

TRANSPORTATION RESEARCH  
**CIRCULAR**

Number E-C238

November 2018

**2018 National Household  
Travel Survey Workshop**

*August 8-9, 2018  
Washington, D.C.*

**TRANSPORTATION RESEARCH BOARD  
2018 EXECUTIVE COMMITTEE OFFICERS**

**Chair: Katherine F. Turnbull**, Executive Associate Director and Research Scientist, Texas A&M Transportation Institute, College Station  
**Vice Chair: Victoria A. Arroyo**, Executive Director, Georgetown Climate Center; Assistant Dean, Centers and Institutes; and Professor and Director, Environmental Law Program, Georgetown University Law Center, Washington, D.C.  
**Division Chair for NRC Oversight: Susan Hanson**, Distinguished University Professor Emerita, School of Geography, Clark University, Worcester, Massachusetts  
**Executive Director: Neil J. Pedersen**, Transportation Research Board

**TRANSPORTATION RESEARCH BOARD  
2017–2018 TECHNICAL ACTIVITIES COUNCIL**

**Chair: Hyun-A C. Park**, President, Spy Pond Partners, LLC, Arlington, Massachusetts  
**Technical Activities Director: Ann M. Brach**, Transportation Research Board

**David Ballard**, Senior Economist, Gellman Research Associates, Inc., Jenkintown, Pennsylvania, *Aviation Group Chair*  
**Coco Brisen**, Deputy Director, Planning and Modal Programs, California Department of Transportation, Sacramento, *State DOT Representative*  
**Anne Goodchild**, Associate Professor, University of Washington, Seattle, *Freight Systems Group Chair*  
**George Grimes**, CEO Advisor, Patriot Rail Company, Denver, Colorado, *Rail Group Chair*  
**David Harkey**, Director, Highway Safety Research Center, University of North Carolina, Chapel Hill, *Safety and Systems Users Group Chair*  
**Dennis Hinebaugh**, Director, National Bus Rapid Transit Institute, University of South Florida Center for Urban Transportation Research, Tampa, *Public Transportation Group Chair*  
**Bevan Kirley**, Research Associate, Highway Safety Research Center, University of North Carolina, Chapel Hill, *Young Members Council Chair*  
**D. Stephen Lane**, Associate Principal Research Scientist, Virginia Center for Transportation Innovation and Research, Charlottesville, *Design and Construction Group Chair*  
**Ram M. Pendyala**, Professor, School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, *Planning and Environment Group Chair*  
**Joseph Schofer**, Professor and Associate Dean of Engineering, McCormick School of Engineering, Northwestern University, Evanston, Illinois, *Policy and Organization Group Chair*  
**Eric Shen**, Director, Southern California Gateway Office, Maritime Administration, Long Beach, California, *Marine Group Chair*  
**William Varnedoe**, Partner, The Kercher Group, Raleigh, North Carolina, *Operations and Preservation Group Chair*  
**Fred R. Wagner**, Partner, Venable, LLP, Washington, D.C., *Legal Resources Group Chair*

TRANSPORTATION RESEARCH CIRCULAR E-C238

# **2018 National Household Travel Survey Workshop**

**August 8-9, 2018  
Washington, D.C.**

**Transportation Research Board  
500 Fifth Street, NW  
Washington, DC 20001  
[www.TRB.org](http://www.TRB.org)**

The **Transportation Research Board** is one of seven programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal.

The **Transportation Research Board** is distributing this E-Circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this circular was taken directly from the submission of the authors. This document is not a report of the National Academies of Sciences, Engineering, and Medicine.

*Sponsored by*  
**National Household Travel Survey Workshop Committee**  
Lisa Aultman-Hall, *University of Vermont, Chair*

Daniel Chatman  
*University of California, Berkeley*

Kate Lawson  
*University of Albany, NY*

Marcus Coenen  
*Des Moines Area Metropolitan Planning  
Organization*

Jennifer Murray  
*Wisconsin Department of Transportation*

Patrick Coleman  
*AECOM Consult*

Wenjing Pu  
*Federal Highway Administration*

Danny Jenkins  
*Federal Highway Administration*

Krishnan Viswanathan  
*Cambridge Systematics, Inc.*

Habte Kassa  
*Georgia Department of Transportation*

Lei Rousseau  
*University of Maryland, College Park*

*TRB Staff*  
Tom Palmerlee, *Associate Division Director, Data and Information Technology*  
Mai Le, *Associate Program Officer*

Transportation Research Board  
500 Fifth Street, NW  
Washington, D.C.  
www.TRB.org

## Preface

Meeting at the Keck Center, in Washington, D.C., on August 8-9, 2018, one hundred and four attendees of the National Household Travel Survey (NHTS) Workshop shared their experiences with the 2017 NHTS dataset. Researchers and practitioners presented preliminary findings from their research for feedback, insights, and suggestions for future NHTS deployments.

Lisa Aultman-Hall, professor at the University of Vermont, chaired the planning committee. The planning committee was solely responsible for organizing the conference, preparing the call for abstracts, reviewing the submitted abstracts, and developing topics for breakout and panel sessions, including guidance for the facilitated discussion sessions. Catherine T. Lawson, associate professor at the University at Albany, State University of New York (SUNY), served as rapporteur and prepared this document as a factual summary of what occurred at the workshop.

This summary report follows the workshop agenda, including session panels, presentations and audience interactions. Special sessions introduced new data dissemination and analytics tools, including using R software and new features on the Oak Ridge National Laboratory website. Several presentations included the future vision of a NHTS that combines traditional surveying efforts with passively collected origin and destination probe data. Additional sessions covered research topics including: new methods used for data collection; travel trends between 2009 and 2017; alternative fueled vehicles; nonmotorized transportation; new sharing modes and underserved populations. A special presentation provided an overview of national statistical programs and recent studies on techniques for blending traditional survey data and administrative data sources.

The views expressed in this summary are those of individual workshop participants and do not necessarily represent the views of all the workshop participants, the planning committee, the Federal Highway Administration, or the Transportation Research Board.

# Contents

<b>Chapter 1 Opening Session</b> .....	1
<b>Introduction and Welcome</b> .....	1
<i>Tianjia Tang</i>	
<b>Why Travel Surveys Matter in the Age of Big Data?</b> .....	2
<i>Patricia Mokhtarian</i>	
<b>Chapter 2 2009 vs. 2017</b> .....	5
<b>The 2017 Method Changes and Implications</b> .....	5
<i>Shawn Mcclosky and Jeremy Wilhelm</i>	
<b>Travel Trends 2009 vs. 2017</b> .....	6
<i>Nancy McGuckin</i>	
<b>Chapter 3 Exploring the NHTS</b> .....	10
<b>Panel</b> .....	10
<i>Anthony Fucci and Alexander Cates</i>	
<b>Chapter 4 The New NHTS Online Tool</b> .....	12
<b>Panel</b> .....	12
<i>Ho-Ling Hwang and Aparajita Benerjee</i>	
<b>Chapter 5 Using the NHTS for Performance Measures</b> .....	18
<b>Panel</b> .....	18
<i>Weijing Pu, Keith L. Killough, and Geena Maskey</i>	
<b>Chapter 6 How the New Methods Affect Me</b> .....	21
<b>Panel</b> .....	21
<i>Stacy Bricka, Soheil Khoil, And Paul Schroeder</i>	
<b>Chapter 7 Alternative Fueled Vehicles</b> .....	25
<b>Revisiting Vehicle Ownership in Activity-Based Models Considering</b> <b>Alternative Fuel Vehicles</b> .....	25
<i>Arash Asadabadi and Lei Zhang</i>	
<b>A Casual Approach to Disentangling Vehicle Technology and</b> <b>Self-Selection Effects on Household Vehicle Use: A Tri-Variate</b> <b>Copula-Based Endogenous Regime Switching Framework</b> .....	26
<i>Behram Wali and Asad Khattak</i>	
<b>Tracking National Household Vehicle Usage by Type, Age, and Area in</b> <b>Support of Marking Assessments for Plug-In Electric Vehicles With Emerging</b> <b>Usage of Shared Services</b> .....	27
<i>Yan Zhou And Vyas Anant</i>	

<b>Chapter 8 Nonmotorized Transportation</b> .....	31
<b>Health Impacts of Nonmotorized Travel Behavior and Built Environment</b> .....	31
<i>Jina Mahmoudi and Lei Zhang</i>	
<b>Trends in Walking and Cycling in the United States, 2001–2017</b> .....	33
<i>Ralph Buehler, John Pucher, Safna Merom, and Adrian Bauman</i>	
<b>Interactions Between the Built Environment and Domain-Specific</b>	
<b>Transportation Physical Activity: Evidence from the 2017 NHTS</b> .....	35
<b>Defining Type of Cyclist Based on Travel and Activity Patterns</b> .....	38
<i>Jim Sullivan and Sarah Howerter</i>	
<b>Chapter 9 The New Sharing Modes</b> .....	40
<b>Trends in Taxi Usage and the Advent of Ridesharing, 1995–2017</b> .....	40
<i>Matthew Wigginton Conway</i>	
<b>Examining the U.S. Ridesourcing Market Using the</b>	
<b>2017 NHTS Data</b> .....	42
<i>Zhenpeng Zou and Sevgi Erdogan</i>	
<b>The Effects of Different Generational Sociodemographic Characteristics on</b>	
<b>Dynamic Ridesharing</b> .....	44
<i>David Donaldson and Arash Ghaffar</i>	
<b>Chapter 10 Underserved Populations</b> .....	48
<b>Food Access for Low-Income Individuals</b> .....	48
<i>Chandra Bhat and Abhilash Singh</i>	
<b>Predicting Daily Trip Frequencies of Vulnerable Households in New York State</b>	
<b>Using Emerging Machine-Learning Approaches</b> .....	50
<i>Bumjoon Bae, Ho-Ling Hwang, Shih-Miao Chin, and Chieh R. Wang</i>	
<b>Travel Patterns of Low-Income Households</b> .....	52
<i>Apara Banerjee</i>	
<b>Persons With Disabilities and Persons Born Outside of the U.S.:</b>	
<b>Demographic and Travel Trends for Transport Planners</b> .....	54
<i>Daniel Chatman, Abigail Cochran, and Nicholas Klein</i>	
<b>Chapter 11 Poster Session</b> .....	58
<b>Using the NHTS as a Daily Activity Pattern Archive in a Passive Data Model</b> .....	59
<i>Greg Macfarlane and Josie Kressner</i>	
<b>Application of NHTS Transferability Statistics to Measure Impacts of</b>	
<b>Transportation Revenue Tools on Household Segments</b> .....	61
<i>Kyle Schroeckenthaler and Stephen Fitzroy</i>	
<b>Modeling Transportation Characteristics: Small Areas Estimation Using the NHTS</b> .....	63
<i>Theresa Firestine and Clara Reschovsky</i>	
<b>A Comparison of the NCTCOG Weighting Method to the NHTS Weighting Method</b> ....	64
<i>Kathleen Yu and Arash Mirzael</i>	

<b>Calibration and Validation of a Micro-Simulation Long-Distance Travel Demand Model Using 2017 NHTS</b> .....	65
<i>Sepehr Ghadr and Aref Darzi</i>	
<b>Insights on Data Quality from a Large-Scale Application of Future Mobility Sensing Technology in Greater Phoenix, Airzona</b> .....	68
<i>Shari Gershenfedl and Shuyao Hong</i>	
<b>A Breakthrough Solution for Statistical Surveys of U.S. Population</b> .....	69
<i>Laurie Wargelin and Chet Bowie</i>	
<b>Rethinking the Survey Part of the NHTS</b> .....	70
<i>Jeremy Wilhelm and Nancy Mcguckin</i>	
<b>Exploring Methods to Increase Efficiency in Generating and Comparing National Travel Behavior Estimates Over Time</b> .....	71
<i>Anthony Fucci and Alexander Cates</i>	
<b>How Built Environment Changes Shape Travel Behavior Over Time: Comparison Between 2009 and 2017 NHTS</b> .....	74
<i>Aref Darzi and Lei Zhang</i>	
<b>Update From 2017: Did the Millennials Really Change Their Travel Behavior in 2009?</b> .....	78
<i>Matthew Coogan and Margaret Campbell</i>	
<b>The Changing Nature of the Activity-Travel Behavior of the Elderly</b> .....	80
<i>Chandra Bhat And Joseph Hutchinson</i>	
<b>Trends in Typical Vs. Actual Travel: An Analysis of NHTS Data Over Time</b> .....	82
<i>Stacy Bricka and Timothy Reuscher</i>	
<b>Travel Trends in Texas: A Comparison of the 2017 NHTS Data to Previous National and Local Travel Survey Analysis Results</b> .....	84
<i>Lisa Green and Michael Martin</i>	
<b>Online Shopping Behavior of U.S. Travelers: The Roles of Internet Usage and Lifestyle</b> .....	85
<i>Huyen Le</i>	
<b>Assessing Equity and Access to Long-Distance Travel Using the 2017 NHTS Data</b> .....	87
<i>Lisa Aultman-Hall, Jonathan Dowds, Greg Thivierge, and Anuarbek Onayev</i>	
<b>Chapter 12 Plenary Session</b> .....	89
<b>National Statistical Programs: Directions and Challenges</b> .....	89
<i>Brian Harris-Kojetin</i>	
<b>Chapter 13 Modelers Panel: Using the NHTS 2017</b> .....	91
<b>Panel</b> .....	91
<i>Habte Kassa, Patrick Coleman, and Arash Mirzael</i>	



<b>Quick Ways to Use NHTS Data to Support Regional Travel</b>	
<b>Demand Model Calibration</b> .....	92
<i>John Miller, Peter Ohlms, and Zulqarnain Khattak</i>	
<b>Chapter 14 Nextgen NHTS</b> .....	99
<b>Introduction of the Nextgen NHTS</b> .....	99
<i>Daniel Jenkins and Wenjing Pu</i>	
<b>Integration and Validation of Passively Collected Data with</b>	
<b>National Household Travel Survey</b> .....	101
<i>Sepehr Ghader and Lei Zhang</i>	
<b>Cross-Validation and Simultaneous Model Estimation from NHTS and</b>	
<b>Aggregate Passive OD Data in Charleston, South Carolina</b> .....	102
<i>Vince Bernardin and Hadi Sadrsadat</i>	
<b>Nextgen Add-On Perspectives</b> .....	103
<i>Zachary Hanson</i>	
<b>Chapter 15 Closing Session</b> .....	105
<b>Great Things Done in 2009 and Moving Forward with 2017</b> .....	105
<i>David Winter and Rolf Schmitt</i>	
<b>Chapter 16 Participants</b> .....	107

# Chapter 1: Opening Session

**LISA AULTMAN-HALL**

*University of Vermont, presiding*

**DANIEL CHATMAN**

*University of California, Berkeley, recording*

**ROLF SCHMITT**

*Bureau of Transportation Statistics*

**TIANJIA TANG**

*Federal Highway Administration*

**KEYNOTE SPEAKER: PATRICIA MOKHTAIAN**

*Georgia Institute of Technology*

As the landscape for travel behavior and data collection changes at unprecedented rates, questions arise regarding the relevance of travel data and what future surveys will look like. The strengths and weaknesses of approaches to acquire the necessary data from households and individuals on to better understand their travel needs remains an active area of research and practice.

## **INTRODUCTION AND WELCOME**

*Tianjia Tang*

Tianjia Tang, Chief of the Travel Monitoring and Survey Division of Federal Highway Administration (FHWA), welcomed participants to the NHTS Data for Transportation Applications Workshop. He expressed a strong need for champions and ambassadorship for the NHTS program to be successful. He thanked Professor Lisa Aultman-Hall, for her leadership, the Planning Committee, the NHTS Task Force Committee members, and others who supported the workshop.

The National Personal Travel Survey (NPTS) program collected data in 1969, 1977, 1983, 1990 and 1995, followed by the NHTS, collected in 2001, 2009 and now, 2017. The vision of this the NHTS is to provide credible, quantifiable information, not anecdotal information, to assist planning and policy-making. When listening to the NHTS community, Tang stressed the importance of FHWA being open to suggestions for improvement. The NHTS program requires professionalism, accountability, expressions of honesty, a caring attitude, a passion for travel data, and concern for the program as a whole. Finally, the NHTS requires users to be detail-oriented, able to deal with hundreds of parameters. Fielding the NHTS is a very complicated undertaking, requiring care to ensure users have instructions on the appropriate use of the data. Adhering to these aspects help guarantee the program will move forward successfully. Using data can be like the six blind men trying to understand what an elephant is. They all think they know what they were experiencing, but they don't have the whole picture unless they work

together. The NHTS community is known for working together. The three workshop goals are: to learn about the new dataset; to exchange early results and ideas; and to look ahead to what travel surveys and data collection will be in the future.

## **WHY TRAVEL SURVEYS MATTER IN THE AGE OF BIG DATA?**

*Patricia Mokhtarian*

Are traditional travel surveys passé? It is rare to find a survey design course in graduate transportation courses, even today. Although the term survey methods is common in psychology or sociology, transportation courses most often offer courses in data acquisition methods. Will survey design completely disappear in the next five years, and be replaced by “Using Machine Learning Methods to Analyze Big Data”? The driving forces for such a change include: Global Positioning Systems (GPS) traces; transit smart cards; clickstreams; radio frequency identification chips and scanner data; twitter feeds and other social media posts; remote sensing; targeted marketing and credit reporting data; and more. We ask ourselves - who needs old-fashioned surveys?

There are three compelling reasons for maintaining travel surveys: no equivalent Big Data source; incomplete Big Data; and combining Big Data with survey data makes both more valuable. It is clear to researchers that Big Data sources are not always available now, or in the near future. Travel behavior research uses a particular set of procedures: qualitative interviews, focus groups and charrettes (which are intensive planning sessions); methods for reliably identifying and measuring small, specialized populations; and hypothetical choices on the impacts of technologies that are currently not available, responses to proposed policies and behavioral intentions.

What Big Data is available is far from perfect. GPS traces have issues: broken trips; urban canyons that cause signal blockage and multi-path interference; poor within-building performance; dead batteries; forgotten phones; and more challenges. In addition, GPS trace data provides no demographic information. Attempts to apply aggregate statistics available in associated geographic units can result in an ecological fallacy – the condition where relationships in aggregate are different (even opposite) of those found at the disaggregate level. GPS trace data is unable to provide an understanding of human behavior, which generally requires the ability to measure the unobservable including: constraints; motivations; intentions; personality characteristics; lifestyle choices; or attitudes.

GPS trace data lacks representativeness, even though some believe the sheer volume of Big Data solves this concern. There are examples of making mistakes with large volumes of data (e.g., 1939 Literary Digest poll that predicted Landon would win in a landslide over Roosevelt). Importantly with respect to Big Data, correlation doesn't equal causation, even though some claim experts are shifting away from causation analysis. However, this eliminates the ability to understand “why” and reduces the ability to predict “what might happen if?” Two causal models illustrate why understanding what causes what is so important. Consider a human giving a dog a treat to stop the dog from attacking, while the dog is attacking precisely because they know that to do so will end in them receiving a treat.

Solving problems requires domain knowledge and methods for understanding relationships and avoid spurious correlations (e.g., using any random data sets that show a close correlation, but actually make no sense). At this point in our capabilities, Big Data can be seen as

enriching survey data by: improving matching for treatments; providing more cases around a regression discontinuity; drilling down to specialized (if identifiable) population segment; experimenting on a large scale, within an ecologically valid settings; and tracking dynamic behavior over time, including panels.

Researchers at TOMNET [Teaching Old Models (and Modelers) New Tricks], a first tier University Transportation Center (UTC), with participation from UTC USF, GATech, ASU, and UW, are incorporating attitudinal information into models, using machine learning. For this effort, the researchers are using the attitudinally rich survey of 3000 Georgia residents, with land use characteristics and targeted marketing data enhancements. These data are being combined with NHTS to create a “virtual survey” of fused data, with additional information from a Georgia subsample of 8000 residents, now enhanced with imputed attitudes. This dual approach overcomes the shortcomings of each data type, while offering a harmonized dataset combines both survey data and passive (Big) data, similar to the direction of NextGen NHTS. To quote Susan Handy, professor in the Department of Environmental Science and Policy at the University of California at Davis speaking on the power of combining qualitative and quantitative methods, “Looking at things in multiple ways creates a richer and more true understanding of the world.”

### **Audience Dialogue Highlights**

In the discussion that followed, questions and concerns focuses on opportunities for Big Data to become more standardized and more responsive to the needs of the travel behavior community. An audience member thought that since the source of Big Data is primarily the private sector, moving forward with structure changes (e.g., data formats and collection methodologies) will require developing a robust dialogue and the cooperation of the vendors of probe data and other related products. One member of the audience suggested asking Big Data providers (e.g., companies that provide shared mobility services) to join the public sector in sharing data.

There was a request for a better understanding of Big Data, with its datasets with billions or trillions of records. This form of data requires different management techniques, particularly with respect to post-processing procedures and data handling strategies. Users of Big Data, and anyone attempting to fuse various types of data together, need to take statistical concerns. According to an audience member, the nature of travel survey data and probe data requires attention to spatial autocorrelation in modeling.

### **REFERENCES**

- Athey, Susan and Guido W. Imbens. (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects. arXiv:1504.01132v1.
- Athey, Susan. (2015). Machine Learning and Causal Inference for Policy Evaluation. <http://dx.doi.org/10.1145/2783258.2785466>.
- Breiman, Leo. (2001). Statistical Modeling: The Two Cultures. *Statistical Science* 16(3), 199-231.
- Grimmer, Justin. (2015). We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together. American Political Science Association, doi:10.1017/S1049096514001784.
- Handy, Susan. (2013). The Power of Mixed Methods: Examples from the Field of Travel Behavior Research. Guest Lecture, TTP 200 (Transportation Survey Methods), University of California, Davis.

- Malokin, Aliaksandr, Patricia L. Mokhtarian, and Giovanni Circella. (2018). An Investigation of Methods to Enrich National Household Travel Survey Data with Attitudinal Variables. Paper presented at the 97th Annual Meeting of the Transportation Research Board, Washington, D.C. Available from the authors.
- Rose, Sherri, Richard J.C.M. Starmans, and Mark J. van der Laan. (2012). Targeted Learning for Causality and Statistical Analysis in Medical Research. In *Statistics: Discovering Your Future Power*. China Statistics Press. <https://biostats.bepress.com/ucbbiostat/paper297/>.
- Sliva, Amy, Scott Neal Reilly, David Blumstein, Steve Hookway, and John Chamberlain. (2017). Modeling Causal Relationships in Sociocultural Systems Using Ensemble Methods. In S. Schatz and M. Hoffman (eds.), *Advances in Cross-Cultural Decision Making, Advances in Intelligent Systems and Computing* 480.
- Van Der Puttan, Peter, Joost N. Kok, and Amar Gupta. (2002). Data Fusion through Statistical Matching. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=297501](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=297501), accessed Aug. 6, 2018.
- Varian, Hal. (2014). Machine Learning and Econometrics. <https://web.stanford.edu/class/ee380/Abstracts/140129-slides-Machine-Learning-and-Econometrics.pdf>, accessed Aug. 6, 2018.
- Wu, Xindong and Vipin Kumar, (2009). *The Top Ten Algorithms in Data Mining*. Boca Raton, FL: Chapman & Hall/CRC.

## **Chapter 2: 2009 vs. 2017**

**LISA AULTMAN-HALL**

*University of Vermont, presiding*

**KRISHNAN VISWANATHAN**

*Cambridge Systematics, recording*

**SHAWN MCCLOSKEY**

**JEREMY WILHELM**

*Westat*

**STACEY BRICK**

*MacroSys LCC*

**DANIEL CHATMAN**

*University of California, Berkeley*

**NANCY MCGUCKIN**

*Travel Behavior Associates*

**JEFF LAMONDIA**

*Auburn University*

**STEVE POLZIN**

*University of South Florida*

In order to start exploring the new NHTS data, survey designers and data collectors shared details on the deployment of the survey and how this method differs from prior surveys. Preliminary survey findings were reported from the recently published, *Travel Trends Summary Report*. Discussants in this session shared their experiences on travel survey efforts to encourage a strong dialogue with the each other and the audience.

### **THE 2017 METHOD CHANGES AND IMPLICATIONS**

*Shawn McClosky and Jeremy Wilhelm*

In national travel surveys deployed from 1969 through 2009, the data was collected using in-person and telephone interviews. The 2001 and 2009 deployments used Random Digit Dialing (RDD), with double-digit response rates. In 2017, the environment for surveying had changed significantly and warranted a new approach. Specifically, households are increasingly forgoing a landline (52%); response rates continue to decrease; and there has been an uptake in internet usage (~90% of households have access to some type of device). These changes prompted the new approaches for the 2017 deployment; an address-based sample (ABS); a Two-stage survey

design; and a new multi-stage incentive structure (e.g., \$2.00 to mail back initial outreach, \$5.00 to provide more detailed information and \$20.00 for a completed survey).

For the modes used in the recruitment phase, 94.8% used mail-back, 4.7% used the web and only 0.5% used the phone. For actual survey, 60.4% used the web, 30.3% used phone and 9.4% used a combination of the two modes. A major difference between the 2009 and 2017 deployments was the definition of a “completed household” (2009 was 50%, while 2017 required 100% response of household members). In addition, the 2017 was a self-reported survey rather than an interactive experience with a professional surveying staff. For 2017, the surveys provided real-time, online geocoding. In addition, in 2009, respondents reported distances, however, in 2017, distances were calculated along a network path, using the Google API.

In 2009, “loop trips” were recorded as two separate trips; the first trip reported using the farthest point out, and a separate trip back. In 2017, such travel was recorded as a loop trip and the respondent was prompted to provide the total distance of the trip. For trips on public transit, in 2009, respondents were asked for the trip roster was provides, however, in 2017, respondents were asked before the places roster was completed. The website provided guidance for the respondent prior to reporting their trips to reduce difficulties. The 2017 deployment had several additional questions including: supplemental walk and bikes questions; travel attitudes and beliefs; ride-sharing services uses; and the general health of the respondent. The design of the website facilitated ease of response, and included an interactive mapping of travel activities self-reported by the respondent.

## **DISCUSSANT RESPONSE**

*Stacy Bricka*

The surveying effort itself used a unique spreadsheet to track all the survey-related activities occurring simultaneously (e.g., mailing out of one million surveys, tracking through all the stages, monitoring pieces, key demographic tracking, etc.). Preparations for the survey occurring in 2016, an election year, faced challenges and required more mailings to achieve the desired sampling goals, particularly among minority communities (e.g., Hispanic populations). Looking at surveying outcomes, while still not perfect, appear much improved over prior deployments. The sample management tool made a big difference, however, especially given hurricane issues that added additional concerns (e.g., starting and stopping the deployments). Data quality also improved with the use of the new tools used for collection phase. The survey produces 600 variables, with only 5% from core questions, 10% from the add-ons, and 10% for multiple-choice questions. The survey itself took only 10 to 15 minutes per person, and on average, 45 minutes per household.

## **TRAVEL TRENDS 2009 VS. 2017**

*Nancy McGuckin*

The major findings from the 2017 NHTS data shows that trip rates by age, sex, urban/rural, income, purpose and other factors, continue to decline. Some of the recorded declining rates could be due to self-reporting and will require more research to explore the true nature of these trends. Since trip distance was collected in a new manner, it may require some adjustment as trip

distance impacts Vehicle Miles Traveled (VMT) and Person Miles Traveled (PMT). The survey results indicate higher transit ridership, while vehicle trip rates are lower – and there is no apparent explanation. FHWA has already funded research to follow this trend. Walk and bike may have experienced a break in their trends with the change in trip definition, and possible an under-reporting of short non-purposeful trips. More research would help understand these impacts.

With respect to the decline in trip rates, trips for shopping and errands are the major contributor to this trend. Immobility rates appear high, as more people reported no travel on their travel day. This could be the result of using self-reporting rather than a professional interviewer (who might have prompted the respondent to remember “forgotten” trips). In addition, zero trip reports differ by survey method used and by age groups. Trips per traveling person were significantly lower for all but older travelers. Some long-term demographic trends may need to be considered to better understand the NHTS results. For example, the number of people living in non-metro areas has remained steady since 1980, while the population in the metro areas has continued to grow.

One of the most prominent aspects of the population is the impact of baby boomers. This cohort continues to age, with experts predicting that within the next few decades, older people will outnumber children. About 61.3% of the 2017 sample reported themselves as “workers” (tracking with the Bureau of Labor Statistics estimate of 60.4% for 2017). While the trends suggest over all household travel is declining, it is of particular note that middle and higher income households reported lower trip making (counter to traditional theory). Comparing 2001 to 2017, there are significant declines in errands and social/recreational trips. VMT for younger people (16 to 34 years of age) remains stable in urban areas between 2009 and 2017.

With respect to gender differences, men and women’s travel rates have declined in similar ways for shopping and errands, social and recreational, school and church, and to and from work. Not surprisingly, the number of home deliveries from on-line shopping has doubled between 2009 and 2017. In fact, all age groups are shopping on-line, with notable growth from those 65 years old and above. The average commute in 2017 was 27.5 minutes, compared to 24.2 in 2009. In addition, the average worker who traveled to and from work, five days a week, spent 33 minutes more commuting per week in 2017 than 2009. Another surprising fact was the lack of evidence of peak spreading, with the 2017 patterns indicating more peaking than previously. With respect to the types of vehicles, autos now make-up half of the residential fleet, with SUVs indicating a growing share of the market. Finally, newer vehicles are less likely to be autos.

## **DISCUSSANT RESPONSE**

*Jeremy LaMondia and Steve Polzin*

More funds would help understand NHTS data quality, particularly issues with representativeness and response rates. It is unclear what the impacts are of moving to a place roster instead of an activity roster. There were concerns in 2009 with respect to the lack of distance reporting for irregular trips. For 2017, the Google API established the shortest path and that distance was recorded for a trip. Primarily very short and very long trips are identified as outliers and using the Google API appears to have generated 10% longer trips than what people were reporting previously. In addition, older vehicles are still part of the vehicle mix.



Declining shopping trip rates and aging population trip rates confirm existing theories (e.g., substitution of on-line shopping for trips, decreasing overall mobility rates due to health). The question is what does it mean in terms of policy and infrastructure planning? Researchers are challenged to explain the lower share of household VMT, with a need to fill the gap with knowledge about other market segments. More research would help understand why the 2017 data indicates a very large per capita increase in transit. This is particularly difficult with the transit agency data indicating decreasing transit market shares.

There continues to be travel behavior differences among urban, suburban and rural populations. For example, urban areas grew by 13%, suburban by 16% and rural by 4%. Rural populations are least likely to move according to Pew Research, while access needs for rural areas increasing. Rural populations make fewer shorter trips and fewer longer trips for work and shopping, which impacts VMT.

### **Audience Dialogue Highlights**

Regarding long distance trips, the sample design would benefit from including this aspect directly. While the national deployment of the 2017 NHTS did not take on this question, several of the add-on deployments focused on long distance travel explicitly (e.g., California high-speed rail questions).

The change from 50% to 100% participation of household members to be considered a complete response, and the change to a web-based survey from a phone interview, both could have impact on the outcomes of survey analysis. For example, the trip rates were different between the web and computer-assisted telephone interviewing (CATI) responses. More research would help understand these differences and to examine the incomplete households to see if there are patterns. It was noted that there were not many partial households. With respect to cohabited households, they were defined as any other household for the purpose of completeness. However, for children living in multiple households (e.g., divorced parents), the data could be recorded as an incomplete tour.

There were questions on the impacts of changes in the approach to asking transit trips in the web survey approach. A review of the differences in transit trips indicated that respondents who were shown how to report trip trips via multiple means reported it more accurately. It isn't clear if there were impacts on walk or bike trips. Overall, there are concerns over the changes in the methodology used, but there appear to be no red flags regarding trip rates for self-reporting household holds. However, a deeper investigation would help understand whether any market segments were impacted differently by method or assumptions of behavioral changes.

Regarding impacts on underrepresented populations (e.g., Hispanic populations) or minority households, responses during the recruitment phase of the survey dipped during the election season and retrieval rates dipped post-inauguration. An additional concern is whether economic growth during 2017 impacts the data compared to the 2009 data. Urban areas have achieved 2007 levels of job growth, while rural areas have not. VMT appears to have responded to the growth in the economic, however, urban trips may be impacted by commercial trips more than household trips.

Smartphone travel survey apps are showing lower immobility rates even accounting for demographic trends. It will be important to understand how to measure non-response over time to ensure accuracy of data and our understanding of travel behavior. A smartphone may be more available than previously thought as 80% of vulnerable populations use smartphones are their

main source for internet services. The American Time Use Survey (ATUS) might be helpful in evaluating immobility. Another concern is the possibility that the decrease in trip rates is due to longer trips.

Additionally, there was recognition that moving forward, methods and techniques that assist the respondent in reporting trips correctly will provide real benefits for the data users and practitioners. Concerns were expressed regarding Margin of Error (MOE) calculations and the need for their inclusion when conducting analyses. VMT is harder to measure than trip rates, such as those provided from household travel surveys. Also, it remains unclear how non-traditional household travel compares with traditional households and what should be done given that personal vehicle commercial trips were not reported in the data and may help explain why VMT at the household level are lower than expected.

## Chapter 3: Exploring the NHTS

**KRISHNAN VISWANATHAN**

*Cambridge Systematics, presiding*

**KRISHNAN VISWANATHAN**

*Cambridge Systematics, recording*

**ANTHONY FUCCI**

**ALEXANDER CATES**

*Westat*

This session explored the 2017 NHTS data using R software developed specifically for the purpose of analyzing the most recent NHTS programs. Included in the demonstration were techniques for loading the most recent NHTS data set into a Windows or Mac environment using RStudio. R has a common development interface and allows users to examine the data structure and variable metadata using R syntax. Methods for analyzing and visualizing the NHTS data using “summarize\_NHTS” were shared.

### **PANEL**

*Anthony Fucci and Alexander Cates*

### **Background**

The NHTS is a complex dataset and requires a set of analytical tools from simple uses (e.g., Excel) to advanced statistical programs [e.g., statistical analysis system (SAS), SPSS, STATA). To facilitate statistical analysis of the NHTS, another option is the use of R, a programming language. R is an open source product, developed over the last thirty years. R is capable of conducting analysis of the national NHTS and the add-on datasets. It has been used to validate the summary of travel trends output and can be applied to a wide variety of analyses for transportation research and planning.

### **How to Use R**

Using the software requires a set of specific steps to set up an R analysis environment. First, the link below is necessary to install the software. The Github repository contains all of the elements and instructions for using the software. In the R analysis environment, users are able to access the NHTS dataset, generate estimates, create new variables, visualize estimates and produce reports that can be used by the transportation community. (<https://github.com/Westat-Transportation/summarizeNHTS/tree/master/inst/install>)

The instructions for downloading the necessary elements including: R, R tools, and RStudio. Table 3.1 provide the required inputs to use after opening the website site.

**TABLE 3.1 Specific Code Inputs to Access NHTS in Rstudio Using Summarize NHTS.**

<b>Required inputs</b>
Install.package("devtools")
Devtools::install_github("Westat-Transportation/summarizeNHTS")
<i>Next, load the software:</i>
Library(summarizeNHTS)

The presentation materials found in the Github file provides guidance on the exact code to use to perform the necessary steps to use the software. Using the software, the NHTS data is accessed directly from the Oak Ridge National Laboratory site, including 2001, 2009 and 2017 data. The software allows users to select a subset of data and access the NHTS codebook. After using "summarizeNHTS", the data can be housed on a local drive. The dataset also includes the replicate weights for use in particular types of analyses.

The software uses the summarize\_data function provides statistical products (e.g., weighted statistics, standard error of the weighted statistics, unweighted surveyed/sampled statistics, and the number of observations). This function produces sums, averages, and medians from numeric variables. The software includes a comprehensive function documentation to aid the user while still within the R environment. The software also allows users to create new variables, including new derived variables. For example, it is possible to create following: flags (e.g., yes/no, is/is-not; has/has-not); collapsed or binned variables (e.g., grouping public transit individual modes into a single "public transportation" variables); categorizations; and mathematical calculations.

The software makes it possible to generate a variety of graphics. For example, visualizations can include tables with simple estimates, one-way, two-way, and three-way tables. Another feature is the ability to produce charts and maps. The mapping geographies includes: Census Regions; Census Divisions; States; and Core-based Statistical Areas (CBSAs). The software formats by digits (e.g., number of decimal places), percentages, scientific notation, and multipliers. Finally, a customizable report is available using simple Markdown syntax capable of producing Word, pdf, and html reports. A template file allows the user to arrange the analysis to meet the needs of the users.

## Chapter 4: The New NHTS Online Tool

**DANIEL JENKINS**

*FHWA, presiding*

**JENNIFER MURRAY**

*Wisconsin DOT, recording*

**HO-LING HWANG**

*Oak Ridge National Laboratory*

**APARA BENERJEE**

*MacroSys LCC*

This session provided an overview of the data analysis tools available from Oak Ridge National Laboratory for all NHTS data users. Audience members received a hands-one demonstration of the analytical tools developed to allow the NHTS add-on partner agencies to analyze their data at different levels of geography (state, Metropolitan Planning Organization (MPO), or county).

### **PANEL**

*Ho-ling Hwang and Aparaj Benerjee*

### **Background**

The NHTS data analysis and dissemination website (available at <https://nhts.ornl.gov/>) developed to accommodate the 1995, 2001 and 2009 NHTS data, is being revamped to provide a new, more modernized, user-friendly interface for the 2017 dataset. In addition to the current advancements, there are plans to add data visualization functionality. The original site is still available under the “legacy” tab for those who want to access the earlier datasets. Future improvements intend to consolidate all the datasets for analysis. The underlying plan to make continuous improvements on the site with new features, particularly responsive to the community of users. The new look is intended to guide users to exactly what they are looking for on the front page, with easy to follow instructions and documentation.

### **How It Works**

The dataset is available for immediate download or for guided analysis using the built in tools. The overall statistics represent the results for the entire nation. The research team examined the sets of analysis performed by data users on the 2009 NHTS dataset to determine the questions most frequently asked, and queries most often performed, as a starting point for the new website. These sets of analyses are now available for use with the 2017 data. The data files are accessed through easy to understand icons under frequently used national statistics. The documentation is available, including the Codebook that includes the data about the data, definitions, variable names, derived

variables, and other valuable information for users. The Codebook is also downloadable as an Excel file so users can reorganize the information for their particular needs. On the bottom of the front page is the access to the Data Explorer (previously the Table Designer). The data analysis steps follow the same assembly approach for a user-defined analysis. The analysis variable, the row variable and the column variable are chosen based on the needs for the analysis. In addition, a new button is now available that will allow for the selection of categories. At this time, the categories are hard-coded, but in the future, the user will be able to make customized categories. Also available on the website is the most recent version of the compendium of uses and the announcement of the NHTS data challenge being hosted by the NHTS Task Force. For add-on State DOTs and MPOs, there is a special login that allows access to their own data, with tool functionality for analysis similar to the national tools. The national dataset uses a set of national weights, while the add-ons have their own calculated weights, with the national weights covering all 7 days, and the add-ons having the choice of 5 days (weekends) or the national 7 day weights. Table 4.1 contains examples of analysis variable options.

**TABLE 4.1 Analysis Variables Available Through Online Software.**

<b>Analysis Options</b>	<b>Analysis Variable</b>	<b>Definition/Calculation</b>
Counts	Households	Total Households from Household File
	Persons	Total Persons from Person Files
	Vehicles	Total Vehicles from Vehicle File
	Drivers	Total Drivers from Person File (where driver = 1)
	Workers	Total Workers from Person File (where worker = 1)
Vehicle Travel Measures	Average Annual Miles per Driver	Means of 'YEARMILE' for Drivers from Person File
	Average Annual Vehicle Miles per Vehicle (self-reported)	Mean of 'ANNMILES' from Vehicle file
	Average Vehicle Age (years)	Mean of 'VEHAGE' from Vehicle File
Travel Day	Annual Person Miles of Travel (Travel Day PMT)	Sum of 'TRPMILES' from Trip File
	Annual Person Trips (Travel Day PT)	Total Trips from Trip File
	Average Person Trip Length (miles Travel Day)	Mean of 'TRPMILES' from Trip File
	Average Person Trip Duration Trips (Travel Day PT)	Mean of 'TRVLCMIN' from Trip File
	Average Vehicle Occupancy	Sum of 'NUMONTRP' for drivers when transportation mode is privately operated vehicles (POV) or rental cars from Trip File
	Annual VMT (Travel Day VT)	Sum of 'TRPMILES' for drivers when transportation mode is POV or rental car from Trip file
	Annual Vehicle Trips (Travel Day PT)	Total Trips for drivers when transportation mode is POV or rental car from Trip File
	Average Vehicle Trip Length (Travel Day VT)	Mean of 'TRPMILES' for drivers when transportation mode is POV or rental cars from Trip File
	Average Vehicle Trip Duration (Travel Day VT)	Mean of 'TRVLCMIN' for drivers when transportation mode is POV or rental cars from Trip File

The research team invited everyone to participate in on-line analysis by following the instructions and choices illustrated below. These examples offer a starting point for any analysis that the users would like to run (see Tables 4.2 to 4.8). Special attention to the codebook definitions, and a clear understanding of the question to be answered, are critical aspects of the analysis process.

**TABLE 4.2 Instructions for Producing the Distribution of Total Trips.**

<b><i>Distribution of Total Trips</i></b>
1) Total Person Trips by Mode-choice
Analysis Variable: Annual Person Trips (Travel Day PT)
Row Variable: TRPTRANS
2) Total Person Trips by Mode- choice for Census Division
Analysis Variable: Annual Person Trips (Travel Day PT)
Row Variable: TRIPTRANS
Column Variable: Census D
<i>Select from the List of Statistics what you want to run:</i>
Sum
MOE
Row Percentage
Exclude Missing

**TABLE 4.3 Instructions for Producing Trip Distance.**

<b><i>Trip Distance</i></b>
1) Share of Person Trips by Trip Distance
Analysis Variable: Annual Person Trips (Travel Day PT)
Row Variable: TRIPMILES
Column Variable: Urburu
<i>Select from the List of Statistics what you want to run:</i>
Sum
Column Percentage
Exclude Missing
<i>To analyze Total Trips by Trip Distance and Mode, use Column Variable: Mode</i>
2) Average Person Trip length by Purpose of the Trip (Process: Mean of Tripmiles by Trip Purpose)
Analysis Variable: Average Person Trip length (miles Travel Day)
Row Variable: WHYTRP1S
3) Average Person Trip length by Purpose of the Trip by Area
Analysis Variable: Average Person Trip length (miles Travel Day)
Row Variable: WHYTRP1S
Column Variable: Mode
Page Variable: Urbrur

**TABLE 4.4 Instructions for the Calculation of Vehicle Trips, Age, and Occupancy.**

<b><i>Vehicle Trips, Age, and Occupancy</i></b>
1) Annual Vehicle Trips by Trip Distance
Analysis Variable: Annual Vehicle Trips (Travel Day PT)
Row Variable: TRIPMILES
2) Annual Vehicle Trips by Age Group & Gender
Analysis Variable: Annual Vehicle Trips (Travel Day PT)
Row Variable: R Age
Column Variable: R Sex
3) Average Vehicle Age by Vehicle Type
Analysis Variable: Average Vehicle Age (years)
Row Variable: Vehtype
Column Variable: HHRAMINCE (optional for further analysis)
Select from the List of Statistics what you want to run:
Mean
Exclude Missing
4) Average Vehicle Occupancy by Vehicle Type
<i>Process: Mean of NUMONTRP – For those trips where respondent drove in POV or Rental Car and Tripmiles is more than 0. The weight is TRPMILES*WTTRDFIN</i>
Analysis Variable: Average Vehicle Occupancy
Row Variable: TRPTRANS
Select from the List of Statistics what you want to run:
Mean
Exclude Missing

**TABLE 4.5 Instructions for the Calculation of Daily Trip Rates by Gender.**

<b><i>Daily Trip Rates by Gender</i></b>
<i>Process: (Total Trips/Total Persons) /365</i>
<i>= (371,151,971,524/301,599,169)/365</i>
<i>= 3.4</i>
Daily Trip rate by Gender
Step 1:
Analysis Variable: Annual Person Trips (Travel Day PT)
Row Variable: R Sex
Step 2:
Analysis Variable: Perons
Row Variable: R Sex
Daily Trip Rate: Step 1/Step 2/ 365



**TABLE 4.6 Instructions for the Calculation of Average Daily Person Miles per Person by Gender.**

<b><i>Average Daily Person-Miles per Person by Gender</i></b>	
<i>Process: (Sum of TRPMILES/Total Persons)/365</i>	
<i>= (370, 904,198,095/301,599, 169)/365</i>	
<i>= 36.07</i>	
Step1	
	Analysis Variable: Annual Person Miles of Travel (Travel Day PMT)
	Row Variable: R_Sex
Step 2	
	Analysis Variable: Persons
	Row Variable: R_Sex
Average Daily Person-Miles per Person by Gender: Step1/Step 2/365	

**TABLE 4.7 Instructions for the Calculation of VMT per Driver by Age Group.**

<b><i>VMT per Driver by Age Group</i></b>	
<i>Process: Sum of Tripmiles by AgeGroup/Total Drivers by AgeGroup (where DRVR-FLG equals '01' and tripmiles &lt;=0 and (TRPTRANS &gt;'02' and TRPTRANS &lt; '07' or TRPTRANS equals '08' or TRPTRANS equals '09' or TRPTRANS equals '18')</i>	
Step 1	
	Analysis Variable: Annual VMT (Travel Day VT)
	Row Variable: R_Age
Step 2	
	Analysis Variable: Drivers
	Row Variable: R_Age
<i>Select from the List of Statistics what you want to run:</i>	
Exclude Missing	
VMT per Driver by Age Group: Step 1/Step 2	

**TABLE 4.8 Instructions for the Calculation of Average Annual Miles Per Driver.**

<b><i>Average Annual Miles per Driver</i></b>	
<i>Process: Mean of YEARMILE by AgeGroup (where yearmile &gt;= 0 and driver = '01')</i>	
Analysis Variable: Average Annual miles per Driver	
Row Variable: R_Age	
<i>Select from the List of Statistics what you want to run:</i>	
Mean	
MOE	
Exclude Missing	

In the future, add-on partner agencies will be able to choose their own region for analysis including county level and customized regions. Most users have chosen the boundaries of the MPO for their customized geography. The geography in the public file is limited to the household state and large Metropolitan Statistical Area (MSA) code.

## **Audience Dialogue Highlights**

Questions on which set of tools should be used to analysis the NHTS 2017 dataset generated a lively discussion. The right choice will depend on the needs and skill level of the user. The key is making sure whatever analysis tool is used, that it is used correctly (e.g., using R software, on-line tools, or commercial statistical packages). There are certain procedures in SAS, for example, with well-established batch code files. Some of the specialized NHTS tools are being designed explicitly with procedures to prevent users from producing erroneous analysis. MOE calculations applied by an experienced user in commercial software is one option, while a relatively new user might opt to use the on-line NHTS tools with this featured included in the processing procedures. A systematic analysis of common errors found users who incorrectly chose the same variable for both the row variable and the column variable, creating a diagonal statistic. Changes are scheduled in the on-line NHTS to reduce this error for occurring.

Variable definitions and reliance on the Codebook are important aspects of using the NHTS. The WHYTRP90 (trip purpose) uses the consistent definition of the trip from 1990, making it possible to compare the NHTS datasets over time. Extended discussions on how variables are defined helped to point out the need for using the Codebook and making certain that the correct variable is included in analyses. There were questions regarding the definition of urban and rural and a need for clarification on the definition that matches the Census definition of Urban (2,500+) or as Urbanized (50,000+). It would be helpful if there was a variable (or flag) to assist in breaking out areas that are Urbanized (50,000+) by themselves for analysis. In addition, State DOTs and MPOs with add-on data could consider developing a summary of methodologies for using various NHTS tools, (e.g., R software or on-line tools) with recommendations for matching desired analyses with available methods, pros and cons, etc.

## **Chapter 5: Using the NHTS for Performance Measures**

**LEI ZHANG**

*University of Maryland, presiding*

**HABTE KASSA**

*Georgia DOT, recording*

**WENJING PU**

*FHWA*

**SUBRAT MAHAPATRA**

*Maryland DOT*

**KEITH L. KILLOUGH**

*Arizona DOT*

**GEENA MASKEY**

*Texas DOT*

NHTS provides a variety of data variables that can directly, or indirectly, support transportation performance management. The NHTS transition to an annual survey format and the integration with Big Data promise more applications in this domain. This panel focused on lessons learned, best practices, and future opportunities on NHTS-based performance measures.

### **PANEL**

*Weinjing Pu, Keith L. Killough, and Geena Maskey*

The new Federal Transportation Performance Management (TPM) requirements, and some additional performance measures, are driving new interest in performance measures. The TPM program has seventeen measures, with explicit equations and calculations for meeting the requirements. The congestion-related measures include: Vehicle Occupancy as a reliability measure; vehicle occupancy in the Peak Hour of Excessive Delay Measure; and the Percent of Non-Single Occupancy Vehicle (Non-SOV) Travel. In the future, there are plans for Vehicle Occupancy and Percent of Non-SOV travel metrics from the NHTS at the State and Urbanized Area levels, rather than the current national metrics.

*Subrat Mahapatra*

VMT in Maryland is at an all-time high (60 billion in 2017). It is one of the most congested regions in the US and has oversaturated conditions that lead to higher unreliability. Maryland has 6.1 million people, with an expectation of an additional million in the next 20 years. In addition, experts expect a 30% increase in overall VMT, with truck VMT expected to double by 2040.

These conditions are extremely challenging for the Maryland Department of Transportation (MDOT). Their customers' needs are more diverse and are increasingly wide-ranging with respect to demographic shifts, an aging population, and particular mobility needs. At the same time, travel behavior and decisions are changing rapidly as people choose where and how they will live, work, shop and travel. Even more challenging are the impacts of the new technologies (e.g., transportation network companies [TNCs], CAVs ATM/ATDM, freight movement). The infrastructure in the region is aging and maintenance must compete with other funding needs. MDOT has made a decision to focus now on system efficiency and reliability rather than the traditional "average day" planning paradigm. Their focus is on providing a safe, reliable and efficient door-to-door travel experience, fostered by TSM&O implementation to provide customers with surety that they can depend on the system for their choices and needs (e.g., destinations, departure time, route, mode). Communication is key to being able to tell the important story of MDOT.

MDOT uses a three-tiered decision-making context, with Level I aimed at strategic planning, Level II covering the transportation system management and operations, and Level III for operations. MDOT already has a set of performance measures that they calculate and communicate to the public and decision-makers. MDOT reports traditional measures that include: ADT/VMT; Vehicle Person Delay/Congested Miles; Vehicle/Person Throughput; Average Incident Clearance Times' and Annual User Savings. Newer measures include: Accessibility/Connectivity for all modes; Reliability for autos, transit and trucks; Market Segments (commuters, businesses, freight); Quality of Life/Sustainability; Economic Indicators; and Greenhouse Gas (GHG) Emissions. At this time, MDOT uses NHTS data for travel behavior and characteristics by using certain variables. They use a set of parameters including: trip making (O/D, commuting patterns), mode choices, vehicle ownership, trip lengths by purpose, travel costs, preferences, etc. They also use NHTS to estimate and validate traveling public's responses to policy changes (e.g., tolls/fares, capacity improvements, incentives, services). And finally, they use the NHTS data to calibrate and validate their travel demand models.

Recently, they have moved forward with using Big Data applications that estimate O/D matrices for a variety of geographies including: state, county, zip code, and Traffic Analysis Zone (TAZ). These data are used to calibrate and validate statewide travel demand models and newer integrated models. The data can be used to compute trip-based performance measures [Texas A&M Transportation Institute (TTI)/PTI] and to determine the distributions of traffic along major corridors, routes and activity centers. In addition, these data are used to estimate traffic volumes and to develop accessibility measures (e.g., peak/off-peak by mode). MDOT are convinced that conventional data sources and traditional surveys aren't sufficient for their needs and see Big Data as a window into their customer's needs. At the same time, Big Data applications for transportation decision-making are in the infancy and need champions, investment and commitment in order to become mainstream options. The future will require collaboration between agencies (DOTs, MPOs), private sector data providers, and the research community, including TRB and university talent. TSM&O and Performance Management are expected to be the first areas to demonstrate progress. There is a need for advanced research to combine traditional data sources with Big Data for travel behavior analysis, requiring a demonstration of the value added and sustainability of such an approach.

**Audience Dialogue Highlights**

The home-based non-work trip rate was noted as declining between 2009 and 2017 and it was echoed that more investigation would be helpful to find out why. It was also suggested that maybe more people are getting their shopping delivered directly to their homes.

Regional travel modelers use NHTS data to feed their travel demand model for performance measures as well as project prioritization. One of the issues for using the NHTS is related to the frequency of the NHTS data availability and the performance measure requirements calculated comes every two years.

## Chapter 6: How the New Methods Affect Me

**DANIEL CHATMAN**

*University of California, Berkeley, presiding*

**PATRICK COLEMAN**

*AECOM, recording*

**STACY BRICKA**

*MacroSys LCC*

**SOHEILA KHOIL**

*California DOT*

**PAUL SCHROEDER**

*EurekaFacts, Inc.*

This session covered professional reactions to the new data collection methods implemented in the 2017 NHTS. Experts discussed how they used NHTS for different purposes with or on behalf of transportation agencies. Discussions included the use of new methods and data structure that may affect agencies and other users.

### **PANEL**

*Stacy Bricka, Soheil Khoil, and Paul Schroeder*

### **What Are the Impacts of the New Methods?**

*Soheila Khoil:* The new method for calculation of trip distance reduces outliers. It also is now possible to include households without a landline in the sampling frame. Using the NHTS, statewide mode shares will be determined and trip rates calculated. There is interest in understanding the build environment effects across places in California.

*Paul Schroeder:* The changes were long overdue, to move from RDD using landlines to ABS, in time for the 2017 deployment. Back in 2006 (prior to the 2009 deployment), we didn't know enough yet to try to using a dual frame telephone sampling method and didn't know how to weight it. Only 23% of the households were wireless and they were younger, renters, and living with unrelated roommates. To preserve trends now, we had to change the method as no one answers the phone and response rates would be in the single digits, regardless). In addition, ABS reduces coverage bias.

*Stacey Bricka:* Who downloads data versus using the online tool because the online tool will tell us who is summarizing what and how can improve the dataset. Only a few people have downloaded the user's guide and tech memo and questionnaire. People should use these resources when they use the tools on-line or download the data directly. It is important to use

other data sources (e.g., HPMS, ridership counts, ATUS) to help validate/calibrate your findings from the NHTS.

### **What Are the Important Differences Between 2009 and 2017?**

*Soheila Khoii:* The difference in the definition of trips for walking and biking (loop versus two directional trips) may impact walking and bike trips. Using the Google API for trip distances is an improvement over self-reported trip length. The new database is rich overall, but we need more demographic research to establish and confirm trends.

*Paul Schroeder:* Overall, the data is better, the coverage bias is minimal, and weighting procedures are cleaner. Using ABS for the sampling strategy alone is leaps and bounds over an RDD approach, making the data more trustworthy and reliable.

*Stacey Bricka:* Response rates have been falling for a long time. Over time, we have equated response rates to data quality. Yet, today, response rates may no longer be the appropriate indicator of survey goodness. For our purposes, the data needs to meet key demographic groups compared to the Census and needs to have sufficient transit trips for modeling. Representativeness is important to make sure we have the right population, separate from just getting a high response rate. We need to note that when we now mail out one million letters and get 100,000 responses, it is a smaller response rate than if we called 30,000 phone numbers. The data quality is now higher, the item number of non-responses is better, and although demographically, it isn't perfect, it is much better in its representativeness. We need to rethink the measures that we use for determining success. These measures would be more comprehensive and flexible than just response rates.

*Paul Schroeder:* Following this line of thought, if we have better data, does that impact how we conduct our analysis? If the data is now more representative, more accurate, should we change our research questions?

*Stacey Bricka:* It is important to understand demographic representativeness and behavioral representativeness. Often, we have a sample that is a good cross section with regards to demographics. Nationally, we don't sample for behavior (e.g., transit riders, walking and biking). The sample is not purposeful and we know that not everyone with the same demographics behaves in the same way because they live in different places (with different travel options). Other dimensions of representativeness that needs to be considered in a representative survey. "Who takes the survey?" "Who travel a particular way?" Both of these questions are important to travel surveying efforts.

*Paul Schroeder:* Caution should be used if you are using 2009 NHTS to research young travelers. The survey methodology was well suited for persons over 50, but not younger individuals based on what we know about the deployment.

*Group Panel Discussion:* Concerns were expressed regarding unobserved heterogeneity when using conventional demographic measures when we can't control for behavior, lifestyle or motivations. These factors are outcomes, and can't be associated with Census data. This issue is also complicated with the way people use a particular mode (e.g., transit trips for transit dependent or choice riders, numerous shopping behaviors). What are we doing about randomness? Using it would increase representativeness, making unobserved heterogeneity less of an issue. Although it has always been difficult to gather data with a truly random sample, it is even more difficult today. It might be more appropriate to use total survey error as a measure of survey goodness than response rates.

## What Changes to Data Collection Methods Do You Think Will Be Necessary for NHTS?

*Soheila Khoii:* In the NextGen program there should be more samples taken more frequently. The survey should include TNCs, thought to impact short walking and biking trips. Weighting factors specific to California and land uses and place types will need to be considered.

*Paul Schroeder:* ABS frame is still the strongest approach for a sample, but Big Data streams could be used before going into the field rather than being appended afterwards.

*Stacey Bricka:* It would be better to have a shorter survey, supplemented with Big Data. While it is difficult to use oversampling for the national sample, future surveys need to be more aggressive in targeting hard to reach populations. If we want to use the data to do planning for these populations, we need make sure we include them in the sample. We need to remain flexible, using smartphone apps and traditional telephones so we still capture the elderly in the survey, as well as younger members of the population. Our surveying options should make sense for society.

## Audience Dialogue Highlights

*If you shorten the survey, how many questions should be included?* Currently, there are 600 variables, including 27 “core” questions. We need to better understand the usage of each of the variable to determine its essential nature. The online analytics tool keeps track of which variables are accessed automatically. Scans of the literature and reports provide information on NHTS uses, including the NHTS compendium produced every 6 months. It is important for research and transportation planners to make sure their uses are recognized. Although it is assumed that respondents experience a survey burden, in fact, individual participants are unaware of other survey options and may not perceive a burden. The number of questions is not necessarily an obstacle. However, the Office of Management and Budget (OMB) is concerned about respondent burden and complicated questions require justification or risk being removed. The survey could be designed to include only core questions.

*If high resolution sensor data were being considered (e.g., smartphone travel survey apps), should we consider a middle ground between big data and traditional surveys?* There are concerns that anonymized Big Data can't be associated with any socio-demographic detail, and traditional/web-based travel surveys aren't able to capture probe data directly. Smartphone survey apps generate high resolution spatial data normally collected along with a demographic profile of the respondent. The smartphone/GPS method reduces the burden of reporting trips and attitudinal questions can be combined in the dataset. NHTS has been compared to smartphone/GPS data for validation. Regional travel survey efforts are using this technology. The NHTS 2017 deployment didn't include a GPS component.

*Could more be done to analyzing historical data (1990, 1995) to identify problems, cautions, and pitfalls to avoid?* One approach would be to look for other benchmark data and see how they compare over these time periods. However, it all depends on what questions you are trying to answer. Researchers and practitioners can look at HPMS, American Community Survey (ACS), etc., to see how well they match findings from the NHTS and other reports. The 1990 deployment was a one-stage recall methodology, while the 1995 deployment used a two-stage approach, using a diary to assist recall efforts.

*Are there long trips in the add-ons and how do you measure days you are out of town?* The only reference mentioned was the California add-on that included two questions regarding



long distance travel in order to better understand travel behavior and high speed rail travel (e.g., poster session of outcomes conducted by Caltrans). Capturing long distance travel and other idiosyncratic trips is difficult. First, we have people who will respond to any survey, then we have those who can be convinced to participate, and finally we have those who won't ever be survey respondents. To get a representative sample for all our questions may require making special outreach efforts to community representatives who can help encourage their populations to respond. For example, to increase participation of the Hispanic and Latino populations, community leadership could announce an upcoming mailing or phone contact. Social media might be a strategy for reaching out to special populations.

*What is known about small areas and small samples? Are there any changes in weighting methods we should be aware of?* For the national sample, weights were developed at the division level. Applicable to both add on/not add on. The weights are based on demographics and geography, not behavioral (e.g., no mode-specific weights).

*Are any adjustments in weighting to turn loop trips into round trips needed?* There are concerns over how to deal with differences in the 2009 and 2017 walking and bike trips. Recommended practice suggests not dividing 2017 data, but rather to go back 2009 and combine the two legs of the trip, turning them into a loop trip.

## **Chapter 7: Alternative Fueled Vehicles**

**LEI ZHANG**

*University of Maryland, presiding*

**HABLE KASSA**

*Georgia DOT, recording*

**ARASH ASADABADI**

*University of Maryland*

**BEHRAM WALI**

**ASAD KHATTAK**

*University of Tennessee, Knoxville*

**YAN ZHOU**

**VYAS ANANT**

*Argonne National Laboratory*

The last decade has seen significant growth on the alternative fuel, hybrid, and electric vehicle market. This session explores the NHTS 2017 for information on ownership and usage patterns of these greener vehicles; who owns them; are they used in different ways compared to traditional vehicles; and what is the future market demand?

### **REVISITING VEHICLE OWNERSHIP IN ACTIVITY-BASED MODELS CONSIDERING ALTERNATIVE FUEL VEHICLES**

*Arash Asadabadi and Lei Zhang*

#### **Background**

The number of motor vehicles available to a household has a major impact on the travel behavior of the members of the household. As a result, many MPOs have incorporated models of household vehicle “availability” or vehicle “ownership” into their activity-based travel model systems. These models relate the number of vehicles available to a household to explanatory household person, zonal, and transportation variables. Conventionally, vehicle ownership models used in Activity-based Models (ABMs) do not take differences in vehicle characteristics into consideration. This practice can be misleading for future scenario models applied to future when the fleet pattern could be dramatically different.

Emerging alternative fuel vehicles in the auto market has been rapidly changing the vehicle fleet, changing the household’s vehicle ownership pattern. As of early 2017, global cumulative electric vehicle sales had surpassed 2 million units (1). The latest Bloomberg energy finance forecast says by 2040, the electric vehicles sales will outpace internal combustion engine vehicles (2). Vehicle ownership models are usually estimated based on household travel surveys

lacking sufficient information for alternative fuel vehicle's ownership. Therefore, these models do not take the variety of vehicles into account. Alternative fuel vehicles have various differences comparing to conventional vehicles such as price, operation cost, and maintenance cost. Therefore, in addition to model the number of vehicles each household owns, it is essential to find out the type of each vehicle.

## Methodology

NHTS 2017 data is the first national-level survey that has valuable and sufficient information on alternative fuel vehicle ownership. The main objective of this research is to use this dataset and develop a vehicle ownership model, which can account for the fuel type, demonstrating how this model can be applied to an ABM as a complementary model for current vehicle ownership models.

## Preliminary Research Findings

NHTS 2017 is the main dataset used for this research. It has valuable information regarding the fuel type of vehicles. Table 7.1 shows the significant change of alternative fuel vehicles from 2009 to 2017 in NHTS data.

**TABLE 7.1 Comparison of Different Fuel Type Vehicles in NHTS Data.**

Fuel Type	Total Number of Vehicles in US	
	2009 NHTS	2017 NHTS
Gasoline	208,744,732	211,899,839
Diesel	2,648,210	5,441,249
Electricity	7,458	4,766,108

## A CASUAL APPROACH TO DISENTANGLING VEHICLE TECHNOLOGY AND SELF-SELECTION EFFECTS ON HOUSEHOLD VEHICLE USE: A TRI-VARIATE COPULA-BASED ENDOGENOUS REGIME SWITCHING FRAMEWORK

*Behram Wali and Asad Khattak,*

## Background

With the continuous advances in vehicle technology and increasing market penetration of alternative fuel vehicles (AFV), there exists a need to understand the travel behavior implications of advanced technology vehicles. In particular, what are the mechanisms/factors leading to households purchasing AFVs (electric vehicles in particular)? In addition, what are the vehicle-use patterns of AFV households, compared to their non-AFV counterparts? That is, do AFV households travel more or less than their non-AFV counterparts in terms of daily miles traveled? This is a complicated issue. On one hand, from an infrastructure and technology standpoint, it seems reasonable to expect that AFV households may travel less owing to limited charging infrastructure and vehicle range. Contrarily, AFVs have significantly higher miles per gasoline equivalent (and lower fuel costs), and which may lead to AFV users traveling more than

conventional vehicle users. This is further complicated by the inclusion of behavioral factors. A key hypothesis in this regard is that households may base their vehicle choice based on their travel needs, and which indicates that observed vehicle-use differences (if any) between AFV and non-AFV households cannot be attributed entirely to the vehicle technology. To summarize, how much of the observed differences in vehicle-use (once quantified) between AFV and non-AFV households can be attributed to self-selection effects and how much to “true” vehicle technology effects?

## **Methodology**

This study will use 2017 NHTS data, a nationally representative sample that includes new information on electric vehicles and other AFVs make it highly relevant to the research questions identified. Methodologically, answering the questions posed in this study requires capturing the complex interdependencies between household vehicle choice decisions and subsequent vehicle use in terms of daily miles traveled. As such, we will develop a new and innovative copula and mixture marginal modeling based tri-variate endogenous regime switching framework to model simultaneously household vehicle choice (AFV/Non-AFV), daily miles traveled given an AFV household, and daily miles traveled given a non-AFV household. The dominant but rather restrictive and inappropriate approach in the self-selection literature to joint-modeling is the assumption of bivariate (or multivariate normality), and which implies a linear and homogenous form of stochastic dependencies between household vehicle choice decisions and subsequent vehicle use. The researchers employed a broad spectrum of innovative elliptical, Archimedean, and rotated survival copulas for the complex dependence structures and different data-driven plausible distributions for the marginal.

## **Preliminary Research Findings**

The study enhances our understanding about the travel behavior implications of AFV. The methodology used highlights the power of new copula approaches to quantify vehicle technology effects on household travel behavior, and to help understand if the differences in AFV and non-AFV household vehicle-use are truly due to vehicle technology itself, or self-selection, or a combination of both. Compared to the proposed approach, the inferences/effects obtained through traditional self-selection modeling approaches can be misleading, and can in fact indicate absence of self-selection effects altogether.

## **TRACKING NATIONAL HOUSEHOLD VEHICLE USAGE BY TYPE, AGE, AND AREA IN SUPPORT OF MARKING ASSESSMENTS FOR PLUG-IN ELECTRIC VEHICLES WITH EMERGING USAGE OF SHARED SERVICES**

*Yan Zhou and Vyas Anant*

## **Background**

Plug-in electric vehicle (PEV) technology promises to provide a reduction in oil use, improvements in local air quality, and/or possibly reductions in GHG emissions to support a sustainable transportation system. However, the higher purchase price of PEV requires high usage rates to pay off the investment in the technology. According to the 2009 NHTS, about 40%

of household vehicles were not used on the survey travel day. In recent years, another emerging technology, ride sharing or car sharing, is booming, which may further reduce the usage of private vehicles.

## Methodology

The authors analyzed why household vehicles were not used in a given travel day to assess the market potential for Plug-in hybrid electric vehicles published results in 2013. Because 2017 NHTS added questions about usage of rideshare or car share service in the past 30 days, besides of PEV technology, this time we are able to investigate the correlation between “vehicle not used” and “usage of shared service”. Vehicles used on survey day with or without a reported travel time and distance in the survey are considered “vehicles used.” All others are referred to as “vehicles not used.” We divided the “vehicles not used” into three categories/reasons: (1) left at home while other household vehicles were used, (2) not used because travelers used other modes, and (3) no household trips. Within each category, we will conduct further analysis by separating used and not-used vehicles by vehicle type, age group for both weekdays and weekends. In addition, vehicle usage are compared in both MSAs and non-MSAs.

Our earlier research results using 2009 NHTS indicate that most vehicles — especially pickups— are not used because the households own and use other vehicles. Moreover, SUVs — especially newer SUVs ( $\leq 10$  years) — are the most utilized vehicle type and should be strongly considered as a primary vehicle type for PEVs, in addition to cars. We would expect this time using 2017 NHTS the share of second category (not used because travelers used other modes) increases from 2009 survey results due to use of ridesharing or car sharing, especially in MSAs.

## Preliminary Research Findings

We analyze the 2017 NHTS data to assess: 1) the reasons vehicles are not used on the survey day, and 2) usage rates of four major vehicles types (cars, SUVs, vans, pickup trucks). We further analyze how vehicle usage varies by vehicle type, age group, population density, MSA/non-MSA for both weekdays and weekends. In the end, we analyzed the frequency of taking taxi and rideshare for each vehicle usage category. The purpose of this research is to identify the market potential of AFV which rely on high usage to pay off the high purchase price and also identify the correlation between vehicle not used and likelihood of using shared ride service. Study conclusions include the following:

### Vehicle Usage

- Overall vehicle usage increased in the 2017 NHTS, compared to the 2009 NHTS.
- The main reasons for this increase are: 1) the usage rate of new vehicles increased more than that of older vehicles, especially cars, 2) % of vehicles “Left at home” and household members used “Other Modes” decreased sharply, especially in MSAs, 3) a sharp increase in pickup usage.
- People in MSAs tend to own more new cars than other vehicle types. Rural areas do not show significant difference between ownership of cars and other vehicle types.
- Vehicle not used because household members used “Other Modes” declined significantly from 2009 survey results, across all vehicles types in both MSA and Non-MSA.

- SUVs and vans are the most-used vehicle. Pickup trucks are most common as an extra vehicle; their everyday use is least probable.
- The usage rate and ownership of all vehicle types decrease as the density increases.
- More cars and SUVs were used on weekdays than weekends among all of the population density groups. However, such difference is minimal in the two highest density categories.
- For both cars and SUVs, the probability of daily use in the most densely populated zones is much lower than in the rest of the areas.
- MSA shows much higher usage of ride/car share than non-MSA. Those who used “Other Modes” on the travel day have much higher probability of using taxi/ride share at different frequency level.
- The “Left at home” group has the lowest probability of using taxi/ride share further indicates that such households own multiple vehicles and are less likely to use taxi or ride share service.

### Vehicle Age

- Most vehicles, especially older ones, not used in the survey are “left at home” because household members own other vehicles or because multiple household drivers ride together in one vehicle on a given day.
- Newer vehicles are used much more, especially SUVs in an MSA. For any of the four vehicle types, usage rates within MSAs are higher than those outside MSAs.
- Cars, especially those  $\leq 10$  years in age, are the most-owned vehicle type, because of their much higher ownership rates in MSAs.

Pickup trucks are used much less intensively and last much longer. Their longer lifetime raises their share of the >10-year age group. Pickups are also far more likely to be “left at home” while household members use other vehicles to travel, particularly in rural areas. Thus, period required to pay off a battery pack is the longest for pickup trucks, and pickup trucks are the most likely vehicle type to outlast the calendar life of and warranty period for plug-in electric hybrid vehicle battery packs. Pickup truck longevity makes the odds of needing a battery replacement much higher than those of any other class of vehicles.

Given the findings of this analysis, considering only probability of vehicle use each day, the greatest market potential for PEVs lies in cars and SUVs in low-density areas of MSAs (i.e., suburbs). Interpreting the density results, at locations where high-rise multi-family residences result in high average population density (and the possible viability of public transit), the daily use rates of vehicles are much lower than elsewhere, reducing the probability of payback of battery packs through frequent charging.

Since plug-in vehicles are relatively more economically attractive in MSAs, the small share of pickup trucks in MSAs shows that pickup trucks are a considerably smaller market than the more numerous cars and SUVs. Design priorities for pickups would logically carry a higher priority based on the operation characteristics outside of MSAs. Even when pickup trucks are found in MSAs, their rate of use is much less than that for other vehicle types.

### **Audience Dialogue Highlights**

Due to an increase in using alternative fueled vehicles travel demand modelers should revisit vehicle ownership in ABMs. The data indicates a large increase in alternative fuel vehicle ownership between 2009 and 2017. The total share of AFV in 2017 is 2.1% and it is expected to increase to 33% by year 2040.

The high purchase value of alternative fueled vehicles requires a high usage rate to pay off the investment. Generally, there is a correlation in states where there is some sort of incentive for alternative fueled vehicle and the number of alternative fueled vehicle ownership. Only higher income household can afford alternative fueled vehicles. There is a significant correlation between the higher educated households and alternative ownership. SUVs and vans are the most used vehicles in the survey and SUV are the second biggest market share for plugin vehicles.

With regards to the distribution of the use of different types of vehicles among households, the middle ground between the very densely and scarcely populated areas is where there were more vehicles available.

## **Chapter 8: Nonmotorized Transportation**

**JANICE MACHADO**

*Westat, presiding*

**LISA AULTMAN-HALL**

*University of Vermont, recording*

**JINA MAHMOUDI AND LEI ZHANG**

*University of Maryland*

**RALPH BUEHLER**

*Virginia Tech*

**JOHN PUCHER**

*Rutgers University*

**DAFNA MEROM**

**ADIRAN BAUMAN**

*University of Sydney*

**THEODORE MANSFIELD**

*ORISE-USDOT*

**JIM SULLIVAN**

**SARAH HOWERTER**

*University of Vermont*

Every NHTS cycle brings improvements in how we measure all types of nonmotorized transportation, including bicycling and walking. This session highlights the 2017 bicycle and pedestrian survey data, with early analysis related to health, built environment, and affinity groupings.

### **HEALTH IMPACTS OF NONMOTORIZED TRAVEL BEHAVIOR AND BUILT ENVIRONMENT**

*Jina Mahmoudi and Lei Zhang*

#### **Background**

Research on the role of travel behavior in public health has been often limited to the use of multiple and different databases on travel behavior and health outcomes. This has been due to the lack of a database—particularly at the national level—that provides travel-related and health-



related data, collected simultaneously from the same survey respondents. Consequently, researchers probing the impact of travel behavior on health have resorted to either: a) conducting their own surveys to collect data on both travel behavior and health outcomes; or b) fusing travel survey and health survey datasets to create combined databases. Both of these approaches are disadvantageous; the former is costly, and the latter does not provide consistent data on travel choices and health status collected from one survey sample. Therefore, research on the link between travel behavior and health outcomes can greatly benefit from a database that provides data on both trends.

The 2017 version of the NHTS offers a tremendous opportunity for resolving the issues of data limitation and data inconsistencies in such research efforts. For the first time in the history of the NHTS, questions regarding health and physical activity levels have been included in the NHTS survey questionnaire. This innovative and much-needed addition allowed for collection of concurrent and consistent individual-level data on travel behavior, health-related behavior and health status. These data can be examined to detect interrelationships between individuals' travel choices, health behavior and general health status.

## **Methodology**

This study employs ordered probit modeling techniques to relate the health status of the NHTS 2017 respondents to their levels of nonmotorized travel behavior and physical activity, as well as the built environment of their place of residence. Since the health status indicators in the NHTS 2017 are in an ordinal format with various categories (i.e., poor, fair, good, very good, excellent), the ordered probit regression model is an appropriate statistical tool to be used in this analysis.

## **Preliminary Research Findings**

Literature suggests that the health of individuals is related to their nonmotorized travel behavior, their health-related behavior and the built environment attributes of their surroundings. Results of the ordered probit models developed in the present study corroborate those prior findings. The results indicate that health status of the NHTS 2017 respondents is related to their socioeconomic and demographic characteristics, their levels of nonmotorized travel and physical activity as well as the built environment characteristics of their neighborhood of residence.

All person-level attributes including the person's age, gender, race, educational attainment, employment status, levels of nonmotorized travel and physical activity show statistically significant correlations with health status of the individual respondents. As expected, older age is associated with lower health rankings (from very good/excellent to good to fair/poor). Further, according to the results, being male is also associated with lower health rankings. The race of the respondents also plays a role in their health status; compared with being White, being of a non-White race is associated with lower health rankings. On the other hand, higher education and being employed seem to be associated with better health rankings (from poor/fair to good to very good/excellent). Individuals' general health status is also linked to their nonmotorized travel behavior (i.e., levels of walking and bicycling trips) and other health behavior such as levels of physical activity. Not surprisingly, higher levels of nonmotorized trips and being more physically active are associated with better health status.

Household-level composition and socioeconomic factors such as household's size, vehicle ownership and level of income also exhibit statistically significant correlations with

health status of individuals. Living in a larger household is associated with lower health rankings, whereas higher levels of household vehicle ownership and income are correlated with higher health rankings. In addition, neighborhood-level control factors are also significant in the models. Better health status is associated with the percentage of neighborhood population that is of working age; however, living in neighborhoods with higher percentages of households that own no private vehicles and higher percentages of low-income workers is associated with lower health rankings. Consistent with past findings, the results of the ordered probit models also confirm the statistically significant correlations between health status of individuals and the characteristics of their neighborhood's built environment.

Higher levels of mixed land use within the neighborhood and higher levels of regional diversity (i.e., deviation of neighborhood ratio of jobs/population from regional average ratio of jobs/population) are associated with higher health rankings. Further, as expected, pedestrian-oriented network density is correlated with better health rankings, whereas automobile-oriented intersection density exhibits a negative correlation with health rankings. In addition, local transit accessibility proves to be an important factor in residents' health status. The coefficient of the variable representing the proportion of neighborhood employment within ¼ mile of transit stops indicates a positive correlation with higher health rankings—meaning higher transit accessibility to employment within the neighborhood is associated with a better health status. Frequency of transit service within the neighborhood does not exhibit statistically significant associations with health status of residents. Also, the coefficient of the variable representing neighborhood compactness (i.e., higher activity density) generally does not reach a statistical significance threshold in the models; nonetheless, it shows a significant negative correlation with higher health rankings in one of the models developed.

With respect to regional accessibility, increased automobile accessibility to employment opportunities (i.e., jobs within 45 minutes automobile travel time) shows a positive and statistically significant correlation with higher health rankings. The coefficient of the variable representing regional transit accessibility to employment opportunities (i.e., jobs within 45 minutes transit commute) does not reach a statistical significance threshold in the models. Better health status is also associated with living in an urban area (vs. in a rural area).

Transportation planners and policy decision-makers can use the results of this study to better understand the roles of walking and bicycling as well as built environment characteristics in human health. Knowledge gained from this research can also help policy arguments for interventions aiming at promoting public health through changing individuals' travel choices and the built environment. As a result, more suitable transportation planning policies and implementation strategies can be developed, and more effective solutions can be implemented to promote the health of individuals, the livability of communities and ultimately, the wellbeing of the society.

## **TRENDS IN WALKING AND CYCLING IN THE UNITED STATES, 2001-2017**

*Ralph Buehler, John Pucher, Safna Merom and Adrian Bauman*

### **Background**

This research extends our analysis of walking and cycling based on NHTS 2001 and 2009 (Pucher, Buehler, Merom, and Bauman 2011) to include 2017 NHTS data. The objective is to

assess changes in walking and cycling in the United States between 2001, 2009, and 2017—at both the trip and person level for persons five years and older.

## Methodology

First, we compute the frequency and duration of walking and cycling per capita using NHTS 2001, 2009, and 2017. In order to do so, we adjust ‘trips’ in the NHTS 2017 to account for the changed definition of ‘loop trips’ (compared to the 2001 and 2009 surveys). For the trip-based analysis, the daily frequency and duration of walking and cycling per capita are calculated by dividing the daily totals for trips by the number of persons, yielding average rates.

Second, we assess the prevalence of walking and cycling using two different measures: (1) any walking or cycling on the assigned travel day and (2) 30 minutes or more of walking and cycling per day. Both thresholds of physical activity have important implications for health benefits. For the person-based analysis, trip duration information for walk and bike trips is aggregated for each trip maker, and then merged with the person data set. The person data set includes individuals who did not make any trips during the travel day (i.e., stayed at home). To include individuals who stayed at home in the walking and cycling prevalence estimates, we assign them to the “no walk trip” and “no bike trip” categories.

For both, the trip-based and person-based analysis, statistical significance is determined by calculating differences in weighted proportions or means between independent samples ( $p < .05$ ). To control for the effects of covariates and possible confounders, logistic regression is used to calculate the likelihood of walking and cycling in 2017 compared with 2009 and 2001, after adjusting for the impacts of other variables.

## Preliminary Research Findings

Preliminary results show that in 2017 the average American made 13 more walk trips than in 2009 and 30 more walk trips than in 2001. In 2017, Americans walked for 37 hours per year (+4hrs since 2001). There was no difference in the average number of bike trips and time cycled in 2001 and 2017. At the population level, the prevalence of “any walking” was 18% in 2001, but then declined to 16% in 2009 and 2017; whereas walking at least 30 minutes per day increased from 7.2% in 2001 to 8.0% in 2009 and 2017. The prevalence of “any cycling” per day was stable between 2001 and 2009 (1.7%) and then declined to 1.5% in 2017. The prevalence of “30 minutes of cycling” remained unchanged and was 0.9% in 2001, 2009, and 2017.

Aggregate trends hide variability by age group—particularly declining trends for children and teenagers 15 years and younger. For example, between 2009 and 2017 the prevalence of “any cycling” increased slightly for those 16 and older (1.2% to 1.3%, not significant at  $p < .05$ ) while it fell significantly ( $p < .05$ ) from 4.1% to 2.3% for those 15 and younger. Similarly, between 2009 and 2017, the prevalence of “30 minutes of cycling” increased slightly from 0.8% to 0.9% (not significant at  $p < .05$ ) for those 16 and older, while it declined significantly ( $p < .05$ ) among those 15 and younger 1.5% to 0.9%. The analysis for “any walking,” shows a similar trend with a statistically significant ( $p < .05$ ) decline in the prevalence of “any walking” for children and teenagers 15 years and younger between 2009 and 2017 (19.8% to 16.5%).

## REFERENCES

Pucher, J., Buehler, R., Merom, D., Baumann, A. 2011. "Walking and Cycling in the United States, 2001-2009: Evidence from the National Household Travel Surveys," *American Journal of Public Health*, December 2011, Vol. 101, No. S1, pp. S310-S317.

## INTERACTIONS BETWEEN THE BUILT ENVIRONMENT AND DOMAIN-SPECIFIC TRANSPORTATION PHYSICAL ACTIVITY: EVIDENCE FROM THE 2017 NHTS

*Theodore Mansfield*

### Background

In 2016, nearly 100,000 premature deaths in the United States were associated with low physical activity. Modest increase in population physical activity could substantially reduce premature mortality and morbidity. Recently, the role of the built environment in potentially increasing physical activity by supporting increased walking and biking has received considerable attention. Built environment measures such as population density, land use diversity, and physical design have been correlated with transportation physical activity. However, less is known regarding how different built environment and individual characteristics shape walking and biking behaviors for specific trip purposes. Further, different surveys collect information about different types of transportation physical activity. For example, the ACS asks respondents to provide information only for work-related transportation physical activity (i.e., walking or biking to work) while travel diaries collect information on walking and biking for all trips types but are expensive to administer and often have relatively small sample sizes. Additionally, the potential for self-selection and activity substitution to attenuate relationships between built environment factors and transportation physical activity is an issue of debate in the literature.

The 2017 NHTS provides a unique opportunity to better understand how the built environment shapes domain-specific transportation physical activity. Compared to the 2009 NHTS, the 2017 version introduced several key changes relevant to capturing walking and biking trips. First, the survey asks respondents to separately identify all trips and recreational trips when recalling walking and biking trips over the previous week. Second, the 2017 NHTS asks respondents to record trips with an origin and destination in the same place ('loop trips') more concisely than previous surveys, potentially capturing recreational walking and biking trips that may have been missed in previous surveys. Third, walking to trips to and from public transit are collected in a more structured manner to enable better estimation of walking trips to access public transit. Finally, individuals were asked about their perceived health status and physical activity levels. In sum, these changes support more detailed characterization of walking and biking trips by purpose in both the person and trip files and provide an indication of self-reported health and physical activity status. To fill the gap in understanding how the role of the built environment and individual characteristics in shaping active transportation behaviors may vary by trip purpose, this work leverages health and physical activity questions in the 2017 NHTS to explore these questions. Improved understanding of how different populations may respond to changes in the built environment for different trip purposes would support more targeted transportation physical activity interventions and improved estimation of population health impacts of such interventions.

## Methodology

From each individuals' daily trip file, walking and biking trip times were aggregated into utilitarian and recreational trip purpose categories. To define trip purpose, the tour-based purpose derived variable (WHYTRP90) was used. All trips to/from work, work-related business trips, shopping trips, other family/personal business trips, trips to school/church, medical/dental trips, and trips to visit friends/relatives, and other trips were considered utilitarian trips. All trips classified as other social/recreational were considered recreational trips. Loop trips for which no purpose was assigned (i.e., WHYTRP90 equaled 'other') were also assumed to be recreational trips. Finally, walking time to and from public transit were added to daily walk time for each individual, using the purpose of the transit trip to define trip purpose.

Once travel day utilitarian and recreational walking and biking time was calculated for each individual, categorical variables were derived based on whether or not that individual's observed walking and/or biking time met Centers for Disease Control and Prevention (CDC) physical activity recommendations—at least 30 minutes per day—on their travel day. For walking time, individuals were assigned into categories based on whether or not they accrued 30 minutes of walking during their assigned travel. Because biking for transportation is a higher intensity activity than walking for transportation, individuals were similar assigned for biking time, using a lower (15 minute) threshold.

To test the effects of built environment and individual characteristics on the likelihood of meeting CDC physical activity recommendations via utilitarian and recreational walking and biking alone, a series of logistic regression models were estimated. The first set of models (naïve recreational and utilitarian walk/bike models) included built environment and individual characteristics but did not include the respondents' self-reported health. A second set of models (adjusted recreational and utilitarian walk/bike models) included self-reported health included as an explanatory variable. Finally, a third set of models (interacted recreational and utilitarian walk/bike models) additionally included an interaction term between population density and self-reported health.

To explore possible evidence of activity substitution, a multinomial logistic regression model was first estimated to test the relationship between self-reported vigorous physical activity and built environment and individual characteristics (naïve physical activity model). A set of models were then re-estimated including a variable indicating whether the individual met CDC recommendations via observed utilitarian and recreational walking and biking met CDC recommended level alone (adjusted physical activity models). Finally, interaction terms between population density and utilitarian and recreational walking and biking at CDC recommended levels and population density were introduced (interacted physical activity models).

## Preliminary Research Findings

In the naïve utilitarian walking model, individuals living in denser built environment are much more likely to walk more than 30 minutes for utilitarian purposes on a given day (OR: 1.29, 1.44, and 1.91 for 1,000-3,999, 4,000-9,999, and >10,000 persons/mi<sup>2</sup> relative to <1,000 persons/mi<sup>2</sup>, respectively). These relationships are consistent after including self-reported health in the adjusted utilitarian walking model (OR: 1.29, 1.44, and 1.92) while individuals reporting fair/poor health are much less likely to walk than those reporting excellent health (OR: 0.78 for good/very good and 0.74 for fair/poor relative to excellent reported health). In the interacted

utilitarian walking model, the interaction term for individuals reporting fair/poor health is consistently positive and significant (OR: 1.51, 1.42, 1.35 for 1,000-3,999, 4,000-9,999, and >10,000 persons/mi<sup>2</sup> interacted with fair/poor self-reported health, respectively), indicating that population density has a positive association with utilitarian walking even in populations reporting fair/poor health.

The effects of population density on meeting CDC recommended physical activity levels are weaker in the naïve recreational walking model (OR: 1.09, 1.10, and *n.s.* for 1,000-3,999, 4,000-9,999, and >10,000 persons/mi<sup>2</sup>) compared to the utilitarian model. These relationships are stable in the adjusted recreational walking model while the relationship between recreational walking and self-reported health is slightly stronger (OR: 0.72 for good/very good and 0.50 for fair/poor) compared to the utilitarian walking model. Interaction terms are not significant for recreational walking; indicating that higher population density is not associated with higher likelihood of walking for recreational purposes at CDC recommended levels for those reporting fair/poor health.

The naïve utilitarian biking model reveals strong associations between population density and the likelihood of meeting CDC physical activity recommendations (OR: 1.49, 1.71, and 1.68 for 1,000-3,999, 4,000-9,999, and >10,000 persons/mi<sup>2</sup>). These relationships are strengthened slightly in the adjusted utilitarian biking model (OR: 1.49, 1.74, and 1.71) and an especially strong relationship is found between self-reported health and utilitarian biking (OR: 0.57 for good/very good and 0.40 for fair/poor). Interaction terms in the utilitarian biking model are not significant.

In the naïve recreational biking model, similar associations are found between population density and meeting CDC physical activity recommendations (OR: 1.26, 1.24, and 1.99) while even stronger associations are found for self-reported health in the adjusted recreational biking model (OR: 0.49 for good/very good and 0.22 for fair/poor). Once more, interaction terms in the recreational biking model are not significant.

Finally, mixed evidence of activity substitution was found: individuals walking for utility at least 30 minutes during their travel and living in the least and most dense group are roughly as likely to report engaging in some vigorous physical activity (OR: 3.5 and 3.0, respectively). Conversely, individuals who walked for recreation for at least 30 minutes in the least dense group are much more likely to report engaging in some vigorous physical activity than those in the densest group (OR: 11.7 and 4.3, respectively). This suggests that population density does not impact the relationship between utilitarian walking and vigorous physical activity; however, activity substitution for recreational walking may be more likely to occur in denser environments.

Transportation physical activity is increasingly seen as an important determinant of health. This work finds that utilitarian walking and biking are more responsive to increases in population density than recreational walking. Interestingly, this work also finds recreational biking quite responsive to increases in population density, potentially reflecting the importance of bicycle infrastructure that may be more common in denser urban environment in encouraging bicycling of all types. For utilitarian walking, the relationship between population density and meeting CDC recommended physical activity levels through walking alone holds among those with fair/poor self-reported health. Additionally, this work finds preliminary evidence that activity substitution may be more likely in denser environments for recreational, but not utilitarian, walking.

In sum, these findings suggest that built environment interventions to increase transportation physical activity should focus on different domains (recreational vs. utilitarian) in different contexts. An improved understanding of how the built environment affects walking and biking behaviors for different trip types and for different sub-populations may support improved quantitative estimates of built environment interventions to increase physical activity and support robust health impact assessment of transportation projects.

## **DEFINING TYPE OF CYCLIST BASED ON TRAVEL AND ACTIVITY PATTERNS**

*Jim Sullivan and Sarah Howerter*

### **Background**

Typical cyclist classifications have been based on driver characteristics (e.g., commuters, non-commuters). Efforts to promote cycling also commonly focus on personal and household characteristics of cyclists. However, cycling is by nature more social than driving, with cyclists exhibiting travel characteristics that transcend typical personal or household characteristics. As a result, travel behavior, by time of day and purpose, can be utilized to classify cyclists using methods common to marketing and website analytics (e.g., behavioral cohort analysis).

### **Methodology**

Person-trip data for persons who recorded cycling in their 2017 NHTS travel day data were classified as cyclists. For each person, a binary matrix was constructed with trip purposes and times of day on the axis, and a “1” in each cell of the activity occurred. The similarity of every pair of these binary matrices were calculated (Jaccard similarity coefficient) and presumed linkages were drawn between respondents based on cells in the binary matrices that match. The strength of the linkage between any two respondents was proportional to the number of matches in their binary matrices (*see* Cui et al., 2018). Scores below a certain threshold for Jaccard similarity coefficient were set to 0, effectively disconnecting the respondents with this condition. Induced networks were build and network clusters were inferred (*see* Blondel et al., 2008) using the concept of Modularity. The characteristics of the new clusters were examined as a new behavior-based classification.

### **Research Findings**

From the 2017 NHTS data, 3,606 persons made 8,034 bike trips on their travel day. The cluster groupings in the induced network for a variety of Similarity Thresholds revealed nine cluster groups. The demographic characteristics of these clusters were compared based on household size, age, vigorous activity in the past week, and average age by home ownership.

This method allows for the classification of cyclists characteristics without regard to who they are or what their household looks like. This allows for the development of new classification categories that represent social associations. This approach is used in marketing where classifications are made based on purchasing behavior and web use, resulting in “Super Clusters” that help businesses market to special market segments (e.g., Upper Crust, High Fidelity, Net Worth & Networks, Picket Fences, Maintaining a Balance, Ways & Means, Golden

Years, Debt Builders, and Hardscrabbles. The next challenge for this research is to name the emerging clusters.

## REFERENCES

- Cui, Yu, Qing He and Alireza Khani, 2008. Travel Behavior Classification: An Approach with Social Network and Deep Learning. *Transportation Research Record* 2018, pp 1-13, Publish online June 18, 2018.
- Blondel, Vincent D., Jean-Poul Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*: 10008.



## **Chapter 9: The New Sharing Modes**

**GUY ROUSSEA**

*Atlanta Regional Commission, presiding*

**WEIJING PU**

*FHWA, recording*

**MATTHEW WIGGINTON CONWAY**

*Arizona State University*

**ZHENPENG ZOU**

**SEVGI ERDOGAN**

**DAVID DONALDSON**

**ARASH GHAFAR**

*University of Maryland*

The convergence of information and communication technologies with transportation has been growing for decades. In the time between 2009 and 2017 NHTS, the use of sharing has impacted mode choices in numerous ways in many parts of the country. This session included presentations from university researchers who are tracking usage levels and modeling ridesharing with the new data.

### **TRENDS IN TAXI USAGE AND THE ADVENT OF RIDESHARING, 1995-2017**

*Matthew Wigginton Conway*

#### **Background**

Ridehailing services, such as Uber and Lyft, have transformed for-hire vehicles (taxis as well as ridehailing) from a rarely used mode concentrated in a few cities into a mode used by many across the US. Given this newfound prevalence, cities across the country need to incorporate for-hire vehicles into their transportation planning. In particular, ridehailing presents equity concerns for people without access to smartphones and credit cards.

#### **Methodology**

The NHTS asked two questions about for-hire vehicle use. The first is trip mode; ridehailing was added to the taxi mode in the 2017 survey. The second question, new to the 2017 NHTS, asks how many times the respondent used ridehailing in the last month. In this research, the first question is used to analyze trends, while the second is used to analyze the demographics of ridehailing users specifically. Trends are analyzed using descriptive statistics, while ridehailing user demographics are explored through logistic regression.

## Preliminary Research Findings

For-hire vehicle use has exploded since 2009. In 2009, for-hire vehicles were used for  $0.19\% \pm 0.04$  (95% confidence interval) of trips in the US; that number had remained quite steady since 1995. By 2017, that number had more than doubled to  $0.50\% \pm 0.10$ . While the growth is large, a modeshare of half a percent may not seem particularly remarkable. However, for-hire vehicles are infrequently used; their modeshare understates their popularity among the population. In any given day,  $0.94\% \pm 0.11$  of Americans used for-hire vehicles, up from  $0.43\% \pm 0.08$  in 2009, much higher than the mode share. Over the past month, the 2017 NHTS indicates that  $9.81\% \pm 0.44$  of Americans 16 and over have used a ridehailing service (specifically, not including traditional taxicabs). While these modes are used infrequently by many, their market penetration suggests that they are an important part of the national transportation system, and should be included in planning and modeling.

Geographic distribution of for-hire vehicle usage has also changed over time. In 2009,  $40.84\% \pm 8.39$  of respondents reporting a for-hire vehicle trip on the travel day lived in the New York metropolitan area; by 2017, only  $20.35\% \pm 2.97$  did. During the same period, the concentration of for-hire vehicle users in major Eastern metropolitan areas (New York City; Boston; Chicago; Washington, D.C.; and Philadelphia) declined from  $59.82\% \pm 9.63$  to  $38.86\% \pm$

While the geographic distribution of for-hire vehicle users overall has deconcentrated with the introduction of ridehailing, the logistic regression shows that residents of larger MSAs are still more likely than residents of smaller MSAs to use ridehailing. Like other modes of transport, ridehailing usage is related to the built environment. People living in denser block groups are more likely to use ridehailing, even when controlling for auto availability. Denser areas likely have shorter wait times for ridehailing services, more parking problems, and shorter trip distances (and thus lower fares), all making ridehailing more attractive.

Ridehailing presents a number of equity concerns. Individuals from low-income households have historically been overrepresented among for-hire vehicle users, and continue to be. However, they are *underrepresented* among ridehailing users specifically, suggesting that they use traditional taxicabs at high rates. This may be due to a relative lack of access to credit cards and smartphones among this demographic. These individuals would likely be better off if they had access to ridehailing, which provides a more geographically distributed and potentially lower-cost service. Additionally, if ridehailing services undermine traditional taxicabs, it could have a negative impact on the nation's poor, eliminating the types of for-hire vehicle service they rely on. To counter this concern, ridehailing services could accept cash. Amazon has recently implemented a system wherein customers can pay in cash at a number of brick-and-mortar retail outlets (*Amazon Cash* 2018); ridehailing companies could implement similar programs.

In the logistic regression, daily smartphone use by the household member responding to the NHTS recruitment survey is strongly associated with an increased probability of ridehailing use. This is unsurprising, of course; these services require a smartphone. However, it also means that households unable to afford a smartphone are left out of ridehailing, likely contributing to the income trends described above. The federal government already provides "lifeline" subsidies for mobile phone service (*Lifeline Program for Low-Income Consumers* 2018); this program could be expanded to subsidize smartphones and allow access to mobility services. Alternately, the ridehailing providers themselves could subsidize smartphone ownership to increase their market, or relax the requirements for smartphone usage.

The racial breakdown of for-hire vehicle users on a given day more closely matches the general population than it has in years past. Previously, African-Americans were significantly overrepresented among for-hire vehicle users, whereas in 2017 they are represented roughly in line with their share in the population overall. This is likely due to the changes in the geographic concentration of for-hire vehicle use described previously. The same, however, does not hold true for ridehailing users specifically: African-Americans are *underrepresented* among ridehailing users, again creating equity concerns. This may be partially due to the differences in income distribution described above. Discrimination may also play a role; Ge et al. (2016) found racial discrimination in ridehailing. The logistic regression confirmed that African-Americans use ridehailing less than whites do. However, the logistic regression found that Asians used ridehailing *less* than their other sociodemographic would predict—probably due to their higher income.

The growth, geographic diversity, and equity concerns related to ridehailing all suggest that for-hire vehicles need to be integrated into planning. Some cities are entering into partnerships with ridehailing firms to complement public transit (for instance, through subsidized fares to transit stops, or by providing paratransit service). Others are seeing city revenues decline as fewer people pay for parking in their downtowns. Yet others are designating ridehailing pickup and dropoff zones. It is unclear, however, which of these efforts will be successful. What is clear is that we must plan for these growing modes, and more research is needed to understand how exactly to do it. It is also important to integrate for-hire vehicles into travel demand models. The largest cities already include taxis as a mode in their travel demand models, but many more cities will need to follow suit. Furthermore, it is important to include ridehailing not only as a travel mode in mode-choice models, but also to simulate effects of ridehailing on vehicle ownership, and the potentially large amount of deadhead travel produced by ridehailing vehicles.

## REFERENCES

- Amazon Cash*. 2018. Accessed May 13. <https://www.amazon.com/b?ie=UTF8&node=14583169011>.
- Ge, Yanbo, Christopher R Knittel, Don MacKenzie, and Stephen Zoepf. 2016. *Racial and Gender Discrimination in Transportation Network Companies*. Working Paper 22776. National Bureau of Economic Research. <http://www.nber.org/papers/w22776>.
- Lifeline Program for Low-Income Consumers*. 2018. Accessed May 13. <https://www.fcc.gov/general/lifeline-program-low-income-consumers>.

## EXAMINING THE U.S. RIDESOURCING MARKET USING THE 2017 NHTS DATA

Zhenpeng Zou and Sevgi Erdogan

### Background

From horse power to horsepower, the evolution of urban transport focused on technology advancement in vehicle mechanics. Nevertheless, the rise of ridesourcing is about the information and communication technologies that make the “sourcing” in ridesourcing possible. With the readiness in hardware (e.g. smartphone), software, and peer- to-peer (P2P) marketplace, TNCs like Uber and Lyft now serve over 200 metropolitan areas in all 50 states and the District of Columbia. Yet, despite being praised for offering convenience and mobility, TNCs are

constantly under public pressure for the unconventional business conducts (Moran and Lasley, 2017).

With the lack of user and trip data from TNCs, research on ridesourcing in the U.S. primarily relies on survey data collected by individual researchers (Rayle et al., 2016; Crewlow and Mishra, 2017). These studies are the go-to research for transportation planners and agencies to attain preliminary understanding of travel behaviors associated with ridesourcing. However, sample representativeness is a common issue with these studies: Survey respondents are likely to be younger, more educated and tech-savvy ridesourcing users than the representative population.

## **Methodology**

This research uses the 2017 NHTS data to examine the impact of sociodemographic and household characteristics on the usage of ridesourcing. The authors produced descriptive statistics for the usage of ridesourcing across all fifty states and the District of Columbia based on weighted average ridesourcing frequency. In addition, the authors cross-tabulated ridesourcing frequency by individual socioeconomic statuses, travel characteristics, as well as household characteristics weighted at the population level. Then, principal component analysis and Pearson's correlation analysis were applied to tease out the significant factors in explaining variances in ridesourcing frequency without running into multicollinearity issues. After filtering the significant factors associated with ridesourcing frequency, the authors applied regression analysis on the D.C.-Maryland-Virginia (DMV) subsample to investigate how sociodemographic and travel characteristics could explain the possibility of using ridesourcing and the number of rides conditional upon the usage. A zero-inflated negative binomial model was adopted to capture the two-stage behavior (usage and frequency conditional on usage).

In Stage One, a binomial logit was specified to differentiate ridesourcing users from non-users. This standard approach has been effectively implemented in previous travel behavior research using the NHTS data (Shin, 2017). In Stage Two, a logarithm linear regression was specified to address how underlying socioeconomic characteristics could affect how frequent a user rides with a ridehailing service. Beside regression analysis, the authors also explored how machine learning technique, such as K-Means clustering, could reveal potential clusters of ridesourcing users (and non-users) in the DMV region, which offered insights on ridesourcing travel demand management for different subpopulations. While traditional regression models are derived from the classical microeconomic theory of utility maximization on ridesourcing decisions, machine learning techniques rely heavily on the similarity and distinctness of data and a rich panel of "features" to be used in classification/clustering.

## **Preliminary Research Findings**

Ridesourcing is still an emerging mobility with less than 10% adoption nationwide. Nevertheless, the adoption rate is higher in more populated states, such as New York, California, Massachusetts, and Illinois. Furthermore, ridesourcing has become a common mode for certain subpopulations, such as millennials, urbanites, and carless individuals. Through the regression analysis, we identified factors that significantly influenced the possibility of ridesourcing usage, such as education level, household income, age, household vehicle count, and the usage of public transit. We also found that the use of alternative modes (walk, public transit, home delivery) were significantly associated with the frequency of ridesourcing. In addition, size of the

metropolitan area also determined how frequent a user hired a ridehailing service. The preliminary clustering identified two distinct groups in the DMV area: a wealthy, educated, millennial urbanite group who ride with ridesourcing frequently and a retired suburban group who barely use ridesourcing at all. While the clusters don't surprise us, the analysis exhibits the potential of using machine learning techniques in market segmentation.

This research fills the gaps in survey study on ridesourcing with a nationally representative sample – NHTS 2017. The outcomes will confirm speculations we have had in years about travel behaviors associated with this emerging type of mobility. In particular, it is to transportation planning agencies' interest to understand how different subpopulations use ridesourcing and whether ridesourcing disrupts the existing transportation market. Our preliminary results suggest that previous studies on ridesourcing correctly characterized users in terms of their socioeconomic statuses. In addition, we found that the use of public transit is associated with both the possibility of using ridesourcing and the frequency of usage. We believe it is necessary to conduct trip-level analysis on a fine-grained geographic scale to confirm whether ridesourcing acts as a substitute, a complement, or both to public transit. From a policy perspective, understanding the specific travel demand for ridesourcing users and non- users is essential to regulate this new mode of mobility.

## REFERENCES

- Clewlow, R. R., & Mishra, G. S. (2017). Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States. Report. Davis: Institute of Transportation.
- Moran, M., & Lasley, P. (2017). Legislating transportation network companies. *Transportation Research Record: Journal of the Transportation Research Board*, 2650, 163-171.
- Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, 45, 168-178.
- Shin, E. (2017). Ethnic neighborhoods, social networks, and inter-household carpooling: A comparison across ethnic minority groups. *Journal of Transport Geography*, 59(2), 14-26.

## THE EFFECTS OF DIFFERENT GENERATIONAL SOCIODEMOGRAPHIC CHARACTERISTICS ON DYNAMIC RIDESHARING

*David Donaldson and Arash Ghaffar*

### Background

Previous studies show that travel behavior differs significantly between age groups, specifically millennials and their aging counterparts. Younger adults, who by 2019 are expected to comprise the largest portion of the US population (1), are less likely to get behind the wheel of a personal vehicle. As a result, the declining number of licensed drivers and car owners (2) leaves many experts searching for justification. One possible explanation arises from the emergence of popular dynamic ridesharing services, including the major ride-hailing services. These digital services are seen by some as the future of travel as users become more accepting of sharing services with strangers and forgoing the hassle and costs associated with car ownership.

The recent emergence of dynamic ridesharing services has prompted a number of academic studies related to system optimization, social behavioral responses (3), and the challenge of rideshare integration into the transportation realm (4). However, the authors believe further research into the individual and household characteristics of typical rideshare participants is warranted to improve rideshare trip generation forecasting, and help planners and business developers incentivize different age groups accordingly. A couple previous studies have specifically looked at ridesharing and “e-hailing” among select generations including Vivoda’s et al. 2017 study (5) which identified male gender, college education, and e-hail awareness as significant predictors of future e-hailing among seniors. However, there is little comparison of ridership trends among age groups. Therefore, this study investigates the effects of a variety of sociodemographic and land-use characteristics on persons’ real-time ridesharing behavior across different generations.

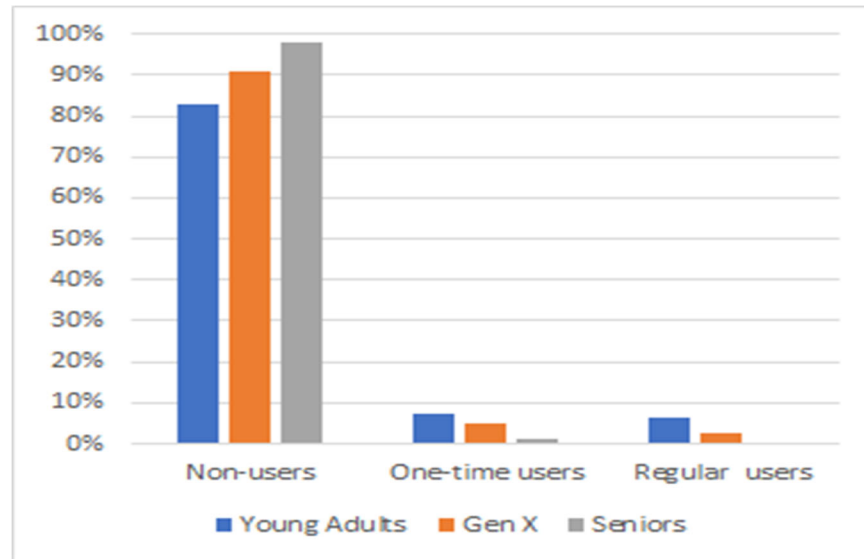
## Methodology

This research encompassed descriptive statistics to highlight significant relationships between the number of rideshare service purchases and select personal, household, and land use characteristics as well as a multinomial logit (MNL) model to estimate leading factors that prompt different generations to participate in ridesharing. Three age cohorts are identified as follows: young adult (18-30 y/o), Generation X (39-53 y/o), and seniors (+65 y/o). The entire analysis is grounded on traveler data made available in the 2017 NHTS. The survey sample comprised of 236,089 people who responded with the number of times they used a rideshare app to purchase a ride in the last 30 days prior to the survey interview.

The MNL model is a discrete choice model that explains and predicts choices between two or more distinct alternatives. The model’s alternatives are classified into three groups: individuals who did not purchase a ride, one-time users (1-2 service purchases), and “regular” users (+5 purchases). The chosen alternative selected should yield the greatest utility to the individual. The utility function is the sum of both systematic and random utility components as shown:  $U_{i,n} = V_{i,n} + \varepsilon_{i,n}$  where individual  $n$  chooses alternative  $i$ . The systematic component is explained by a variety of individual characteristics or behaviors that become  $X_i$  parameters in the equation  $V_{in} = \beta_0 + \beta_1 X_{1i,n} + \beta_2 X_{2i,n} + \dots + \beta_k X_{ki,n}$

## Preliminary Research Findings

Using 2017 NHTS-supplied weights to account for nonresponses and oversampling, the population-level estimated proportion of rideshare consumers among age cohorts reveals 17% of young adults had used a rideshare service within a month prior to the NHTS interview compared to Generation Xers (9.3%) and Seniors (2.2%) (see Figure 9.1). Furthermore, the highest rideshare users per capita among all age groups from the survey were Asian, attained at least some college degree, and come from a household earning over \$100,000 a year. A higher proportion of Generation X and senior males used a rideshare service at least once where females made up the majority of young adult riders. Surprisingly, about the same percentage of rideshare consumers had a license versus no license for all age groups.



**FIGURE 9.1. Distribution of rideshare consumer alternatives by age group.**

The MNL model estimated significant effects of driver status, home ownership, household income, and housing density for all three age groups. According to the model, a driver and homeowner are less prone to rideshare while an increase in the latter two characteristics increases the chance the person will use a rideshare service. For the youngest cohort, individuals with children are less likely to use rideshare services than single adults. Among Generation Xers, as the population density of the census block they reside in increases, the higher the probability middle-aged persons will pay to share a ride. Interestingly, the eldest age group are likely to try rideshare once, but not necessarily regularly as the housing density in their census block rises. Overall, almost all sociodemographic variables had a significant effect on ridesharing usage for young adults versus only a few for seniors.

The increasing demand pressed upon the ridesharing network puts planners and policy-makers in a crucial position to strategize the continuous growth of this evolving technology. The results above support many findings by other researchers that suggest higher education, wealth, and an urban setting do have a significant effect on rideshare service consumption. However, now planners and business developers will be able to understand the influence certain characteristics have on different age groups. Hence, the opportunity emerges for society's decision-makers to allocate resources in ways that improve rideshare equity and promote ridesharing and its benefits to individuals of all backgrounds.

## REFERENCES:

- Fry, R. *Millennials projected to overtake Baby Boomers as America's largest generation*. Pew Research Center, 2018. [www.pewresearch.org/fact-tank/2018/03/01/millennials-overtake-baby-boomers/](http://www.pewresearch.org/fact-tank/2018/03/01/millennials-overtake-baby-boomers/). Accessed July 22, 2018.
- Kuhnimhof, T., Armoogum, J., Buehler, R., Dargay, J., Denstadli, J.M., Yamamoto, T., 2012. Men shape a downward trend in car use among young adults—evidence from six industrialized countries. *Transp. Rev.* 32, 2012, 761-779. <http://dx.doi.org/10.1080/01441647.2012.736426>.

Sarriera, J. M., Álvarez, G. E., Blynn, K., Alesbury, A., Scully, T., & Zhao, J. To Share or Not to Share: Investigating the Social Aspects of Dynamic Ridesharing. *Transportation Research Record*, Jan 2017, Vol. 2605, pp. 109-117.

Clewlou, R. R., G. Shankar Mishra, R. Clewlou, R. Affiliate, and S. Kulieke. Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States 2017. [https://its.ucdavis.edu/research/publications/?frame=https%3A%2F%2Fitspubs.ucdavis.edu%2Findex.php%2Fresearch%2Fpublications%2Fpublication-detail%2F%3Fpub\\_id%3D2752](https://its.ucdavis.edu/research/publications/?frame=https%3A%2F%2Fitspubs.ucdavis.edu%2Findex.php%2Fresearch%2Fpublications%2Fpublication-detail%2F%3Fpub_id%3D2752).

Vivoda, J. M., A. C. Harmon, G. M. Babulal, and B. J. Zikmund-Fisher. E-hail (rideshare) knowledge, use, reliance, and future expectations among older adults. *Transp. Res. Part F Psychol. Behav.*, Vol. 55, 2018, pp. 426–434. Available at: <https://doi.org/10.1016/j.trf.2018.03.020> [Accessed July 16, 2018].



## **Chapter 10: Underserved Populations**

**KRISHNAN VISWANATHAN**

*Cambridge Systematics, presiding*

**JENNIFER MURRAY**

*Wisconsin Department of Transportation, recording*

**CHANDRA BHAT**

**ABHILASH SINGH**

*University of Texas, Austin*

**BUMJOON BAE**

**HO-LING HWANG**

**SHIH MIAO CHIN**

**CHIEH R. WANG**

*Oak Ridge National Laboratory*

**APARA BANERJEE**

*MacroSys LLC*

**DAVID CHATMAN**

**ABIGAIL COCHRAN**

*University of California, Berkeley*

**NICHOLAS KLEIN**

*Cornell University*

The importance of having safe, reliable, efficient transportation options are critical for access to jobs, food, and healthcare and for the general uplift of all sections of society. In this session, methods were identified to understand the travel patterns of immigrants, persons with disabilities, and low-income households.

### **FOOD ACCESS FOR LOW INCOME INDIVIDUALS**

*Chandra Bhat and Abhilash Singh*

#### **Background**

Ease of access to a healthy food environment (FE) is important for individual health and wellbeing, as has been demonstrated in earlier studies through the positive effects of healthy FEs on the lower incidence of chronic disease rates in the general population and the higher school performance of individuals in the school-going population. Historically, in the context of FE, as

land use consolidated, large grocery stores serving a greater area replaced small specialty markets serving local neighborhoods. These large grocery stores, although typically centrally located, can be prohibitively burdensome to reach for low income neighborhoods with limited mobility. On the other end of the land use spectrum, the increasing sprawl of many urban cities, without the corresponding proportionate establishment of stores in non-core areas, has also led to a deterioration in the ease of healthy food access, a situation further exacerbated by gentrification that has pushed out many low income residents from urban cores to non-core areas.

To assess the (in-)ability of such low income individuals to access healthy food, the USDA (2017) defines food deserts (FDs) as low income census tracts in which either at least 500 people or 33 percent of the population reside more than one mile (radial distance) from a supermarket or large grocery store. To be sure, high-income census tracts can also be a geographic food desert; however, high-income households typically have means of mobility to reach stores or can afford alternatives, such as grocery/meal delivery.

The concept of an FD is a useful one to characterize food access for low income individuals. However, as currently defined, it can provide an overly optimistic picture of food access for such individuals, because it does not consider mobility tools (such as car ownership or transit-access) and also uses radial distances that can be substantially shorter than actual network access distances. This study seeks to develop an improved metric of food accessibility that can help refine the definition of an FD. We propose to do so by developing a model of food shopping location for low-income individuals. Through the revealed choice of low income individuals in the 2017 NHTS on where they undertake their food shopping from the many available food shopping locations around them, we will tease out the influence of demographic, mobility tools, and FE factors in terms of their tempering or enhancing effects on the attractiveness of food shopping locations. This will then enable us to develop a multivariate metric for food accessibility. Information on the location of each out-of-home activity episode (including food shopping) is collected, in the 2017 NHTS, at the level of the place name (such as HEB) as well as its address/closest intersection. Based on this information, we developed our location choice model at the most disaggregate spatial level using the Dallas-Fort Worth region sub-sample and supplemented it with land-use and transportation network GIS data for the region.

## **Methodology**

We have estimated the spatial location choice model for 1,005 food-shopping trips by low-income individuals in the Dallas/Fort Worth (DFW) metro area to determine the Traffic Survey Zone (TSZ)-level attributes and sociodemographic interactions that influence the choice of food shopping location. The most important results from our analysis are as follows: 1) The number of wholesale and supercenters in a zone has a strong positive association with food shopping choice for low-income individuals. 2) Households may not be influenced in the same way by food store proximity based on a wide range of factors besides household income. Therefore, the interactions between socio-demographic factors with built environment attributes give us an insight of how households value the choice of food shopping locations and how this shapes a low-income household's FE. 3) Trip location decision may not always be influenced by trip-related variables such as travel-time.

## Preliminary Research Findings

Our results indicate that low-income households much prefer to shop for food at wholesale stores and supercenters. Rather than increase the access and attractiveness of grocery stores to low-income individuals, their FEs may be best improved if they have better access to supercenters and wholesale stores. Additionally, these stores should market healthy food options to low-income individuals that are similar or the same that could be purchased at typical grocery stores such as fresh produce, grains, and proteins. Though we see that low-income families prefer to shop for food at locations closer to home, many low-income families do not have access food shopping locations that are relatively close to home. The average distance from home to food shopping location in our sample was 6.64 miles. This is also evident in the literature as households may be differentially affected by food store proximity (or lack thereof) based on a wide range of factors besides household income including access to a vehicle, transit networks, culturally specific foodways, broader social networks, and already existing daily mobility (Glanz et al. 2005).

Ideally, the sustainable solution to food access would involve the increase of the use of public transportation or active transportation to complete food-shopping trips. However, our descriptive analysis of the sample households indicates that few low income families are using the public transportation system in a car-dominated metro area such as Dallas Fort Worth.

A key contribution of our study is that it does not measure the overall food accessibility of a geographic region, which under the typical FD definition may include a large share of rich households with excellent FEs. We focus on the *actual* food shopping behavior of households to understand what influences their decisions to undertake food-shopping trips at different locations. The departure from the typical food desert concept leads to a more disaggregate approach that can be applied to *any* low-income household or individual shopping for food. By using a more disaggregate approach, we can help food researchers, policymakers, and planners to better formulate solutions to improving the FE of low-income people.

## REFERENCES

- Glanz, K., Sallis, J. F., Saelens, B. E., & Frank, L. D. (2005). Healthy Nutrition Environments: Concepts and Measures. *American Journal of Health Promotion*, 19(5), 330–333.
- USDA Food Access Research Atlas (2017). US Department of Agriculture, Economic Research Service.

## PREDICTING DAILY TRIP FREQUENCIES OF VULNERABLE HOUSEHOLDS IN NEW YORK STATE USING EMERGING MACHINE-LEARNING APPROACHES

*Bumjoon Bae, Ho-Ling Hwang, Shih-Miao Chin, and Chieh R. Wang*

### Background

This research compares the performance of trip generation models including traditional regression-based models and emerging machine learning methods. Trip generation is the first step of the four-step travel demand modeling process. Despite its frequent uses, the limitations of trip generation models with linear-regression structure have also been discussed in the literature.

There have been various comparative studies on different types of a trip generation model by other researchers. Interestingly, the alternative models in the literature do not show better performance or significant improvement in general, compared to linear regression for predicting the number of trips of individuals. The comparative studies in the literature have mostly focused on econometric models, however, there have been insufficient efforts to expand the comparison scope to other machine learning methods.

The objective of this study is to compare the trip frequency prediction performance of supervised learning models with the traditional linear regression approach. The alternative models to be considered in this study include: regularized regression (i.e., Ridge, elastic net, Lasso), negative binomial regression, ordered logit, logistic regression, k-nearest neighbor (kNN), classification trees, bagging trees, and random forest.

## **Methodology**

This study uses the 2017 NHTS data to train and test the trip generation models. A total of 1,731 travelers from low income households within the New York State (NYS) are selected. Twenty explanatory variables related to the person, household, and regional characteristics for corresponding subjects are extracted from the NHTS data. The 1,212 traveler records (70% of 1,731) are randomly selected from the data and used for model estimation. The remaining 519 records (30%) are used for model validation. This validation process is replicated 100 times to ensure reliable comparison of prediction results. As a base model, the linear regression model is specified using all 20 variables.

The differences in the prediction performance of the base model and three regularized regression models are not easily visible. Although the kNN has a little higher root mean squared error, its correlation coefficient (0.473) is noticeably larger than that of the base model (0.331). This implies that kNN's predictions are more linearly correlated with the observed trip frequencies, even though they are biased.

## **Preliminary Research Findings**

To compare the model performance at the aggregate level, coincidence ratios (CR) are compared. CR is a value obtained by dividing the intersection area of the observed and predicted trip frequency distributions by the union area. The distributions of the linear regression models are prone to be concentrated at the mean value. Therefore, the CRs of these models are in a low range between 0.347 and 0.360. The negative binomial and ordered logit models have higher ratios (0.376 and 0.395, respectively). However, they do not overcome such central tendency and do not fit the long right tail of the observed distribution. In contrast, it is obvious that the classification models are superior to the regression-type models for fitting the true distribution. The multinomial logistic regression model has the highest CR of 0.873.

This study compares several trip generation models for individuals from low-income households in NYS. A set of 20 variables on person, household, and regional characteristics were selected from the 2017 NHTS data. A total of 11 supervised learning methods, including regression- and classification-type models, were explored to overcome or alleviate the limitations of a linear regression model.

The validation results from the 100 runs show that advanced methods do not necessarily provide significant improvements in trip frequency prediction. The regularized linear regression

models, including Ridge, elastic net, and Lasso are useful to identify the best subset of variables. Therefore, as more variables are added to a model, the benefit of them will be inclined. The negative binomial regression and ordered logit models showed only marginal improvements at the aggregate level. This result is consistent with the findings in the literature. The other classification models, including logistic regression, kNN, classification trees, bagging tree, and random forest, showed higher errors at the disaggregate level. However, they outperformed the linear regression models in replicating the observed trip frequency distribution.

The results of this study can vary depending on the selected population group, variables considered, and/or use of different data sources. Additional adjustments, such as adjusting tuning parameters, setting categorical variables in a different manner, and adding more observations, can be made to possibly improve the results of classification models.

## **TRAVEL PATTERNS OF LOW INCOME HOUSEHOLDS**

*Apara Banerjee*

### **Background**

According to the ‘Consumer Expenditures–2016’ Report (1) Housing and Transportation are the two most expensive categories where households spend 33% and 16% of their average annual expenditures respectively. A large part of this transportation cost can be attributed to the high cost of owning/leasing, insuring and maintaining a vehicle for travel (2). High auto-dependency along with rising costs of car ownership and inconvenient transit systems (3) hinder mobility of the low-income population. With 12.7% of the U.S. population under poverty level (4), the 2017 NHTS data shows that 60% of the Households Below Poverty Level (HBP) feel travel is a financial burden as compared to around 35% of the Households Above Poverty (HAP). This difference in the financial burden of transportation makes it important to understand how the HBP level tend to travel and how their travel behavior differs from those who are financially better off.

### **Methodology**

Though relationship between transportation and poverty is not a new concept, most prior literature in this area has analyzed mobility challenges for the low-income population either from the perspective of employment accessibility or for a specific region (urban or particular city). In this study we aim to look at the full array of travel behaviors of the HBP in comparison to the HAP for the entire country using the most recent NHTS data of 2017 (and 2009 for some trends). The Federal Poverty Guidelines (5) for the year 2009 and 2017 were considered to classify households in NHTS 2009 and 2017 as above or below poverty thresholds, based on a combination of reported household income and size. The variances in travel between the two types of households are explored, focusing on mode choice, vehicular ownership patterns, reported trip purposes and distance traveled.

## Preliminary Research Findings

2017 NHTS data suggests a daily trip rate for members of HBP level was 2.9, in contrast to 3.5 for those hailing from HAP. Apart from making fewer trips per day, members of HBP are also likely to travel shorter distances than their richer counterparts. Average person trip length for those living in HAP was 11.3 miles, which is 1.5 times more than the average person trip length of HBP members (7.2 miles). Additionally, analysis of the average distance traveled by trip purpose shows that an average work trip length for people belonging to HBP is about 10 miles while for those from HAP it is 14 miles. HBP members are also found to report trip lengths that are on an average 2 miles less than HAP members for trips related to shopping/errands or social/recreation.

Along with trip rate and distance, mode-choice also differs between HBP and HAP. According to 2017 NHTS, HBP were found to walk, bike or use public transit for 21% of their trips, which is almost twice the percentage of trips made by HAP (13%). A comparison with previous NHTS data shows, mode-choice for both HBP and HAP have remained consistent over the years with POVs as the dominant mode of transportation in both the years and across all households. In 2017, HBP used POVs for almost three-quarters (74%) of their trips while POV usage accounted for 84% of the trips made by HAP.

The modal share pattern was found to have a direct correlation with the vehicle ownership structure in case of the HBP. Just as HBP used POVs for three-fourth of their trips, 2017 NHTS shows approximately three-fourth of the HBP reported owning at least one car. 27% of the HBP did not own any vehicles in 2017, as compared to only 4% HAP. Among the total walking, biking and public transit trips made by HBP, 52% were made by people belonging to zero-vehicle households while 6% of these trips were made by members of households with 3-5 vehicles. Percentage of total walk, bike and transit trips undertaken by HBP were thus found to decline as household vehicle ownership increased. On the contrary, among the total walking, biking or using public transit trips made by HAP, the highest percentage (33%) of trips were undertaken by households with 2 vehicles.

Mode-choice pattern was also found to vary between HBP and HAP by purpose of the trips. Work trips were the only exceptions when mode-choice for both HBP and HAP remained comparable. As per 2017 NHTS data, HBP were found to travel in POVs for 83% of their work trips while 15% walked, biked or used public transit to work. The numbers are similar for HAP, with 85% traveling to work in POVs and 13% walking, biking or availing public transit. The reported mode usage differs between HBP and HAP when it comes to shopping/errand or social/recreation. 20% of the shopping trips made by HBP were done by walking, biking or public transit as opposed to HAP, who only walked, biked or used public transit for 8% of their shopping trips. Similarly 28% of the social/recreational trips of HBP were by walking, biking or using public transit as compared to 21% for HAP.

Furthermore, mode-preference is highly impacted by location. Reported travel by nonmotorized transportation and public transit increases in a big urban area with better availability. HBP, living in a MSA with more than 1 million population and connected by heavy rail, reported using public transit more than double of that of the HAP (13% vs. 6% respectively). POV usage for trips made by HBP within this same MSA category (56%) was also reported to be over 15% less than HAP (74%) though POV remained the dominant mode of transportation for both. On the other hand, percentage of total walking, biking and public transit trips decline as one considers areas with less density. Absence of alternative transport option, in

counties not in any MSA, resulted in comparable POV usage for both HBP (85%) and HAP (90%).

This analysis of the travel pattern of the economically disadvantaged households as reported in the 2017 NHTS is an attempt to build on to previous research and to strengthen the understanding of the travel needs of the poor. Evidence from 2017 NHTS indicates, people hailing from HBP tend to walk, bike or use public transit more than their richer counter parts, with non-auto travel reports higher for social-recreational or shopping purposes. In large urban areas with rail, HBP use public transit more but POV still remains the dominant mode. Further research in these areas can lead to a more comprehensive understanding of how income influences travel behavior.

## REFERENCES

1. Bureau of Labor Statistics, 2017. *Consumer Expenditures-2016*, USDL-17-1175. U.S. Department of Labor, Washington D.C. Available Online: <https://www.bls.gov/news.release/pdf/cesan.pdf>
2. Miller, G.E., 2018, April 1. Transportation Costs: *The Second Highest Expense Category is Clearly the Most Ridiculous*. Significance.com. Available online: <https://20something.com/transportation-costs/>
3. Pucher, J. and Renne, J.L., 2003. *Socioeconomics of Urban Travel: Evidence from the 2001 NHTS*. *Transportation Quarterly*, 57(3), pp.49-77.
4. Semega, J., Fotenot, K.R., and Kollar, M.A., 2017 September 12. *Income and Poverty in the United States: 2016*, Report Number: P60-259. United States Census Bureau. Available Online: <https://www.census.gov/library/publications/2017/demo/p60-259.html>
5. Federal Poverty Guidelines 2017 and 2009. The poverty guidelines updated periodically in the Federal Register by the U.S. Department of Health and Human Services under the authority of 42 U.S.C. 9902(2). Available Online: <https://aspe.hhs.gov/2017-poverty-guidelines> and <https://aspe.hhs.gov/2009-hhs-poverty-guidelines>

## PERSONS WITH DISABILITIES AND PERSONS BORN OUTSIDE OF THE U.S.: DEMOGRAPHIC AND TRAVEL TRENDS FOR TRANSPORT PLANNERS

*Daniel Chatman, Abigail Cochran, and Nicholas Klein*

### Background

We describe the travel patterns of immigrants and people with disabilities, with a focus on how new app-based ride hailing services provided by TNCs are used. Ride hailing is a potentially important mode for people in both groups. Is the take-up of ride-hailing services higher or lower among people with disabilities? What factors, like low income and location within urban areas, affect whether ride hailing can be used to help overcome the transport barriers that people with disabilities face? In the case of immigrants, is ridehailing used at a greater or lower rate than people who are born in the US, given the fact that immigrants, particularly recent immigrants, are also much more likely to use public transportation and to own fewer cars? Is there any evidence that app-based ride-hailing services will enable lower dependence on autos among recent immigrants?

According to recent data from the U.S. Census Bureau, 12.8 percent of the approximately 318 million people in the United States reported having one or more disabilities, while 13.5

percent of the US population was foreign-born. Both populations are important submarkets for public transport agencies specifically, and for transport planners more generally, particularly as planning agencies. The concern is any shifting from exclusively publicly provided shared transport services, to notions of “mobility on demand” including hybrid forms of paratransit and privately-provided supplementary services such as TNCs. Immigrants make up a very large share of public transport ridership in the US and serving their needs has been an important part of retaining or growing that ridership. Meanwhile persons with disabilities receive special accommodations and services to ensure that public transport is accessible to them, as well as the provision of paratransit services. TNCs may serve an important role in either competing with or augmenting public transport or paratransit services among both groups.

## Methodology

The NHTS has historically collected limited data on both people born outside the US, and people with disabilities. In the case of immigrants, the 2009 and 2017 NHTS asked whether the respondent was born outside the United States, and the year the respondent arrived in the US (in surveys prior to 2009 the country or region of origin was also collected). The NHTS does not ask directly about disability status. The only relevant question collected in both 2009 and 2017 is whether the respondent has a medical condition that limits their ability to travel. The 2017 questionnaire collected additional data about medical conditions including whether the respondent uses an assistive device like a wheelchair or a white cane. This paper looks only at the 2017 dataset.

The research focuses on trends, while contrasting large metropolitan areas with high levels of transit service and more readily available TNC services, to smaller metros and non-metro areas: (1) the use of shared ride modes among these populations; (2) the distribution of mode choices for work and non-work purposes; and (3) the household characteristics that contribute to these patterns. This research focused mainly on two data points from different NHTS 2017 data files. The person file includes a question about the number of app-based ride-sharing trips in the past month, which we call “monthly TNC use.” The trip file includes the travel mode for each trip reported on the designated travel day, including a category called “Taxi / Limo (including Uber / Lyft),” which we refer to here as “Taxi/TNC.”

## Preliminary Research Findings

Nationwide, only 10% of respondents reported using TNCs at least once during the previous month, and only one-half of one percent of daily trips were in the taxi/TNC category. About 20% of recent immigrants—defined as those who arrived in the US within the previous decade—report using TNCs, twice the rate in the general population as well as twice the rate for immigrants who arrived prior to 2007. Recent immigrants also report much higher rates of daily taxi/TNC use, at 0.05 trips per day compared to 0.02 for the US-born. For persons with disabilities, the rate of monthly TNC use is much *lower* than that of the general population, at about 3%. The rates are particularly low among respondents who report relying on manual or motorized wheelchairs or scooters, and people with sight problems who use white canes or seeing-eye dogs. Among those same groups public transportation use is much higher (as is paratransit use, as would be expected). Interestingly, *daily* taxi /TNC use is not very different among those with disabilities and those without.



TNC services are currently more widely available in larger cities, which could play an important role in these differences. The data show that TNC use is significantly higher as city size increases, and is consistently higher for recent immigrants regardless of metropolitan area size. The disparity is smaller for larger cities where TNC services are more widely available, reaching a high of 24 percent for recent immigrants living in cities of greater than 3 million, compared to 18 percent for US-born respondents living in those cities. For people reporting disabilities, monthly TNC use increases somewhat as city size increases, reaching 5.7% in the largest cities. But the increase is not nearly as much as it is for people without disabilities, whose share of monthly TNC use reaches 18% in those cities. The *daily* reporting of taxi/TNC trips is higher for people with disabilities across all cities, except the largest. In cities over 3 million, the number of taxi/TNC trips is lower for people without disabilities.

Some of the differences between immigrants and the US-born in ride-sharing use are likely due to differences in their age, employment status, and location in cities with public transport, rather than their immigrant status as such. The same can be said for people with disabilities versus those who do not report disabilities. Regression models controlling for these factors still find that recent immigrants have a higher propensity to use TNCs on a monthly basis or taxis/TNCs on a daily basis. Meanwhile people with disabilities remain less likely to report monthly TNC use even when controlling for their lower income, lower probability of employment and more advanced average age. Daily taxi/TNC use by people reporting a medical condition remains higher when controlling for those factors, but it is lower in the largest cities.

For recent immigrants, TNCs may play an important supplementary role in enabling multimodal travel, and could even serve to delay auto ownership among the recent cohort in comparison to settled immigrant cohorts. For people with disabilities, TNCs do not appear to be playing as important a role as policy makers might hope. In the largest cities with the greatest number of TNC users, taxi/TNC trips are now dominated by TNCs which are much more heavily used by people without disabilities.

### **Audience Dialogue Highlights**

Audience members suggested several additions to the research including examining transit access, zero car households, considering food trucks bringing goods to neighborhoods, and including wholesale stores.

The New York dataset wasn't included in the research because each method requires 100 iterations to capture the range of errors. Statistical considerations, especially when considering the use of Big Data, included: random selections; validation of plan; variable selections; model validation; and model optimization for the machine techniques. There were 20 variables were used based on the literature.

The researchers used a national perceptives rather than stratifying across regions. There was concern expressed over whether the 2009 recession and changes in work force participation would have impacted mode share. Researchers indicated this would not be the case, but that households have challenges affording a car based on average household incomes, which could explain the decline in trip rates.

Regarding accessible for populations in needing special accommodations, TNC apps are being designed with visual impairment considerations. Taxis have accommodations for people with disabilities, while others in the TNC category do not appear to provide such service. Currently, there are different regulations in different urban areas addressing this issue.

Accessibility and service level guarantees are important and perhaps can be addressed using performance metrics related to accessibility.

## Chapter 11: Poster Session

**GREG MACFARANE**  
**JOSIE KRESSNER**  
*Transport Foundry*

**KYLE SCHROECKENTHALER**  
**STEPHEN FITROY**  
*EDR Group*

**THERESA FIRESTINE**  
**CLARA RESCHOVSKY**  
*Bureau of Transportation Statistics, USDOT*

**KATHLEEN YU**  
**ARASH MIZAEI**  
*North Central Texas Council of  
Governments (NCTCOG)*

**SEPEHR GHADER**  
**AREK DARZI**  
*University of Maryland*

**SHARI GERSHENFELD**  
*Mobile Market Monitor*

**SHUYAO HONG**  
*Maricopa Association of Governments*

**GARY JORDAN**  
**PANAGIOTIS ANASTASOPOULOS**  
*University at Buffalo, SUNY*

**LAURIE WARGELIN**  
**CHET BOWIE**  
*University of Chicago*

**JEREMY WILHELM**  
*Westat*

**NANCY MCGUCKIN,**  
*Travel Behavior Analyst*

**ANTHONG FUCCI**  
**ALEXAMDER CATES**  
*Westat*

**AREF DARZI**  
**LEI ZHANG**  
*University of Maryland*

**MATTHEW COOGAN**  
*Independent Consultant*

**MARGARET CAMPBELL**  
*RSG*

**CHANDRA BHAT AND JOSEPH HUTCHINSON**  
*University of Texas, Austin*

**STACEY BRICKA**  
*MacroSys LCC*

**TIMOTHY REUSCHER**  
*ORNL*

**LISA GREEN AND MICHAEL MARTIN**  
*Texas A&M Transportation Institute*

**HUYEN LE**  
*Virginia Institute of Technology*

**LISA AULTMAN-HALL**  
**JOHNATHAN DOWDS**  
**GRE THIVIERGE**  
**ANUARBEK ONAYEV**  
*University of Vermont*

Workshop participants enjoyed meeting with poster authors during the early evening reception. Fifteen posters were on display and covered a wide range of NHTS research and applications.

## **USING THE NHTS AS A DAILY ACTIVITY PATTERN ARCHIVE IN A PASSIVE DATA MODEL**

*Greg Macfarlane and Josie Kressner*

### **Background**

Travel modelers commonly use passive transportation-related data, such as cellular or GPS origin-destination matrices, to calibrate or validate their planning models. The passive data are generally used, in this case, to expand and adjust behavioral models estimated from a small-sample local household survey. Recently, there has been interest in deriving synthetic records of individual travel directly from the passive data, with the support of other datasets. Kressner (2017) previously developed a method to build synthetic daily activity and travel patterns for a complete population. In this method, simulated individuals used the 2009 NHTS as a basis for tour patterns while passive origin-destination matrices spatially locate the simulated tours within a specific region.

In this research, we update our tour patterns to the 2017 NHTS and comment on observed differences in the resulting simulated travel patterns. Specifically, we consider the temporal distribution of weekday trips. We also examine the consequences of these differences in the synthesized daily schedules and in their traffic assignment. The results show that the NHTS has a similarly high proportion of mid-day, off-peak travel in both 2009 and 2017, and that this proportion is higher than other comparable local household surveys. The results also show that the change from the 2009 to the 2017 dataset does not substantively affect the simulated demand or assignment in a data-driven travel model using NHTS data.

In a conventional travel modeling paradigm, planners estimate econometric models of travel behavior — mode choice, destination choice, etc. — and expand the survey to match the total population distribution. Under this paradigm, a local dataset on which to estimate the models is important to ensure the models capture region-specific behavior. Passively collected travel data, such as cellular phone traces and highway counts, are used to validate and calibrate the models, but are not useful in estimation because they do not contain information on the behavior and choices of individuals.

The general NHTS (U.S. Department of Transportation 2017) is not typically suitable for a conventional travel modeling exercise, because the local sampling rate is too small and the geographic resolution is too large for model estimation or calibration (with some modifications, state or MPO-specific NHTS add-ons may be used).

In the past several years, there have been efforts in the travel demand modeling community to explore the use of so-called “data-driven” travel demand models. In contrast to conventional models, a data-driven model uses the large-scale passive data explicitly in the model process. Kressner (2017) previously described a methodology for such a model; in this methodology, synthetic individuals generate tours by referencing the NHTS for individuals living in similarly-sized cities and then draw locations for their tour activities from origin-destination datasets specific to the model region. The resulting synthetic tours can then be

assigned to highway and transit networks to understand route and mode choice. In this research we use an open-source transport network simulator, MATSim.

The most recent two iterations of the NHTS were in 2009 and 2017. Though the methodologies between successive survey collections did change somewhat, enough similarities exist that a comparison of travel trends between is useful and important. In this research we consider the diurnal distribution of tour start times and trips-in-motion reported in the NHTS and other related household surveys, and we compare the results to the data-driven model developed by Transport Foundry.

## Methodology

We developed a data-driven model for Asheville, North Carolina, which is a medium-sized city in western North Carolina. The CBSA in Asheville has a population of approximately 400,000 people. We used the public data files for the 2009 and 2017 NHTS for respondents living in a CBSA with a population between 250,000 and 500,000 people. In both years, there are more than 20,000 respondents in cities of this size. We compare the distribution of tour types and trips-in-motion in the data-driven simulation against the public NHTS schedules data from which the model is derived.

## Preliminary Research Findings

*Tour Start Distribution.* For the distribution of tour types, we calculate the probability of archetype tours beginning in each 90 minute period for workers only, accounting for person weights in the NHTS data. Non-workers are excluded in this plot for simplicity. An archetype tour is a means of tour classification: a “simple work” tour archetype includes a single work activity, a “multi-part work” tour includes more than one work activity, and a “composite” tour includes both work and non-work activities.

The distribution of probabilities for each archetype of tours by time of day and a distribution of the difference between the distributions by purpose and period were produced. Workers in the 2017 survey are somewhat less likely to make tours that involve a work activity, and somewhat more respondents take multi-part work tours. There is also less of a peak in the late afternoon in 2017. Overall however, the distributions are comparable. That is, the largest increase in probabilities are for multi-part and simple work tours, and the decrease has come primarily from simple work and composite from-work tours. This may indicate a true trend in travel behavior, or a discrepancy in how respondents classified the purpose of their activities, or simply a random variation. That said, there are over 20,000 respondents to the NHTS in cities of this size in both survey years, so a random variation is unlikely. With this information as a background, we now turn to understanding how this change in tour information affects the simulated trips and tours in the data-driven model. To do this, we consider the temporal distribution of trips-in-motion.

*Trips-in-Motion.* For a distribution of trips-in-motion, we calculate the number of trips moving during each 15 minute interval throughout a 24-hour weekday period. To be counted in a 15 minute interval, a trip either needs to be currently underway or completed entirely within the next 15 minutes. We generated separate distributions for home-based work (HBW) trips and all trips. We also obtained comparable distributions from two other data sources: a local household travel survey for the Hickory, North Carolina region and a 2009 NHTS add-on for the

Blacksburg, Virginia region. Though these regions are both smaller than Asheville in terms of population, they share common elements in culture, geography, and infrastructure.

## REFERENCE

Kressner, J. D. (2017). *Synthetic Household Travel Data Using Consumer and Mobile Phone Data* (No. NCHRP Idea Project 184).

## APPLICATION OF NHTS TRANSFERABILITY STATISTICS TO MEASURE IMPACTS OF TRANSPORTATION REVENUE TOOLS ON HOUSEHOLD SEGMENTS

*Kyle Schroeckenthaler and Stephen Fitzroy*

### Background

Alternative revenue mechanisms are a key topic for transportation professionals today due to the decline in fuel excise tax revenue resulting from increasing fuel economy and greater use of alternative fuels, especially the expected future electrification of the passenger automobile fleet. The research examines the distributional effects of transitioning from a fuel excise tax-based revenue regime to a mileage-based road usage charge ecosystem.

This research was carried out on behalf of the Western Road Usage Charge Consortium (RUC West) under two consecutive contracts to evaluate the household fiscal impacts of transitioning from a gasoline excise tax to a road usage charge (RUC) at a revenue neutral rate. A major goal of this project was to develop a methodology that could be applied across a variety of states – it has been employed for 10 so far. States include Arizona, California, Colorado, Hawaii, Idaho, Montana, Oregon, Utah and Washington.

There is a significant body of work on the distributional effects of RUCs, especially with regards to urban and rural populations. However, this research is the only work that the authors are aware of that provides a multi-state framework rather than a national or state-specific analysis. This allows states to understand differences and similarities between their revenue environments, which might affect a future multi-state implementation of RUCs.

### Methodology

The Transferability Statistics work by the Bureau of Transportation Statistics was one of the foundational sources for this work, which allowed estimates of household VMT a fine-grained level in all states that varied across urban, suburban, and rural areas and different regions of the country. The authors combined the regression coefficients from the Transferability Statistics with current data on household characteristics from the ACS and full detail on vehicle registrations from each state. The NHTS Transferability Statistics allowed all analysis to be completed at a census tract level, allowing mapping of results within and across states and comparisons of the range of expected impacts between rural tracts (and many other comparisons).

By utilizing state registration databases, the authors were able to link travel data from the NHTS with detailed census tract by census tract estimates of vehicle fuel types and fuel

efficiency rather than relying on the survey responses. Previous studies of RUCs have utilized NHTS samples for a specific state and decoded the vehicle make and model information to estimate fuel efficiency and fuel type. However, our data fusion approach allowed us to identify significantly finer grained geographic differences and produce maps and charts that were helpful for communicating household impacts with decision makers. Most states have insufficient samples within the NHTS dataset to truly explore effects in different regions of the state.

In the second project, the authors identified additional parameters of travel behavior and vehicle characteristics that could be captured in a RUC rate formula for testing. After considering 13 potential parameters, vehicle fuel types and fuel efficiencies were chosen for further analysis. The data fusion methodology provided a framework for tracking the distribution of fuel efficiencies for vehicles within each census tract and how they differed from statewide averages. Mileage allowances could be assessed in terms of household impacts using relationships from sociodemographic data such as average vehicle ownership.

All analysis was scripted in R to allow scenario testing, comparisons between proposed formulas and updating if new ACS data or registration information was available. The work has focused on differences between urban and rural households but could easily be applied to examine other equity or efficiency concerns such as income variables captured in the NHTS, Transferability Statistics, and ACS data sources.

### **Preliminary Research Findings**

The research found that rural households could potentially benefit from transition to a RUC, contrary to conventional wisdom. Outer suburban households with long commutes into urbanized areas are most likely to pay more under a RUC. The analysis also examines the distribution between households in the urban, mixed, and rural regions of each state and finds that less than 5% of rural tracts are likely to see household payments rise on average, while there are a meaningful number of urban tracts that can expect savings even though statewide urban households contribute a higher share of revenue.

The changes in the distribution of payments are very small on average, in many cases less than \$10 per year. Because urban households represent a greater share of population in almost all states, their payment increases are even smaller in dollar and percentage terms than the savings for rural households. Including additional parameters of a RUC allow agencies the possibility of pursuing other policy objectives through this revenue instrument while still improving the sustainability and resolving some perceptions of inequality under the fuel excise tax. The authors found that the tested fuel efficiency parameters could reduce the distributional changes caused by shifting from an excise tax to a RUC, while reducing the range in payment rates for vehicles of different efficiencies. Fuel type parameters had negligible impacts for the current fleet because alternative fuel penetration remains relatively low (especially for electric vehicles not covered by the fuel excise tax).

This research has not assessed the range of payment changes within tracts – focusing only on tract level averages and averages for different parts of the state (statewide there is no change in payments due to intentionally revenue-neutral policy design). There is of course significant variation in the change of payments at a vehicle level and at the individual household level – which the research did identify when examining how payments changed for different fuel types. However, for making statewide policy decisions, average impacts on different population groups and geographic regions provides valuable information.

Future research could apply these methods to forecasts of the vehicle fleet and travel demand to test the impact of different electrification policy targets or revenue goals. The methods developed for application of the NHTS and Transferability Statistics could also be beneficial to transportation planning and policy analysis on topics other than revenue considerations.

## **MODELING TRANSPORTATION CHARACTERISTICS: SMALL AREAS ESTIMATION USING THE NHTS**

*Theresa Firestine and Clara Reschovsky*

### **Background**

NHTS is an excellent source of travel information on how, why and by what modes people are traveling in the United States. While these data provide a reliable picture of trips for the nation, by states, and for selected 'add-on areas,' the NHTS is not considered a reliable data source for most smaller geographic areas due to sample size limitations. 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 ACS data.

This approach breaks out the NHTS data into six geographical areas and uses urban/suburban/rural classifications to estimate average weekday household person trips, vehicle trips, PMT, and VMT. These estimates are then transferred to individual census tracts using the household and demographic data from ACS for each census tract. The resulting estimates provide potentially 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 estimate from one Census tract can be compared to another even when not in the same geographic area. This would not be possible when comparing local surveys with differing methodologies.

### **Methodology**

BTS initially produced Census tract level estimates using the 2009 NHTS and the 2007-2011 five year ACS estimates. BTS divided the NHTS data into six census region/division groups and then subsetted into three density groups, for a total of 18 separate categories. BTS developed a regression model for each to predict average weekday household person trips, vehicle trips, person miles, and person trips. In looking at average weekday travel, BTS observed considerable variation both across geographical divisions and between urban groups.

The geographic disaggregation therefore ensures more homogenous groupings of households for the regression equations and thereby produces more accurate estimates. BTS included the following explanatory variables: number of vehicles in the household, number of workers in the household, median household income, homeownership, presence of a child less than 18, and several lifecycle indicators (1 person household less than 65; 2+ person household with no persons 65+; and 2+ person household with 1 or more persons 65+). BTS evaluated the prediction accuracy of the regression equations using the non-public NHTS files. BTS additionally identified and flagged statistical and spatial outliers. The final dataset contains estimates of average weekday household person trips, vehicle trips, PMT, and VMT for all Census tracts passing statistical and spatial checks by number of vehicles available and number



of persons in the household. BTS is now working to produce estimates using the 2017 NHTS and the 2012-2016 five-year ACS. As part of this work, BTS is refining the model to improve the quality of the data. Refinements include combining subgroups, where possible, to handle instances where the ACS MOE exceeds the estimate and otherwise flagging the data to indicate its lower reliability.

### **Preliminary Research Findings**

This research presents the initial model and results using the 2017 NHTS and the 2012-2016 five-year ACS, identifying the issues and challenges of fusing data and producing estimates for small geographies. Challenges include validating the results. Validation requires an indication of what is truth or a reasonable estimate. This task, at the national level, is difficult. As with any model, its output is only as good as its inputs and assumptions. In constructing the assumptions, BTS attempted to eliminate problems that may skew the results such as flagging estimates produced from ACS estimates where the MOE exceeds the estimate. BTS also excluded NHTS records where the respondent either moved during the survey process or was on a trip to a different location. Trip rates and travel behavior for these persons may not be representative of that person's NHTS tract of residence.

Other challenges involve analyzing the impact of methodological changes in collecting the survey data, for both NHTS and ACS between 2009 and 2017, along with changes to the processing of that data versus real demographic and socioeconomic changes. Further work would help to tease out changes and understand the impacts of these changes on trip rates and travel behavior.

## **A COMPARISON OF THE NCTCOG WEIGHTING METHOD TO THE NHTS WEIGHTING METHOD**

*Kathleen Yu and Arash Mirzaei*

### **Background**

This research explains the NCTCOG's method of expansion of the NHTS 2017 add-on records in the DFW Metropolitan Planning Area.

### **Methodology**

The NCTCOG method used a reduced recordset of weekday households, different socio-economic cross tabulations at the county level, and consistent weights in all survey tables. The reduced recordset was due to the consideration of national holidays and local holiday periods in the DFW Area. In addition, records with incomplete values for variables used in the expansion were excluded. The NCTCOG expansion method incorporated different household variables than were used in the standard NHTS 2017 expansion. The NCTCOG household variables were based on the number of workers, household size, household income, and number of vehicles in the households. Control totals were taken from the 2016 ACS at the county level, whereas NCTCOG data was originally expanded using 2015 ACS data at the regional level. The main NHTS 2017

used a waterfall method to assign weights for households and persons of the same household; this method relies on separate external target values for households and persons.

### **Preliminary Research Findings**

The waterfall method results in different weights assigned to households and the persons in those households, and different weights between persons in the same household. NCTCOG's method for weighting the household travel survey records was based only on household characteristics. The product of this method are consistent weights for each record throughout different tables of the survey. In order to consider persons in the weighting, some of the weighting variables included categorizing households by person characteristics including age and sex. This presentation will compare the results of using the original NHTS 2017 5-day weights against NCTCOG's expansion of the 2017 data. In addition, a temporal comparison of these weights will be done by comparing NCTCOG's expansion of the NHTS 2009 dataset versus NCTCOG's expansion of the 2017 dataset. These comparisons will identify if the finer geographic level for expansion is necessary. This study also investigates the use of consistent household weights versus traditional waterfall weights in measuring travel behavior. The temporal comparison could measure the changes in quantitative travel behavior such as trip rate and trip length distribution.

## **CALIBRATION AND VALIDATION OF A MICRO-SIMULATION LONG-DISTANCE TRAVEL DEMAND MODEL USING 2017 NHTS**

*Sepehr Ghadr and Aref Darzi*

### **Background**

The increasing interest in national transportation policies—from strategic infrastructure investment to infrastructure operation and management with regard to efficiency, sustainability, and safety has incited researchers and decision makers to call for advanced and policy-sensitive analysis tools<sup>1</sup>. Highway infrastructure investment, high-speed rail, and airport development all depend on national travel markets. Any infrastructure investments or operational and management improvements should be evaluated through a capable national travel analysis tool instead of regional-level, corridor-level, or state-level models, which are mainly used in these types of analyses.

After the Intermodal Surface Transportation Efficiency Act (ISTEA)<sup>2,3</sup> was enacted in 1991, many state departments of transportation began developing statewide travel demand models as critical analysis tools in addressing legislative requirements in statewide planning. However, the statewide models are weak in external trips, which are usually generated with information from federal and neighboring states. A national long-distance travel demand model can provide external trips for statewide models in base-year and future-year. According to Giaimo and Schiffer's review<sup>4</sup> of statewide travel demand modeling developments, most statewide travel demand models in the U.S. do not consider long-distance travel. However, this situation has started to change. Today more than 35 states have active modeling efforts to meet statewide policy and legislative development needs.

National long-distance passenger travel demand analysis has been an understudied area in transportation planning. The nation and various states are engaging in funding transportation infrastructure improvements to meet future long-distance passenger travel demand including: interstate highway tolling/expansion, high-speed rail, and next-generation passenger air transportation system that relies more on smaller airports and aircrafts. In addition, development in underway for a national Multimodal Travel Analysis System and an American Long-distance Personal Travel Program have become the priorities and fundamental work for planners when conducting national travel analysis. One of the main reasons for insufficient efforts in modeling long-distance national travels in the U.S. is the lack of recent data resources. American Travel Survey (ATS) was supposed to be the survey dedicated to long-distance travels, but it was only conducted in 1995. This 1995 ACS is the only source of long-distance travel survey data with sufficient sample size to estimate all components required in a long-distance travel demand model. Later NHTS included long-distance trips, but the long-distance sample size was considerably smaller than ACS, not enough to estimate a comprehensive long-distance travel demand model.

### **Methodology**

The 1995 ACS is an appropriate dataset for building a long-distance travel demand model. The research team at the University of Maryland took advantage of this dataset, and estimated sets of econometric models on long-distance travels. The estimated models were utilized in a micro-simulation travel demand model that simulates the entire U.S. population and their national long-distance trips. This model was the first activity-based micro-simulation effort of modeling national long-distance travels of U.S. residents. The model is described in more details in the next section of the paper. Even though the micro-simulation model is able to generate result using the up-to-date population, land use, and skim inputs, the results might not be able to fully represent the current travel patterns since the model was estimated with a dataset collected in 1995. Even though the micro-simulation model can use the up-to-date population, land use and skim inputs, it is still not sufficient for us to trust the results of a model estimated with a dataset collected in 1995. Travelers' preferences may have been changed due to new technologies, new lifestyles, and travel trends. Therefore, the estimated values might be changed. This highlights the probable need for a calibration effort. The NHTS surveys provide a unique opportunity to validate and calibrate the long-distance model.

### **Preliminary Research Findings**

Table 11.1 compares the number of long-distance observations in the new 2017 NHTS and compares it with 2009 NHTS. Model calibration involves the optimization of model coefficients to match statistics obtained from an independent ground-truth dataset. The optimization is a step-by-step procedure of changing parameters to improve an objective function. The direction of change in the parameter space in each step is obtained from the gradient. In many applications, such as the case of this paper, a mathematical representation of the problem to calculate the gradient is not available. Consequently, information on the direct measurement of the gradient vector is not available. In such cases, the gradient can be approximated from measurements of the objective function <sup>5</sup>.

**TABLE 11.1 Number of Long Distance Observations.**

Survey	Number of Trips	Number of Long Distance Trips
NHTS 2009	1,167,321	27,718
NHTS 2017	923,572	25,461

In the calibration of the micro-simulation national long-distance model, due to the many sub-models embedded in the model, there is no simple closed-form mathematical relationship between model coefficients and model results. As a result, gradient needs to be approximated. The gradient approximation requires measuring the objective function at two adjacent points. Here, the objective function is the measure of the difference between simulation results and independent ground-truth data (statistics obtained from up-to-date travel surveys). The challenging part is the amount of computation required for running the simulation model. The micro-simulation model simulates about 350 million individuals, which requires considerable computational resources and time. Therefore, measuring the objective function for each step of the optimization is very costly. This highlights the need for a very efficient optimization algorithm. The efficient algorithm should require only a few simulation runs to get close to the optimal value.

Simultaneous Perturbation Stochastic Approximation (SPSA) is an efficient method used in many complex multivariate optimization applications<sup>6,7</sup>. The power of SPSA is in its efficiency and ease of implementation. The algorithm is very efficient, in that its gradient approximation only requires two measurements of the objective function. SPSA has been applied in transportation to calibrate traffic simulation models (See<sup>8,9</sup>), calibrate traffic assignment models (See<sup>10,11</sup>), Origin-Destination calibration (See<sup>12,13</sup>), and demand model estimation using machine-learning (see<sup>14</sup>).

This work contributes to the travel demand modeling literature by proposing an efficient way to calibrate large-scale travel demand models, and presenting the calibration results. The value of model calibration is significant, considering time, cost, and data requirements for estimating new models. In the age of big data, with advancements in new mobile technologies and improvements in data size and data quality of passively collected data, passively collected data can serve many of the travel demand modelers' needs. The next generation NHTS will include a national passively collected data component, which can be a valuable source of data for calibration of large-scale travel demand models such as statewide and national models. This research showcases the calibration methodology by calibrating the destination choice component of the model using the most recent NHTS 2017).

## REFERENCES

1. Lundqvist L, Mattsson L-G. *National transport models: Recent developments and prospects*. Springer Science & Business Media; 2002.
2. Dilger RJ. ISTEA: a new direction for transportation policy. *Publius: The Journal of Federalism*. 1992;22(3): 67-78.
3. Gage RW, McDowell BD. ISTEA and the role of MPOs in the new transportation environment: A midterm assessment. *Publius: The Journal of Federalism*. 1995;25(3): 133-154.
4. Giaimo GT, Schiffer R. Statewide Travel Demand Modeling: A Peer Exchange, Longboat Key, Florida, September 23-24, 2004. *Transportation Research E-Circular*. 2005(E-C075).

5. Spall JC. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins apl technical digest*. 1998;19(4): 482-492.
6. Spall JC. *Introduction to stochastic search and optimization: estimation, simulation, and control*. Vol 65: John Wiley & Sons; 2005.
7. Spall JC. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control*. 1992;37(3): 332-341.
8. Ma J, Dong H, Zhang H. Calibration of microsimulation with heuristic optimization methods. *Transportation Research Record: Journal of the Transportation Research Board*. 2007(1999): 208-217.
9. Antoniou C, Azevedo CL, Lu L, Pereira F, Ben-Akiva M. W-SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. *Transportation Research Procedia*. 2015;7: 233-253.
10. L, Xu Y, Antoniou C, Ben-Akiva M. An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models. *Transportation Research Part C: Emerging Technologies*. 2015;51: 149-166.
11. R, Ben-Akiva M, Koutsopoulos H. Offline calibration of dynamic traffic assignment: simultaneous demand-and-supply estimation. *Transportation Research Record: Journal of the Transportation Research Board*. 2007(2003): 50-58.
12. Tympakianaki A, Koutsopoulos HN, Jenelius E. c-SPSA: Cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin-destination matrix estimation. *Transportation Research Part C: Emerging Technologies*. 2015;55: 231-245.
13. Kattan L, Abdulhai B. Noniterative approach to dynamic traffic origin-destination estimation with parallel evolutionary algorithms. *Transportation Research Record: Journal of the Transportation Research Board*. 2006(1964): 201-210.
14. Balakrishna P, Ganesan R, Sherry L. Airport taxi-out prediction using approximate dynamic programming: intelligence-based paradigm. *Transportation Research Record: Journal of the Transportation Research Board*. 2008(2052): 54-61.

## **INSIGHTS ON DATA QUALITY FROM A LARGE-SCALE APPLICATION OF FUTURE MOBILITY SENSING TECHNOLOGY IN GREATER PHOENIX, AZ**

*Shari Gershenfedl and Shuyao Hong*

### **Background**

In 2016-17, the Maricopa Association of Governments (MAG) conducted its first 100% GPS-based survey using the Mobile Market Monitor (MMM), an automated travel survey platform. The MMM platform uses Future Mobility Sensing (FMS) capable of leveraging sensing technologies and machine learning techniques to obtain high resolution, multi-day data.

### **Methodology**

Previous studies comparing these new technologies to traditional survey data, conducting in Singapore and Israel, found the FMS data to be more accurate, complete and richer. The research team examined weighted data distributions from MAG's surveying efforts to NHTS surveys.

## **Preliminary Research Findings**

The comparison of the two datasets revealed similar overall patterns (e.g., travel time distribution for direct home to work vehicle trips). NHTS respondents, reporting travel from recall, tend to round travel times, resulting in jumps in travel time the distribution curve. The FMS methodology detects travel time directly from an app, giving it a much smoother distribution. The comparison between FMS and NHTS with respect to intermediate stops during home to work trips found the NHTS reporting 83% compared to 72% reported by FMS. The main types of intermediate stops in the NHTS included: drop-off/pick-up; eating meals out; exercise; and shopping. For the FMS, the reported stops included: drop-off/pick-up; change of mode/transfer; shopping; errands; and eating meals out. The FMS/MAG Household Transportation Survey revealed a higher percent of passengers making intermediate stops during home to work travel and with a greater variety of activity sequences.

While the NHTS indicates the vast majority of work activities preceded and/or were preceded by home activities. The MAG /FMS HTS showed a much richer context. This difference suggests these types of activities are under-reported in the NHTS. Since FMS is sensitive to location changes, a long work stop may appear to be broken into several smaller ones. As a result, if the traveler moved a significant distance during a particular time period (e.g., moving around in a campus setting), it could impact analysis.

The FMS platform application produced high quality, multi-day data, enhanced with strong respondent support. A richer activity context was observed in the MAG/FMS HTS, with a large variety of activities before and after work. More intermediate stops were captured with FMS, in addition to the main trip destination. Finally, there are significant rounding effects in reporting travel times in the NHTS, whereas the FMS captures travel time much more accurately.

## **A BREAKTHROUGH SOLUTION FOR STATISTICAL SURVEYS OF U.S. POPULATION**

*Laurie Wargelin and Chet Bowie*

### **Background**

Surveys and Big Data are complementary data sources, not competing sources. There are differences between approaches and uses, but these should be seen as an advantage. For travel planning and travel modeling needs, Big Data (particularly cellular data) will increasingly be important for calibrating journey to work data and time of day travel patterns at the geographic aggregation level. Big Data will be needed to augment and update trends in travel, special generators and external travel, and for defining visitor and long distance markets. At the same time, surveys will be important for differentiating travel by trip purpose and by market segment, for resolution of non-work travel purposes, and for inference of tours and joint travel. “While Big Data is the new paradigm; individual-level analyses of travel behavior will still be needed. The behavioral paradigm has staying power.” The problem is that more efficient ways of collecting national travel survey data are needed now. Continued reliance on repeated custom recruitment of households from full coverage sample frames such as ABS is resulting in low response rates with prohibitive costs.

## Methodology

This research proposes the establishment of a new survey approach where large organizations could invest resources to establish sizable representative panels of individuals/households, recruited and maintained from ABS samples, with oversamples of desired populations (e.g., millennials). Standard profiles for these panel would include gender, age, educational attainment, race and ethnicity and for household characteristics renter/owned, household income, Internet/cell access, household size, and age of householders. National panels would be maintained with on-going procedures to update participation and demographic characteristics using continuous protocol for retention and replacement. National panel sizes range from 30,000 to 50,000 households.

For NextGEN NHTS, a panel would conduct a survey expected to have a 50% participation rate from panel households with a stage-two response rate estimated at 67% for travel inventories collected with preferred smartphone app passive and active data capture. A panel based multi-modal recruit interview for NextGEN NHTS would be reduced to 5-8 minutes and re-contacts for stage-two travel data would diminish due to pre-agreements to participate. A national panel would provide ready samples for cognitive and usability testing and large-scale pretests. Panel oversamples would ensure representativeness. Panels would have defined programs for incentive awards at all stages and are fully conducive with conduct of the NextGEN NHTS as a continuous survey over an extended period. Finally, panels would allow for robust behavior survey hypernetwork access to rare travel populations such as those who use ridesharing, bike commuters, and frequent online shoppers.

## Preliminary Research Findings

The proposed methodology should provide new outcomes for transportation planners. The approaching future will likely be characterized by a fusion of new and traditional data sources, with new reports and national outlets going forward.

## REFERENCES

1. Kimon Proussaloglou, Cambridge Systematics, Inc., NCHRP 08-95, Presentation to ABJ40 Travel Survey Methods, 1/10/2018.

## RETHINKING THE SURVEY PART OF THE NHTS

*Jeremy Wilhelm and Nancy McGuckin*

### Background

The NextGen NHTS will include the integration of passive or ‘big data’ with smaller, more frequent surveys to generate datasets yearly with the goal of providing a more dynamic and relevant snapshot of current travel trends. ABS has been the industry standard since the declining coverage of RDD because it both provides ideal coverage and allows for the development of probability-based, representative samples. However, low response rates result from address-based surveys (in regional surveys, generally below 10% in recruitment and below 7% overall).

## Methodology

The 2017 NHTS sought an improvement to the initial and overall response rates by utilizing a legacy survey method of paper mail-back forms. This approach succeeded yielding an average initial response rate of 27.1% and final rate of 13.9%, which, while representing a major improvement on other, current approaches, still falls short of standards and guidelines set forth by the OMB and the United States Standards and Guidelines for Statistical Surveys(1). The later guideline suggests an average response rate of 85% and notes that most surveys achieved better than 70% response rates. Barring rates at or above that level, OMB requires that survey administrators conduct a rigorous non-response bias analysis and state their case for why lower response rates will be adequate.

Low response rates in HTS appear to be a reality irrespective of the methods selected, the firms conducting the research, or the region of the survey. This may call into question the reliability and representativeness of the data produced. However, it is clear that the need for these surveys will persist so long as forecasting and demand modeling are part of the planning process. The question then is two-fold. Can ABS surveys remain affordable with response rates declining, and are there sufficiently equivalent but more cost-effective options? This paper focuses on the second part of the question and examines the merits and limitations of non-probability approaches to sampling, with a review of publications related to best practices in sampling, weighting, and expansion of non-probability produced datasets. We then discuss the merits and challenges of utilizing several non-probability sample frames with the NHTS in mind. We conclude with thoughts for how best to incorporate non-probability sample in the NextGen NHTS toolkit.

## REFERENCE

1. Office of Management and Budget. (2018, April 4). statistical-programs-standards. Retrieved from [www.whitehouse.gov: https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/assets/OMB/inforeg/pmc\\_survey\\_guidance\\_2006.pdf](https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/assets/OMB/inforeg/pmc_survey_guidance_2006.pdf).

## EXPLORING METHODS TO INCREASE EFFICIENCY IN GENERATING AND COMPARING NATIONAL TRAVEL BEHAVIOR ESTIMATES OVER TIME

*Anthony Fucci and Alexander Cates*

### Background

Generating estimates to make statistically representative claims about the population from a NHTS dataset is typically conducted with statistics-focused software, usually within SAS, SPSS, STATA or R. Like other complex sample surveys, a significant task the software assists one with is correctly performing variance estimation for a desired statistic, needed to make statistically significant claims about the population. This research demonstrated a framework using R software that simplifies generating national travel estimates across the three most recent NHTS programs, covering 2001, 2009 and 2017.

The principal aim is to increase efficiency and quality of travel trend analysis by reducing programming burden to increase focus on exploring statistics and interpreting data. The



presentation will explore automating reports with side-by-side faceting of multi-year estimates in tables, charts and maps, and it will automate some summarization of the differences between the NHTS years of interest and the provided estimate. The resulting report will demonstrate the method's potential for increasing our ability to study national travel behavior over time using a case example analysis topic: national regional geographies and travel behavior to access food.

## Methodology

The components required for generating estimates across NHTS programs may be grouped into high level categories. To increase efficiency in generating estimates for multiple years, each of the components listed in Table 11.2 are configurable at the program year level within the analysis software.

**TABLE 11.2 High Level Components Required for Generating Estimates.**

Item	Description
Core data	Data files with survey variables
Replicate weights	Data files with replicated sample weights needed to compute error
Standard error formula	The formula needed to compute error according to the weighting method
Analysis variables	Normalized variable definitions for comparing across programs
Statistical query	The definition of the statistic
Variable metadata	Meanings of coded values
Visualization object	Visualization preference for statistic

Using R software that has reduced each of these components to functions with input parameters, one can manage them to develop even higher level functions that focus on multi-year analysis. By having the capability to generate a statistic within a software environment that already understands the data domain, variable metadata, and weighting method for multiple NHTS years, we can codify and automate statistic requests across studies.

This method of increasing throughput takes advantage of these capabilities by developing an R software function that processes a text file of configured statistic definitions. The function processes the statistic for each year, compiles the multi-year results, and draws them as specified into a single document. Each item in Table 11.3 is an attribute named in the configuration file.

**TABLE 11.3 Components that Define a Statistic for Analysis.**

Item	Description
Title	The title of the configured statistic being defined
Base Statistic	Desired base statistic: count, proportion, mean, sum, median
Variable	Variable to use for numeric statistic, if selected
Grouping Variable(s)	Variable(s) by which to group the statistic
Proportion	Display count statistic as percent, if selected
Subset	Include or exclude certain observations

The burden of programming and organizing the components of generating an estimate is now being reduced to a single configuration file. At this point, we can summarize the analyst workflow using this method in four steps:

1. Define statistics in the statistic configuration file
2. Compare variables in statistics' definitions across years
3. Process configuration file
4. Analyze results

In step 1, the analyst provides values for the components in Table 11.3 that define their statistic. Aside from the Title for this example, the analyst will only need to input "person\_trip\_rate" for the Base Statistic and "CENSUS\_R" for the Grouping Variable. In step 2, the analyst will need to evaluate any variables used in their statistic definition for NHTS program transferability or likeness. That is, are the variables being included directly comparable across the NHTS years being analyzed? This is easy in the case example because the only variable specified, "CENSUS\_R," is a census classification variable that has remained consistent over the NHTS programs. However, in this important step, analysts will often need to normalize values or categories by creating new derived variables so that cross-year comparison is valid. This task is assisted by a text configuration file that allows the analyst to define new derived variables at the year level.

Analysts may focus on the likeness of their statistic's inputs, and what it means from year to year, and redefine or code them as necessary for more comparable or better analysis. In step 3, the analyst processes their configured statistics using a function that calculates each statistic for each year and renders their visualized output in an HTML report. In addition to rendering charts side-by-side, the processing function uses the confidence level intervals from each year's statistic to compute and then visualize the statistical differences from year to year. In the final step, analysts will review their results in the functionally-generated report.

### **Preliminary Research Findings**

Generating estimates from three complex surveys for trend analysis is reduced to four steps, requiring almost no computer programming by the analyst. The report output for our case example statistic, "Daily Person Trips by Census Region," visualizes all of its 12 computed interval estimates (four estimates for three years) for the analyst in two ways. They are first presented at the year level in independent bar charts with lines plotting the 95% confidence area for each estimate. Daily person trip estimates for each region can be compared side-by-side at the year level as they are presented or blown-up individually for review. The second part of the report presents the 12 estimates in a single crosstab table where the cells contain nested line charts that draw each statistic's change from year to year as increasing or decreasing based on significant changes according to the 95% confidence intervals of the estimates. On a typical desktop computer, all of this information fits and is presented on one web page. This convention also worked for more complex statistics and statistics with additional analysis variables.

Timeliness and efficiency of analysis is a regular challenge that is further complicated when looking across multiple survey program years. The method implemented in this presentation removes the data analysis bureaucracy of loading files, merging tables, labelling values, and programming statistics. Analysts focus on only two configuration files: 1) where they

define the statistic they would like estimated; and 2) where they normalize variable codes across years for comparability. By simply altering a single-field parameter to break apart an analysis group or change an aggregation level or statistic, transportation analysts are encouraged by automation to look at an increasing number of statistics to expand their knowledge in their topic of interest.

## REFERENCES

- Dippo, C. S., Fay, R. E., Morganstein, D. H. (1984) Computing Variances from Complex Samples with Replicate Weights. 1984 Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, D.C., 489-494
- Cohen, S. B. (1997). An evaluation of alternative PC-based software packages developed for the analysis of complex survey data. *The American Statistician*, 51(3), 285-292.
- G.M. Oyeyemi, A.A. Adewara and R.A. Adeyemi, 2010. Complex Survey Data Analysis: A Comparison of SAS, SPSS and STATA. *Asian Journal of Mathematics & Statistics*, 3: 33-39.
- Fucci, A., & Cates A., summarizeNHTS, (2017), GitHub repository, <https://github.com/Westat-Transportation/summarizeNHTS>

## HOW BUILT ENVIRONMENT CHANGES SHAPE TRAVEL BEHAVIOR OVER TIME: COMPARISON BETWEEN 2009 AND 2017 NHTS

*Aref Darzi and Lei Zhang*

### Background

Urban sprawl, traffic congestion, increase in oil consumptions and climate change are some of today's most worrisome concerns in most countries. Higher speed of growth for passenger VMT rather than population, personal income, and developed land for several decades and over 3 trillion vehicle miles traveled and 176,000 million gallons of fuel consumption on all US roads are just a few samples of today's vexing problem that societies are dealing with. Besides that, acknowledging the fact that urban transportation produces about 30 percent of total carbon emissions produced in the U.S. (EPA, 2006) shows the inevitable needs of land use planning and urban design to control the problem. Urban transportation researchers and planners, as well as land use policymakers, have been trying to decrease VMT as a mean to mitigate traffic congestion, reduce energy consumption and GHG emission by understanding the relationship between land use and travel behavior in the past decades.

The objective of this research is threefold: (1) to reexamine the impacts of built environment on travel behavior, in particular, VMT; (2) to analyze the temporal trend of built environment within the same study area; and (3) to compare the variability of the built environment effects on the travel behavior based on the geographical characteristics of each observation.

The case study area includes the entire state of Maryland which includes the Baltimore metropolitan area as a large urban area and many small cities within the Maryland state representing small to medium-sized urban areas. Regarding the datasets, this study utilizes 2009NHTS and 2017NHTS to understand the temporal changes of built environment effects on

VMT and also use Smart location database (SLD) as the reference for the built environment variables.

There are statistically significant linkages among a wide range of built environment variables. These variables include measures of density, diversity, design, and destination accessibility, and travel behavior as trip generation, VMT, and mode choice (Cervero and Kockelman 1997; Ewing and Cervero 2001; Frank et al. 2007; Handy et al. 2002; Nasri and Zhang 2012; Zhang et al. 2012; Ewing and Cervero 2010). On the other hand, some studies found insignificant effects of certain built environment variable on travel behavior, such as travel speed and distance. (Boarnet and Crane 2001).

**Methodology**

The 2009 NHTS and 2017 NHTS datasets have been used in this study. The 2009 NHTS data has trip information for 355 households within the State of Maryland and the 2017 NHTS contains information for 1475 households. After going through the consistency check for both 2009 NHTS and 2017 NHTS, 589 and 2380 individuals are used in this study, respectively.

For the built environment variables, SLD has been employed in this research. SLD incorporates several measures of land-use for the entire State of Maryland at census block group level. The gross activity density, household workers per job equilibrium index, road network density, and job accessibility has been used as a measure for density, diversity, design, and destination accessibility, respectively. Table 11.4 contains the definitions of the variables and Table 11.5 presents the statistical summary of variables used in this research.

**TABLE 11.4 Definition of Variables.**

Measures	Definition	Expected Effect on VMT
edu	Binary variable: 1 for some college or higher, 0 Otherwise	Positive
work	Binary variable: 1 if a person's work status is yes, 0 Otherwise	Positive
HHFAMINC	Ccategorical variable, includes 11 level from less than \$10,000 to \$200,000 or more	Positive
R_sex	Binary variable: 0 for male, 1 for female	Negative
R_Age	Age of the respondent	Positive
age2	The square of the age of the respondent	Negative
D1d	Gross activity density (employment + housing units) on unprotected land	Negative
D2c_wremix	Household workers per job equilibrium index;	Negative
D3a	total road network density	Negative
D5ar	Jobs within 45 minutes auto travel time, time- decay (network travel time) weighted	Negative

The diversity equilibrium index is computed with the following equation:

$$\text{Diversity equilibrium index} = \exp - \left| \left( \frac{\# \text{workers}}{\text{total employment}} \right) - 1 \right| \quad (1)$$

where a number of workers and number of total employment are calculated at the census block group and closer values to one show more balanced the resident workers and jobs in the CBG are.

Two models have been considered in this paper to empirically measure the impacts of the built environment variables on VMT.

The multiple linear regression model has been used as a global prediction model which tries to model the relationship between a set of explanatory variables and a response variable by fitting a linear regression equation. In the linear regression model, every value of independent variables is associated with a value of the response variable (Gujaratti, 2009). Multiple linear regression model is shown in the following equation:

$$Y_i = \beta_0 + \sum_k \beta_k X_{ik} + \varepsilon_i \quad (2)$$

where  $Y_i$  is the respond variable,  $X_{ik}$  are the independent variables, and  $\varepsilon_i$  is the error term that captures the unobserved variables that have impacts on the respond variable in the  $it$  observation.

As discussed in the literature review section, spatial autocorrelation is one of the concerning methodological in this topic. The spatial autocorrelation happens when the independence assumption of observation does not hold and nearby observations show similar characteristics. This phenomenon is a common issue in geographic analysis and in presence of spatial autocorrelation the ordinary least square estimators are highly prone to biased estimations (LeSage, 1997; Hong et al., 2014; Nowrouzian and Srinivasan, 2013). In this study, geographically weighted regression (GWR) has been used not only to overcome the spatial autocorrelation issue but also to provide more locally estimates that can help urban planners and policymakers to gain more information about the sensitivity of individuals to built environment variables in different locations.

GWR model relaxes the constraint of the fixed coefficients over the space. The mathematical formation of GWR is presented in the following equation:

$$y_i = \sum x_{ij} \beta_j(p_i) + \varepsilon_{ij} \quad (3)$$

where the only fact that changes in this equation in compare to the abovementioned model is  $\beta_j(p_i)$  which denotes that the estimated coefficient in this model is a function of  $p_i$  which is denotes the observations the observation location in GWR model.

**TABLE 11.5 Socioeconomic and Built Environment Variables.**

<b>Socioeconomic variables</b>				
Measures	2009		2017	
	Mean	SD	Mean	SD
edu	0.57	0.49	0.65	0.48
work	0.53	0.50	0.55	0.50
HHF AMINC	\$60,000 to \$64,000		\$50,000 to \$74,000	
R_sex	0.52	0.50	0.51	0.53
R_Age	44.99	21.63	47.54	21.44
age2	2491.79	1874.64	2720.15	1954.92
<b>Built environment variables</b>				
Measures	Mean		SD	
D1d	7.4		14.6	
D2c wremix	0.2		0.31	
D3a	15.5		44.86	
D5ar	133043.7		87318.87	

## Research Findings

Global regression model reconfirms the significant impacts of socio-economic and built environment variables on the VMT. Furthermore, the GWR model shows a better descriptive power with a lower residual sum of squares and lower AIC. Also, the signs of the estimated parameters were consistent with the literature for both global regression model and the GWR model. Moreover, the built environment parameters estimated by GWR model show significant spatial variation over the study area which shows that relaxing the constraint of constant estimation for the entire study area can improve the modeling efforts of finding the linkage between built environment and travel behavior. Besides that, the variation on built environment parameters can also help urban planners and decision makers to analyze land use scenarios more accurately.

## REFERENCES

- USEPA. "Greenhouse gas emissions from the US transportation sector 1990–2003." (2006).
- Cervero, Robert, and Kara Kockelman. "Travel demand and the 3Ds: density, diversity, and design." *Transportation Research Part D: Transport and Environment* 2.3 (1997): 199-219.
- Ewing, Reid, and Robert Cervero. "Travel and the built environment: a synthesis." *Transportation Research Record: Journal of the Transportation Research Board* 1780 (2001): 87-114.
- Frank, Lawrence Douglas, et al. "Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity?" *Social science & medicine* 65.9 (2007): 1898-1914.
- Handy, Susan L., et al. "How the built environment affects physical activity: views from urban planning." *American journal of preventive medicine* 23.2 (2002): 64-73.
- Nasri, Arefeh, and Lei Zhang. "Impact of metropolitan-level built environment on travel behavior." *Transportation Research Record: Journal of the Transportation Research Board* 2323 (2012): 75-79.
- Zhang, Lei, et al. "How built environment affects travel behavior: A comparative analysis of the

- connections between land use and vehicle miles traveled in US cities." *Journal of transport and land use* 5.3 (2012): 40-52.
- Ewing, Reid, and Robert Cervero. "Travel and the built environment: a meta-analysis." *Journal of the American planning association* 76.3 (2010): 265-294.
- Boarnet, Marlon, and Randall Crane. "The influence of land use on travel behavior: specification and estimation strategies." *Transportation Research Part A: Policy and Practice* 35.9 (2001): 823- 845.
- Gujarati, Damodar N. *Basic econometrics*. Tata McGraw-Hill Education, 2009.
- LeSage, James P. "Regression analysis of spatial data." *Journal of Regional Analysis and Policy* 27 (1997): 83-94.
- Hong, Jinhyun, Qing Shen, and Lei Zhang. "How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales." *Transportation* 41.3 (2014): 419-440.
- Nowrouzian, Roosbeh, and Sivaramakrishnan Srinivasan. "Modeling the Effect of Land Use on PMT by Using Geographically Weighted Regression." *Transportation Research Record* 2397.1 (2013): 108-116.

## **UPDATE FROM 2017: DID THE MILLENNIALS REALLY CHANGE THEIR TRAVEL BEHAVIOR IN 2009?**

*Matthew Coogan and Margaret Campbell*

### **Background**

This research explores new and developing data concerning two policy areas: 1) auto ownership rates by age group; and, 2) VMT estimates by age group. First, if indeed a good portion of the population has turned away from the preference to owning private cars, this would support a hypothesis that very significant changes in behavior have taken place over the past two decades. Second, if such a change were to be manifest as significantly lowered VMT, the implications would be powerful. In both cases, this paper seeks to understand the difference between the younger half of the population and the older half.

### **Methodology**

Our prime concern is the difference in economic and travel conditions between 2009 and 2016. The per-capita rate of auto ownership tumbled in the period before, and even during the 2009 survey year. The economic slowdown in the United States flattened in 2011, and the economic recovery was discernible in transportation patterns by the year 2013. The decrease in auto ownership revealed in the 2009 survey results has been reversed, with present day levels of the highest in history. But the use of aggregate data can hide valuable patterns, as in fact the patterns of auto ownership vary considerably by age category. This, of course, underscores the need for the kind of information we rely upon from the NHTS. In fact, the growth in the level of auto ownership between the 2009 survey period and the 2016 period is almost entirely attributable the older age groups, and not the under 35 at all.

### **Preliminary Research Findings**

The question of the increase in auto ownership that occurred between the survey years of 2009 and 2016. For those under 46 years of age, there is a slight but quite consistent decrease in auto

ownership between the 2009 bar and the 2016 bar. By contrast, for every age category over 47, the 2016 ownership levels are higher than the 2009 levels; and this pattern of increase over time gets *stronger* with increasing age. Thus, there are two patterns to observe. For each age category under 45, auto ownership levels could be characterized as either stable of the 15-year period or declining modestly over that 15- year period. At the same time, a historical pattern in which auto ownership declines in the early 60s is being replaced by a pattern where the same age group simply does not surrender their vehicles with increasing age.

This cursory examination of trends in auto ownership is largely consistent with work undertaken outside of the NHTS context, including TRB's National Cooperative Highway Research Program and its Transit Cooperative Research Program. Significant differences exist between younger and older Americans on the subject of need for owning a private automobile. For example, in response to the statement "I feel I am less dependent on cars than my parents are/were," those under 35 expressed modest levels of agreement, while those older expressed levels of disagreement which become stronger by increase in age. Similarly, those over 35 express clear disagreement with the statement "Because of new services helping me make trips, I feel less need to own a car," those under 35 express little or no disagreement with the premise. It is important to note, that no age group strongly agrees with this statement, with a strong pattern of rejection from those over 35. *All age groups* would prefer to own their own car, rather than participating in a program to share or rent a car when needed. Consistent with other patterns, those above 35 years of age reject the sharing concept more strongly than those under 35. In general, the attitudinal studies do NOT support any hypothesis of a mass rejection of private automobile ownership on the part of the millennials, or any other age group.

Arguably, the results of these attitudinal studies are consistent with the pattern of the level of auto ownership which for those under 40 years of age shows a small decrease; a decrease which is consistent over the disaggregated age categories. The attitudinal studies tend to support the concept that auto- ownership is not questioned by the older age categories, which is generally consistent with their very significant increase in auto ownership rates over the past 15 years. This supports the observation that VMT for the entire sample ("All") is up somewhat from 2001, and up more significantly from the economic doldrums of 2009. However, the experience varies distinctly by age group. Those under 30 seem to be driving less than they did 15 years ago. Those in their 40s and 50s seem to be driving about the same and those above 60 are driving significantly more than before. As has been extensively documented before, it was those under 35 who suffered the greatest loss of auto-based mobility in the 2009 timeframe.

The release of the 2017 NHTS allows further examination of the role of income level in the significant loss of auto mobility. For this socio-economic group, unlike the sample as whole, this group has not returned to the level of VMT generation they had in 2001. As shown in the "All" column on the right-hand side of the chart, a minor-bounce back is reported with the use of the adjustment factor recommended for detailed VMT observations.

The 2017 data can be used to finish observations left unresolved in 2009; auto travel per person is again increasing, fueled by increase use by those over 45 years of age. Auto travel rates for those in their 30s and under continues to be significantly less than in 2001. The portrait offered is one in which those under 40 have not rejected auto travel but are simply not increasing personal travel rates the way the older groups are. In response to the question "I need my car to get where I need to go," *all age groups agree with this premise*, while those under 35 report less agreement than the older age groups.



## THE CHANGING NATURE OF THE ACTIVITY-TRAVEL BEHAVIOR OF THE ELDERLY

*Chandra Bhat and Joseph Hutchinson*

### Background

Studying the mobility choices and needs of the elderly is increasingly important as the elderly population continues to grow. As people age, health-related issues, cessation of driving, and fear of uncomfortable travel situations further contribute to their decreased mobility (Marin-Lamellet and Haustein, 2015), putting them in a vulnerable position. Current seniors are, however, more active, retire later, and may even work full-time or part-time well after traditional retirement age when compared to earlier generations (Goulias et al, 2007). The elderly have higher rates of health and mobility issues than their younger counterparts, which renders driving a dangerous and sometimes even impossible task even as driving remains the preferred mode of transport.

A key objective of this study is to investigate differences among subgroups of the elderly population to recognize the heterogeneity in activity-travel behaviors that exist within this cohort. The study of the elderly's activity-travel has often treated individuals aged 65 or older as a monolithic group. These studies often use age as an explanatory variable, not focusing on explicit differences among disaggregate groups of the elderly.

Despite a growing body of literature that addresses heterogeneity within the elderly population, the need to analyze differences in behaviors among different elderly subgroups with respect to emerging activity-travel behaviors remains very high. With advances in medicine and the emergence of new mobility options, smartphone technologies, and online services, it would be of value to examine the extent to which activity-travel choices differ among elderly subgroups in the current context. The 2017 NHTS data set collected in the United States offers an opportunity to conduct such an investigation and explore heterogeneity that prevails today within the elderly population.

The analysis in this research focuses on three key aspects of activity-travel behavior. The first is action space, which represents the spatial extent of activity engagement outside the home. The second is mode usage by activity purpose, to capture differences in mode usage patterns that may exist among different subgroups of the elderly. The third dimension of interest in this study is that of time use allocation for various activities, with a view to investigate differences among elderly subgroups. Essentially, these three measures capture a spatial dimension (action space), a temporal dimension (time use), and a travel dimension (mode use). By examining these three diverse measures of behavior, this study aims to offer a rich set of insights into differences that exist (or not) among the subgroups of the elderly using the latest version of the NHTS data set.

### Methodology

We conducted a comprehensive exploratory analysis of the activity-travel of elderly persons. The study considers the ages of 65 to 74 years, 75 to 84 years, and 85 years or older. In addition, the age group 55 to 64 years is included, both as a basis for comparison and because some people in this age group retire early. These individuals may exhibit activity-travel patterns similar to those of persons aged 65 to 74.

A descriptive comparison shows that heterogeneity continues to exist; those in the older elderly subgroups are less likely to have driver's licenses, depict lower levels of mobility and trip

making, exhibit lower levels of car ownership, and undertake fewer activities and spend less time outside the home. In other words, the fading of activity-travel engagement with age continues to be an issue confronting individuals as they advance into the later stages and ages of life.

### **Preliminary Research Findings**

The results suggest that there is heterogeneity among age groups even after controlling for various other effects; however, the degree of heterogeneity varies among the choice dimensions considered and the age-specific dummy variables are selectively significant in different choice alternatives. For example, the 65-74 age variable contributes negatively to work/school action space while the 85+ year age variable contributes negatively to the social/recreational/health action space. With respect to activity time allocation, age variables are selectively significant; for example, the variable representing 75-84 years has a negative influence on activity time allocation to work/school, while the variable representing 65-74 years has a positive influence on activity time allocation to shopping/eating. The results suggest that efforts aimed at enhancing density and diversity of opportunities and amenities in space would help the elderly continue to engage in activities even as their action spaces shrink. Providing convenient mobility options (virtually on par with the private automobile) would enable seniors to travel even when their physical and cognitive capabilities diminish and they are no longer able to drive or use transit on their own.

It appears that activity-travel choices are determined to a significant degree by the physical and cognitive abilities of the individual; as these abilities generally diminish with age, the age variables serve as surrogates to represent the diminished capacity of individuals to undertake activities and travel. As medical advances allow individuals to remain active into later years of life, age may become less of a factor in explaining activity-travel behavior in the future; rather the lifecycle stage and medical capacity of the individual will govern and drive activity-travel choices. Transport surveys should explicitly collect such information, and variables representing these dimensions should be explicitly incorporated in travel forecasting models. Transport policy interventions should be aimed at assisting and providing options to all who have disabilities and medical needs, regardless of age, so that mobility and equal opportunity is truly provided to all. Future research should aim to unravel the relative size effects of different explanatory factors; i.e., to what extent is the lower mobility among the elderly due to age, medical conditions and devices, lifecycle stage, and household structure? Such an exercise will help policy makers target interventions strategically and assess the extent to which lower mobility among the elderly is representative of social exclusion or merely an artifact of their lifecycle stage representing a voluntary choice.

### **REFERENCES**

- Marin-Lamellet, C, and J. Haustein. Managing the Safe Mobility of Older Road Users: How to Cope with Their Diversity? *Journal of Transport and Health*, 2(1), 2015, 22–31.
- Goulias, K. G., Blain, L., Kilgren, N., Michalowski, T., and Murakami, E. Catching the Next Big Wave: Do Observed Behavioral Dynamics of Baby Boomers Force Rethinking of Regional Travel Demand Models? *Transportation Research Record*, 2014(1), 2007, 67–75.

## **TRENDS IN TYPICAL VS. ACTUAL TRAVEL: AN ANALYSIS OF NHTS DATA OVER TIME**

*Stacy Bricka and Timothy Reuscher*

### **Background**

The NHTS series is a rich source of data regarding travel trends. Like most regional and state level travel surveys, the NHTS documents the “who, what, where, when and how” of daily household and personal travel for an assigned 24-hour period. The NHTS also includes a battery of questions regarding typical travel patterns. These typical questions obtain details about usual mode and commute time to work and regularity of teleworking. In addition, the 2017 NHTS included questions about walking and biking in the past 7 days, and use of public transit, rideshare and carshare in the past 30 days.

### **Methodology**

The information provided in response to these questions provides insights into who travels by these alternative modes and how often, details that are difficult to glean from a one-day diary.

### **Preliminary Research Findings**

Examining differences between the “typical” commute compare to the “actual” commute found that in 2001, the reported work mode matched the actual work mode for 97% of the cases. By mode, 98% of commutes reported “last week” by auto were also reported made by auto on the travel day and 92% of those who commuted by transit “last week” also reported commuting by transit on the travel day. The match rates for bike and walk commutes were lower: 76% for bike and 65% for walk respectively.

In 2009, the match rate for auto commutes was again 98% (“last week” as compared to actual travel day). The transit match rate decreased slightly to 90%, while the bike and walk match rates increased slightly to 88% and 70% respectively. For 2017, the match rates remained fairly consistent for auto and transit commutes (98% and 91% respectively), while those for walk and bike decreased over prior years (84% and 63% respectively). In general, the lower the match rate, the higher the possibility that the travel survey captured commuting by an occasional travel mode. The lower match rate for walk and bike travel could also indicate variations in weather from week to week. However, without a specific question asking for why there was a difference in reported vs. actual travel modes, it is not possible to draw firm conclusions regarding these differences or the trends over time.

A similar comparison was made for the reported vs. actual travel time to work. The 2017 NHTS had three questions where travel time to work was captured. These include a question about the time to get to work “last week”, a question about time to work if not congestion, and then the actual commute time on the travel day. The research team found that: the commute time “last week” generally tracked the actual commute trip times on the travel day; travel time for last week’s commute was higher than the actual commute for trips between 16 and 20 minutes in length and those between 30 and 50 minutes. Travel time for last week’s commute was lower than the actual commute for trips between 26 and 30 minutes. As expected, if there were no congestion, the commute trips would be shorter.

With respect to how well reports of “usually” working from home compare to actual working from home on the travel day, researchers found that: 12% of all workers indicated “yes” they usually work from home. On the actual travel day, 6% of workers who indicated they usually work from home reported at least one instance of doing so as compared to 0.4% of their non-home based worker counterparts. When considering weekday travel days only, the difference increases to 7.7% and 0.5%, respectively. This analysis also illustrated that working from home does not necessarily translate into no travel. Of workers who usually work from home, only 19% reported having no travel on the travel day. For those reporting no travel, 44% gave the reason for non-travel as “working from home (for pay).” Comparatively, 8% of workers who indicated they do not usually work from home, did not report travel on the travel day. Of these, only 3% gave the reason for non-travel as “working from home (for pay)”. The remaining 81% of workers who reported they usually work from home and reported travel on their travel day reported an average daily person trip rate of 3.8 trips. This trip rate matched that of the 92% of workers who reported they usually do not work from home.

A review of reported trip purposes for weekday travel shows that the main reasons for travel reported by those who work from home are for family/personal business (25%), social/recreation (22%) and shopping (18%). For this group, only 13% of trips are for travel to/from work. Their counterparts, who do not usually work from home, report 40% of their weekday trips to be to/from work, followed by 17% for family/personal business and 15% for social/recreation.

What about typical travel for non-auto modes? Respondents were asked about how often they walked or biked in the past 7 days, and how often they used transit, rideshare, or carshare in the past 30 days. The wording of these questions did not allow for a direct comparison to the actual travel reported on the travel day, but the responses provide insights into who travels by these modes and where (geographically) the travel is more or less likely to take place.

Across all mode usage questions, urban dwellers reported the highest average usage rates. Suburban dwellers reported the lowest weekly walk and bike trip rates, while those living in rural areas reported the lowest monthly transit and rideshare usage rates. Respondents who were ages 75+, retired, or in poor health reported the lowest usage rates across all four travel modes. Reports of highest mode usage varied by age, with those ages 18-34 reporting the highest rates of walking and rideshare usages, while those ages 0-13 reported the highest levels of bike trips and those ages 14-17 reported the highest transit usage.

With respect to status, workers reported the highest levels of rideshare usage, students reported the highest levels of walk and transit trips, and those in the “other” category (homemakers, those looking for work, etc.) reported the highest level of bike usage. Respondents in excellent health reported the highest usage rates across all four travel modes as compared to those with an opinion of their health being less than excellent. Characteristics of those with higher usage rates also tend to be associated with higher level of travel in general (particularly as compared to those ages 75+, retired, and in poor health).

Understanding daily travel behavior patterns and choices are an important underpinning for transportation planning and prioritization. Considerable efforts are expended in providing and promoting the use of alternative modes of transportation. By comparing typical to actual travel modes, the mode usage captured through a one-day travel diary data can be put into a broader perspective and provide important insights into the use of all travel modes, such as the insights presented here regarding daily travel by those who usually work from home. This analysis also

identifies the importance for improved wording in typical use questions if the purpose of those questions is to aid in more direct comparisons of typical vs. actual travel.

## **TRAVEL TRENDS IN TEXAS: A COMPARISON OF THE 2017 NHTS DATA TO PREVIOUS NATIONAL AND LOCAL TRAVEL SURVEY ANALYSIS RESULTS**

*Lisa Green and Michael Martin*

### **Background**

The TTI has an interagency contract with the Texas Department of Transportation (TxDOT) that includes two aspects of NHTS analysis. The first component involves completing a travel trends report using Texas NHTS data—including results obtained as part of the Texas Add-on surveys. TTI has previously completed a travel trends report using the 2001 and 2009 NHTS datasets, and the report follows the same format as the FHWA National Travel Trends report. In the coming months, TTI will produce an updated Texas travel trends report that incorporates the 2017 NHTS data. Where survey methodology continuity allows, analyses is performed to replicate trends developed for previous NHTS years (2001 and 2009). Changes in survey methodology that preclude further trend development are noted. This trends report will be especially valuable to Texas transportation planners and policy makers. It will help decision makers understand how Texans are traveling and better anticipate how to address their transportation concerns and travel needs. These insights will help justify the cost associated with paying for Texas Add-on data, and paint a historical picture of Texas travel trends across time.

The second aspect of the interagency contract related to the NHTS involves developing a methodology that will allow 2001, 2009 and 2017 NHTS data to be meaningfully compared to household travel survey data collected as part of the TxDOT Travel Survey Program (TSP) during comparable time periods. The TxDOT TSP is robust and includes data from household travel surveys performed for study areas across the state. Over the past decade-and-a-half, TTI has analyzed over 20 household travel surveys.

### **Methodology**

The trip rates developed from these analyses are provided to transportation modelers developing or updating models used in the planning process. Using this large repository of TxDOT TSP data (ranging from 2002 to 2017), TTI will be developing a method of how to meaningfully make comparisons to the NHTS.

### **Research Findings**

The process provides insights into how national and local survey results compare. It provides additional justification and shows value in continuing to perform household travel surveys as part of the TxDOT TSP. The process of developing a comparison methodology may also provide insights to TxDOT on potential changes to their household travel surveys that may improve results and/or allow for more direct comparisons with future NHTS datasets.

## **ONLINE SHOPPING BEHAVIOR OF US TRAVELERS: THE ROLES OF INTERNET USAGE AND LIFESTYLE**

*Huyen Le*

### **Background**

Shopping has been one of the most common purposes for personal travel, accounting for 20% trips in the US. In the past decades, shopping has been evolving into many different forms, including the emergence of online shopping. However, the issues surrounding online shopping are still underexplored. For example, how does online shopping affect other aspect of travel? Is it complementary for current trips, or a substitute for shopping trips to stores, or a realization of latent shopping needs that were not possible before (Mokhtarian, 2009; Suel & Polak, 2017)? To gain a better understanding of these issues, it is necessary to first identify the online shopping propensity and predict the proportion of online shoppers in the general population. A few studies on this topic found that online shoppers tended to be younger, highly educated, and have higher incomes (Lee, Sener, & Handy, 2015; Lee, Sener, Mokhtarian, & Handy, 2017; Suel & Polak, 2017). Other studies using attitudinal measures found that online shoppers were also more tech savvy and tended to use active travel modes (Lee et al., 2017). The effects of gender and the built environment on online shopping were mixed (Lee et al., 2015; Suel & Polak, 2017).

However, most previous studies were conducted at a small scale with non-probability samples within small geographical areas. To the best of our knowledge, the only study that employed a representative sample is Zhou & Wang (2014) which investigated the on-line and in-store shopping patterns using the NHTS 2009. However, the online shopping pattern might have changed in 2017 compared to that in 2009. Therefore, we aim to address these shortcomings by employing the full NHTS 2017 dataset to sketch a profile of online shoppers. The study looks to answer these following questions: (1) Can internet and technology predict the demand for online shopping; and (2) What are the other reliable predictors of the online shopping tendency?

### **Methodology**

We employed two machine learning techniques (random forest and gradient boosting) to identify the most important drivers of online shopping behavior. We split the dataset into a training set contained of 80% of the sample, and reserved the rest (20%) for testing (i.e., cross validation). Random forest and boosting are tree-based methods, that is, decision trees ( $n=5,000$ ) were repeatedly grown from subsamples of the training sets then averaged the probability (or took the majority vote) to get the final results. The process of growing trees involved stratifying or segmenting the predictor space into various regions, then using the mode of the training observations in its own region. The two methods have been applied widely in the transportation research, and are able to provide robust predictions for the variable of interest, which is online shopping frequency in our case.

We regrouped the outcome variable (DELIVER) into our groups: (1) Did not purchase online in the past month, (2) online shopped less than 4 times (less than once a week), (3) online shopped less than 8 times (less than twice a week), and (4) online shopped for more than 8 times (more than twice a week). This approach mitigated the effect of outliers that skewed the distribution of the online shopping frequency. Input variables to the models included internet usage, travel patterns and lifestyle (e.g., physically activeness, mode use, trips by modes, car and

bike share membership, vehicle ownership), socio-demographic and other variables. Cross-validations were performed to alleviate model over-fitting.

### **Preliminary Research Findings**

We found a number of factors associated with online shopping from the models. The use of ridesharing, PC, internet, smartphones, and tablets were among the strongest predictors of online purchasing. In other words, the tech savvy group was more likely to shop online.

Household income was the strongest predictor of this tendency in random forest model, whereas education was the strongest predictor in the gradient boosting model. Age also played a big role in predicting online shopping frequency. Other factors that had smaller effects were gender, life cycle (i.e., household composition), number of adults in the household, worker density, and population density around the home locations. The last two variables suggest the role of the built environment on predicting online purchasing. In short, part of the results aligns with findings from previous online shopping studies using smaller sample, suggesting that practitioners can identify a market segmentation of online shoppers in order to predict the online shopping patterns in travel demand models.

Cross validation results suggested that the random forest model had a poor predictive power, with only 6% of the test set were correctly predicted. The boosting method provided a better results, with 33% of the test set being correctly predicted. We also tested the original continuous DELIVER variable in our predictive models, however the results were unsatisfactory. These results imply future improvement using other prediction methods, as well as the needs for other variables that are currently absent in the NHTS dataset.

The results provided a preliminary step to predict online shopping patterns to be incorporated into travel demand models. More studies are needed to replicate the results as well as improve the predictive power of the model in order to identify the segments of travelers who tend to purchase more goods online. From that, modelers may be able to estimate potential delivery truck traffic, as well as enhance the connection between household travel demand models and freight demand models.

The link between online and in-store shopping cannot be well-illuminated using the current dataset. Further research is needed to identify whether online shopping substitutes, complements, or modifies in-store shopping travel and the associated change in VMT. Once this relationship is clear, modelers could estimate the actual demand for shopping travel based on online purchasing frequency. Practitioners may devise strategies to manage in-store shopping travel demand and modify parking requirement for commercial zones.

### **REFERENCES**

- Lee, R. J., Sener, I. N., & Handy, S. L. (2015). Picture of Online Shoppers. *Transportation Research Record: Journal of the Transportation Research Board*, 2496, 55–63. <https://doi.org/10.3141/2496-07>
- Lee, R. J., Sener, I. N., Mokhtarian, P. L., & Handy, S. L. (2017). Relationships between the online and in-store shopping frequency of Davis, California residents. *Transportation Research Part A: Policy and Practice*, 100, 40–52. <https://doi.org/10.1016/j.tra.2017.03.001>

- Liao, Z., & Cheung, M. T. (2001). Internet-based e-shopping and consumer attitudes: an empirical study. *Information & Management*, 38(5), 299–306.  
[https://doi.org/10.1016/S0378-7206\(00\)00072-0](https://doi.org/10.1016/S0378-7206(00)00072-0)
- Mokhtarian, P. (2009). If telecommunication is such a good substitute for travel, why does congestion continue to get worse? *Transportation Letters*, 1(1), 1–17.  
<https://doi.org/10.3328/TL.2009.01.01.1-17>
- Suel, E., & Polak, J. W. (2017). Incorporating online shopping into travel demand modelling: challenges, progress, and opportunities. *Transport Reviews*, 0(0), 1–26.  
<https://doi.org/10.1080/01441647.2017.1381864>
- Zhou, Y., & Wang, X. (Cara). (2014). Explore the relationship between online shopping and shopping trips: An analysis with the 2009 NHTS data. *Transportation Research Part A: Policy and Practice*, 70.

## **ASSESSING EQUITY AND ACCESS TO LONG-DISTANCE TRAVEL USING THE 2017 NHTS DATA**

*Lisa Aultman-Hall, Jonathan Dowds, Greg Thivierge, and Anuarbek Onayev*

### **Background**

Long-distance trips are a growing component of travel. The last national travel survey to collect long distance travel was the 2001 NHTS. In this deployment, long distance trips were those longer than 50 miles, one-way. The 2009 NHTS did not include an explicit long distance module, but using the person miles for daily travel with FHWA total vehicle miles, average vehicle occupancy, and the Federal Aviation Administration's domestic air miles, long distance trips accounted for 40% of passenger miles in 2009. Despite the magnitude of long distance travel in the United States, the last one-year dedicated long distance travel survey was conducted over 20 years ago, in 1995, in the ATS deployment.

Access to long distance travel has profound economic and social implications for individuals. Not only does long distance travel provide access to a broader set of opportunities for education, employment, healthcare, and cultural experiences, but long distance travel is increasingly necessary to maintain relationships with family and friends as social networks expand geographically. Consequently, inequitable access to long distance travel may reinforce a range of equity issues.

### **Methodology**

Because of its very large sample size, inclusion of geocoded origin and destination information, and its documentation of participants who were away from home on their assigned travel day, the 2017 NHTS offers the potential to study important aspects of long distance travel even though a distinct long distance module was not undertaken. The research questions addressed in this effort included: 1) Does participation in long distance travel vary by socioeconomic group suggesting potential access or equity issues? 2) Can methods to estimate the air access between long distance trip ends be developed to facilitate mode choice models? Three aspects of long distance travel are considered here, two descriptive and one methodological.



**Research Findings**

There is significant long distance travel (with significant variation between individuals). While in general, there is little difference in long distance travel by age, race and sex, long distance travel increases with income and education (with exceptions in California add-on recall data). There are more personal, than business, long distance travel. Choice of distance threshold is important for analysis. Next steps include the use of multivariate modeling. In addition, methods to estimate attributes of the unchosen air alternatives are viable, and will be needed for long distance mode choice models. Surface modes dominate < 500 miles as expected, but some still drive longer distances. Those with higher incomes appear more likely than those with lower incomes to fly, including trips under 500 miles.

## **Chapter 12: Plenary Session**

**KRISHNAN VISWANATHAN**

*Cambridge Systematics, presiding*

**LISA AULTMAN-HALL**

*University of Vermont, recording*

**BRIAN HARRIS-KOJETIN**

*Board Director, Committee on National Statistics  
National Academies of Sciences, Engineering and Medicine*

Large-scale probability sample surveys have been a primary method for collecting data for federal statistics for decades, including for the transportation community. However, continued reliance on sample surveys as a principal means of collecting national statistical data faces challenges of increasing difficulties with participation and cost. These surveys with consequent threats to data quality, are challenged by the increasing demand for more geographically detailed and timely information. Keynote Speaker, Brian Harris-Kojetin provided an overview of recent studies by the Committee on National Statistics, exploring a possible shift in federal statistical programs. Our current approach of providing users with the output from a single Census, survey, or administrative records source, to a new paradigm of combining data sources with state-of-the-art methods will provide users with richer and more reliable datasets, leading to new insights about policy and socioeconomic behavior.

### **NATIONAL STATISTICAL PROGRAMS: DIRECTIONS AND CHALLENGES**

*Brian Harris-Kojetin*

The Committee on National Statistics (CNSTAT) was established in 1972, as a standing unit of the National Academies of Sciences on the recommendation of the President's Commission on Federal Statistics. The purpose of the Committee is to provide an independent, objective resource for the evaluation and improvement of federal statistical methods and operations. There are approximately 50 standing units in CNSTAT in the Academies. In addition to improving statistical methods, it serves as a coordinating force in the highly decentralized U. S. federal statistical system. During its 46 years, CNSTAT has produced over 270 consensus, interim, letter, and workshop reports, including some in conjunction with TRB. Its flagship publication is *Principles and Practices for a Federal Statistical Agency*. Twice a year, in October and May, CNSTAT holds a public seminar on a topic of broad interest to the federal statistical and research communities, including holding a lunch for heads of major statistical agencies. CNSTAT has a broad portfolio that spans issues with the Decennial Census Coverage and Quality and the ACS to Statistical Methods and Estimates for Policy Use.

As with other surveying efforts, the Federal statistical agencies face threats from declining participation rates and increasing costs. Demands are being made for more timely data, with greater geographical detailed information. At the same time, increasingly alternative data

sources are available that can offer solutions for these shortcomings. Costs for surveys have increased due to response rates and other factors at a time when agency budgets are limited due to many forces. In a non-Census year, it is about 0.15% of the federal budgets. In these fiscally challenging times, budgets have affected survey execution. Priority issues for these new sources of data include:

- How can agencies use administrative records and private-sector data sources to enhance federal statistics?
- What privacy issues need to be addressed when combining data sources?
- What quality frameworks and metrics are appropriate for these new sources and for blended estimates?
- How can agencies collaborate to access, evaluate, and use these new data sources?

A new ad hoc panel of nationally renowned experts has been assembled to take on the task of exploring the opportunity to foster a new paradigm in federal statistical programs. Rather than providing users with a single course of data (e.g., census, survey, or administrative records), a new paradigm, blending public and private data sources, is expected to provide users with richer and more reliable statistics.

The panel's first report is *Innovations in Federal Statistics* (available at <https://www.nap.edu/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy>). The panel recommendations included: the creation of collaborative research programs to address the challenges associated with blended data; consideration of private-sector data sources and annual reports on progress; the establishment of a new entity or an existing entity to facilitate secure access to data; and the encouraging maximum use of such data while protecting privacy with modern techniques.

Their second report, *Federal Statistics, Multiple Data Sources; and Privacy Protection*, (available at <http://mitsloan.mit.edu/shared/ods/documents/?DocumentID=4439>), includes a number of recommendations for next steps for combining data sources; data processing issues; appropriate personnel staffing; privacy issues; and a broader framework for assessing data quality that includes both public and private data sources. Key to the success of such a new entity would be: the legal authority to access data; strong authority to protect the privacy of the data; the authority to permit appropriate uses of the data; and qualified staff with sufficient skills to meet the needs of modern processing. It is important to avoid black boxes and make it public how data are processed and to provide access to the data. Quality and timeliness are both very important, in addition to accuracy, relevancy, and understandability. To implement the vision of a new entity capable of providing blended data. This effort will need a carefully structured strategic plan, in phases, with sufficient detail to ensure success.

The transportation data community's sense that data, including survey data, are critical to policy and decision-making, recognizing that new data sources offer opportunity as well as many challenges, we have no choice but to move forward and advance the science. We must move beyond traditional data sources and bring our detailed careful expertise to the issues of accuracy and appropriate use of new data. All data programs and all surveys are facing the same growing pains and challenges that we face in the travel behavior realm.

## Chapter 13: Modelers Panel—Using the NHTS 2017

**JENNIFER MURRAY**

*Wisconsin DOT, presiding*

**NANCY MCGUCKIN**

*Travel Behavior Analyst, recording*

**HABTE KASSA**

*Georgia DOT*

**PATRICK COLEMAN**

*AECOM*

**ARASH MIRZAEI**

*NCTCOG*

**JOHN MILLER**

**PETER OHLMS**

**ZULQARNAIN KHATTAK**

*Virginia DOT*

Modelers discussed the NHTS data for travel demand modeling. The discussion highlighted the use of the NHTS for trip rates, travel mode choices, and thoughts on improving NHTS data for travel demand modeling purposes. An innovative pedestrian trip calibration method was also presented.

### **PANEL**

*Habte Kassa, Patrick Coleman, and Arash Mirzaei*

Panel members represented transportation modelers from Georgia Department of Transportation (GDOT), Virginia Department of Transportation (VDOT), NCTCOG and a private consulting firm. GDOT uses their NHTS Add-on data for smaller MPO models and their statewide model. Atlanta Regional Council models the Atlanta region with their own data. VDOT uses the NHTS to update elements of their forecasting models. NCTCOG participated as an Add-on with TxDOT in both the 2009 and the 2017 NHTS.

Panel members agreed that the NHTS is an important source for both default data for areas that don't have local estimates. It is also a source of understanding for the context of travel behavior, market segmentation, and the nature of changes in travel behavior. There is need for a greater understanding of the changes in the 2017 NHTS deployment and previous NHTS efforts. One emerging issue for the modeling community is the presence of TNCs. The panel expressed satisfaction with the inclusion of questions on the use of TNCs in the 2017 NHTS. However,

other new modes (e.g., scooters) are such a very small proportion of trips that their impacts on modeling are insignificant.

According to the private sector panel members, mode choice is still an important matter, but no large urban areas use the NHTS for mode choice models. If there is a mode choice model (many smaller MPOs do not have mode choice models), it is calibrated through on-board transit survey data and transit counts from transit agencies in the region. It should be noted that the NHTS can be used to check assumption on mode choice. Panel members pointed out more than once that household travel surveys are just one of the many data sources used for model inputs. Other sources include additional surveys, the Journey-to-Work data from the ACS. With regards to declining trip rates, GDOT and VDOT were not as concerned about the lower trip rates as NCTCOG. The “typical” household trip rate has been approximately 10 trips per households for many decades. The NCTCOG Add-on data from the NHTS estimates about 9 trips per household, a 10% drop. The HBW trips and Non-home Based (NHB) are the same as previous estimates, but the Home-based Non-Work has declined, indicating some trip purposes have declined. Other measures, such as congestion, VMT counts are not showing declines, so more research is needed to understand these differences. Models can’t be updated with lower trip rates until there is a better understanding of this discrepancy.

### **Audience Dialogue Highlights**

There was a question about merging the detail of the American Community Survey Journey-to-Work (ACS JTW) with the NHTS. The panel agreed that this is an important research area given the ACS reports work trips only for a very large sample, while NHTS has this detail on all travel for a very small sample. Another issue to be addressed is whether the inclusion of toll roads in current models as a mode choice or a route choice. Further research is needed on transit, trip chaining and other related topics that will help explain the difference in the 2017 data from previous surveying efforts.

## **QUICK WAYS TO USE NHTS DATA TO SUPPORT REGIONAL TRAVEL DEMAND MODEL CALIBRATION**

*John Miller, Peter Ohlms, and Zulqarnain Khattak*

### **Background**

This research examines the extent to which the NHTS data can support better estimation of pedestrian trips in order to rapidly update a region’s urban travel demand model and then, using as a case study the regional model from Charlottesville, Virginia, how this update potentially affects investment decisions. In the Charlottesville regional model, mode split is based on utility functions which incorporate variables such as wait time and fare (for local bus), driving time (for auto), and of interest to this paper, a “pedestrian environment” variable (for walking) (1). The researchers used NHTS data to quickly update the model’s mode split utilities based on four steps.

## Methodology

First, at the zone level, the researchers ascertained whether density is correlated with pedestrian environment factors as well as the proportion of pedestrian trips in the base year of the model—an analysis that proceeds independently of the NHTS data. Second, based on these findings, they used the updated NHTS data to relate density to the proportion of pedestrian trips for each zone, ensuring that the NHTS densities replicated those found in the Charlottesville area. Third, to retain the structure of the travel demand model, the target shares (in the regional model) were updated based on the NHTS results, which leads to an adjustment of the mode-specific constant in the appropriate utility expressions. Fourth, we compare forecast trips (from the model) with observed trips (from ground counts). Finally, we demonstrate the impact that the updated model may have on a subset of pedestrian-oriented investment decisions.

## Preliminary Research Findings

The first step in the research tested the density to forecast proportions of pedestrian trips. In the Charlottesville regional model, the utility functions for walking trips include a “pedestrian environment” variable based on four factors: sidewalk availability, ease of street crossing, nonmotorized connectivity, and presence of minimum building setbacks ( $I$ ). For the base 2010 model, these four factors are scaled from 0 to 3 for each zone. For instance, a zone is given a sidewalk availability score of 0 if it has no sidewalks, 1 if less than 10% of streets have sidewalks, 2 if between 10% and 90% of streets have sidewalks, or 3 if more than 90% of streets have sidewalks. Similar percentage thresholds are used for the other three factors. For instance, for minimum building setbacks, a score of 1 is obtained if some buildings but fewer than 10% have minimum setbacks ( $I$ ). These four factors are then summed to yield a numerical pedestrian environment variable that can range from 0 to 12. For example, we can contrast two zones: zone 109 (with a relatively low pedestrian environment score of 4) and zone 66 (with a relatively high pedestrian environment score of 12).

After removing zones with zero population, the correlation between population density (people/mile<sup>2</sup>) and the pedestrian environment (which the model defines as the sum of the aforementioned four factors) remained low at 0.30. However, there was a much stronger correlation (0.74) between population density and the proportion of trips that are pedestrian. (This proportion of trips that are pedestrian refers to all trip purposes, which include HBW, home-based other, non-home-based, trips by students living on campus, and trips by students living off campus; it was obtained from the production/attraction matrices in the model). To be clear, density only explains about 55% of the variation in the percentage of trips made by pedestrians, so there are other factors that influence mode and trip choice decisions. That said, the coefficient for density in Equation (1) is highly significant ( $p < 0.01$ ) and the intercept is marginally significant ( $p = 0.09$ ). These results suggest that based on the original model, density has the potential to forecast the proportion of pedestrian trips, which leads to an interest in using the NHTS data for this purpose.

$$\text{Percent of Trips that are Pedestrian} = 0.00001227 * (\text{Population Density}) + 0.00736896 \quad (\text{Eq. 1})$$

Data from the 2017 NHTS (2) were used to relate pedestrian trips to density for each zone in the regional model in the second step of the research. The 2017 population densities in

the Charlottesville area were estimated from the model via interpolation between the 2010 observed and 2025 forecast density using Equation 2. (The results suggest NHTS data are appropriate for this purpose: population densities in the range of 0 to 29,460 people/mile are observed in Charlottesville area, which correspond with the NHTS categories of 50, 300, 750, 1,500, 3,000, 7,000, 17,000, and 30,000 people/mile<sup>2</sup>.)

$$2017 \text{ Density} = 2010 \text{ Density} + ((2017-2010)/(2025-2010)) * (2025 - 2010 \text{ Density}) \text{ (Eq. 2)}$$

The analysis was stratified by purpose (HBW, home-based other, and non-home-based), and, for the first two purposes, whether a vehicle was available. For all purposes combined, the proportion of pedestrian trips is nonlinear, with sharp changes around 6,000 and 17,000 people per square mile. However, with the limited number of density categories in the NHTS sample, care must be taken to avoid overfitting models to the data. For instance, Table 13.1 shows the development of two separate equations for the proportion of pedestrian trips as a function of density; the coefficients for density appear reasonable (e.g., the elasticity of pedestrian trips to density increases at higher densities.) However, this coefficient is not significant ( $p = 0.18$ ). Another concern for nonlinear models is that when the data are further stratified by trip purpose, the sample for each density value becomes relatively small; for instance, for HBW trips where a car is available, there are only 158 walk trips at the highest density category of 30,000 people/mile<sup>2</sup>.

**TABLE 13.1 Pedestrian Trips as a Function of Density Stratified by Low-High Density.**

Household Type	Equation for Calculating Proportion of Pedestrian Trips	<i>p-value</i>		R <sup>2</sup>
		density	constant	
Low Density (<7,000 people/mile <sup>2</sup> )	0.0604 + 0.000009*(density)	<0.001	<0.01	0.93
Medium-High Density (≥7,000 people/mile <sup>2</sup> )	-0.058 + 0.000018*(density)	0.18	0.69	0.82

Accordingly, the linear expressions relating density to pedestrian trip with regards to trip purpose were developed as shown in Table 13.2. For most purposes, density explains some of the variation in the proportion of pedestrian trips: higher density is correlated with higher proportions of trips being pedestrian. For the case of HBW trips where no vehicle is available, however, density is not significant ( $p = 0.62$ ), with the equation only explaining about 4% of the variation in the data. For the remaining purposes, however, the coefficient of density is highly significant at the 99% confidence level (as is the intercept), with the equations explaining 90% of the variation in the data. Thus, population density appears to be a decent predictor of the proportion of pedestrian trips for four of the five purposes examined in Table 13.2.

**TABLE 13.2 Percent of Trips that are Pedestrian as a Function of Density.**

Household Type	Trip purpose	Equation for Calculating Proportion of Pedestrian Trips	<i>p-value</i>		R <sup>2</sup>
			density	constant	
With a vehicle	HBW	0.01435 + 0.000003*(density)	<0.001	0.011	0.92
	Home-based other	0.09030 + 0.000006*(density)	<0.001	<0.001	0.96
Without a vehicle	HBW	0.2056 + 0.000002*(density)	0.624	0.002	0.042
	Home-based other	0.3144 + 0.000007*(density)	<0.001	<0.001	0.90
All	NHB	0.0398 + 0.000011*(density)	<0.001	<0.001	0.94

In the third step, the regional model as modified by updating the proportion of pedestrian trips by zone. Table 13.3 shows the expected 2025 pedestrian shares for the Charlottesville region for the trip purposes of HBW (with and without a vehicle), home-based other (with and without a vehicle), and non-home-based. These percentages were calculated by applying the equations in Table 13.2 to each zone (to obtain a percentage of pedestrian trips by zone) and then weighting each zone by its population to get an average percentage. For comparison purposes, the original target shares (based on the 2025 model) are provided.

**TABLE 13.3 Expected Percent of Trips that are Pedestrian.**

Household Type	Trip purpose	2025 (original)	2025 (revised)
With a vehicle	HBW	1.83%	3.08%
	Home-based other	3.27%	12.31%
Without a vehicle	HBW	3.66%	no change <sup>b</sup>
	Home-based other	6.55%	39.80%
All	Non-home-based	2.66%	10.00%

<sup>a</sup> For the Charlottesville model, these exclude internal-external and external-external trips which are drive alone.

<sup>b</sup> Because of the low coefficient of variation in Table 13.2, this mode share was not updated.

The Charlottesville model may then be recalibrated for year 2025 with these new mode shares as targets in the model. This entails performing a procedure where, in the model, a mode-specific constant in the utility function is updated. For example, for home-to-work (and work-to-home) trips where a vehicle is available, the target proportion of pedestrian trips is changed from 1.83% to 3.076% and the calibration procedure is executed. As a result, in the mode specific constant for pedestrian trips being adjusted from -1.7 (in the original mode choice utility function) to -1.595 such that individuals are more likely to take walking trips. Execution of the model’s “self-calibration” feature altered mode shares to a lesser extent than shown in Table 13.3. For instance, the model documentation (*I*) notes that trip-generation rates overall were increased to align observed and forecast vehicle volumes. The authors thus chose not to override such constraints when recalibrating the model. Overall, after execution of the equations in Table 13.3, the number of pedestrian trips increased by about 4.98% compared to what would have resulted without recalibrating the model.

The fourth step involved comparing forecast trips (from the model) with observed trips using ground counts. Because a modeled pedestrian trip is a line whereas a counted pedestrian



trip is a point, there is an inherent mismatch between modeled and observed pedestrian trips. The model forecasts nonmotorized trips between *zones*: e.g., the 2010 Charlottesville model forecast the number of daily trips between zones 59 and 60 as roughly  $7.06 * 2 = 14.12$ . The model does not place pedestrian trips on the network, and may not even have smaller facilities (where pedestrians like to walk) on the network. By contrast, a pedestrian counter is a specific point. Thus there will not be an exact linkage between the volume observed at a pedestrian counter (a given volume at a specific location) and the volume forecast by the model (a matrix of trips between zone interchanges). However, performing a GIS-based analysis allows one to loosely compare the modeled pedestrian trips to the observed pedestrian trips. (This method entails four sequential operations: determine the centroid of each zone, create a “near table” that lists, on each row, every possible combination of two centroids, create a desire line connecting each pair of centroids, and then overlay these desire lines with the zone containing the pedestrian counters.)

The GIS-based analysis has two simplifications: first, the desire lines are based on the shortest aerial distance between two centroids (a simplification whose veracity depends on the street network’s connectivity); second, the desire lines presume that all travel passing through the zone where the counter is housed will also pass through the counter. With this method, it is possible to compare modeled trips that pass through zones 59 or 92 with the observed trips at one specific location, which in this case is on Emmet Street just north of Ivy Road, where pedestrian counters are located in zones 92 and 59 on opposite sides of the street. The results suggest that for year 2017, for the Emmet Street count locations in zones 59 and 92, there were a total of 4,336 such trips, or, if one eliminates any modeled trips longer than two miles (a rough 95<sup>th</sup> percentile based on the NHTS data), there are a total of 4,259 modeled trips. This differs from the observed number of trips, which was either 895 weekday trips or 1,064 weekend trips (3). Thus, the counter only captures a portion of all the trips that would pass through these two zones. Based on the two aforementioned simplifications, an area for future research might be to determine if, for a given study area, there is some consistency between the ratio of observed trips at the counter to the forecast trips from the model. (For example, the data point herein suggests a ratio of about 25%.)

This research addresses the question: *to what extent may better pedestrian forecasts alter investment decisions?* Although recalibration of the model based on NHTS data increased the total number of 2025 pedestrian trips by almost 5%, this increase was not uniform across the study area. As a way of representing the geographic variation in how forecast pedestrian trips are altered, the net change in the total number of pedestrian trip ends—e.g., origins plus destinations—in each zone. (Some zones saw a sizable decrease in the number of pedestrian trip ends, and as these are near the University of Virginia, this may be a sign that the calibration procedure can be revised in future efforts to account for special generators.)

This research can be used to prioritizing neighborhoods for small-area planning. Communities create small-area plans to identify specific opportunities such as needed sidewalk and crosswalk improvements. In the Charlottesville area, Albemarle County recently completed small-area planning efforts at two urban nodes north of the City of Charlottesville (4,5) and, as of June 2018, was revising a plan for an urban node to the east of the City (6), among other efforts. A plan for Crozet (7), a developing area west of Charlottesville, was completed in 2004 and updated in 2010. These results suggested that the Crozet area may experience particularly large increases in forecast pedestrian volumes compared to previous forecasts, lending support to local arguments in favor of prioritizing the Crozet area for the county’s next small-area plan update.

Another application is the evaluating projects for funding allocations. SMART SCALE, VDOT's process for prioritizing transportation projects for funding, includes a congestion metric that considers total person throughput for existing and new users (8). Thus, differences in forecasts for pedestrian volumes could affect total person throughput and a project's score for this metric. By way of example, the City of Charlottesville prepared four SMART SCALE project applications in 2018. Two of these were in locations where our forecasts indicated that pedestrian volumes would be substantially higher than previously forecast; the other two were near zones with a mix of smaller increases, decreases, and minimal changes. Thus, if these forecasts were used to estimate total pedestrian throughput for each of these scenarios, the former two projects could receive additional points relative to the latter two. While pedestrian volumes may not be actually included in this metric for most projects, and the score for this metric is a portion of the score for the congestion measure, which is one of several measures that constitute the overall project score. In contrast to the results presented in Table 13.2, veteran planners typically wish to see mode choice forecast as a function of more, rather than fewer, variables. Further, the results of step 4 show the limitations of using ground counts to assess the validity of model forecasts. As regional models are infrequently updated (e.g., a sample of Virginia models showed an average age of about eight years [9]), techniques to quickly update such models based on new data, without altering the model structure, can be an effective way for transportation agencies to make use of limited resources. For this particular case study, two lessons appear transferrable to other areas. Density is generally a significant predictor of the proportion of trips made by pedestrians ( $p < 0.01$ ) for most purposes examined. This finding suggests the NHTS data (2) has promise for incorporation into other regional models. Recalibration of the Charlottesville regional model, which resulted in higher pedestrian volumes in some locations, has the potential to affect investment decisions that target the pedestrian mode. This suggests that in regions with older models, the effort of using NHTS data may be of value depending on the types of decisions for which model results are used.

## REFERENCES

1. The Corradino Group. *Technical Methodology Report: Charlottesville Model Base 2007 Version 1.0*, for the Virginia Department of Transportation, Richmond, 2009.
2. FHWA. 2017 National Household Travel Survey, Oak Ridge, Tennessee, 2018. <https://nhts.ornl.gov/>. Accessed May-June 2018.
3. City of Charlottesville. Bicycle & Pedestrian Counts: Emmet Street. Thursday, September 22, 2016 to Monday, January 09, 2017. 2018. <http://www.charlottesville.org/departments-and-services/departments-h-z/neighborhood-development-services/transportation/bicycle-and-pedestrian/bike-ped-data>. Accessed May 22, 2018.
4. Albemarle County Department of Community Development. Rio+29 Small Area Plan, Charlottesville, 2018. <http://www.albemarle.org/department.asp?department=cdd&relpage=20580>.
5. Virginia Department of Transportation. Hydraulic Planning Advisory Panel Documents, Richmond, 2018. [http://www.route29solutions.org/panels\\_meetings/hydraulic\\_planning\\_advisory\\_panel\\_documents.asp](http://www.route29solutions.org/panels_meetings/hydraulic_planning_advisory_panel_documents.asp). Accessed June 8, 2018.
6. Albemarle County Department of Community Development. Pantops Master Plan. Charlottesville, 2018 <http://www.albemarle.org/department.asp?department=cdd&relpage=3734>. Accessed June 8, 2018.

7. Albemarle County Department of Community Development. Crozet Master Plan Home. Charlottesville, 2014. <http://www.albemarle.org/departement.asp?departement=cdd&repage=3733>. Accessed June 8, 2018.
8. Commonwealth Transportation Board. *SMART SCALE Technical Guide*. 2018. [http://vasmartscale.org/documents/2018documents/ss\\_technical\\_guide\\_nov13\\_2017\\_revised\\_feb2018\\_for\\_posting.pdf](http://vasmartscale.org/documents/2018documents/ss_technical_guide_nov13_2017_revised_feb2018_for_posting.pdf). Accessed June 8, 2018.
9. Virginia Department of Transportation. Virginia Transportation Modeling Program. Richmond, 2017. <http://www.virginiadot.org/projects/vtm/vtm.asp>. Accessed February 12, 2018.

## **Chapter 14: NextGen NHTS**

**WENJING PU**

*FHWA, presiding*

**STACY BRICK**

*MacroSys LLC, recording*

**DANIEL JENKINS**

*FHWA*

**SEPEHR GHADER**

**LEI ZHANG**

*University of Maryland*

**VINCE BERNARDIN**

**HADI SADRSADAT**

*RSG*

**ZACHARY HANSON**

*Virginia DOT*

The FHWA is establishing the NextGen NHTS program through a pooled fund effort (<http://pooledfund.org/Details/Solicitation/1466>). This new program will include both travel surveys and passive data to provide annual multimodal data products. It also offers opportunities for State DOTs, MPOs and other entities to join the program as Add-On partners to gain insights and purchase additional data products.

### **INTRODUCTION OF THE NEXTGEN NHTS**

*Daniel Jenkins and Wenjing Pu*

The NHTS has been collected for the past 50 years (as the NPTS, 1969 through 1995, followed by the National Household Travel Survey (NHJTS) from 2001 through 2017. These deployments overcame signification challenges including budget constraints and lack of timeliness. Emerging opportunities now exist for the use of passive data. FHWA has undertaken Next Gen NHTS with two components: the survey core data (National and Add-On samples) and passive (origin and destination) data (National and Add-On).

The objective of the new approach is to obtain the national survey core travel behavior data, with FHWA responsible for national survey data on an annual basis from 2019 to 2023. Add-on partners will create a sub-national level core dataset from 2019 to 2023. The National Survey Core Data will use a probability-based sampling approach, with independent surveys in each year. The data collection conducted on an annual basis will have fewer samples than previous NHTS deployments. In addition, it will have fewer, but more targeted questions,

including Long Distance travel questions. The data products expected from this effort will be similar to the existing NHTS, with both national summaries and disaggregated records.

The national travel behavior summaries will include: average trip rate and trip length (daily person/vehicle trip rate; average person/vehicle trip length; by purpose, gender, income, age, time of day, urban or rural); model share (by mode and by purpose); along with other summaries of available data. In addition to data collected directly from the respondent, the dataset will contain derived variables. The Add-On program will allow states and MPOs to purchase additional surveys, as in the past. The validation process will include the use of other USDOT data including the Airline Origin and Destination Survey (DB1B), the Air Carrier Statistics database (T-100) and the National Transit Database (NTD). In addition, special attention will be paid to rural travel.

The passive component of the Next Gen NHTS Program will be developed using a national, regional, state, and local travel origin/destination (OD) data. FHWA will be responsible for the National OD data on an annual basis from 2019 to 2023. Add-on partners will create a sub-national level OD dataset for specific year(s) between 2019 and 2023.

The geographic resolution for the National OD will be state-specific MSAs and the remainder of the state, in 1 or more zones. FHWA will provide the zonal system. The Add-on partners OD data will be specified by the Add-on agency, with options to include TAZ, TAD, County, Block Group, Census Tract, etc. The cost for the data could potentially be based on population and/or area covered. The temporal resolution for the National OD dataset will be the entire calendar year, with each month of the year, and broken out by weekday and weekend. The Add-on partners OD data will also be for the entire calendar year, with each month of the calendar year (or other seasonal specifications). In addition, the OD data will be available for weekdays and weekends, or for each day of the week. The time periods of the day will include AM peak, midday, PM peak, and night).

The travel modes to be included will be air, rail, vehicle (including buses and passenger vehicles), and other modes (e.g., bike, pedestrian, etc.). Trip purposes will be work and non-work. Trip distances will be binned and assigned an alphabet (e.g., A:  $\leq$  10 miles, B:  $>$  10 miles but  $\leq$  25, up to M:  $\leq$  850 miles. Trip duration will be binned in a similar manner (e.g., A:  $\leq$  15 minutes, B:  $>$  15 minutes but  $\leq$  30 minutes, up to J:  $>$  10 hours). A number of datasets will be used for quality control including: NHTS; ACS; DB1B and T100; the NTD; HPMS; and others. The expected products include: data files; technical and supporting documents; and interactive analytics and visualizations.

The Add-on partners will have two levels for participation. First, they can join without a data purchase (\$5,000/year). They will be able to join the Working Group (annual in-person meeting), have access to the latest national data and information, and be able to gain centralized technical and peer support with data applications. The second option is the join with a data purchase, with a minimum of \$25,000/year. These agencies will be able to purchase additional survey data and/or passive OD data for any year(s) between 2019 and 2023. It is expected that they will have lower data costs due to economies of scale. They will be able to gain nationally consistent data so that data controls and comparisons can be efficiently enabled. It is expected they will have no or minimum contracting or oversight hassles.

The Add-on partners will be able to opt in/out at 1-year increments. They will be able to join with a data purchase for survey data 6+ months before data collection begins and anytime for OD data. Those who join without data can do so at anytime. Partners are encouraged to make a pledge using the pooled fund website or contacting FHWA directly.

## **INTEGRATION AND VALIDATION OF PASSIVELY COLLECTED DATA WITH NATIONAL HOUSEHOLD TRAVEL SURVEY**

*Sepehr Ghader and Lei Zhang*

Traditional surveys such as NHTS have been the most important source of data for studying travel behavior for many years. Travel surveys provide useful information on travel behavior necessary for estimating various travel demand models. However, they are usually very costly to conduct in large scale. They also suffer from small response rates. In the age of big data, transportation field is also starting to benefit more from all sorts of new datasets such as passively collected data. Passively collected data from mobile devices such as GPS devices and cellphones can complement traditional surveys to improve traditional surveys' shortcomings. The size of passively collected data is increasing exponentially, and collecting them is often less expensive than collecting traditional surveys. However they usually lack some useful information since many attributes such as mode, purpose, and socio-demographic characteristic of traveler is unobserved in the passively collected data. Even though researchers in the fields of transportation, machine learning, and computer science have developed algorithms to impute the non-observables in passively collected data; this type of data best shows its capabilities when integrated with a traditional travel survey. FHWA has recently announced their plan for the next generation of NHTS which has three components: Probability-based samples for the national survey; National OD tables obtained from passively collected data; and Partner add-ons.

This plan offers a great opportunity for the transportation field to study the best practices for integrating probability-based traditional surveys with OD data obtained from the passively collected data. Various companies are now capable to provide OD tables from passively collected data, and a user might be overwhelmed with all the available data out there. The source of the passively collected data can be GPS devices, cell-phone towers, location-based services, or combination of these. Temporal and geographical coverage of passively collected data varies by its source, therefore OD obtained from different sources might not be comparable. Standards and guidelines should be set for OD tables from different sources of passively collected data to make them comparable. OD tables should also be validated with some sort of ground truth before they can be used by the transportation community. Traditional surveys can be used as a ground truth for validating OD tables obtained from passively collected data. Once validated or possibly calibrated, these OD tables can produce results comparable with traditional surveys. They will also be more comparable if they are calibrated/validated against the same dataset such as NHTS. This research seeks to use the most recent household travel survey, NHTS 2017, to validate OD products with different components of NHTS using the following validation efforts:

- Trip generation validation at national level: The number of long-distance trips (Longer than 50 miles) generating from each state will be compared between NHTS and an OD product obtained from the combination of cell-phone and GPS passively collected data.
- Trip attraction validation at national level: The number of long-distance trips (Longer than 50 miles) attracted to each state will be compared between NHTS and an OD product obtained from the combination of cell-phone and GPS passively collected data.
- Trip length distribution at national level: trip length distribution for the long-distance trips (Longer than 50 miles) will be compared between NHTS and an OD product obtained from the combination of cell-phone and GPS passively collected data.

- Trip generation validation at metropolitan level: The number of trips generating from each county within Baltimore Metropolitan Council MPO area will be compared between NHTS and an OD product obtained from the combination of cell-phone, GPS, and LBS passively collected data.
- Trip attraction validation at MPO level: The number of trips attracted to each county within Baltimore Metropolitan Council MPO are will be compared between NHTS and an OD product obtained from the combination of cell-phone GPS, and LBS passively collected data.

## **CROSS-VALIDATION AND SIMULTANEOUS MODEL ESTIMATION FROM NHTS AND AGGREGATE PASSIVE OD DATA IN CHARLESTON, SC**

*Vince Bernardin and Hadi Sadrsadat*

### **Background**

The Charleston, South Carolina MPO is updating its regional travel model with information on travel patterns and behavior from both a 2017 NHTS Add-On Sample and aggregate passive origin-destination (OD) data from cellular signaling. These two data sources are being used together in two ways, for cross-validation of the two datasets, and for the estimation of destination choice models from both datasets. The research developed comparisons from the cross-validation of trip-rates, trip length frequency distribution, and highly aggregated district-to-district flows (to ensure adequate sample from the NHTS). The NHTS will be compared to both the raw (vendor expanded) and final expanded passive data, before and after expansion to traffic counts used as control data. The aggregate data fusion to identify and withhold truck trips from these comparisons is briefly discussed. The trip rate comparisons will look for residual correlation of the errors with demographic and socioeconomic variables to detect any biases not corrected for by the vendor's preliminary expansion. Trip rates were examined by district to detect any effects related to varying cellular coverage levels within the region.

The focus of this research, however, is the use of the NHTS and passive data together to estimate destination choice model parameters. Each data source provides valuable information for destination choice that the other cannot. The NHTS data provides rich traveler characteristics and trip purpose information, while the passive data provides adequate coverage of the OD solution space to estimate a robust set of constants. This investigation has the potential to provide insight on the value of having both types of data for destination choice modeling and the potential pitfalls of relying on either type of data alone. In total, five destination choice models are estimated using a genetic algorithm. The models vary by data sources (NHTS vs passive data vs combined), and whether or not they include constants in the model. These tests will examine the effect of the inclusion or exclusion of constants on other parameter estimates, particularly the disutility of travel time. The five models are as follows: Model 1 - The estimation is based on NHTS data alone; Model 2 -The estimation is based on passive OD data alone without constants in the model (same utility specification as in Model 1); Model 3 - The estimation is based on passive OD data with constants in the model; Model 4 - The estimation is based on combined NHTS and passive OD data without constants in the model; and Model 5 - The estimation is based on combined NHTS and passive OD with constants in the model.

The combination of NHTS and passive OD is weighted according to the ratio of the NHTS sampling rate to the passive OD data market penetration rate. The constants included in

two destination models are called “shadow prices”. The term shadow price is an additional component in the utility function of the demand models, which can be inferred from the actual observed travel patterns (the cell-phone data in this case). The shadow prices are defined in district-to-district level rather than zonal level for two reasons. First, as repeated findings in the research on cell-phone based OD data attest, this particular type of passive data can be accurately validated against other data sources using larger districts but are less reliable at the level of smaller zones. Second, a destination choice model with zone level shadow would be over-specified and could not therefore properly represent sensitivity to key variables.

Enriched models such as this destination choice model are capable of supporting the development of synthetic fused data sets, and this approach has been successfully applied to produce synthetic trip diary data based on both survey and passive data for Chattanooga, Tennessee. This approach to data fusion can produce disaggregated data while avoiding the privacy considerations of direct or true disaggregate data fusion and/or serve as a method for producing disaggregate data when only aggregate passive data is available. Although the Charleston example does not include the production of synthetic diary data, it helps further clarify the benefits of this type of data fusion.

## **NEXTGEN ADD-ON PERSPECTIVES**

*Zachary Hanson*

VDOT is an add-on agency for the NextGen NHTS pooled fund study. Previously, VDOT participated as an add-on agency for the 2009 NHTS, purchasing an additional 14,584 samples. In addition, VDOT supplemented the 2009 Add-on effort by partnering with the Weldon Cooper Center and four universities to better understand student travel behavior. For the Next Gen study, VDOT expects to support metropolitan and rural transportation planning efforts statewide; increase sample size at the local level for more reliable analysis; adapt add-on questions to meet agency needs; and compare Virginia results against NHTS data from other states, the nation, and past NHTS efforts. VDOT sees the potential uses for NextGen include: assisting future VTrans (Virginia’s long range, multi-modal policy plan) updates by supplementing regional or statewide travel surveys; Travel Demand Model development; Strategically Targeted Affordable Roadway Solutions (STARS) and the Arterial Preservation Programs; SMART SCALE; and to analyze regional and local multimodal travel trends. NextGen uses for VTrans updates will be useful for VTrans 2040, Virginia’s current statewide long-range multimodal policy plan. The Virginia Multimodal Transportation Plan Needs Assessment component identifies transportation and safety needs for corridors, introregional networks, and designated urban development areas. The purpose of the assessment is to screen projects that will then be considered for SMART SACLE/HB2 prioritization. As travel behavior changes, the Needs Assessment will need to be updated in future VTrans.

NextGen uses for Travel Demand Models include estimates, calibrates, and validations of regional and statewide travel demand models. There is a need to develop trip generation models, determine average trip lengths, trip length frequency distributions for trip distribution estimation, and aggregate mode shares. VDOT has data-driven programs that will benefit from NextGen. For example, the STARS. STARS identifies innovative and cost-effective multimodal measures aimed at improving safety and reducing congestion. SMART SCALR prioritizes transportation funding, with measures that include: safety, congestion, mitigation, accessibility, environmental



quality, economic development, and land use. NextGen NHTS is also expected to meet agency needs with the passive OD data at high geographic resolution, by time of day and multiple trip purposes: home based work, NHB, and home based non-work trip purposes. VDOT will have the flexibility to design their survey area and the time of their survey implementation. They also will have the ability to add several agency-specific survey questions related to value of time, preferences, and opinions. VDOT is looking forward to obtaining vehicle occupancy by jurisdiction and time of day with route attributes; percent of single and non-single occupancy vehicle information by jurisdiction. Of particular interest is the ability to have add-on questions that address mode choice, travel time, travel time reliability, distance to work, and travel to non-work destinations that support performance measurement, data-driven programs and the development and evaluation of SMART SCALE project applications.

## **Chapter 15: Closing Session**

**LISA AULTMAN-HALL**

*University of Vermont, presiding*

**LEI ZHANG**

*University of Maryland, recording*

**DAVID WINTER**

*FHWA*

**ROLF SCHMITT**

*BTS, USDOT*

### **GREAT THINGS DONE IN 2009 AND MOVING FORWARD WITH 2017**

*David Winter*

The NHTS is a critical data program for FHWA and USDOT and it is used for a wide variety of purposes, including to study major issues and to inform different policies and programs that are already underway. FHWA briefed the Office of the Secretary of Transportation (OST) on a number of important topics using the NHTS including: teen drivers; older drivers; retail trips; pedestrian travel. Another briefing is scheduled, illustrating continued interest in the NHTS results. Previous deployments of the NHTS have been used to inform other policies and programs, such as the National Travel Demand Assessment and the National Travel Condition Assessment. The NHTS has been used to better understand infrastructure needs, financial needs (e.g., tolling, high-occupancy vehicle and high-occupancy toll lanes), Safe Routes to School, Ladders of Opportunity, SAFETLU and the Fixing America's Surface Transportation Act. NHTS is at the core of programs important to transportation community and the country.

Another example of the importance of the NHTS is the safety data initiative underway in the OST. The purpose of the initiative is to drive down fatalities at both the national and local levels. Issues include determining who the people are who are being killed on the side of the road, their gender, age and purpose for traveling. The NHTS is the only data source available to answer these questions. These factors aren't reported in the Fatal Accident Reporting System or local police reports. It is hoped that using the NHTS will provide insights into the context of these accidents considering most pedestrian fatalities are not occurring at intersections. These fatalities are predominantly male between 51 and 55 years of age.

Funding for the NHTS has been a challenge for FHWA as it competes with all other research needs for funding within FHWA. FHWA is attempting to make NHTS funding a line item in the budget. Currently, FHWA has \$10M using Section 6028 funding, with NHTS just one of the many competing uses for these funds.

FHWA is looking forward to the NextGen program and will continue to strive to improve: the timeliness of the data; the dissemination of the data; techniques for gathering data; data quality; maintain of comprehensive measures of local and long distance travel; weekday and weekend travel data; and multimodal travel behavior. At the same time, there are major challenges ahead in the

NextGen approach that will use different sources and techniques, particularly those associated with Big Data. To meet these challenges, FHWA sees workshops like this one as the best way to bring together members of the transportation community to provide guidance and expertise. Clearly, a new approach is needed, and as mentioned, those at this workshop can be the champions and ambassadors for the NHTS as it belongs to the transportation community, not to FHWA or USDOT. We will all need to work together to make this our program.

### *Rolf Schmitt*

BTS tries to paint a statistical mosaic of all transportation and the consequences of transportation and NHTS is a major part of mosaic. The storyline so far: household travel is down while overall travel is up and household transit use is up while transit ridership is down. Excluding transit, the story sounds right: for example, we have anecdotal evidence of home delivery substitution that could explain drop in trips for shopping and errands. Are households a harbinger of the future and overall travel will decline? Or do we have a measurement problem? Are we in overlapping error between NHTS and VMT estimates? We have error estimates for NHTS but not for VMT or transit ridership.

NHTS doesn't cover freight, commercial use of personal vehicles, or long distance travel. We can assume that long distance passenger vehicle travel is up because travel on commercial aviation is up. A Vehicle Inventory and Use Survey would help measure the areas missed by the NHTS. NextGen NHTS or integrated an estimation system for passenger travel parallel to the Freight Analysis Framework might help cover the missing pieces of the storyline. The combination of cell phone and survey data for NextGen NHTS brings us back to opening keynote: Surveys vs other data sources:

- Need to continue surveys to get at phenomenon that cannot be observed; however, cost, timeliness, and availability of new sources push us toward Stats Canada shift to surveys last rather than surveys first
  - Administrative records, sensors, imagery, web scrapping, model outputs plus surveys, often integrated in estimation systems such as travel demand models
  - All data have error; survey has advantage of methods of measuring and adjusting for error; some concepts apply and others do not
    - Integrated or blended data: do errors compound or offset?
    - Historical note: National Travel Survey collected long distance and commuting data including transit access and parking costs in 1963; long distance travel measured in 1995 ATS and unsuccessfully combined with local travel in 2000
    - Data science is finding ways to compromise confidentiality and may eliminate the availability of public use microdata files from statistical agencies, increasing the importance of giving the research community better access to Federal Statistical Research Data Centers
    - Causality is central to evidence-based decisions and is a high statistical quality bar for the transportation community

We are in the biggest data revolution since 1930s when surveys became the foundation for our world. We need to be leaders rather than casualties of the revolution. Thank you for helping make the NHTS a foundation of our understanding of passenger travel, particularly thanks to Susan Lisa for her dedication to earlier generations of the NHTS. We look forward to your help in improving our understanding in the future.

## Chapter 16: Participants

Saad Alhussein, *King Saud University*  
Sulabh Aryal, *Richmond Regional Planning District Commission*  
Lisa Aultman-Hall, *University of Vermont*  
Charles Baber, *Baltimore Metropolitan Council*  
Apara Banerjee, *MacroSys*  
Vince Bernardin, *RSG*  
Cynthia Bland, *RTI International*  
Stacey Bricka, *MacroSys*  
Ralph Buehler, *Virginia Tech*  
Charlynn Burd, *US Census Bureau*  
Alex Cates, *Westat*  
Laura Chaney, *Oklahoma DOT*  
Yohan Chang, *Connectics Transportation*  
Daniel Chatman, *UC Berkeley*  
Shih-Miao Chin, *Oak Ridge National Laboratory*  
Soon Chung, *North Carolina DOT*  
Patrick Coleman, *AECOM*  
Andrew Collings, *Des Moines Area Metropolitan Planning Organization*  
Matthew Conway, *Arizona State University*  
Christopher Coy, *RSG*  
Aref Darzi, *University of Maryland*  
Michael Dennis, *NORC*  
Nimish Dharmadhikari, *INCOG*  
David Donaldson, *University at Maryland*  
Carlos Figueiredo, *AARP*  
Theresa Firestine, *US DOT*  
Lean Flake, *Resource Systems Group*  
Anthony Fucci, *Westat*  
Shari Gershenfeld, *Mobile Market Monitor*  
Sepehr Ghader, *University of Maryland*  
Arash Ghaffar, *University of Maryland*  
Jodi Godfrey, *Center for Urban Transportation Research at the University of South Florida*  
David Gohike, *Argonne National Laboratory*  
Charlie Han, *MacroSys*  
Zachary Hanson, *Virginia DOT*  
Brian Harris-Kojetin, *National Academy of Sciences*  
Joseph Hutchinson, *University of Texas at Austin*  
Ho-ling Hwang, *Oak Ridge National Laboratory*  
Daniel Jenkins, *US DOT*  
Kenneth Joh, *Metropolitan Washington Council of Governments*  
Naveen Juvva, *StreetLight Data*  
Habte Kassa, *GDOT*  
Soheila Kholi, *CALTRANS*  
Keith Killough, *Arizona DOT*  
Tae-Gyu Kim, *North Carolina DOT*  
Woon Kim, *AAA Foundation*  
Ulric Kwan, *Palo Alto, CA*  
Jeffrey LaMondia, *Auburn University*  
Catherine Lawson, *University at Albany*  
Huyen Le, *Virginia Tech*  
Mai Le, *TRB*  
Codie Leseman, *INRCOG*  
Chi Li, *US EPA, Office of Transportation & Air Quality*  
Susan Liss, *Consultant*  
Kelly Lynn, *RTI International*  
Gregory Macfarlane, *Transport Foundry*  
Janice Machado, *Westat*  
Jina Mahmoudi, *University of Maryland*  
Tom Mank *Ithac-Tompkins County Transportation Council*  
Theodore Mansfield, *ORISE/USDOT*  
Michael Martin, *Texas A&M Transportation Institute*  
Geena Maskey, *TxDOT*  
Shawn McCloskey, *Westat*  
Mancy McGuckin, *Travel Behavior Associates*  
Brian McKenzie, *US Census Bureau*  
Jasmy Methipara, *Federal Highway Administration*  
John Miller, *Virginia DOT*  
Arash Mirzaio, *North Central Texas Council of Governments*  
Patricia Mokhtarian, *Georgia Institute of Technology*  
Jennifer Murray, *Wisconsin DOT*  
Anuar Onayev, *University of Vermont*  
Thomas Palmerlee, *TRB*  
Alan Pisarski, *Alan Pisarski Consultancy*  
Steven Polzin, *Center for Transportation, University of South Florida*  
Wenjing, Pu, *Federal Highway Administration*  
Mark Radovic, *Maryland State Highway Administration*  
Clara Reschovsky, *US DOT/Bureau of Transportation Statistics*  
Timothy Reuscher, *Oak Ridge National Laboratory*  
Shelley Roth, *Westat*  
Guy Rousseau, *Atlanta Regional Commission*  
Hadi Sadrsadat, *RSG*  
Rolf Schmitt, *Bureau of Transportation Statistics/OST/USDOT*  
Kyle Schroeckenthaler, *EDR Group*  
Paul Schroeder, *EurekaFacts*  
Lisa Shemer, *MDOT- State Highway Administration*  
Chowdhury Siddiqui, *South Carolina DOT*  
Amy Smith, *Uber*  
Jeremy Sudweeks, *Virginia Tech Transportation Institute*  
Tianjia Tang, *Federal Highway Administration*  
Greg Thivierge, *University of Vermont*  
Miguel Vescovacci, *MTGG, Inc.*  
Krishnan Viswanathan, *Cambridge Systematics, Inc.*  
Chieh Wang, *Oak Ridge National Laboratory*  
Yudan Wang, *UNC HSRC*

Laurie Wargelin, *NORC at the University of Chicago*  
Penelope Weinberger, *AASHTO*  
Jeremy Wilhelm, *Westat*  
Rik Williams, *Uber*  
David Winter, *Federal Highway Administration*  
Kathleen Yu, *North Central Texas Council of Governments*

Heather Zaccaro, *Smart Growth America*  
Lei Zhang, *University of Maryland*  
Yan Zhou, *Argonne National Laboratory*  
Zhenpen Zou, *National Center for Smart Growth*

# *The National Academies of* SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, non-governmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at [www.national-academies.org](http://www.national-academies.org).

---

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to increase the benefits that transportation contributes to society by providing leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied committees, task forces, and panels annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

Learn more about the Transportation Research Board at [www.TRB.org](http://www.TRB.org).



TRANSPORTATION RESEARCH BOARD  
500 Fifth Street, NW  
Washington, DC 20001

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

[www.national-academies.org](http://www.national-academies.org)