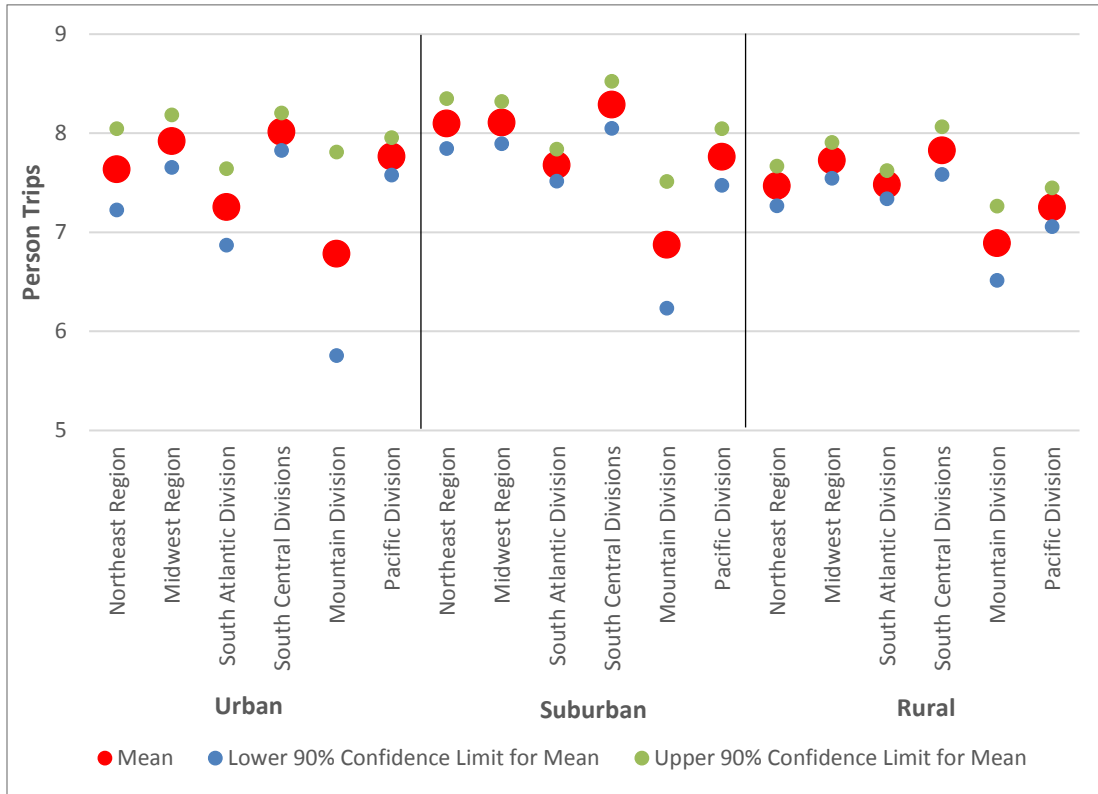


Figure 3. NHTS Mean Vehicle Miles Traveled and 90% Confidence Intervals

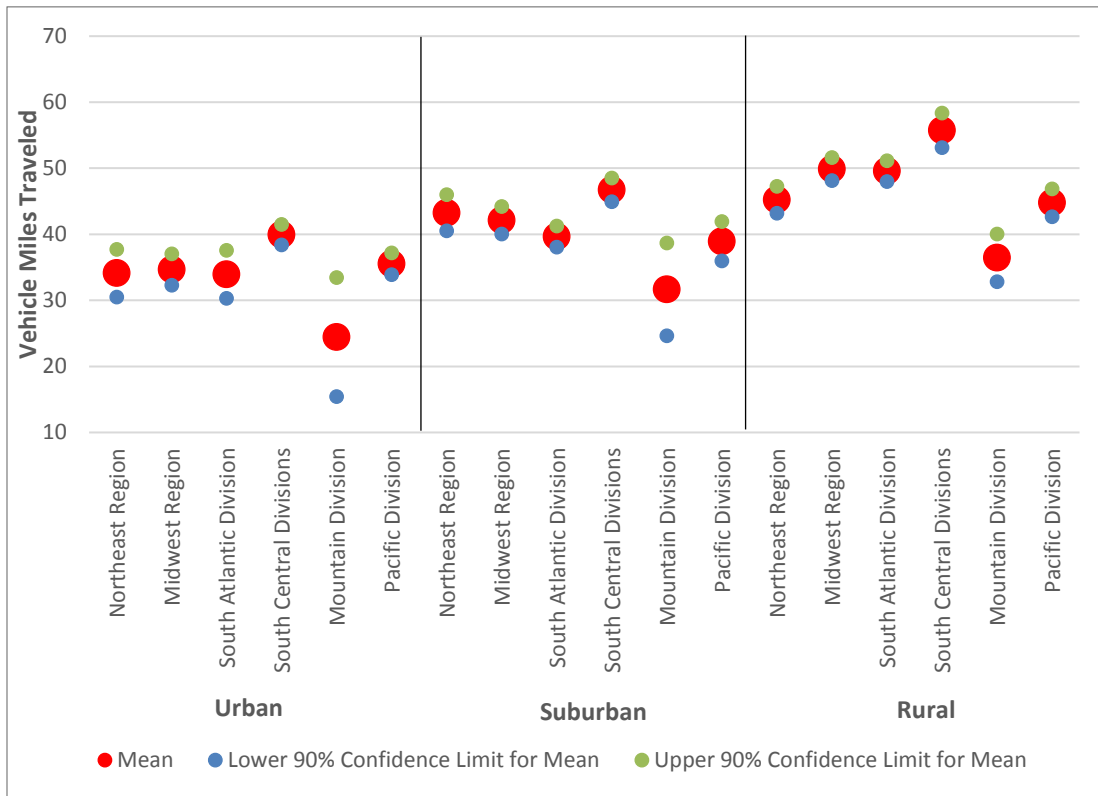


Figure 4. NHTS Mean Vehicle Trips and 90% Confidence Intervals

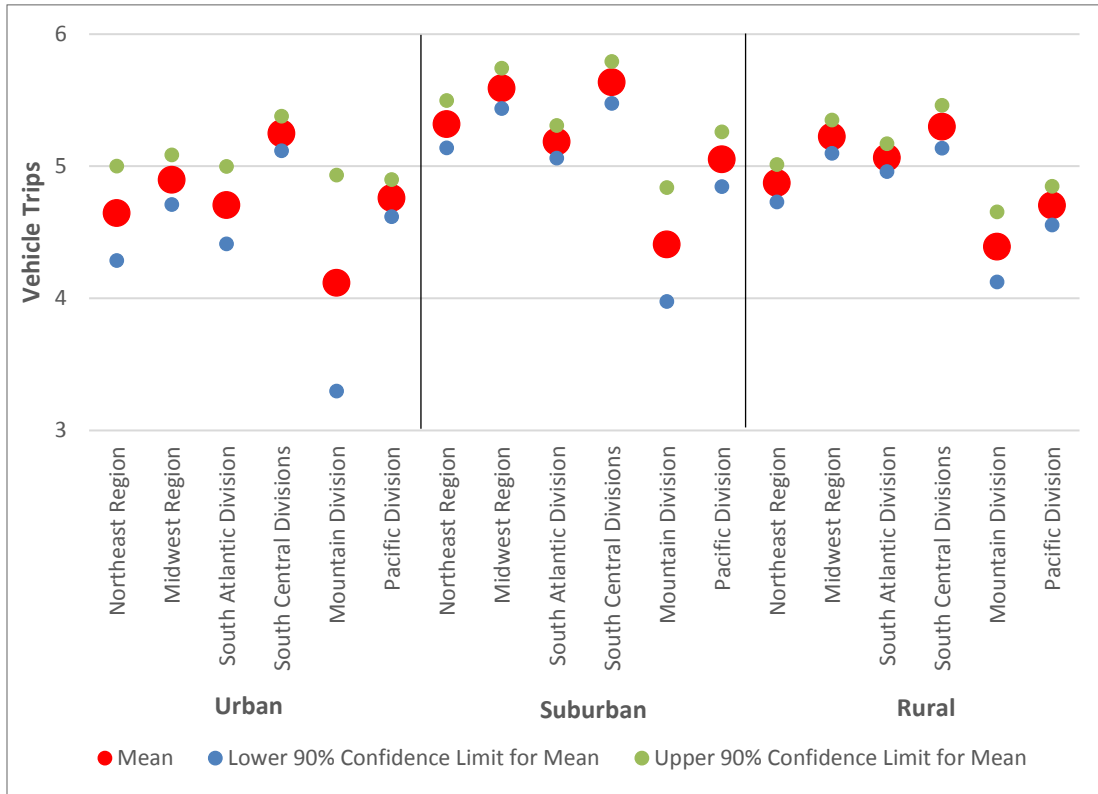


Table 3. 2017 NHTS Add-on Partners

Census Region/Division	NHTS Add-on Partners	Target Number of Households
Northeast Division	New York State	15,851
Midwest Region	Wisconsin	11,000
Midwest Region	Des Moines Area MPO in Iowa	1,200
Midwest Region	Iowa Northland Council of Governments in Iowa	1,200
South Atlantic Division	Georgia	8,000
South Atlantic Division	Maryland	1,000
South Atlantic Division	North Carolina	8,000
South Atlantic Division	South Carolina	6,500
East South Central & West South Central Divisions	Texas	20,000
East South Central & West South Central Divisions	Indian Nations Council of Governments in Oklahoma	1,000
East South Central & West South Central Divisions	North Central Texas Council of Governments in Texas	2,917
Mountain Division	Arizona	2,444
Pacific Division	California	24,000

SOURCE: Add-on partners and target number of households from *2017 NHTS User Guide*, available at: <https://nhts.ornl.gov/assets/2017UsersGuide.pdf> as of November 2018.

Definition of Travel Variables and Exclusion of Certain Households

The model includes only weekday travel. This follows the methodology used in the previous 2001 NHTS Transferability Project (Hu et al. 2007) and is a common assumption in urban planning models. As in the previous study, the model excludes all households sampled in the NHTS in Manhattan, New York, due to the unique travel patterns of that area. The model also excludes suspected outliers - approximately 1 percent of the trips in the upper tail of each distribution.

Person trips include all trips, except that by airplane. Vehicle trips include trips using cars, vans, SUVs, pickup trucks, other trucks, RVs, motorcycles, and light electric vehicles. It includes only trips taken by the driver of the vehicle. Household trips represent the sum of all trips taken by members of the household.

Four travel variables were estimated from the 2017 NHTS:

1. Total household person miles traveled, excluding outliers defined as > 500 miles
2. Total household number of person trips, excluding outliers defined as > 30 trips
3. Total household vehicle miles traveled, excluding outliers defined as > 310 miles
4. Total household number of vehicle trips, excluding outliers defined as > 20 trips

Total household vehicle miles traveled were calculated using NHTS trip distance (in miles) derived from route geometry.

For each, observations exceeding the 95-99th quantile were identified as suspected outliers and excluded. The NHTS data showed a natural break at the 95-99th quantile. For example, the maximum total vehicle miles traveled for a household was 11,147 miles while the 99th quantile was 313 miles. The large gap between the two quantiles led to identifying and removing observations exceeding 310 miles.

Explanatory Variables

The selection of explanatory variables used in the analysis relied partially on previous work in the 2001 NHTS Transferability Study (Hu et al. 2007) and partially on an examination of other NHTS household variables available. The examination included the requirement that comparable data be available in the Census American Community Survey (ACS) public data tables at the Census tract level. This became a significant constraint in developing the life-cycle household variables. The NHTS defined life-cycle variables do not have equivalent counterparts in the ACS data tables. As a result, BTS developed alternative life-cycle variables that could be used with the available ACS data tables. The final set of explanatory variables used in the model include:

1. Household income [**HH Income**] BTS developed this variable by converting the household income categories in the NHTS data to a point estimate, using the mid-point of each category range. For the last category, household income above \$200,000, more detailed Census household income tables were used to derive a weighted average of \$250,000 for that category.⁴ Household income is the best available proxy for household wealth, which is assumed to be the primary driver of discretionary travel expenditure.
2. Household vehicles, 0 available [**HH Vehicles (0 available)**] BTS developed this variable from the count of vehicles owned by a household in the NHTS data. BTS converted the NHTS count into a series of categorical variables to match the data available in the ACS.

⁴ U.S. Bureau of Census, Current Population Survey, Table HINC-01, Selected Characteristics of Households, by Total Money Income in 2016. <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.2016.html>

3. Household vehicles, 1 available [**HH Vehicles (1 available)**] BTS developed this variable from the count of vehicles owned by a household in the NHTS data. BTS converted the NHTS count into a series of categorical variables to match the data available in the ACS.
4. Household vehicles, 2+ available [**HH Vehicles (2+ available)**] BTS developed this variable from the count of vehicles owned by a household in the NHTS data. BTS converted the NHTS count into a series of categorical variables to match the data available in the ACS.
5. Workers in household, 0 workers [**Workers in Household (0 workers)**] BTS developed this variable from the count of workers in a household in the NHTS data. BTS converted the NHTS count into a series of categorical variables to match the data available in the ACS.
6. Workers in household, 1 worker [**Workers in Household (1 worker)**] BTS developed this variable from the count of workers in a household in the NHTS data. BTS converted the NHTS count into a series of categorical variables to match the data available in the ACS.
7. Workers in household, 2+ workers [**Workers in Household (2+ workers)**] BTS developed this variable from the count of workers in a household in the NHTS data. BTS converted the NHTS count into a series of categorical variables to match the data available in the ACS.
8. Life-Cycle, 1 or more children in household, less than 18 years old in the NHTS data but greater than or equal to 5 years old [**Life Cycle (5<=1+C<18)**] BTS developed this variable from the age of the respondent in the NHTS data. The NHTS collects trip information for persons greater than or equal to 5 years old.
9. Life-Cycle, 1 person household, less than 65 years old in the NHTS data [**Life Cycle (1P hh<65)**]
10. Life-Cycle, 2 or more person household, all less than 65 years old in the NHTS data [**Life Cycle (2+P hh, 0 65+)**]
11. Life-Cycle, 2 or more person household, at least one 65 or more years old in the NHTS data [**Life Cycle (2+P hh, 1+65+)**]

The model used to develop Census tract estimates from the 2009 NHTS also included the count of household members and an indicator of an owner-occupied household. The model that uses the 2017 NHTS data does not include the count of household members. BTS dropped the variable, because the life-cycle variables already capture the size of the household. Inclusion of the count of household members introduces multicollinearity from two variables measuring the same attribute. The model that uses the 2017 NHTS data also does not include the indication of an owner-occupied household. In reviewing the model and examining the literature, BTS found that home-ownership does not directly affect travel behavior. Rather, homeownership is a proxy for household income/wealth, which

significantly affects travel behavior. Because the model already includes household income, BTS saw no reason to include homeownership.

The model that used the 2009 NHTS data additionally used counts of household vehicles and workers from the NHTS and the average number vehicles and workers per household from the ACS. BTS calculated the average by dividing the aggregate number of vehicles owned by households in a Census tract by the total number of households, and likewise, dividing the aggregate number of workers in a Census tract by the total number of households. These averages potentially hide variation within a Census tract. For instance, a Census tract may have many households without a vehicle and the same number of households with two or more vehicles. The average would be one vehicle per household, even though all households own either no vehicle or two plus vehicles. Given that the average potentially hides household variation, the model presented here converts the NHTS counts to categorical variables. The categorical variables are those available in the ACS and better capture potential household variation.

REGRESSION ESTIMATION

The model uses multiple linear regression to estimate the relationship between each dependent variable (person miles traveled, person trips, vehicle miles traveled, and vehicle trips) and the aforementioned explanatory (independent) variables from the 2017 NHTS (see table 4). Households are the unit of observation.

Table 4. Independent and Dependent Model Variables from 2017 National Household Travel Survey

Dependent Variable	Independent Variables		Type
		Name	
Transportation variable			
Person miles traveled,	∝	Household income	Continuous
Person trips,		Household vehicles available (1 available)	Y/N
Vehicle miles traveled, or		Household vehicles available (2+ available)	Y/N
Vehicle trips		Workers in household (1 worker)	Y/N
		Workers in household (2+ workers)	Y/N
		Life cycle (5<=1+C<18)	Y/N
		Life cycle (1P hh<65)	Y/N
		Life cycle (2+P hh, 0 65+)	Y/N
		Life cycle (2+P hh, 1+65+)	Y/N

NOTE: Households are the unit of observation. Household vehicles available (0 available); workers in household (0 workers); life cycle (0 C<18); and life cycle (1P hh 65+) are the omitted indicator variables.

The model created using the 2009 NHTS data dropped insignificant independent variables. Independent variables may become insignificant when correlated with other independent variables. Dropping the insignificant ones can result in omitted variable bias. Because the literature shows each of the selected independent variables as predictors of travel behavior, the model presented here keeps them all, and notes in the results, the insignificant variables.

The results of the model (the regressions) are in Appendix A.

BTS explored principal component analysis as a refinement to the model when using the 2009 NHTS. Principal component analysis offered no improvement to the model using the 2009 NHTS. As a result, BTS did not attempt to use principal component analysis for the 2017 NHTS.

Validation

BTS evaluated the prediction accuracy of the models in two ways.

1. BTS evaluated the overall prediction accuracy of the regression models by comparing the mean household person miles traveled, person trips, vehicle miles traveled, and vehicle trips in each Census region/division and urban group to the estimated value. To calculate the estimated value, BTS calculated the mean value for each of the independent variables from the NHTS. For example, BTS calculated median household income in the Northeast urban area. BTS then inserted the means calculated for each independent variable into the regression equations. Table 5 compares the estimated value to the mean value calculated directly from the NHTS sample living in each Census

region/division and urban group. Most regression estimates are within a 90% confidence interval of the mean values. The model predicts higher mean person miles traveled for households in rural areas in most Census region/divisions than estimated from the NHTS. This may be a result of the low explanatory power of the model for person miles traveled. Specifically, the independent variables in the regression model for person miles traveled explains the least amount of variability in person miles traveled, while the independent variables in the regression models for person trips tends to explain more of the variability in person trips (i.e., the models for person trips have a higher r -squared than the models for person miles traveled).

Table 5. Comparison NHTS Mean Values and Regression Estimates

Census Region/Division Category and Urban Group		Household Person Miles Traveled				Household Number of Person Trips			
		Mean	90 percent confidence interval		Mean Regression Estimate	Mean	90 percent confidence interval		Mean Regression Estimate
			Lower	Upper			Lower	Upper	
Northeast Region	Urban	46.0	41.4	50.6	38.7	7.1	6.7	7.5	7.1
Northeast Region	Suburban	59.2	55.5	62.9	58.0	7.6	7.3	7.8	7.7
Northeast Region	Rural	62.4	59.6	65.3	67.7	7.0	6.8	7.2	7.0
Midwest Region	Urban	47.7	44.5	51.0	42.5	7.4	7.2	7.7	7.2
Midwest Region	Suburban	54.5	51.7	57.4	54.5	7.6	7.4	7.8	7.8
Midwest Region	Rural	67.0	64.4	69.5	72.9	7.2	7.0	7.4	7.5
South Atlantic Division	Urban	49.0	43.8	54.3	42.1	6.8	6.4	7.1	6.5
South Atlantic Division	Suburban	52.5	50.5	54.6	52.8	7.2	7.0	7.3	7.2
South Atlantic Division	Rural	66.6	64.4	68.9	67.2	7.0	6.8	7.1	7.0
South Central Divisions	Urban	52.6	50.5	54.7	52.7	7.5	7.3	7.7	7.9
South Central Divisions	Suburban	62.4	59.5	65.3	63.7	7.8	7.6	8.0	8.2
South Central Divisions	Rural	72.2	68.8	75.7	80.6	7.3	7.1	7.6	7.8
Mountain Division	Urban	38.2	23.9	52.5	38.6	6.3	5.3	7.3	6.5
Mountain Division	Suburban	43.3	34.7	51.9	49.0	6.4	5.7	7.0	6.9
Mountain Division	Rural	49.9	44.7	55.1	61.2	6.4	6.0	6.8	7.5
Pacific Division	Urban	48.9	46.4	51.4	48.7	7.3	7.1	7.5	7.1
Pacific Division	Suburban	51.7	47.6	55.8	55.9	7.3	7.0	7.5	7.4
Pacific Division	Rural	61.0	58.2	63.8	60.8	6.8	6.6	6.9	7.2

NOTE: Values in bold are outside of the 90 percent confidence interval. Mean regression estimate calculated by inserting the mean values for each of the independent variables in the regression equation for the specified Census region/division and urban group.

Table 5. Comparison NHTS Mean Values and Regression Estimates (Continued)

Census Region/Division Category and Urban Group		Household Vehicle Miles Traveled				Household Number of Vehicle Trips			
		Mean	90 percent confidence interval		Mean Regression Estimate	Mean	90 percent confidence interval		Mean Regression Estimate
			Lower	Upper			Lower	Upper	
Northeast Region	Urban	34.1	30.5	37.7	25.1	4.6	4.3	5.0	3.9
Northeast Region	Suburban	43.3	40.5	46.0	43.1	5.3	5.1	5.5	5.5
Northeast Region	Rural	45.2	43.2	47.3	50.6	4.9	4.7	5.0	4.8
Midwest Region	Urban	34.7	32.3	37.1	30.6	4.9	4.7	5.1	4.8
Midwest Region	Suburban	42.1	40.1	44.2	42.2	5.6	5.4	5.7	5.4
Midwest Region	Rural	49.9	48.1	51.6	52.6	5.2	5.1	5.3	5.5
South Atlantic Division	Urban	33.9	30.3	37.6	31.5	4.7	4.4	5.0	4.4
South Atlantic Division	Suburban	39.6	38.0	41.3	39.0	5.2	5.1	5.3	5.1
South Atlantic Division	Rural	49.6	48.0	51.1	51.3	5.1	5.0	5.2	5.0
South Central Divisions	Urban	39.9	38.4	41.5	40.0	5.2	5.1	5.4	5.4
South Central Divisions	Suburban	46.7	44.9	48.5	47.6	5.6	5.5	5.8	5.9
South Central Divisions	Rural	55.7	53.1	58.3	61.4	5.3	5.1	5.5	5.7
Mountain Division	Urban	24.4	15.4	33.4	26.8	4.1	3.3	4.9	4.3
Mountain Division	Suburban	31.6	24.6	38.7	33.0	4.4	4.0	4.8	4.6
Mountain Division	Rural	36.4	32.8	40.0	40.6	4.4	4.1	4.7	5.0
Pacific Division	Urban	35.5	33.9	37.2	34.2	4.8	4.6	4.9	4.5
Pacific Division	Suburban	38.9	36.0	41.9	39.7	5.1	4.8	5.3	4.9
Pacific Division	Rural	44.8	42.7	46.9	44.1	4.7	4.6	4.8	5.0

NOTE: Values in bold are outside of the 90 percent confidence interval. Mean regression estimate calculated by inserting the mean values for each of the independent variables in the regression equation for the specified Census region/division and urban group.

2. BTS evaluated the linear regression models for their prediction accuracy at the Census tract level. BTS compared the mean number of person miles traveled, person trips, vehicle miles traveled, and vehicle trips to the number calculated from the corresponding regression model. BTS used the non-public 2017 NHTS files to calculate the mean value of the four household travel variables in each Census tract with data. BTS calculated the predicted values in two ways:

(1) By calculating the mean value for each independent variable⁵ from the non-public NHTS dataset, and inserting them into the appropriate regression equation for that Census tract, and

(2) By inserting the value extracted from the ACS dataset for each independent variable into the appropriate regression equation for each Census tract (see Table 6 for an example).

⁵ The NHTS final household weight was used as a weight in the linear regression models to predict household travel but was not used to weight the mean household characteristics for a given Census tract since the NHTS final household weight was not intended to make households within a Census representative of the Census tract itself.

Table 6. Validation Process: Prediction of Person Miles Using American Community Survey Data (ACS)

Variable	Regression parameter (A)	Value for Census Tract 36001001400 (from 2012-2016 ACS) (B)	(A) * (B) (C)
Intercept	8.20	NA	8.20
Median household income (thousands)	0.13	44.6350	5.70
1 vehicle available	3.17	0.4737	1.50
2 or more vehicles available	21.06	0.1669	3.51
1 worker household	2.49	0.5986	1.49
2 or more worker household	12.50	0.2142	2.68
Life Cycle (1+C<18)	6.60	0.0354	0.23
Life Cycle (1P hh < 65)	2.49	0.6375	1.59
Life Cycle (2+P hh, 0 65+)	11.00	0.2472	2.72
Life Cycle (2+P hh, 1+65+)	10.21	0.0531	0.54
Person miles traveled			28.16

NOTE: Column B is a proportion from 0 to 1 for all except median household income

BTS compared the predicted values to the NHTS values in all Census tracts where at least eight or more households were surveyed for the NHTS. This size requirement, developed in conversations with some researchers of the previous NHTS study, provides greater confidence in representing a given Census tract. However, requiring more than eight households reduces the number of Census tracts that can be evaluated for their prediction accuracy. See Tables B1 to B4 in Appendix B for the count of Census tracts in each Census region/division and urban group that had eight or more households and the necessary data for making an accuracy assessment.

The predicted values came from a model using NHTS data in each Census region/division. Individual households have a small influence on the predicted values. In other words, the predicted values are sufficiently independent from the values estimated directly from individual households in a Census tract. Independence is a requirement for data comparisons when attempting to perform an accuracy assessment. Complete independence could be obtained by splitting the NHTS sample into two parts, with one part used to calculate the predicted values and the other part used to calculate values directly from individual households. This requires re-weighting of the NHTS sample. BTS did not split or re-weight the NHTS data given sufficient independence of the predicted values from those estimated directly from individual households.

For the Census tracts where a comparison could be made, BTS calculated the absolute percent error between the NHTS value and the predicted value in each Census tract and then calculated the mean of these errors to arrive at the mean absolute percent error (MAPE) in each Census region/division and urban group (see Tables B1 to B4 in Appendix B). The models for vehicle trips and person trips tend to predict better than the models for vehicle miles and person miles, as they show much smaller MAPEs across all Census region/divisions and urban groups. The MAPE tends to be larger in Census

region/divisions and urban groups in all models where the medians for the independent variables from the NHTS dataset are significantly different from the medians from the ACS dataset for all Census tracts included in the evaluation (see Tables B5 to B8 in Appendix B). In particular, the average percent of households with a child calculated from the NHTS is lower in all Census regions/divisions than the average percent calculated from the ACS. This results from the NHTS counting children greater than or equal to 5 years of age (the NHTS collects trip information only for respondents greater than or equal to 5 years of age). The ACS tables provide a count of households with a child less than 18 but no further age-subdivision to match the data to the NHTS data. These differences in the mean values are a major contributor to the larger MAPEs.

Travel Variable Estimates by Census Tract

Examination of the MAPEs revealed no issues with the models. The MAPEs were small in most areas. Areas where the MAPEs were larger are areas where NHTS and the ACS characteristics for households are different. For these areas, it is unclear which data accurately characterize the households.

After validating the model through an examination of the MAPEs, BTS produced estimates of the four household travel variables for all Census tracts in the U.S., with the exception of Census tracts in Manhattan, using the 2012-2016 ACS.⁶

BTS identified the urban group for all Census tracts in the 2012-2016 ACS dataset per the method described in the above section on the development of the urbanicity index using 2010 Census boundaries and population information. A list of the data pulled from the 2012-2016 ACS data files can be found in Table 7. BTS used the ACS data to predict household travel. BTS evaluated the estimates of household travel spatially and non-spatially for reasonability. The following describes the reasonability tests.

Table 7. Independent Variable Derivations from 2012-2016 American Community Survey

Variable	Formula (ACS Table ID Variable)
Household Income (Thousands \$)	B19013e1 HD01_VD01
1 vehicle available in household	B08201 HD01_VD04
2 or more vehicles available in household	B08201 HD01_VD05, HD01_VD06, HD01_VD07
1 worker household	B08202 HD01_VD04
2 or more workers in household	B08202 HD01_VD05, HD01_VD06
Life Cycle (1+ child <18)	B11005e2 HD01_VD01
Life Cycle (1 person household, <65)	B11007e8 HD01_VD08
Life Cycle (2+ person household, 0 65+)	B11007 HD01_VD09
Life Cycle (2+ person household, 1+ 65+)	B11007 HD01_VD04

NOTE: Data pulled from American Fact Finder Download Center.

Spatial identification of unreliable estimates

Prior to testing for spatial reasonableness, BTS examined the estimates of household travel, since the quality of the spatial analysis can be compromised by extreme values. BTS defined extreme estimates as those less than the 1st percentile and greater than the 99th percentile of the NHTS mean value in a given Census region/division and urban group. No estimates were identified as extremes per these criteria.

BTS examined spatial reasonableness by comparing household travel estimates for a Census tract to those of its neighbors. Spatial theory suggests locations near to each other are usually more similar than data from locations far away from each other. In analyzing trip distribution in urban areas, Mazzulla and

⁶ Census tracts in Manhattan were identified from the New York City Department of Urban Planning: (<http://www.nyc.gov/html/dcp/html/bytes/dwndistricts.shtml#cbt>) and were suppressed given the significant difference in travel behavior between those living in Manhattan and those outside of Manhattan but still in the same Census region/division.

Forciniti (2012) found clusters of similar values among neighboring census parcels and spatial dependence in the data.

BTS defined neighboring tracts as those sharing an edge or a corner with the Census tract being evaluated for spatial reasonableness. If the household travel estimate for one or more of the four variables was significantly lower or higher than that of neighboring tracts, the significantly different estimate was considered spatially unreliable. This was performed by first using ArcGIS to identify the neighbors of each Census tract and then by calculating the Moran's I statistic for each Census tract in SAS.

The Moran's I statistic is a spatial statistic that tells how much a feature is similar or dissimilar to its neighbors. Features that are similar to their neighbors have large, positive Moran's I statistics. Features that are dissimilar to their neighbors have large, negative Moran's I statistics. Here, being dissimilar is used to mark spatially unreliable estimates as it is assumed that features near one another should exhibit similar travel patterns. The formula for the Moran's I statistic used to identify these dissimilar values can be found in figure 6.

Figure 6. Moran's I Statistic

$$I_i = \frac{(x_i - \bar{x})}{s^2} * \sum_j w_{ij}(x_j - \bar{x})$$

Where x_i is the target feature value, s^2 is the variance, w_{ij} is the weight for the target feature and neighbor pair, x_j is the neighbor value, and \bar{x} is the mean. The variance and the mean are calculated from the estimated values for the same Census region/division and urban group as the target feature.

As shown in the formula (figure 5), the Moran's I statistic involves the use of the overall mean (\bar{x}) in comparing a Census tract (x_i) to its neighbors (x_j). Since the models were developed specific to a Census region/division and urban group, the mean for the Census region/division and urban group were used in the formula rather than the overall mean. The Moran's I statistics were evaluated for statistical significance by calculating the z-score. Negative Moran's I statistics with statistically significant z-scores at the 99% confidence interval belong to estimates that are dissimilar from surrounding values. These estimates are marked as being spatially unreliable. In each Census division/region and urban group, 13 or less Census tracts were marked as such and suppressed (Appendix C).

Identification of non-spatial outliers

After testing for spatial reasonableness, BTS evaluated the estimates further for reasonableness by examining the distribution of the estimates. All estimates for all variables were found to be within the range of values for the same travel variable used to create the regression model. The bottom and top 0.5 percent of the distribution of the estimate were suppressed to tighten the range of the estimates and eliminate possible outliers.

Identification of less precise estimates

As a final quality check, BTS created a variable to identify Census tracts where the ACS margin of error for one or more of the independent variables in the model exceeds the ACS estimate. To reduce the number of Census tracts where the margin of error exceeds the estimate, BTS combined several ACS categories to create a new variable with a smaller margin of error. Households with 2 or more vehicles available and households with 2 or more workers both are variables created by BTS. BTS calculated the number of households with 2 or more vehicles by combining the following ACS categories: (1) households with 2 vehicles available, (2) households with 3 vehicles available, (3) households with 4 vehicles available, and (4) households with 5 or more vehicles available. Likewise, BTS calculated the number of households with 2 or more workers by combining the following ACS categories: (1) households with 2 workers and (2) households with 3 or more workers. BTS calculated the margin of error for these combined variables using the Census Transportation Planning Products Margin of Error Toolkit.⁷ BTS performed the margin of error calculations in SAS using the formula for the weighted generalized variance function presented in the toolkit. BTS includes the margins of error as part of the final dataset.

See Appendix C for final counts and distribution of the estimates after completion of all reasonableness checks and identification of less precise estimates.

Estimates by Household Size and Number of Vehicles Available

After producing estimates and identifying and removing outliers, BTS estimated average household person miles traveled, person trips, vehicle miles traveled and vehicle trips by household size and number of vehicles available in each Census tract. Estimates by household size and number of vehicles available could not be calculated from the model using the NHTS and ACS data because of insufficient detail in the ACS data. The ACS data are summary tables. Therefore, we know the percent of households with zero vehicles available in a Census tract, but we do not know the characteristics of those zero vehicle households, such as median income of those zero vehicle households, the number of workers, etc.

BTS developed a new linear regression model to calculate household travel by household type using the estimated average household person miles traveled, person trips, vehicle miles traveled and vehicle trips by household size and number of vehicles available and the summary ACS data for each Census region/division and urban group. Census tracts were the unit of observation in the new model. The

⁷ See [http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36\(135\)_TutorialCTPPMOEToolKit.pdf](http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(135)_TutorialCTPPMOEToolKit.pdf)

dependent variable in the model was the estimated person miles traveled, person trips, vehicle miles traveled, and vehicle trips. The independent variables were the household types for which BTS estimated household travel - the percent of households by household size (e.g., percent of two person households in a Census tract) and the percent of households by number of vehicles available (e.g., percent of households with only one vehicle available) (see table 8). Because the goal is to predict household travel by household size and number of vehicles available, BTS included no other independent variables. Data on household size and number of vehicles available by Census tract come from the 2012-2016 American Community Survey 5-year estimates.

Table 8. Independent and Dependent Model Variables

Dependent Variable (as estimated from model with NHTS and ACS data)	∝	Independent Variables (from 2012-2016 American Community Survey 5-year estimates)
Estimated		
Person miles traveled,	∝	Percent of 2 person households
Person trips,		Percent of 3 person household
Vehicle miles traveled, or		Percent of 4 or more person household
Vehicle trips		Percent of 1 vehicle available
		Percent of 2 vehicles available
		Percent of 3 vehicles available
		Percent of 4 or more vehicles available

NOTE: Census tracts are the unit of observation. Percent of one person households and percent of households with zero vehicles available are the omitted indicator variables.

Using the new regression models, BTS predicted average household person miles traveled, person trips, vehicle miles traveled, and vehicle trips for each of the following household types:

1. One person households with zero vehicles available
2. One person households with one vehicle available
3. One person households with two vehicles available
4. One person households with three vehicles available
5. One person households with four or more vehicles available
6. Two person households with zero vehicles available
7. Two person households with Two vehicle available
8. Two person households with two vehicles available
9. Two person households with three vehicles available
10. Two person households with four or more vehicles available
11. Three person households with zero vehicles available
12. Three person households with Three vehicle available
13. Three person households with two vehicles available
14. Three person households with three vehicles available
15. Three person households with four or more vehicles available
16. Four or more person households with zero vehicles available

17. Four or more person households with Four or more vehicle available
18. Four or more person households with two vehicles available
19. Four or more person households with three vehicles available
20. Four or more person households with four or more vehicles available

BTS obtained an estimate for each of the above household types directly from the results of the new regression model for each Census region/division and urban group. Table 9 shows how to estimate average household person miles traveled by various household types in the urban areas of the Northeast Region. Appendix E contains the regression model results for all Census region/division and urban groups.

Table 9. Estimation Process by Household Size and Number of Vehicles Available by Census Region/Division and Urban Group for Selected Household Types

Variable	Northeast Region-Urban Area				
	Regression parameter (A)	One person household zero vehicles available * (A)	One person household with 1 vehicle available * (A)	Two person household with 1 vehicle available * (A)	Two person household with 2 vehicles available * (A)
Intercept	7.27	7.27	7.27	7.27	7.27
2 person household (percent)	32.91	0 * 32.91	0 * 32.91	1 * 32.91	1 * 32.91
3 person household (percent)	25.19	0 * 25.19	0 * 25.19	0 * 25.19	0 * 25.19
4 or more person household (percent)	27.41	0 * 27.41	0 * 27.41	0 * 27.41	0 * 27.41
1 vehicle available (percent)	3.93	0 * 3.93	1 * 3.93	1 * 3.93	0 * 3.93
2 vehicles available (percent)	28.98	0 * 28.98	0 * 28.98	0 * 28.98	1 * 28.98
3 vehicles available (percent)	30.31	0 * 30.31	0 * 30.31	0 * 30.31	0 * 30.31
4 or more vehicles available (percent)	33.14	0 * 33.14	0 * 33.14	0 * 33.14	0 * 33.14
Person miles traveled		7.27	11.2	44.11	69.16

To calculate Census tract level estimates, BTS inserted the percent of each household type in each Census tract into the appropriate regression equation (see Table 10 for an example). Data on the percent of each household type (e.g., percent of one person households with zero vehicles available) are from the 2012-2016 American Community Survey 5-year estimates (the same data used as independent variables in the model). Insertion of the ACS data results in an estimate of average household person miles traveled, person trips, vehicle miles traveled, and vehicle trips for each Census tract. BTS used this estimate and calculated the percent difference from the estimate obtained from first set of linear regression models (Appendix A) that used the NHTS data. BTS used the percent difference to transfer

the Census region/division and urban group estimates by household type to the Census tract level. For example, if the percent difference for person miles traveled was 19.2 percent in Census tract A in the urban area of the Northeast Region, then average person miles traveled for one person households with zero vehicles available in the urban Northeast Region was increased by 19.2 percent to obtain the value for Census tract A. Table 11 shows this process.

Table 10. Process for Obtaining Average Household Person Miles Traveled from Model by Household Size and Number of Vehicles Available

Variable	Regression parameter (A)	Value for Census Tract 36001001400 (from 2012-2016 ACS) (B)	(A) * (B) (C)
Intercept	7.27	NA	7.27
2 person household (percent)	32.91	0.2595	8.54
3 person household (percent)	25.19	0.0269	0.68
4 or more person household (percent)	27.41	0.0138	0.38
1 vehicle available (percent)	3.93	0.4737	1.86
2 vehicles available (percent)	28.98	0.1461	4.23
3 vehicles available (percent)	30.31	0.0161	0.49
4 or more vehicles available (percent)	33.14	0.0046	0.15
Person miles traveled			23.61

NOTE: Column B is a proportion from 0 to 1

Table 11. Estimation Process for Average Household Person Miles Traveled by Selected Household Types

Estimated person miles traveled from model using LATCH estimates and ACS data (Table 10) = 23.61

Estimated person miles traveled from model using NHTS and ACS data (Table 6) = 28.16

Difference in estimated person miles traveled = $(28.16 - 23.61) / 23.61 = 19.2$ percent

Difference calculated for each Census tract and used to transfer Census region/division estimate for each household type to the Census tract.

	Value for Census tract 36001001400				
	Calculated difference (proportion)	One person household zero vehicles available (Table 9) * (1 + (A))	One person household with 1 vehicle available (Table 9) * (1 + (A))	Two person household with 1 vehicle available (Table 9) * (1 + (A))	Two person household with 2 vehicles available (Table 9) * (1 + (A))
Variable	(A)				
Person miles traveled	0.192	7.27 * 1.192 = 8.67	11.20 * 1.192 = 13.35	44.11 * 1.192 = 52.58	69.16 * 1.192 = 82.44

RESULTS, CHALLENGES, AND CONCLUSIONS

Census tract estimates of average household person miles traveled, person trips, vehicle miles traveled and vehicle trips are available in state-by-state flat files, as well as in a SAS data file (the format of the SAS file is given in Appendix D). The estimates are a per day average. The files can be found on the BTS website: <https://www.bts.gov/statistical-products/surveys/local-area-transportation-characteristics-households-latch-survey>. Maps of the four travel variables are in Figures 7 to 10 and available on the BTS website.

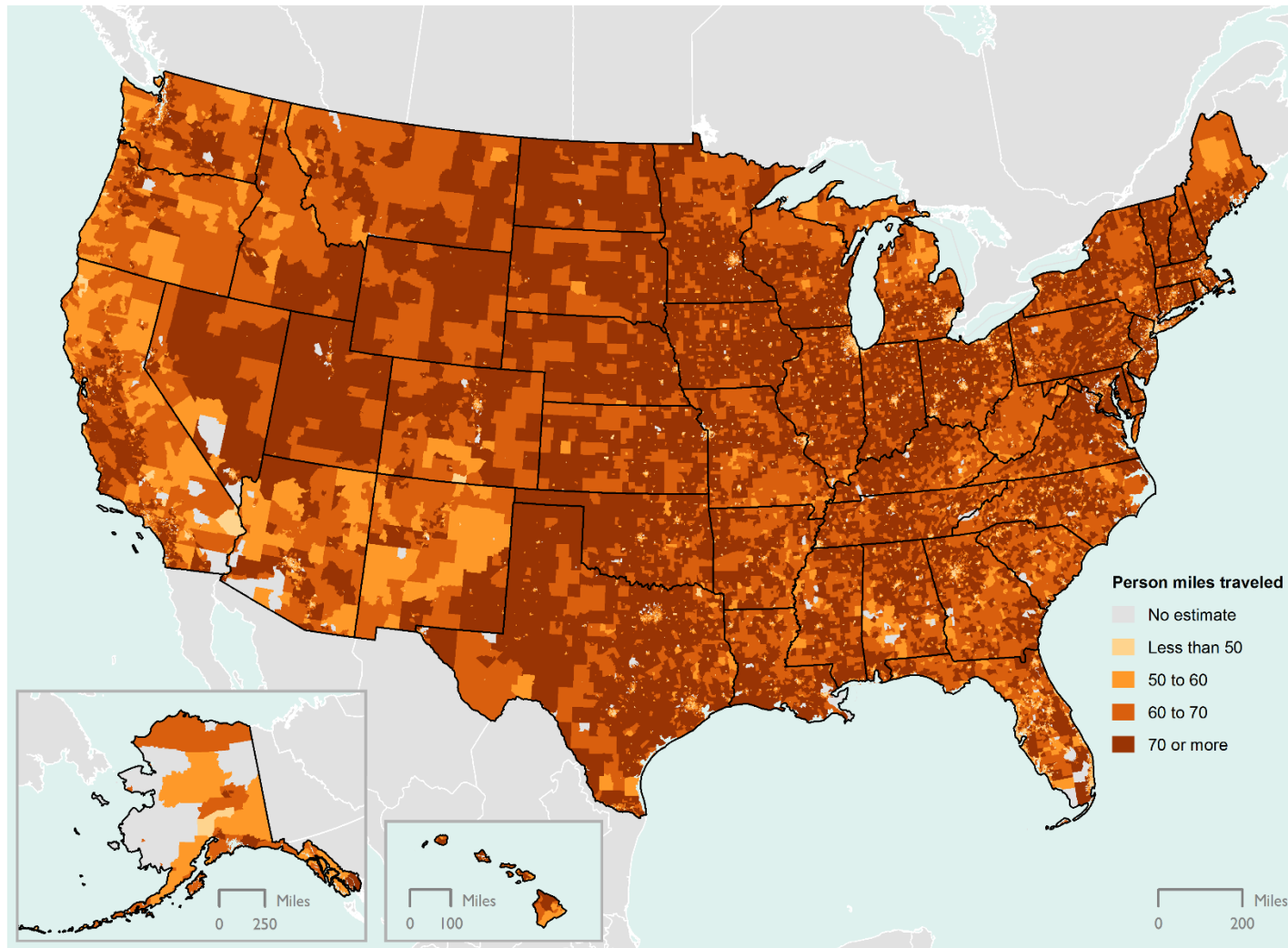
The resultant household per day averages of the key transportation measures, by region and by urbanicity, provide assurance as to the quality of the regressions employed. For example, as expected for the mean of person miles traveled (Appendix C1), the urban person miles are the lowest (as compared to suburban and rural) for each region, and the order of mileage for each region is consistently urban lowest, then, suburban, and finally rural with the highest. The Northeast Region has the smallest urban person miles at an average of 39.6 miles. The Pacific Division, not surprising, has the longest urban person miles – at an average of 54.9, but not the longest average of rural (the Pacific Division had a rural average of 68.2). The longest rural person miles are in the Northeast and East/West South Central Divisions, at an average of 73.8 miles.

Since the results for passenger trips are represented by a count variable, the averages are not as dispersed as the person miles traveled results (Appendix C2). The trips range from 7.5 (for South Atlantic Division urban) to 9.0 (Mountain Division rural). For the Mountain Division, rural trips are the highest (9.0); but, for the remaining regions, the suburban trips are the greatest.

Looking at vehicle miles (Appendix C3), the lowest average mileage of 22.0 is again for the Northeast urban; the highest, 53.2 miles, is for the Northeast rural. For all regions/division, the average urban mileage is less than the suburban, which is, in turn, less than the rural mile average.

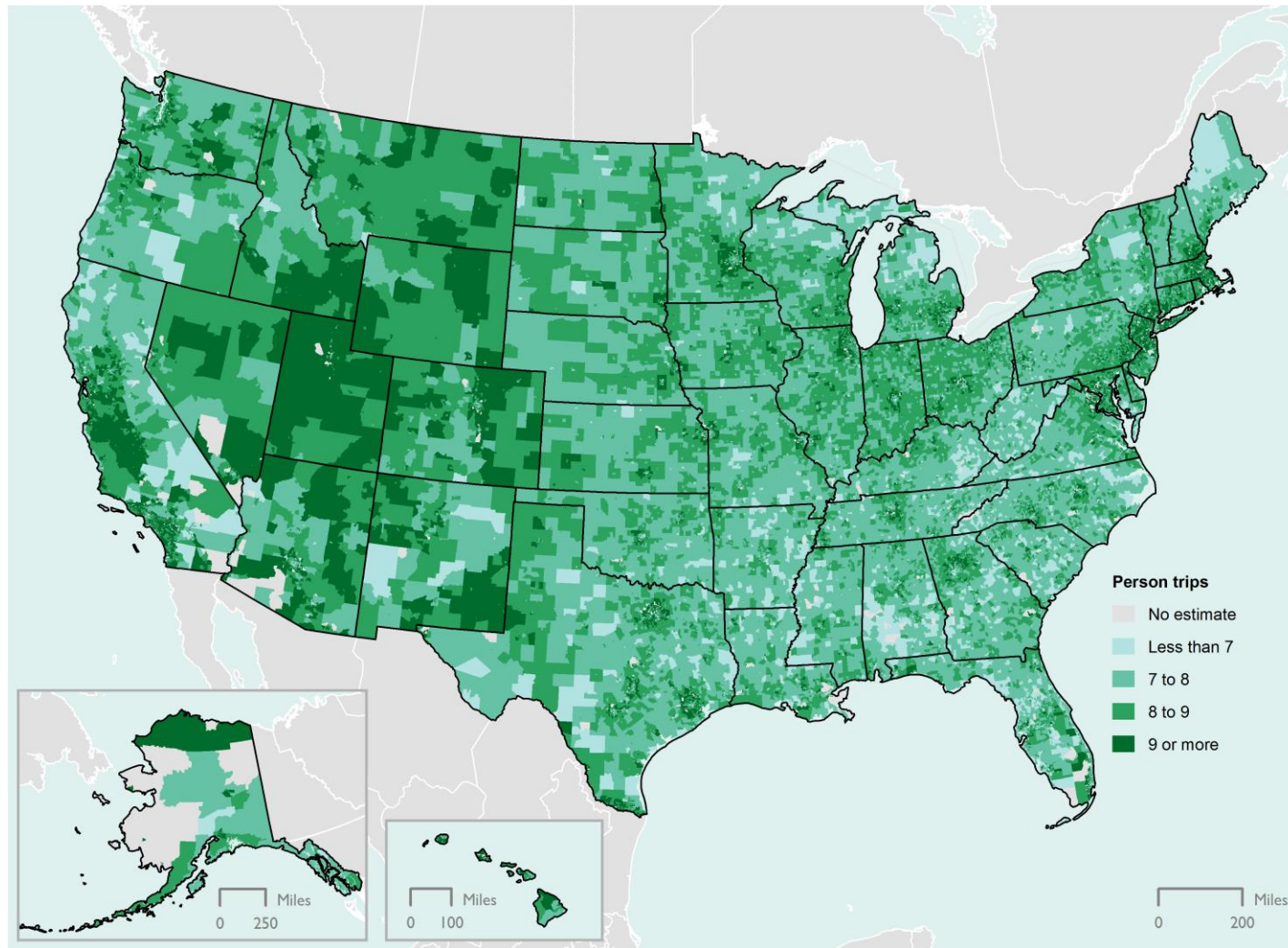
Lastly, the vehicle trip estimates (Appendix C4) differ slightly from the person trips. The lowest average number of trips is for the urban area in the Northeast Region (3.6), whereas the highest vehicle trips average is for the rural Mountain Division (5.9). In all cases, urban areas have the lowest average number of vehicle trips, but in most regions/divisions, the number of suburban vehicle trips is higher than the number of rural trips. Only two rural areas were higher than the suburban areas, the Midwest Region and the Mountain Division (in the Pacific Division, rural vehicle trips equaled suburban).

Figure 5. Average Weekday Household Person Miles Traveled by Census Tract (per day)



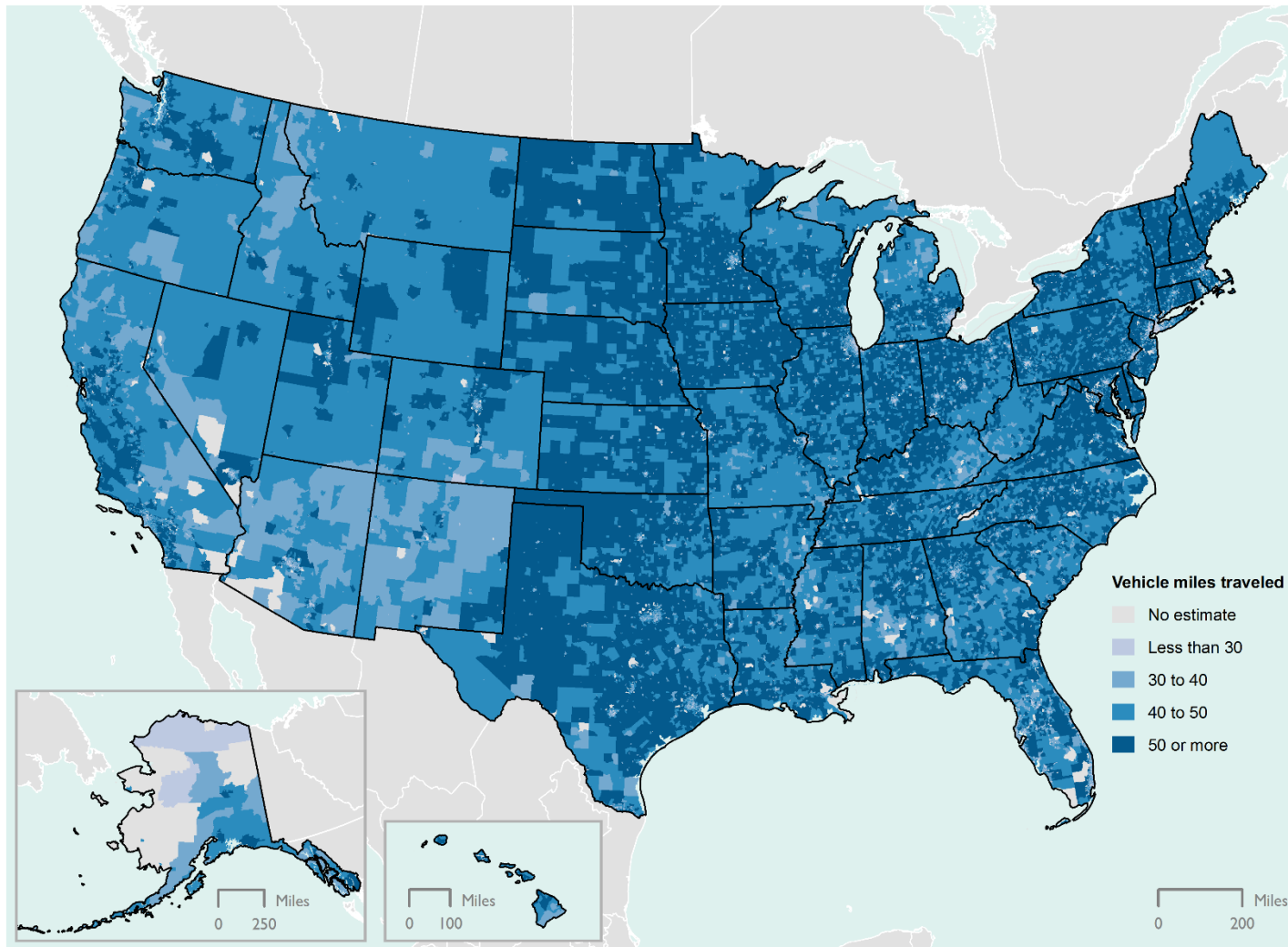
NOTE: Areas with no estimate are areas where there is no population or areas where one or more of the model parameters are not available.

Figure 6. Average Weekday Household Person Trips by Census Tract (per day)



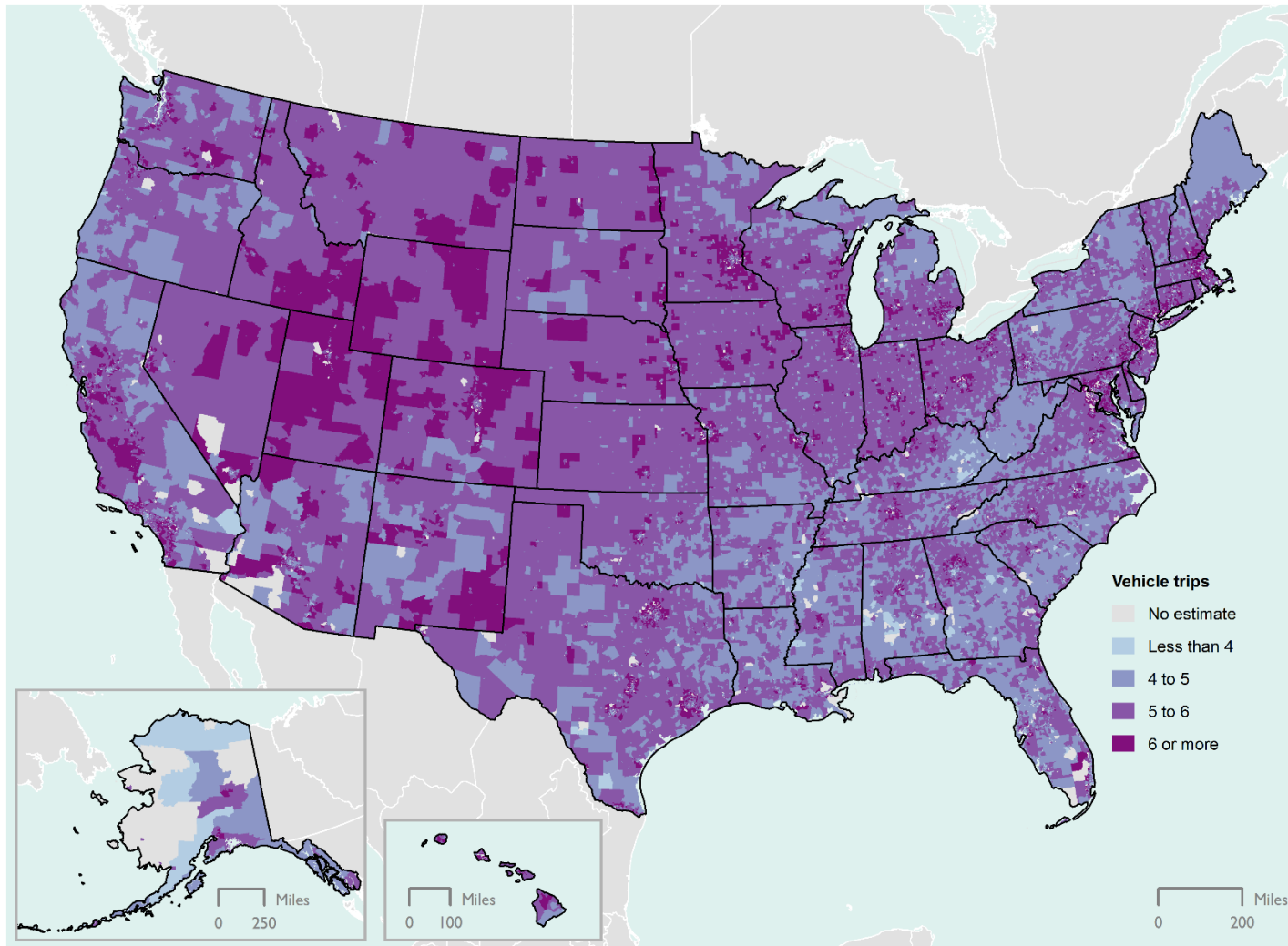
NOTE: Areas with no estimate are areas where there is no population or areas where one or more of the model parameters are not available.

Figure 7. Average Weekday Household Vehicle Miles Traveled by Census Tract (per day)



NOTE: Areas with no estimate are areas where there is no population or areas where one or more of the model parameters are not available.

Figure 8. Average Weekday Household Vehicle Trips by Census Tract



NOTE: Areas with no estimate are areas where there is no population or areas where one or more of the model parameters are not available.

Comments on Data

There are a few challenges associated with the data. The accuracy of the Census tract estimates could not be measured directly as there are no Census tract data to compare against the model results. Because the models explained only a limited amount of the variation in person miles traveled, person trips, vehicle miles traveled and vehicle trips at the region level, the models are likely to explain even less at smaller geographies where statistical variability is expected to be higher. A limited comparison was made against NHTS data, where a reasonable number of households were sampled in a Census tract. These NHTS estimates proved similar to the estimates made by transferring the ACS data.

The ability to produce sub-national estimates is limited by the NHTS sample design. NHTS data are collected through address based sampling for a national sample and for select 'add-on' or oversampled geographic areas. The oversampled geographic areas are the areas where subnational level estimates can be best measured because of the larger sample size. The regions created here to estimate tract level person miles traveled, person trips, vehicle miles traveled and vehicle trips include these oversampled geographic areas with areas covered by a much smaller sample. The characteristics of the areas with a smaller sample may be different from the oversampled areas and as such, the estimates of person miles traveled, person trips, vehicle miles traveled and vehicle trips may be less accurate in these sparsely sampled areas, for example in the Mountain Division.

The address based sampling design itself poses challenges in coverage and nonresponse bias. The response rate for the NHTS is relatively low (approximately 20 percent), which suggests the potential for nonresponse bias. This may be more extensive among demographic groups that are difficult to reach because they are highly mobile.

Finally, there are a few challenges associated with using the ACS data. ACS census tract level data are multi-year estimates. This adds variation to the data when changes occur over time. Additionally, the ACS data are less precise in some Census tracts than others. In some cases, the margin of error is larger than the ACS estimate. This affects the precision of the final estimates of person miles traveled, person trips, vehicle miles traveled, and vehicle trips, because the model uses the ACS data to produce estimates of household travel. ACS data are available for geographies larger than Census tracts. The larger geographies tend to have smaller margins of error, which would increase the precision of the final estimates. However, estimates at geographies larger than Census tracts tend to hide variation within the geography (e.g., areas of high and low vehicle miles traveled), which makes them less useful.

The margins of error associated with the ACS data also pose a challenge in measuring change in household travel in a Census tract. An observed change, e.g., average vehicle miles traveled, may be due to a non-statistically significant change in one or more of the ACS estimates. There is no clear method to aggregate the statistically significant ACS changes to mark the observed change in travel as significant (or not).

Despite the above challenges, the models still provide useful travel data, by Census tract, that can be employed by planners and researchers alike.

REFERENCES

Henson, K. M. and K. G. Goulias. 2001. Travel Determinants and Multiscale Transferability of National Activity Patterns to Local Populations. *Transportation Research Record: Journal of the Transportation Research Board* No. 2231, pp. 35-43.

Hu, P. S., T. Reuscher, R. L. Schmoyer and S. M. Chin. 2007. *Transferring 2001 National Household Travel Survey*. Oak Ridge, TN: Oak ridge National Laboratory. May 2.

Mazzulla, G. and C. Forciniti. 2012. Spatial Association Techniques for Analysing Trip Distribution in an Urban Area. *European Transport Research Review*, No. 4, pp. 217–233.