TRANSPORTATION RESEARCH

CIRCULAR

Applying Census Data for Transportation

50 Years of Transportation Planning Data Progress

November 14–16, 2017
Kansas City, Missouri

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

TRB
TRANSPORTATION RESEARCH BOARD

Applying Census Data for Transportation

50 Years of Transportation Planning Data Progress

November 14–16, 2017
Kansas City, Missouri

Organized by
Transportation Research Board

Supported by
American Association of State Highway and Transportation Officials
Census Transportation Planning Products
Federal Highway Administration

Catherine T. Lawson
State University of New York, Albany
Rapporteur

Transportation Research Board
500 Fifth Street NW
Washington, D.C.
www.trb.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 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 E-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.

Preface

Meeting in Kansas City, Missouri, November 14–16, 2017, 115 participants attended the 2-day Census Conference—Applying Census Data for Transportation: 50 Years of Transportation Planning Data Progress—reflecting on past accomplishments, current lessons learned, and the future of the Census and related data products. The Transportation Research Board organized the event, with support from the American Association of State Highway Transportation Officials Census Transportation Planning Products (CTPP) program and the U.S. Department of Transportation. A pre-conference CTPP workshop focused on insights and techniques for using the CTPP. During the opening reception, 13 researchers presented their posters. Data sets covered during the conference included the Decennial Census; the American Community Survey; the Longitudinal Employer–Household Dynamics (LEHD); LEHD Origin–Destination Employment Statistics; Public Use Micro Samples (PUMS); and the National Household Travel Survey (NHTS).

Ed Christopher, Federal Highway Administration (retired, now an independent consultant), chaired the planning committee. Committee members provided expertise in transportation planning, data analysis, Census data, private- and public-sector data, and education and training. The planning committee was solely responsible for organizing the conference, preparing the call for abstracts, assisting in the solicitation of four commissioned papers to address topics in conjunction with the CTPP Oversight Board, reviewing the submitted abstracts, and developing topics for breakout and panel sessions, including guidance for the facilitated discussion sessions. Catherine T. Lawson, from the State University of New York, Albany, served as the conference rapporteur and prepared this document as a factual summary of what occurred at the conference.

The conference provided a forum for participants to share experiences with the use of Census data, including new techniques for integrating different data sets for use in transportation planning and decision making. Participants also learned about recent and forthcoming Census products (e.g., updates in the CTPP software). The conference further provided an opportunity for participants to discuss opportunities, limitations, and challenges involved in using Census data, data available from the private sector, and data from global positioning systems and other technologies. Finally, participants were able to discuss research and training needs associated with applying Census data and data from other sources to transportation planning and decision making.

This conference summary report follows the conference agenda. The presentations made in each session are summarized. The opening panel included Deborah Stempowski, from the Census Bureau, sharing the progress being made with the planning and preparations for the 2020 Decennial deployment. The first breakout session included a retrospective on original Census data products (e.g., Urban Transportation Planning Package), an introduction to the CTPP and a presentation of the first commissioned paper and facilitated discussion on the use of the CTPP for performance measures. The evening reception provided ample opportunity for the authors of the 13 posters to share their findings with conference participants, exploring a wide range of Census data topics.

Additional sessions covered activities at the Census Bureau, the use of Census data for equity analysis, and advanced data analysis. The second commissioned paper and facilitated discussion focused on transportation analysis zones and the current issues surrounding the geographies for the 2020 Census and beyond. The third commissioned paper and facilitated

discussion explored emerging "Big Data" opportunities and challenges associated with the use of private-sector data and how to keep the Census data relevant. Additional breakout sessions discussed PUMS and Public Use Micro Sample Areas, transportation modeling, and applications for alternative modes. The fourth commissioned paper and facilitated discussion compared various options for workplace data including: LEHD, NHTS, and private-sector options. The two final breakout sessions looked at the future of data for transportation planning and methods for comparing Census data sets.

According to Catherine T. Lawson, the plan for this conference is to provide attendees with lessons learned from the past, information on what is happening today and in the future, and how best to take advantage of data assets to inform transportation planners and decision makers. Key topics include methods and opportunities for complementing data products and putting together new ways of thinking, particularly with the CTPP. When the 2020 Census data is made available, using their Application Programming Interface at the Census, researchers and planners will be able to start answering their questions. There is time between now and when that data arrives to make plans for new visualizations, data analytics, data combinations, and data fusions products. Cloud computing will be a big part of the next data deployment.

Contents

CHAPTER 1

Opening Session

ED CHRISTOPHER
Independent Transportation Planning Consultant, Chair

TRACY LARKIN
Census Transportation Planning Package Oversight Board

ROLF SCHMITT
Bureau of Transportation Statistics

DEBORAH STEMPOWSKI
U.S. Census Bureau

CHARLES PURVIS
Metropolitan Transportation Commission (retired)

CATHERINE T. LAWSON
State University of New York, Albany

WELCOME AND CONFERENCE OVERVIEW
Ed Christopher

This conference would never have happened without the financial support and backing of the American Association of State Highway and Transportation Officials (AASHTO) Census Transportation Planning Products (CTTP) Technical Services Program.

This program has several themes including looking backwards at the past; examining the present; and then exploring future uses of Census data for transportation planning. Peppered into the program are four papers commissioned by the CTPP Oversight Board. The topics of the papers chosen will help inform the discussion the board is having as it plans for its future.

One of the themes running though the conference that is not stated anywhere is that we are all here with one goal in mind. To make our special tabulation, and the American Community Survey (ACS) the best quality that it can be. Quality data is key to our business. We need to know how our data was collected, what warts it has, and when not to use it. Just because it is on the Internet, or someone has it in an app, doesn't mean it is useful for our needs in transportation.

THE NEED FOR THIS CONFERENCE: WHY IS THE
CTPP PROGRAM INTERESTED?
Tracy Larkin

The CTPP has always been interested in research, along with data, making the data more useable, more understandable, and more accessible. We continually look to the Census data to ensure that the data is relevant to users. There have been eight conferences, the last one was held in 2011, and the one before that was in 2005. This conference is aimed at informing our decisions going forward. There are challenges facing our data programs. States do recognize the value of the CTPP and they have been supporting the program. They are the ones that subsidize and pay for the program.

The data quality is an issue. Despite the proliferation of big data, what data do we really need? How do we parlay this data into support for decision making? We must show value and accountability for the data that we show to the public. Along with research and training, funding outreach, we also have some critical issues moving forward. Concerns for ACS and the CTPP include how to work with and interpret period estimates, the margins of error, and changes over time. We want to increase in the relevance and utility of the data, squeezing out all the value we can. The four commissioned papers assist in this goal, covering topics that include using CTPP for performance measures; what to do about transportation analysis zones (TAZs), keeping the Census data relevant, and the importance of workplace data.

THE CTPP AND NATIONAL TRANSPORTATION STATISTICS:
A 40-YEAR PERSPECTIVE
Rolf Schmitt

This perspective begins with a meeting in 1978, with Alan Pisarski, who was focused on the development of special tabulations for what was then the Urban Transportation Planning Package (UTPP) (becoming the CTPP of today). The UTPP provided special tabulations of the 1970 and 1980 Censuses for states and metropolitan planning organizations (MPOs) purchasing the package. Subscribers chose their own geographies, which was a pretty radical idea in those days. The UTPP inspired part by the Highway Research Board (HRB) annual conference held in Washington in 1970, documented in *Special Report 21*. The HRB was the precursor to the Transportation Research Board (TRB). The 1980, and later work, was guided by the TRB Conference on Census Data and Transportation Planning held in Albuquerque in 1973 and documented in *Special Report 145*.

With the 1990 Census, the UTPP went nationwide as CTPP, an AASHTO pooled-fund project. The expanded version of the CTPP was inspired in part by TRB Conference on the Decennial Census Data and Transportation Planning held in Orlando in December of 1984. This effort was documented in *Special Report 206: Proceedings of the National Conference on Decennial Census Data for Transportation Planning* and *Transportation Research Record 981*. Subsequent TRB meetings on the Census for transportation planning were held in Irvine in 1994, 1997, 2005, and 2011.

While the Census provided special tabulations on a cost-reimbursable basis for others, the UTPP of four decades ago was a big breakthrough product in several ways. It provided statistics on the workforce by place of work, not just by place of residence. The 1980s UTPP was likely

the only product that Census was willing to reprocess when local planners identified data problems in the place of work tabulations and provided effective ways to correct the problems.

The 1990 CTPP was a pioneering product using a new technology: the CD ROM. Like the 1980 UTPP, the 1990 package was initially distributed on nine-track computer tapes and was only usable by the big MPOs and the state departments of transportation (DOTs) with big computers. AASHTO allowed the U.S. Bureau of Transportation Statistics (BTS) to distribute freely the CTPP CD-ROMS with software for preparing simple maps. This is noteworthy as these CD ROMs placed enormous amounts of data in the hands of groups that frequently opposed projects of AASHTO members. Frank Francois, then-Executive Director, believed in democratizing data and asked only that BTS handle customer support for CTPP users who were not states and MPOs.

The UTPP and CTPP grew out of the traditional world of transportation planning, with a focus on capital investment, driven by peak-hour demand for commuting. UTPP and CTPP data were major inputs for the four-step travel demand models, providing key information for the determination of trip generation, distribution, and mode split components. Local surveys frequently supplemented Census products to cover trips for purposes other than commuting. UTPP and CTPP also provided a wealth of information for analysis of accessibility and economic linkages among localities, as well as the three volumes of *Commuting in America*.

The world of transportation planning has certainly changed in the last four decades. The focus on new capital projects has been replaced by reconstruction and operations. Planners are increasingly concerned with congestion related to trip purposes other than commuting, and to incidents rather than recurring peak period volumes. Local planners must deal with freight movement as well as passenger travel. The world of transportation data and analysis have also changed in the last four decades. The four-step process no longer dominates travel demand models. Local surveys are becoming prohibitively expensive to conduct and suffer from declining response rates. New data sources as such as cellphone traces, provide huge amounts of observations at little cost, but they cover much narrower aspects of transportation activity and have limited, or no information on traveler characteristics. Unlike local surveys, new data sources cannot be tied directly to CTPP and other Census data. The CTPP itself is no longer a 1-day picture taken through the long-form of the Censuses every 10 years. It's a 5-year aggregation, under the guidance of AASHTO, as part of the ACS.

The CTPP is no longer the only Census Bureau picture of the relationship between residences and place of work for detailed geography. It competes with the Longitudinal Employer–Household Dynamics (LEHD) program. Is time past for the CTPP or does it still have a vital role to play? How does it need to evolve to remain worthwhile source of information? There remains a vital role for the CTPP as a key element of the complete picture of passenger travel throughout the United States.

The National Household Travel Survey (NHTS) provides a comprehensive picture of local passenger travel by all modes and trip purposes, but its geography is very limited. Only the CTPP relates local travel by all modes to local geography for all localities. The NHTS and CTPP must be used together, to create a consistent picture of small-area travel throughout the country. They could be combined with distance travel, which is currently limited to trips on commercial aviation, to complete the picture of passenger movement.

CTPP has been the bedrock from understanding local travel for more than four decades. This data resource resulted from combined efforts of many individual over the years, but no one was more central to those early days of the UTPP and the CTPP than James J. McDonnell.

McDonnell was Branch Chief in the Federal Highway Administration (FHWA) who built a relationship between the Census Bureau and the transportation community and who pushed the ad hoc TRB Committee to develop a solid and complete specification for the package. The productive history of the CTPP and the many related HRB–TRB conferences, including this one, are his legacy. Thank you, JJ, and thank you all current and past participants in these TRB Conferences who have made the CTPP a pillar of our understanding.

REMARKS FROM THE CENSUS BUREAU

Deborah Stempowski

The purpose of the Decennial Census is to conduct a Census of population and housing and disseminate the results to the president, the states, and the American people. The primary uses of this data is to apportion representation among states as mandated by Article 1, Section 2 of the U.S. Constitution. Funding levels, since 2012, indicate growth over the years in the requested amounts and fluctuation in the enacted budgets (Figure 1.1). Under the current situation with the Continuing Resolution, funding levels are generally being held from the previous year, so for the 2020 program, that would be at the 2017 level. The Department of Commerce Secretary, after completing an audit, announced an increase in funding to $15 billion from a little over $12 billion.

The 2020 Census is being conducted in a rapidly changing environment, requiring a flexible design that takes advantages of new technologies and data sources while minimizing risk to ensure a high-quality population count (Figure 1.2).

The primary goal is to count everyone once, only once, and in the right place. The environment for deployment will be unlike the 2010 environment. A 10-year planning effort is challenging, especially in the first year of the 10-year period.

Response rates are declining. Every 1% not provided as a self-response costs approximately $55 million to send people out to the field. In addition, plans need to consider increasingly complex living arrangements in the population (e.g., people move, children live with both parents).

The method for achieving the best outcome is to use four key innovation areas: reengineering address canvassing; optimizing self-response; utilizing administrative records and third-party data; and reengineering field operations (Figure 1.3).

2020 Census (Dollars in Millions)								
	FY12	FY13	FY14	FY15	FY16	FY17	FY18* Original Request	FY18** Adjusted Request
Congressional Request	$67.0	$131.4	$244.8	$443.2	$662.5	$778.3	$800.2	$987.0
Enacted Budget	$66.6	$94.0	$232.7	$344.6	$625.3	$754.4		

* Currently operating under Continuing Resolution through December 8, 2017 (PL 115-56). Amount references the FY 2018 President's Budget Request.
** Adjustment to FY 2018 President's Budget Request, as referenced publicly in the testimony by Secretary of Commerce to the House Committee on Oversight and Government Affairs on October 13.

FIGURE 1.1 FY 2018 funding update.
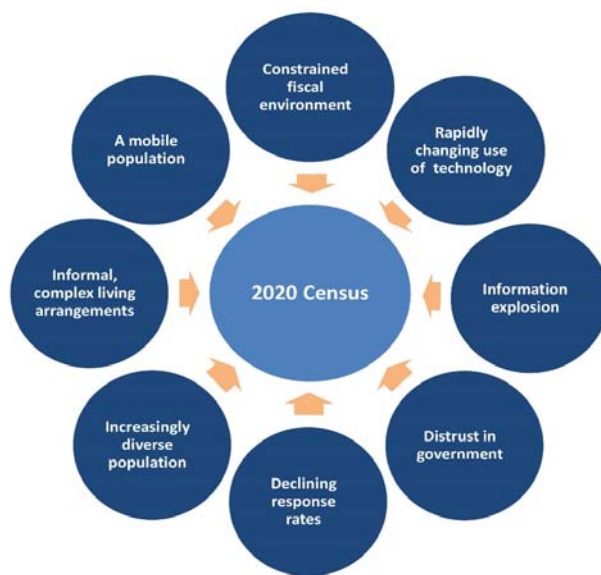
FIGURE 1.2 Rapidly changing environment for the 2020 Census.

FIGURE 1.3 Four key innovation areas.

According to the 2020 estimates, there will be 330 million people living in more than 140 million housing units. The reengineering address canvassing, now in production, is a process that will allow for a reduction in the nationwide in-field address canvassing operations by using methodologies for updating and maintaining our address list throughout the decade (referred to as "in-office address canvassing"). Internet will be the preferred method of response, however, after four attempts, nonrespondents will receive a paper survey. Telephone operators will be available for those preferring to give their Census response over the phone. Administrative records will also be used to reduce the nonresponse follow-up workload (e.g., identification of vacant housing units).

To address reengineering field operations, field staff across the country will be incorporating automation into their tasks. In 2010, automation assisted address canvassing. In 2020, devices will be used for nonresponse follow-up, processing their time, and listing expenses.

Figure 1.4 lists the milestone for the 10-year cycle. The address canvassing aspects of the end-to-end Census test involves exercising a final listing and mapping capability in the field including in-field listing quality control. Participating jurisdictions include Providence County, Rhode Island; Pierce County, Washington; and Bluefield–Beckley–Oak Hill, West Virginia. In

Milestone	Date	
Begin 2020 Census Program	November 18, 2011	✓
Deliver 2020 Census Operational Plan	September 30, 2015	✓
Federal Register Notice of Proposed Residence Criteria and Residence Situations for the 2020 Census	June 30, 2016	✓
Conduct the Second Round of Tribal Consultations	September – November 2016	✓
Begin Local Update of Census Addresses (LUCA) Program (including Appeals Process)	January 2017	✓
Deliver 2020 Census Topics to Congress	March 28, 2017	✓
Conduct the Block Boundary Suggestion Program	December 2015 – May 2017	✓
Conduct 2018 End-to-End Census Test	August 28, 2017 – August 20, 2018*	
Deliver 2020 Census Questions to Congress	by March 31, 2018	
2020 Census Day	April 1, 2020	
Deliver Apportionment Counts to the President	by December 31, 2020	
Complete Delivery of Redistricting Counts to the States	by March 31, 2021	
Complete 2020 Census Program	September 29, 2023	

FIGURE 1.4 2020 Census key activities.
(*Duration represents the timeframe for data collection.)

the test in Providence County, the miniature Census operation data-collection effort will include a prototype data product.

The number of area Census offices are being reduced from approximately 498 to 248, 40 of which will open in January of 2019, to help manage the address canvassing operation. In partnership with ACS, the Census Bureau met their legislative deadline last March by delivering the topics to Congress, to be covered in the Census and the ACS. Another milestone is the delivery of questions to Congress, as required in Title 13, by March 31, 2018. The remaining operational readiness includes the 2020 Census Operational Plan 3.0 (release date of October 2017); completion of Local Update of Census Addresses (LUCA) which began in January 2017; and field infrastructure of space, Decennial logistics management training, recruiting, and onboarding planning now underway.

The 2020 Census Operational Plan reflects and supports evidence-based decision making by describing design concepts and their rationale, identifying any remaining decisions, and describing remaining risks related to the implementation (Figure 1.5).
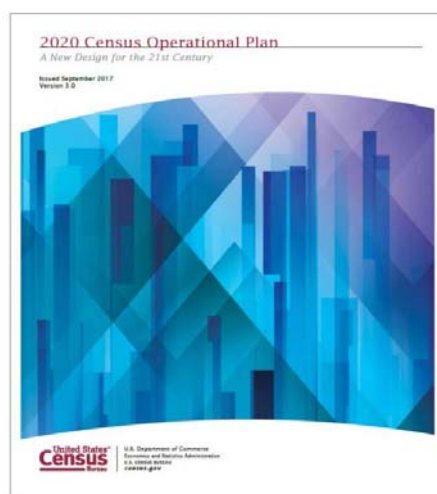
FIGURE 1.5 2020 Census operational plan.

The LUCA operations are underway inviting all levels of government to join in making decisions on local aspects to review and comment on the address list for their jurisdiction prior to the Census. This input assists in creating an address list that is up to date and to form local partnerships with the Census Geography Division. More than 6,100 governments have registered, covering up to about 67% of the population.

The Census has mandated products that include the apportionment counts to be delivered to the president by December 2020, and then the redistricting data files are released by April 1, 2021, followed by the remaining data products.

LOOKING BACKWARD AND FORWARD: PERSPECTIVES FROM AN MPO PLANNER
Charles Purvis

This presentation is based on materials produced in 2002, from the perspective of an MPO. Now, in 2017, it is possible to evaluate these predictions. The expectation of the delivery of the ACS and the Public Use Micro Sample (PUMS) (using the ACS), has been realized. Although the 2005 data was delivered later than originally expected and excluded the Group Quarters data, the ACS has been released annually every year since 2006. While the ACS doesn't provide county-to-county commute patterns, it does include intracounty data (people living and working in the same county).

The 5-year data at the tract and zone level was first released for 2006 through 2010 and thereafter. The 2012 through 2016 data should be released soon. Of course, the warning is to not use datasets that overlap—use the 2008 through 2011 and 2012 through 2016. Also, the ACS questionnaire changed in 2008, including questions on health insurance and Internet usage.

Some issues that weren't anticipated in 2002 include inexplicable year-to-year changes; reduced emphasis on workplace coding and increased emphasis on residence; reconciliation between the Decennial numbers and the ACS numbers; and a continuing need for local involvement.

The annual products from the ACS, in addition to the standard tabulations, include a 1% annual PUMS at the Public Use Microdata Area (PUMA) level at the 100,000 plus level. The 3-year product was discontinued in 2014. The 5-year data products includes a 5% PUMS and information on journey-to-work (JTW). Challenges for the ACS remain including modifying questions, weighting, and expansion factors.

CHAPTER 2

The Greybeards

NANDA SRINIVASAN
Energy Information Administration, presiding

ED CHRISTOPHER
Independent Transportation Planning Consultant

ALAN PISARSKI
Alan Pisarski Consulting

ROLF SCHMITT
Bureau of Transportation Statistics

CHARLES PURVIS
Metropolitan Transportation Commission (retired)

PHILLIP SALOPEK
U.S. Census Bureau (retired)

KEN HODGES
Claritas

This session focused on the history of the Census as it relates to transportation and the evolution of the CTPP, the transportation related questions to the long form ACS, and how the CTPP can survive in this age of resource cuts, reduced response rates, and lack of trust in institutions. The panel session consisted of experienced practitioners and new data users who offered perspectives on the CTPP, its relevance to transportation planning, and its place in the future.

HISTORY OF UTPP–CTPP
Ed Christopher

In 1960, the JTW question was added to the Census questionnaire to meet the requirement of the Office of Management and Budget (OMB) for information on commuting flow patterns for designating metropolitan areas. It was coded with a city or county designation. There were no special tabulations and questions included "What city and county did he work last week?" and "How did he get to work last week?" In 1970, the first transportation tables were assembled as part of the UTPP. FHWA provided the specifications for the tables and limited funding was provided by the U.S. DOT. Forty-three tables were purchased by 112 agencies, with the data available for local TAZs. Users were expected to use caution when they began using the data. The address coding was processed in Dual Independent Mapping and Encoding and placed information at the block level.

In 1980, the UTPP program was expanded and the Census Bureau hired JTW staff to improve geographical quality control. Work trips were included in the UTPP through imputation and allocation. There were 150 purchasers of the UTPP, with additional data including: more modes; vehicle occupancy; and travel times. In 1990, a new era for transportation data began with the decision to assemble the CTPP rather than the UTPP. The new program provided "wall-to-wall" coverage through an AASHTO–NARC pooled fund. The package included more data: departure time and vanpool occupancy. The data was extractable using software on a CD from the BTS.

In 2000, a logo was developed to "brand" the CTPP for transportation planners and researchers (Figure 2.1). The CTPP user community matured with the addition of a TRB Subcommittee (established in 1998) and a series of outreach products including a newsletter, a listserv, and on-call assistance with outreach. Some issues surfaced regarding TAZ delineations. Improvements in software made extraction easier while the first disclosure protection rules were undertaken.

The JTW data products and funders experienced strong growth over the last several decades. The programs have been overseen by ad hoc consortiums of interested individuals and a number of agencies including U.S. DOT (FHWA lead), BTS, Federal Transit Administration (FTA), and the Office of Secretary of Transportation, the Census Bureau (JTW–Migration and Geography Division), TRB Subcommittee, the AASHTO Standing Committee on Planning, and various states and MPOs.

When the Census Bureau made the decision to use a continuous long form approach (the ACS era), the CTPP was renamed the Census Transportation Planning Products and the logo was modified (Figure 2.2). The program evolved as AASHTO took on leadership of the program and formed the Oversight Board to handle all development of the data products. The Technical Service Program includes on-demand technical support, training and capacity building, research, data products, and related activities.

FIGURE 2.1 Original CTPP logo.

FIGURE 2.2 Modified CTPP logo.

OUR HISTORY WITH THE JOURNEY-TO-WORK: WRITTEN IN BLOOD
Alan Pisarski

It is important to remember that the original purpose of the JTW was to assist OMB with sufficient detail to define metropolitan areas, not for meeting the needs of transportation professionals. A series of joint planning meetings were held in Albuquerque, New Mexico (1973); Orlando, Florida (1983); and the Beckman Center in Irvine, California (1994 and 1996) to better understand the uses of the JTW data. One challenge identified was that 10 years was not enough time between Census deployments to deal with the issues. Also, the Census collects data on work trips, with no opportunity to collect additional transportation related data. However, the NHTS, collected since the 1960s, provides additional trip types.

When the UTPP was created, it relied upon the strengths of the Census. It had complete coverage of the United States, using consistent definitions and procedures for all of the collection efforts. However, a great limitation of the data observations is the fact that the data was collected in April every 10 years. In the 1960s, metropolitan areas were conducting large transportation data collection efforts to meet the first round of 1962 mandates for transportation planning. Census data was used to confirm locally collected social–economic variables, acting as a check on the home interviewing process. In late 1960s and early 1970s, there was a need for updates. Figure 2.3 compares data formats from 1960 to 1980.

The first edition of Commuting in America was produced and used for 10 years. Researchers began to realize that the long time horizon of the Census meant that changing public issues were difficult to address (e.g., *w*ill carpooling matter 10 years from now?). Some researchers and practitioners argued for only the most basic information be collected in the 10-year cycle.

At this time, there were critical internal differences within the Census Bureau regarding the use local expertise and local geographic tools to compile coding guides. Specific challenges centered upon the impact of major generators (e.g., Does Union Station have an address?). An address of a location site was a Census protected data item and covered by nondisclosure rules. A serious concern was voiced about the lack information on data quality, fearing the data would be less useful than expected. Recoding was not permitted*;* however, New York did accomplish this and was able to recode data by spending their own resources.

Data Element	1960	1970	1980
Vehicles Owned	Autos (0,1,2,3+)	Autos (0,1,2,3+)	Autos (0,1,2,3,+) Vans, Trucks (0,1,2,3+)
Mode Choice	Eight Basic Categories	Nine Categories Auto Driver Auto Passenger	Twelve Categories Car Truck Van Motorcycle Bicycle Carpool Question Auto Occupancy Question
Travel Mode	Principal means last week	Chief means last day worked last week	Mode usually used in usual work trip
Work Place Identification	Cities, Counties	Block (otherwise 1960)	Block (or place level)
Travel Times	None	None	Travel Time to Work in minutes

Transportation Planning and Census Data: An Evolving Relationship; A. Pisarski Orlando Special Report #206

FIGURE 2.3 Evolving early decennial data.

FEDERAL PERSPECTIVE
Rolf Schmitt

Census economic series are produced using a 5-year cycle, with deployments in years ending in two and seven (e.g., Commodity Flow Survey). In 1963, a survey effort focused on passenger travel, setting the stage for the later NHTS. There was also a home-to-work survey that may have contributed to the improvements between the 1960 and 1970 Census questions. Some of the questions that were included asked about the distance between home and place of work [now calculated using geographic information system (GIS)]; distance to public transit (e.g., number of blocks away); and parking (e.g., parking on street, off street with charge; off street without charge).

Even with changes, researchers and planners continue to struggle with methods for linking the ACS and CTPP with alternative data sources. For example, neighborhood characteristics are used as a surrogate for the individuals when trying to use these alternatives. The LEHD is a possible model. It contains place-of-work and place of residence information that is not collected with a survey, but rather through the use of administrative records. Similar models exist for freight where the Commodity Flow Survey is transformed into a total freight model through the addition of many other data sources. Another approach is to leverage information from the CTPP and ACS about commuting under the assumption that considerations of lifestyle (e.g., where people shop and go to school) could be related to JTW, making it possible to produce a total travel data set.

METROPOLITAN PLANNING ORGANIZATION PERSPECTIVE
Charles Purvis

The computers available in the 1970s and 1980s presented an extreme challenge for transportation planners because of slow speeds and terminals that needed to be connected to a mainframe to use the data. In 1980, JTW cost $51,000, coming to an agency by mail to the executive director. The data arrived on a 6,250 reel-to-reel tapes. By 1995, the Internet had replaced CD-ROMs as a method for receiving data. Interregional commuting flows were difficult in the older data sets. In 1980, county-to-county and neighboring counties could be calculated as zone-to-zone flows. By 1990, tract-to-tract was available within a state. This was a major leap forward in interregional analysis.

The Orlando Conference facilitated changes by supporting the addition of travel by ferry or streetcar, and departure time to work. Perhaps the most profound change occurred when the Census Bureau moved to a continuous measurement approach for the long form questions. This change prompted both the 1994 and the 1996 conferences for the transportation community to adapt to this major change. Fortunately, the 2000 deployment included both the long form and the ACS to ease the transition for transportation professionals. Referred to as the Purple Report, the BTS provided guidance for working with continuous data. Today, changes may need to be pursued, or at least explicit instructions, to clarify whether Transportation Network Companies (e.g., Uber or Lyft) should be indicated as a taxi or a new category.

MAIN ISSUES FOR CENSUS BUREAU IN 1980, 1990, AND 2000: TRANSPORTATION PLANNING PACKAGES—50 YEARS OF TRANSPORTATION PLANNING DATA PROGRESS
Phillip Salopek

In 1980, the JTW Branch's main issue was to improve the accuracy of place-of-work data. This issue was primarily the result of feedback from the transportation agencies who received the 1980 package and also comments voiced at the Albuquerque Conference. The coding process in 1980 was manual. It was revised and rewritten to implement coding procedures for clerks to convert written responses into Census codes. The Census Bureau provided improved reference materials, including telephone books, zip code directories, commercial maps, and major employer lists. Headquarters staff were assigned permanently to each processing office to serve as expert resources for any coding questions or problems. For the UTPP, the first-ever place-of-work allocation was developed and implemented.

In 1990, the JTW Branch, along with the Geography Division, developed a computer coding system. Using manual coding procedures with 1980 as a guide, programmers created, tested, revised, and finalized computer coding algorithms for the first step in data processing. In addition, software and processes were created for the clerks to use to input the data.

For the 2000 CTPP, the main focus was to improve access to the data for the wide range of participants in the program, and to take advantage of more modern data processing and analysis technologies. The first step in this process was to contract with two companies to create software to accompany the distribution of the 2000 CTPP data on CD. *Beyond 2020* and its data browser were chosen as the vehicle for displaying and extracting CTPP data. Digital Engineering Corporation was selected to team with *Beyond 2020* to create mapping software for data selection and data

analysis. More modern methods were also provided for agencies to create their TAZ equivalency files and for examining and updating the workplace files (major employer lists) used in place-of-work coding. FHWA, along with the Geography Division, sponsored an ArcView application using Topologically Integrated Geographic Encoding and Referencing (TIGER) line files that CTPP 2000 customers used to define TAZs. The JTW Branch contracted with ESRI to create an ArcView application called the Workplace Update Extension (WORK-UP). This was a tool that MPOs could use to verify, correct, and add entries for employers in their area to ensure more accurate and complete workplace data. An extended place-of-work allocation location system was also developed and implemented in the 2000 Census.

CONSULTANT PERSPECTIVE
Ken Hodges

The product, PRISM, segments small-area Census data to form block groups with specific lifestyle clusters directly useful for consumer behavior. The concept is that "you are where you live" and thus knowing where people live provides a way to infer future behaviors. This assumption would not necessarily apply for place of work. While there is no ongoing effort to produce a work place PRISM, there is continued interest in the improvement of the CTPP.

CTPP Program 101

PENELOPE WEINBERGER
AASHTO

TOM FAELLA
LaCrosse Area Planning Committee

CHRIS BONYUN
Beyond 2020

PHIL MESCHER
Iowa Department of Transportation

BENJAMIN GRUSWITZ
Delaware Valley Regional Planning Committee

W ith the advent of the ACS, the special transportation tabulation product was brought under a state DOT funded, cooperative program and broadened to include research, technical assistance, and training for the transportation planning community. This session highlighted the various aspects of the program.

PROGRAM OVERVIEW
Penelope Weinberger

Begun in 1970 to provide special tabulations, the CTPP is now under leadership of AASHTO, with funding from state DOTs for their MPOs and their own staff. Ninety-seven percent of states participate in the program and representatives from FHWA, OST, FTA, and TRB provide input. CTPP is guided by an Oversight Board of 17 people, with equal representation from the states and MPOs. The MPOs represent various sizes and locations across the country. They make decisions for the CTPP Program to make sure it is useful. There are also friends of the program who help the board members. The data is free and is made available on the web through special software.

The JTW question first appeared on the Census in 1960. By 1970, transportation planners were well aware of the usefulness of JTW data. The original JTW question allowed for respondents to write in modes. The 1970 tables were purchased by a variety of users, but in 1990, a decision was made to use a nationwide approach as the CTPP. In 2005, the ACS became the data source for the CTPP, and $5.8 million was collected from the states to fund 2007–2013. The Technical Services Program is in its second 5-year period. There was a county-to-county version for 2009–2013 data. The next CTPP will cover 2012–2016. It is free to use as it has already been paid for by states through AASHTO. The CTPP software is being upgraded. A special demonstration of the new software features was held at the Mid-America Regional Council (MARC) facilities for interested conference participants. Information on CTPP activities

is available on the CTPP listserv, the Status Report, and on the FHWA website. TRB has a subcommittee for the Census, through Urban Transportation Data and Information Systems Committee.

NEW DATA
Tom Faella

The Census Bureau asked the CTPP Oversight Board to reduce the proposed 2012–2016 tabulation by two-thirds and to reduce the number of tables produced at all geographies. In May 2014, a CTPP "Tables Subcommittee" set to work producing recommendations for tables for "All Geographies" and "Large Geographies Only." The Large Geographies tables will include nation, state, county, metropolitan statistical area (MSA), principal city, place, PUMA, Municipal Civil Division (MCD, for 12 strong MCD states), and traffic analysis districts (TADs). All Geographies includes data produced for Census Tract and TAZs. Some tables will still include perturbated data ("B" Tables) including most flow tables; means of transportation, aggregate household (HH) income, and carpool tables. One-hundred-seventy-six tables will be deleted from the new tabulations and eight new tables will be added. Some tables will not be requested from the Census Bureau as transportation researchers and planners will be able to create them using the CTPP software (e.g., collapsed tables).

The methodology used to determine the tabulation proposal included using the CTPP access software to analyze how many times each table was accessed since November 1, 2013, not including "Power Users" who routinely download full or full state CTPP data for analysis. The CTPP Oversight Board members, users, and CTPP listserv members were polled for their preference for tables that should be retained and which should be included in All Geographies. Town hall meetings were held in November 2015, with approximately 75 attendees. Figure 3.1 provides the details on the recommended changes in the table elements. The final criteria for the decision to eliminate tables included tables accessed less than 150 times by November 6, 2015; tables with five or more recommendations for removal by committee members; and tables that were not slated for elimination but that had two or fewer recommendations to retain. All

	Number of 2006-2010 CTPP Tables	Number of Tables Deleted	Number of Tables Retained at Large Geography Only	Number of All Geography Tables Retained	Number of Tables Proposed for 2012 - 2016
Overall	344	176	76	99	168
Part 1	187	85	48	56	102
Part 2	116	73	17	30	43
Part 3	41	18	11	13	23

FIGURE 3.1 Table comparison to 2006–2010 tabulation.

Geographies were proposed as Large Geographies Only. The tabulation proposal was approved by the full CTPP Oversight Board on May 10, 2016. The proposal was submitted to the Census Bureau for tabulation. The Census Disclosure Review Board approved the tabulation (August 2016). Data access software with the new 2012–2016 CTPP tabulation will be available in 2018 (or early 2019).

NEW SOFTWARE
Chris Bonyun

Although the 2012–2016 CTPP will be available in 2018, it will not be available in the CTPP software until 2019. The data is available from the software or directly from the Census in large chunks. The software allows users to manipulate the tables and then download the data to a local computer. The software is going to be upgraded, but is now available with the features. There are currently three datasets available in the software: the 3-year data (2006–2008); the 5-year data; and commuting flows (2009–2013). Four more datasets will be added including the 2000 CTPP and the 1990 CTPP. With these additional years, users will be able to see trends, but not for every geography, as the geographies have changed over time. For larger areas and consistent questions, comparisons will be possible.

     To access the data, users need to use the dropdown and choose a dataset. Next, they would pick a geography. Every geography and the tables will be available by residences and workplaces. The flows are between residences and workplaces or workplaces for residences for all counties in a state, or all of the states. A future feature will allow users to upload their own shape files and the data will fill customized geographies. Full-day training is available to learn how to use the software.

     The table data can be displayed on the maps (e.g., county-to-county flows). The data can be downloaded as a CSV, EXCEL, or GIS file. If the dataset requested is small, it will be provided immediately. However, if the dataset is very large, it will be requested in a queue and notification will be made by e-mail with a link when it is ready.

     CTPP data is also available as CTPP profiles. These are available for states or counties (already prepared). It compares 2000 to 2006–2012 data. When the 2012–2016 is available, there will be three points in time to compare for trends. The new software allows the user to customize comparisons, with templates that will automatically populate and produce data in CSV, EXCEL, or PDF formats and to batch file extractions as well.

RESEARCH
Phil Mescher

The Research Subcommittee is an integral part of the improvement path for the CTPP. CTPP Board members and staff create problem statements on a regular basis to seek out research on a wide variety of topics on CTPP and other travel data. CTPP leverages available research mechanisms but also funds its own research efforts. CTPP also supports conferences like this one to share ideas and solutions. The subcommittee works to prioritize the problem statements and pick the ones most likely to meet some immediate need and be successful. Research priorities include *Census Data Guidebook on Analysis*; reporting, presentation and dissemination of CTPP

data; investigate sources for the Non-JTW trip; investigate combining ACS with administrative records to develop O-D matrices; and archive 1980 UTPP. New project ideas include income spent on housing crossed with transportation variables; vehicle sufficiency data crossed with other variables; use of PUMS data and CTPP crosses at PUMA level; primary work role (student or worker?); work-at-home estimates; poverty data at two times the poverty level; is it possible to get at any geography; and disability data. Since 2006, in excess of $1 million of research has been generated or funded by the current CTPP.

Highlights of Recent CTPP-Related Research

National Cooperative Highway Research Program (NCHRP) Project 08-36/Task 127: Employment Data for Transportation Planning produced a guide to using employment data by Cambridge Systematics, begun in July 2015. The initial work created a vision, objectives and a work plan. The project produced a Technical Memo: Synthesis of Key Elements and Characteristics of Common Employment Data Sources.

NCHRP Project 08-36/Task 128 produced a final report, *NCHRP Web-Only Document 226: Data Visualization Methods for Transportation.* The objective of this report is "to evaluate data visualization methods and their applicability to transportation planning and analysis." The focus of the research is to better understand data visualization tools and techniques as data can be hard to understand (Figure 3.2). Questions addressed include what methods can be most effective and what data visualization methods exist, recognizing the need to classify them for best uses.
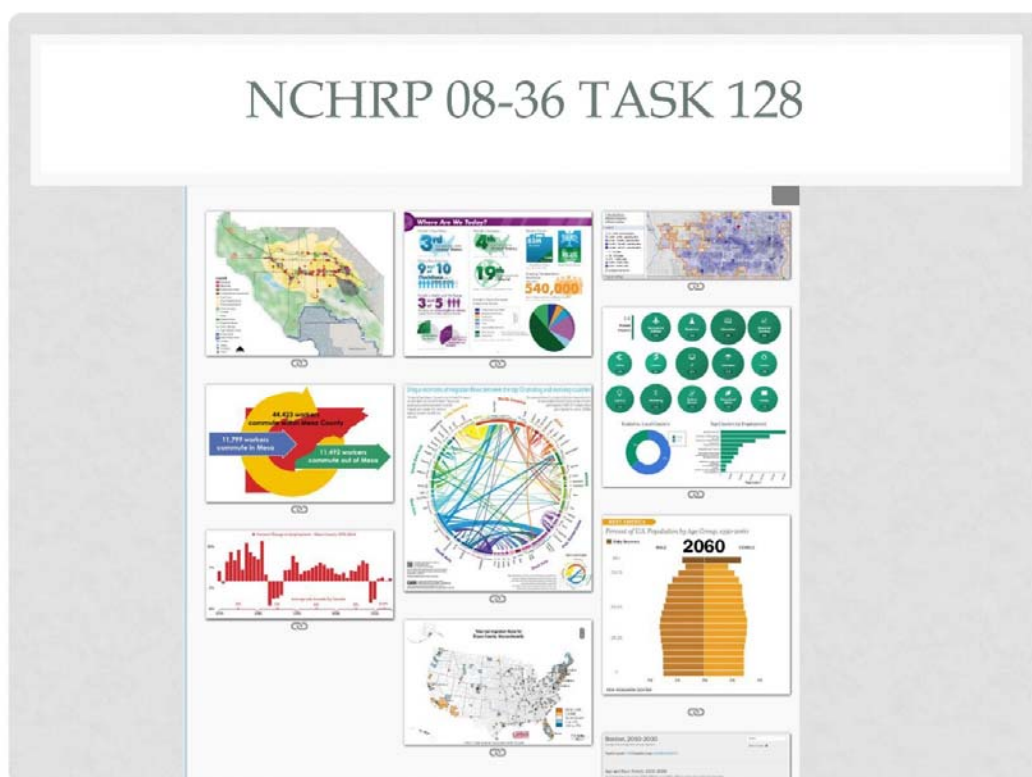
FIGURE 3.2 Sample visualizations.

NCHRP Project 08-36/Task 135: Addressing Margins of Error in Small Areas of Data Delivered through the American Fact Finder or the Census Transportation Planning Products Program produced a final report that provides guidance on how to appropriately handle large margins of errors (MOE) for use in data applications and how to communicate MOE when data are represented visually (e.g., heat maps or pie charts). Also addressed is the concern when ACS data is spread too thin to constitute an appropriate use of the data.

The NCHRP FY2019 Program will include a new research project provisionally titled "Census Transportation Data Use and Application Field Guide." The project is being planned to assist agency staff to effectively use and understand the limitations of the CTPP, ACS, and PUMS data sets in transportation system planning, programming, and project analysis.

Another research essential is the production of the next *Commuting in America* (CIA). The CIA is a national report describing travelers and their commuting behaviors (Figure 3.3). The goal is to provide factual commuting data for transportation professionals to use for decision making. The first CIA was produced in 1984 and now relies on the CTPP for commuting trends. The most recent CIA is available electronically on the CTPP website at http://traveltrends .transportation.org/Pages/default.aspx.

TRAINING AND OUTREACH
Benjamin Gruswitz

The task of the training subcommittee is to oversee training and outreach activities for the CTPP program. This includes in-person trainings (Figure 3.4); assessing the training needs of the user community; conducting in-person training; developing training materials on the CTPP program website; the e-learning modules; creating how-to-videos; conducting webinars; and conducting conferences (Figures 3.5 and 3.6).
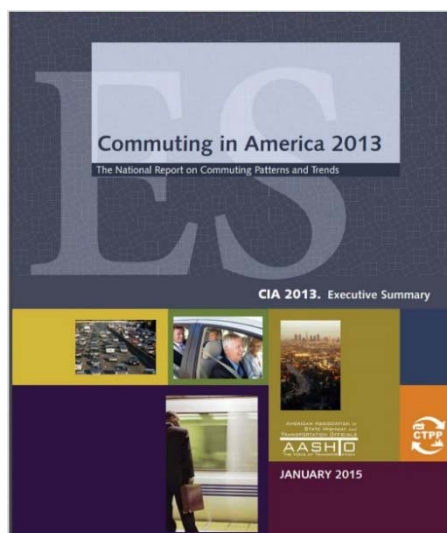
FIGURE 3.3 CIA 2013.

FIGURE 3.4 Training locations since launch of CTPP 2006–2010 data set.
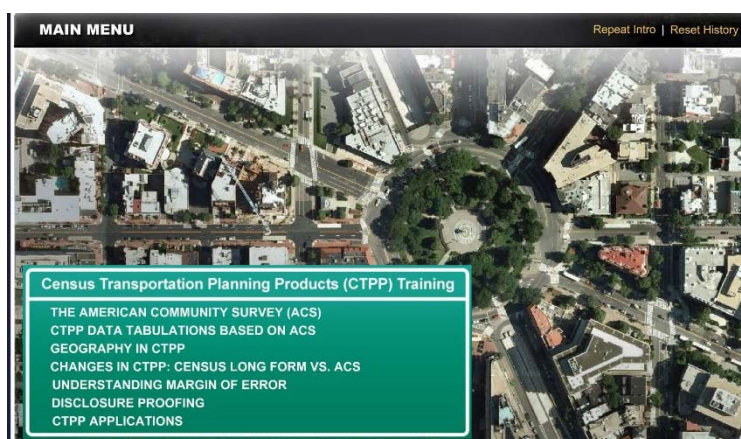
FIGURE 3.5 E-learning modules.
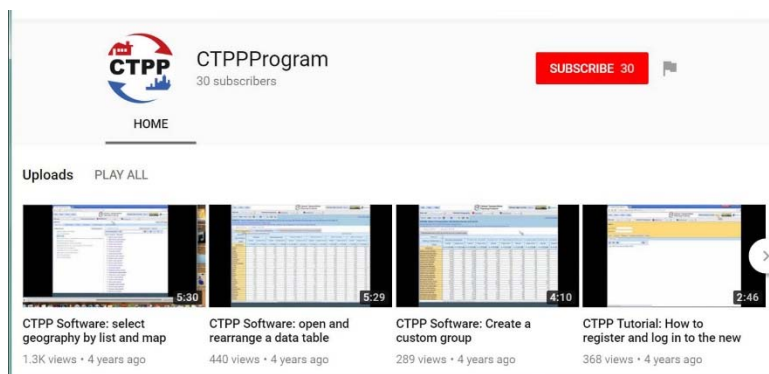
FIGURE 3.6 Software tutorials.

The in-person training program consists of 1- or 1½-day hands-on training courses that include understanding and dealing with data issues; transportation data and how to get it; and what kind of data is collected and where it is. Also covered are topics on the Census, the CTPP geographies and how to understand and use them, and the CTPP data access software.

Other resources include the CTPP Status Report (newsletter) that includes descriptions of applications using CTPP data. Recent issues include a list of tables for environmental justice (EJ) analysis. Another source of news is the CTPP new mailing list (listserv) that keeps the community of users informed on advances and issues. It provides a great forum for asking questions to a diverse user community. More information is available at http://www.chrispy.net/mailman /listinfo/ctpp-news.

Supporting Transportation Performance Management and Metrics with Census Data

IVANA TASIC
Department of Civil and Environmental Engineering, University of Utah

JIM HUBBELL
Mid-America Regional Council, presiding

KAREN MILLER
Missouri Department of Transportation, recording

Transportation Performance Management (TPM) and metrics are an ever-increasing component of our transportation decision and policy processes. As these TPM processes mature, Census data likely will be used to support them. This commissioned paper explored several uses of ACS and CTPP data to support different TPM activities.

ADVANCING TRANSPORTATION PERFORMANCE MANAGEMENT AND METRICS WITH CENSUS DATA
Ivana Tasic

Background

The CTPP program was funded by state DOTs and administered by AASHTO. The purpose of this program as a partnership among all states is to support the development of Census data products and their application in the field of transportation. Since the initial development of CTPP data, a number of transportation projects and studies have benefited from using the data. By introducing CTPP, the early stages of transportation project planning and development are being emphasized. The quality of the data is improved with the new formatting and access capabilities, and the CTPP application is free for the public use.

While previous studies extensively used CTPP data to inform practice and research about the characteristics of JTW traveler behavior dynamics, data sampling issues, the implications for new travel demand models, and improving the data structure (*1–3*), this paper shifts the focus to performance management and metrics. Transportation system performance indicators have been driving decision making for decades, and as data availability improves, the range of metrics is becoming wider to accommodate the variety of users in the transportation system (*4–5*).

This paper is divided into several sections. The introductory sections explain the purpose of CTPP, the research objectives, general approach to performance measurements selection, and the case study that serves as a demonstration here. The core sections focus on three performance metrics: safety, mobility, and accessibility, and demonstrate how these can be developed using the CTPP data. The final sections compare the performance metrics obtained by using CTPP data only, and the potential for fusing CTPP data with currently existing open data platforms, and the

summary of findings. With this type of setting, this research is aiming to identify (1) the currently available data that can serve as the foundation for transportation decision making; (2) the performance metrics that can be developed using the currently available data, while primarily relying on CTPP data; and (3) the way we can use the developed performance metrics to advance current performance management of transportation systems. While considering transportation as a system in the performance analyses conducted in the core sections, this paper also discusses the transferability and potential for future applications and improvements of CTPP data, particularly for the purpose of establishing long-range TPM strategies.

Research Objectives

The main goal of this research is to demonstrate the application of CTPP for the purpose of advancing TPM. The importance of developing transportation system performance measurements that can be adequately implemented in various stages of transportation project, ranging from programming and planning to operations and maintenance, has been increasing over the past two decades. This effort to improve TPM exists on the national level, as a strategic approach towards creating policy and infrastructure investment decisions that aim to achieve nationally established performance targets and goals (*6*). While earlier efforts in transportation research and practice also have been geared towards performance improvements, the current efforts, particularly in long-range transportation planning have taken a much more systematic approach towards identifying transportation performance outcomes that should be prioritized. As previous TPM efforts scarcely consider the application of CTPP, this research is focused on exploring the potential of CTPP for the purpose of TPM development.

In addition to potential CTPP contributions in TPM field, the era of big data and open data has brought tremendous opportunities in terms of the variety of data sources that are now available for transportation stakeholders. Past decision making in transportation has been highly dependent on the data collected and available from transportation agencies. The current decision making has a much broader range of data resources that can be utilized to not only improve transportation project-related decisions, but also reflect higher level of inclusion of various data generating platforms (*3*). For example, transportation agencies and transportation users are becoming more and more equal in terms of data provision, and thus transportation users are becoming more and more invested in transportation decision making. This is very significant, because transportation is primarily a service, and whether a local, a regional, or a state agency provides it, the outcomes and quality of this service need to prioritize and include users as much as possible.

This research brings the existing transportation data resources and performance metrics together, using CTPP data as the foundation, and performance metrics as the target outcome, with the purpose of exploring how CTPP can be used to advance the current TPM efforts.

Performance Measurements

The role of TPM is crucial for transportation decision making and policy formulation. A major shift in TPM begun during the past decade as performance metrics became more inclusive and started to account for the quality of transportation service for all users in the transportation system. In addition to being more inclusive, the metrics we now use are oriented towards enhancing the methodology used to evaluate the transportation service. The main goal of TPM improvement is to develop

performance metrics that are transferable, data-driven, facilitate decision making, and enable communication between decision makers and transportation service users.

The TPM methods today go beyond the traditional metrics, which mostly focused on evaluating traffic congestion. The FHWA has established six target groups of major transportation issues that need to be resolved through the development and implementation of adequate transportation performance metrics in the decision-making process (*6*):

- Improving safety;
- Maintaining infrastructure condition;
- Reducing traffic congestion;
- Improving efficiency of the system and freight movement;
- Protecting the environment; and
- Reducing delays in project delivery.

These six rules for TPM development clearly distinguish six performance metrics for the transportation system evaluation: safety, infrastructure condition, traffic congestion, efficiency, environmental impact, and project delivery. This paper will mainly focus on the metrics related to safety, congestion, and efficiency by demonstrating how CTPP data can be used to develop the following performance metrics:

- Safety,
- Mobility, and
- Accessibility.

These three areas of TPM are selected to capture both traditional and more-recent approaches to performance measurement, with the capability to implement the developed metrics to private vehicle users, public transit users, pedestrians, and bicyclists. The goal is to demonstrate how CTPP data can be used to develop this set of metrics for various transportation users, and then demonstrate how the developed metrics could potentially be improved by fusing CTPP data with other data sources from transportation agencies and publicly available data platforms.

In the area of road safety, target-based and result-oriented decisions towards reducing or eliminating the most-severe crash types are preferred when selecting the most-effective countermeasures. This safety performance-based approach is already used in microlevel road safety analyses related to intersections and road segments. The macroscopic road safety analysis is gaining the momentum with the increasing need to incorporate road safety targets in the long-range transportation plans. This is where Census-based data could play a major role in capturing the areawide effects that are associated with crash frequencies and severities for multimodal transportation users.

Mobility-oriented performance metrics relate to speed and utilization of the available capacity of transportation infrastructure. Mobility usually is linked to intersections or roadway segments, but it is also an important element of long-range transportation planning. Census data have been used for decades to build travel demand models and evaluate the needs to invest in transportation infrastructure improvements. In certain parts of the country, MPOs develop and conduct their own surveys to build travel demand models, and the advantages and disadvantages of using local data with limited sample size versus CTPP need to be further explored.

Accessibility is dependent on the availability of multimodal infrastructure, and its integration with the land use patterns. It describes the ability of transportation users to reach desired destinations within the given time constraints. Accessibility as a transportation performance metric that recently became incorporated in transportation policies, particularly in the regional and city-level long-range transportation plans. The way accessibility is measured highly depends on data availability and the purpose of measurement.

The common thread for all three measures: safety, mobility, and accessibility, in this paper is the demonstration of the development of these metrics based on CTPP data only, and the comparison with the potential improvements that can be achieved when CTPP data are combined with data from alternative sources which are addressed in the following section.

Case Study and Data

City of Chicago is the case study. The most recent efforts that Chicago made to improve urban data collection make it a great candidate for future research efforts in this field. The possibility to transfer the findings of this paper to other cities and regions will be discussed in the final section of the paper. The City of Chicago Department of Innovation and Technology maintains a very detailed database on transportation and urban environment features. Chicago's robust data portal was established in 2010 and hosts over 900 datasets with information on various services in the city, in tabular, GIS, and Application Program Interfaces (API) formats. The portal is developed to enable residents to access government data and utilize them to develop tools that can improve the quality of life in the city. This is currently one of the "largest and most dynamic models of open government in the country" (*7*). In addition to improving the decision-making process by merging various data sources and developing an open data platform, the city of Chicago is also invested into developing new ways to generate and collect urban data.

Apart from the major efforts to develop high fidelity open-source data platforms, Chicago is also known for its extensive multimodal transportation system. The city has developed complete streets design guidelines (City of Chicago, 2013), with "Make Way for People" initiative that converts underutilized "excess asphalt" street spaces into active public spaces with purpose to increase safety, encourage walking, and support community development. Chicago has invested in bicycling infrastructure to become one of the best major U.S. cities for biking with over 200 mi of on-street bike lanes. The city of Chicago is also known for its active safety research not only vehicles but bicyclists and pedestrians as well, and a very extensive transit system. Chicago is the first major city in the United States to adopt a citywide policy for the investments in safety countermeasures that would reduce pedestrian crashes, as a part of the national "Vision Zero Network" initiative. All factors described above made Chicago a valid case study for the purpose of this research.

This study combined data from several sources, including open data and data obtained from multiple transportation agencies, to develop a comprehensive framework for the analysis of the relationship between multimodal transportation features and safety in urban transportation systems. Data collection included crash data, multimodal transportation features, road network features and traffic conditions, land use data, socioeconomic characteristics, and analysis of spatial features to select the adequate spatial units of analysis. The CTPP data packages are developed from ACS data for the designated 5-year periods. Thus, the most recent available CTPP data package is based on ACS data for the period from 2006–2010. The data includes residence tables, workplace-based tables, and flow tables (home-to-work trips) with the capability to extract tabulated data in various formats and visualize them using the available map tool. Tables include means of transportation

univariate and crossed with travel time, household income, vehicle availability, age, time leaving home, and (new) presence of children, minority status, number of workers in household, and median household income. The characteristics of CTPP data formatting, as well as the fact that the data are collected for the 5-year periods, makes the data very flexible for transportation analysis purposes. In addition to CTPP, data were obtained from the Illinois DOT, Chicago Metropolitan Agency for Planning (CMAP), Chicago Transit Authority, and the available open data platform supported by the city of Chicago.

Determining the level of spatial data aggregation is an important step in this study, as the choice of spatial analysis units could significantly affect the outcomes of the study. Census tracts were the most appropriate for spatial analysis in this case due to the data coverage and availability, and the convenient link to socioeconomic characteristics, which have proven to be relevant for safety outcomes. The ranges of spatial units numbers used in the available literature indicated that Census tracts would be appropriate as well. Census tracts are small statistical county subdivisions with relatively permanent geography that are updated each decade under the initiative of the U.S. Census Bureau. Census tracts are supposed to be somewhat homogeneous and ideally have around 1,200 households (perhaps 2,000 to 4,000 people), but, in Chicago, population varies from 0 up to 16,000. Census tracts in the city of Chicago have remained nearly constant since the 1920s, but the numbering system has changed. Census tracts in the suburbs have changed a great deal over the years, in most cases by splitting. There were 876 Census tracts in Chicago according to the 2000 Census. After merging the data needed for the analysis, and eliminating some Census tracts due to missing data in the geocoding process, a total of 801 Census tracts remained in the dataset. Table 4.1 shows the summary statistics of data used to develop performance metrics described in the following sections of this paper.

The following sections of the paper focus on the application of CTPP data combined with other data sources in the city of Chicago, to develop transportation performance metrics of safety, and mobility and accessibility for private vehicle users, pedestrians, and bicyclists.

Applying Census Data for Safety Evaluation

The main purpose of this section is to develop transportation safety evaluation methods based on Census data. The question that safety evaluation metrics are attempting to answer is what the expected frequency of crashes is under the particular areawide set of characteristics that can be described by using Census data. Safety performance functions (SPFs) are developed to predict vehicle-only (vehicular), pedestrian–vehicle (pedestrian), and bicyclist–vehicle (bicyclist) crashes on the Census tract level. SPFs are statistical models developed to estimate the average crash frequency for the selected entity (intersection, segment, area) as a function of exposure measures (traffic volume and road segment length) and, if the data availability allows, other conditions that characterize transportation network design and operations, and its environment. The general formulation of SPFs follows negative binomial regression model form as the most common approach to representing count data with over dispersion. The general form of each SPF is as following (*9*):

$$\theta_i = e^{(\beta_0 + \beta_1 \ln(Exp1_i) + \beta_2 \ln(Exp2_i) + \sum_j \beta_j x_{ij} + \varepsilon_i)}$$

TABLE 4.1 Descriptive Statistics (801 Census Tract Observations)

Variable	Description	Mean	SD	Min.	Max.
DOT Crash Data					
VehCrash	Vehicle-only crashes	375.176	354.534	5	3,920
Veh_KA	Vehicle-only fatal and severe injury crashes	8.004	8.465	0	71
PedCrash	Crashes involving pedestrians	17.750	22.528	0	481
Ped_KA	Fatal and severe injury crashes involving pedestrians	2.131	2.555	0	41
BikeCrash	Crashes involving bicyclists	9.528	13.178	0	172
Bike_KA	Fatal and severe injury crashes involving bicyclists	0.783	1.293	0	12
CTPP Data					
Population	Population size	3.402	1.741	0.000	15.740
Pop_Dens	Population density per mile squared	18.203	20.206	0.000	485.019
Employed	Percent of employed population	6.759	18.955	0.000	86.000
Unemploy	Percent of unemployed civil population	14.970	9.459	0.000	51.000
PerCapInc	Average income per capita	27,786.690	20,029.490	0.000	131,548.000
NoVeh	Households with no vehicles, %	26.537	15.118	0.000	89.400
Veh1	Households with 1 vehicle, %	43.589	9.508	0.000	81.300
Veh2	Households with 2 vehicles, %	22.558	11.544	0.000	59.100
Veh3plus	Households with 3 or more vehicles, %	6.814	5.648	0.000	26.900
DriveAlone	Drive-alone trips to work, %	50.186	15.522	0.000	86.300
Carpool	Carpool trips to work, %	9.511	6.560	0.000	39.500
Transit	Transit trips to work, %	27.506	12.956	0.000	79.100
Walk	Walk trips to work, %	0.603	3.156	0.000	35.000
OtherMeans	Trips to work by other means, %	2.542	2.942	0.000	21.300
WorkHome	Work home, %	4.058	3.296	0.000	21.300
TT_min	Average travel time to work, min	34.019	6.303	0.000	56.500
Open Data					
Road	Total length of roads, miles	6.278	3.910	0.142	30.762
Art	Arterials, % of street network	0.924	0.790	0.000	7.675
BikeLane	Total length of bike lanes, mi	0.679	0.723	0.000	6.163
BusRoute	Total length of bus routes, mi	1.541	2.559	0.000	39.980
Ltrain	Total length of l train lines, mi	0.147	0.353	0.000	4.411
Sidewalk	Total sidewalk area, feet squared	287.382	198.201	0.000	1,131.373
Intersect	Total number of intersections	37.803	27.800	0.000	163.000
Connect	Connectivity index, intersections/mi of road	5.798	1.531	0.000	16.232
Signal_P	Signalized intersections, %	0.123	0.141	0.000	1.333
BusStops	Total number of bus stops	13.104	9.099	0.000	75.000
LStops	Total number of l train stops	0.091	0.325	0.000	2.000
DVMT	Daily vehicle miles traveled	40,563.580	57,246.750	8.057	522,024.400
Ped	Pedestrian trips generated	47.715	103.345	1.191	1581.315
Bike	Bicyclist trips generated	2.511	5.439	0.062	83.227

where

θ_i = expected number of crashes for Census tract i;

β_0 = intercept;

β_i = coefficients quantifying the effect of the j explanatory variables characterizing Census tract i on θ_i;

Exp1 and Exp2 = measures of exposure in Census tract i;

x_i = a set of j explanatory variables that characterize Census tract i and influence θ_i; and

ε_i = disturbance term corresponding to Census tract i.

We compared four SPFs for vehicular, pedestrian, and bicyclist crashes aggregated on the Census tract level:

- Models based on CTPP data only, where exposure to road crashes is based on ACS commute trips;
- Models based on CTPP data fused with open data, where exposure to road crashes is based on ACS commute trips;
- Models based on CTPP data, including exposure characteristics from regional travel demand models; and
- Models based on CTPP data fused with open data, including exposure characteristics from regional travel demand models.

This process of model development resulted in a total of 12 safety evaluation models for all three crash types (vehicular, pedestrian, and bicyclist) for all crash severities. The process enabled us to compare the SPFs based on CTPP data only, to SPFs developed by combining CTPP data with data from regional transportation agencies and open data platforms. The purpose of this process was to ensure that the advantages and disadvantages of using only CTPP data for safety evaluation methods, and to demonstrate the variety of options that transportation agencies may use to develop their own SPFs depending on data availability and desired complexity and level of information required in road safety performance management process.

Results of the statistical modeling process are provided in Tables 4.2 to 4.4. Table 4.2 shows the results for SPFs developed to predict vehicular crashes on the Census tract level. The basic SPF developed using CTPP data to predict vehicular crashes, has the following form:

$$\text{Vehicular crashes}_i = e^{(5.55 + 3.64 \times 10^{-4} \times \text{Workers Driving} + 4.00 \times 10^{-6} \times \text{Income} - 2.99 \times 10^{-3} \times \text{Median Age})}$$

To further improve CTPP data-based SPFs, crash exposure variables from the CMAP's travel demand model were used to replace the commuter trips and see how this change in exposure data reflects on the statistical models. The resulting SPF obtained by including CMAP exposure measure of daily vehicle miles traveled (DVMT) for predicting vehicular crashes on the Census tract level is:

$$\text{Vehicular crashes}_i = e^{(0.23 + 0.54 \times \ln(\text{DVMT}) + 1.00 \times 10^{-6} \times \text{Income} - 1.33 \times 10^{-3} \times \text{Median Age})}$$

The SPFs can be developed using CTPP data only, in the absence of other data resources. However, the primary advantage of using DVMT as the measure of exposure in the developed SPFs is the possibility of making the assumption of the expected zero crashes in Census tracts where DVMT value is zero. Further, SPFs based on CTPP data only show that the expected increase in vehicular crashes is associated with the increase in exposure measures (workers driving or DVMT), increase in income per capita, and decrease in median age. Census tracts with higher average income are expected to have higher vehicle ownership, as cars would be more

TABLE 4.2 Vehicle-Only Statistical Crash Models

CTPP Data						CTPP Data with CMAP Exposure					
Coefficients:	Estimate	Std. Error	Z value	P value		Coefficients:	Estimate	Std. Error	Z value	P value	
Intercept	5.552000	0.144700	38.382	0.000	***	Intercept	0.234100	0.192200	1.218	0.223	
Workers Driving	0.000364	0.000042	8.601	0.000	***	ln(DVMT)	0.545300	0.017760	30.711	0.000	***
Income per Capita	0.000004	0.000001	2.995	0.003	**	Income per Capita	0.000000	0.000001	-0.099	0.921	
Median Age	-0.002998	0.004373	-0.686	0.493		Median Age	-0.001337	0.003114	0.429	0.668	
AIC	10921.63					AIC	10296.14				
CTPP Data + Open Data						CTPP Data + Open Data with CMAP Exposure					
Coefficients:	Estimate	Std. Error	Z value	P value		Coefficients:	Estimate	Std. Error	Z value	P value	
Intercept	4.726000	0.121800	38.790	0.000	***	Intercept	1.215000	0.227300	5.348	0.000	***
Workers Driving	0.000255	0.000033	7.848	0.000	***	ln(DVMT)	0.403200	0.022540	17.886	0.000	***
Income per Capita	0.000004	0.000001	3.807	0.000	***	Income per Capita	0.000000	0.000001	0.472	0.637	
Median Age	-0.005388	0.003269	-1.648	0.099	.	Median Age	-0.001696	0.002899	-0.585	0.559	
Arterial Network	0.249600	0.021620	11.547	0.000	***	Arterial Network	0.033860	0.022240	1.523	0.128	
Intersection Density	0.001062	0.000338	3.143	0.002	**	Intersection Density	0.000907	0.000302	3.006	0.003	**
Bus Stops	0.032590	0.002651	12.297	0.000	***	Bus Stops	0.025420	0.002394	10.618	0.000	***
AIC	10363.79					AIC	10167.14				

affordable in these areas, which could explain the estimated relationship between income per capita and the expected number of vehicular crashes. More complex SPFs developed by combining CTPP data with data from other transportation agencies and Chicago open data platform are also given in Table 4.1, using exposure expressed as the workers driving (from CTPP) and exposure expressed as the measured DVMT (from CMAP). The addition of open data makes the SPFs much more informative as it allows better prediction of the expected number of vehicular crashes on the Census tract level, through association with the increase of arterial roads mileage, intersection density, and the number of bus stops. Arterial roads are characterized by higher speeds than the local roads, and less uniform speeds than the freeways, which could explain the statistical significance of this variable in the SPF developed for predicting the expected number of vehicular crashes. The relationship between the intersection density and conflict points, as well as the presence of speed-changing behavior around bus stops, explains the association between these two variables and vehicular crashes.

Table 4.3 shows the results for SPFs developed to predict pedestrian crashes on the Census tract level. The basic SPF developed using CTPP data to predict pedestrian crashes, has the following form:

$$\text{Pedestrian crashes}_i = e^{(3.42 + 1.47 \times 10^{-4} \times \text{Workers Driving} + 1.76 \times 10^{-3} \times \text{Workers Walking} + \cdots)}$$

The resulting SPF obtained by including CMAP exposure measure of DVMT for predicting pedestrian crashes on the Census tract level is:

$$\text{Pedestrian crashes}_i = e^{(0.98 + 0.17 \times \ln(\text{DVMT}) + 0.41 \times \ln(\text{Ped.Trips}) + \cdots)}$$

TABLE 4.3 Pedestrian–Vehicle Crash Models

CTPP Data						CTPP Data with CMAP Exposure					
Coefficients:	Estimate	Std. Error	Z value	P value		Coefficients:	Estimate	Std. Error	Z value	P value	
Intercept	3.426000	0.290900	11.777	0.000	***	Intercept	0.989200	0.337000	2.936	0.003	**
Workers Driving	0.000147	0.000045	3.243	0.001	**	ln(DVMT)	0.172200	0.025360	6.789	0.000	***
Workers Walking	0.001764	0.000143	12.340	0.000	***	ln(Pedestrian Trips)	0.524300	0.035860	14.621	0.000	***
Income per Capita	-0.000007	0.000002	-4.111	0.000	***	Income per Capita	-0.000008	0.000001	-5.919	0.000	***
Male Population	-0.009973	0.005547	-1.798	0.072	.	Male Population	-0.024110	0.005168	-4.665	0.000	***
Median Age	-0.008215	0.004626	-1.776	0.076	.	Median Age	-0.010170	0.004208	-2.416	0.016	*
AIC	6008.585					AIC	5853.124				
CTPP Data + Open Data						CTPP Data + Open Data with CMAP Exposure					
Coefficients:	Estimate	Std. Error	Z value	P value		Coefficients:	Estimate	Std. Error	Z value	P value	
Intercept	3.130000	0.268100	11.672	0.000	***	Intercept	1.976000	0.340600	5.803	0.000	***
Workers Driving	0.000052	0.000043	1.206	0.228		ln(DVMT)	0.068300	0.028060	2.435	0.015	*
Workers Walking	0.001080	0.000134	8.075	0.000	***	ln(Pedestrian Trips)	0.418600	0.035870	11.670	0.000	***
Income per Capita	-0.000003	0.000001	-1.876	0.061	.	Income per Capita	-0.000006	0.000001	-4.329	0.000	***
Male Population	-0.011200	0.005093	-2.198	0.028	*	Male Population	-0.025210	0.004999	-5.044	0.000	***
Median Age	-0.016850	0.004265	-3.951	0.000	***	Median Age	-0.014870	0.004028	-3.691	0.000	***
Intersection Density	0.000890	0.000435	2.048	0.041	*	Intersection Density	0.001190	0.000416	2.858	0.004	**
L Train Stops	-0.165000	0.072360	-2.280	0.023	*	L Train Stops	-0.229600	0.070390	-3.262	0.001	**
Bus Stops	0.040490	0.002942	13.763	0.000	***	Bus Stops	0.029360	0.003189	9.208	0.000	***
AIC	5854.43					AIC	5773.997				

The SPF developed to predict pedestrian crashes based on CTPP data only uses drive and walk trips to work as exposure measure, with the disadvantage that commuter trips to work make less than one quarter of total daily trips for transportation users in major cities. The SPF estimated using the measured DVMT and pedestrian trips based on CMAP's travel demand model enables to predict pedestrian crashes with the assumption that Census tracts with zero vehicular trips or zero pedestrian trips are likely to have no crashes involving pedestrian, which is a more realistic model specification. Further addition of open data demonstrates that variables such as intersection density, L train stops, and bus stops are statistically significant in pedestrian crash SPFs. Intersection density and bus stops are associated with the increase in pedestrian crashes, due to higher exposure to pedestrian–vehicle conflicts in these areas. The L train stops, as low-speed areas, are associated with the decrease in the expected pedestrian crash frequency. The statistically significant variables in pedestrian crash models show areawide effects that influence pedestrian crash frequency, and capturing the impact of these variables on the corridor-level or intersection-level analysis would be challenging to capture. Whether the SPFs are based on CTPP data only, or CTPP data merged with open data, SPFs developed to predict crashes on the Census tract level in general prove to be informative for road safety managers in the manner that is complementary to microscopic-level statistical models.

Table 4.4 shows the results for SPFs developed to predict bicyclist crashes on the Census tract level. The basic SPF developed using CTPP data to predict bicyclist crashes, has the following form:

$$\text{Bicyclist crashes}_i = e^{(-0.02+4.36\times10^{-4}\times\text{Workers Driving}+5.26\times10^{-2}\times\text{Workers Biking}+\cdots)}$$

The resulting SPF obtained by including CMAP exposure measure of DVMT for predicting bicyclist crashes on the Census tract level is:

$$\text{Bicyclist crashes}_i = e^{(-2.77+0.27\times\ln(\text{DVMT})+0.50\times\ln(\text{Bike Trips})+\cdots)}$$

TABLE 4.4 Bicyclist–Vehicle Crash Models

CTPP Data						CTPP Data with CMAP Exposure					
Coefficients:	Estimate	Std. Error	Z value	P value		Coefficients:	Estimate	Std. Error	Z value	P value	
Intercept	-0.021680	0.394100	-0.055	0.956		Intercept	-2.778000	0.412500	-6.734	0.000	***
Workers Driving	0.000436	0.000049	8.920	0.000	***	ln(DVMT)	0.277500	0.026650	10.411	0.000	***
Workers Biking	0.052640	0.010440	5.044	0.000	***	ln(Bike Trips)	0.506700	0.042970	11.794	0.000	***
Income per Capita	0.000018	0.000001	11.907	0.000	***	Income per Capita	0.000010	0.000001	7.296	0.000	***
Male Population	0.046220	0.007155	6.460	0.000	***	Male Population	0.047710	0.006134	7.779	0.000	***
Median Age	-0.034730	0.005338	-6.506	0.000	***	Median Age	-0.025950	0.004781	-5.429	0.000	***
AIC	4919.062					AIC	4719.006				
CTPP Data + Open Data						**CTPP Data + Open Data with CMAP Exposure**					
Coefficients:	Estimate	Std. Error	Z value	P value		Coefficients:	Estimate	Std. Error	Z value	P value	
Intercept	-0.535500	0.364800	-1.468	0.142		Intercept	-2.200000	0.419800	-5.240	0.000	***
Workers Driving	0.000312	0.000045	7.013	0.000	***	ln(DVMT)	0.177900	0.029970	5.937	0.000	***
Workers Biking	0.045400	0.009210	4.929	0.000	***	ln(Bike Trips)	0.423600	0.043310	9.781	0.000	***
Income per Capita	0.000015	0.000001	11.176	0.000	***	Income per Capita	0.000010	0.000001	6.894	0.000	***
Male Population	0.044540	0.006474	6.880	0.000	***	Male Population	0.045690	0.006099	7.492	0.000	***
Median Age	-0.037880	0.004829	-7.843	0.000	***	Median Age	-0.028440	0.004688	-6.066	0.000	***
Intersection Density	0.002053	0.000440	4.666	0.000	***	Intersection Density	0.002344	0.000424	5.534	0.000	***
Bus Stops	0.028030	0.003338	8.397	0.000	***	Bus Stops	0.011130	0.003662	3.039	0.002	**
Bike Lanes	0.230700	0.041410	5.571	0.000	***	Bike Lanes	0.201900	0.040200	5.021	0.000	***
AIC	4712.349					AIC	4655.8				

Similar to vehicular crash and pedestrian crash models, the SPF for predicting the expected number of bicyclist crashes can be estimated using CTPP data only, and the number of driving and biking commute trips to work as the measure of exposure. When CMAP exposure measures are incorporated in the SPFs for predicting bicyclist crashes, the assumption holds that no bicyclist crashes are expected to occur in Census tracts where either DVMT or the number of bike trips have the value of zero. After fusing CTPP data with open data, additional variables show statistical significance in SPF specification for the expected number of bicyclist crashes: intersection density, bus stops, and bike lanes. In this case, bike lanes serve as a proxy for bicyclist exposure to crashes, so this variable should not be interpreted as the cause for the increase in bicyclist crashes. The presence of bike lanes may be associated with higher volumes of bike traffic, however, bike traffic is expected to be present on the parts of the roadway network where bike lanes are unavailable, so including this variable is a form of a surrogate for bike miles traveled on the Census tract level. Just as in the case of pedestrian crashes, these systemwide effects can be captured easily as the analysis is conducted on the Census tract level. Four SPFs for predicting bicyclist crashes provided in Table 4 demonstrate how combining CTPP data with other data resources can provide relevant information about the expected crash frequency due to investment in multimodal infrastructure.

Further comparison of the developed SPFs for vehicles, pedestrians, and bicyclists can be conducted based on the results from the tables. The CTPP-based variables that serve as the measures of exposure for these three modes of transportation include: workers driving, workers walking, and workers biking. Although the work commute trips represent only a portion of total trips in each Census tracts, in the absence of other exposure variables, commute trips can still provide logical relationship between the increase in travel demand and increase in road crashes. These exposure variables proved to be statistically significant in all three CTPP data-based SPFs. It was important to explore whether some other socioeconomic variables coming from CTPP datasets may be used to estimate the number of crashes for various transportation users. Median age in Census tract is associated with decrease in vehicle only, pedestrian, and bicyclist crashes. This finding could be the consequence of lower driving populations in Census tracts with higher

percentage of seniors. Variable describing gender (percent of male population) was not statistically significant in vehicle crash models, while it was negatively associated with pedestrian crashes, and positively associated with bicyclist crashes. This could lead to further exploration of the expected vulnerability levels of vehicle–pedestrian crash victims, or recently explored gender gap in biking studies. The authors however emphasize that further research is required before gender-related variables are used to develop Census tract level SPFs, and that the presented models are stable even without these variables. Another important socio-economic characteristic presenting income per capita, was associated with increase in vehicle-only crashes, potentially due to higher level of driving affordability in Census tracts with higher income. Similar is the finding for bicyclist crashes and it could be explained by higher investments in biking infrastructure in higher-income neighborhoods. The income per capita is associated with decrease in pedestrian crashes, potentially indicating that people are more likely to walk in lower-income neighborhoods.

Statistical models are validated using bootstrapping method. Ordinary nonparametric bootstrapping allowed to fit the model repeatedly by selecting data subsets randomly with replacement (*8*). The bootstrapping was conducted for 2,000 resamplings of the given dataset. After reaching the final model specifications the model goodness-of-fit is assessed using the Akaike information criterion (AIC), calculated as:

$$AIC = 2k - 2\ln(\hat{L})$$

Where k is the number of estimated parameters in the model, while $\hat{L}$ is the maximized value of the likelihood function of the estimated model.

The comparison of SPFs that use commuter trips as exposure variables and SPFs that use travel demand estimates as exposure variables show a better model with a lower AIC value for CTPP data-based model with CMAP exposure. The new model that uses CMAP exposure information also shows that the SPFs that are based on CTPP data only tend to overestimate the association of socioeconomic variables with estimated crash frequencies in the case of vehicle-only crashes. This could indicate that socioeconomic variables are more influential in Census tracts with higher percentage of population (including workers) walking and biking.

As previously indicated, some transportation agencies will have only CTPP data at their disposal, while others have more extensive transportation data availability, including open data platforms. The SPFs for vehicle-only, pedestrian, and bicyclist crashes were developed based on CTPP data combined with data from Chicago transportation agencies and city's open data platform. These SPFs were developed based on CTPP exposure variables and CMAP exposure variables. They show how characteristics of roadway network (e.g., presence of arterials and intersection density), and multimodal transportation infrastructure (bus stops, L-train stops, and bike lanes) are associated with multimodal crashes.

Applying Census Data for Mobility Evaluation

Performance metrics that describe mobility are developed to primarily indicate how congested the transportation system is. Metrics traditionally used to evaluate mobility–congestion level rely on the fundamental traffic flow theory characteristics, including volume, speed, and density (*10*). The resulting indicators of mobility usually are expressed as level of service and travel time. The main question these metrics are aiming to answer is how efficient travel is under a particular set

of areawide characteristics. In order to remain consistent with the safety evaluation metrics described in the previous section, mobility metrics in this section are also developed on the Census tract level, for the city of Chicago case study, using both CTPP data and data from alternative sources.

The simplest mobility metric that can be extracted from CTPP data is the average commuter travel time for each Census tract. Further, CTPP data provide the information on the mode of transportation used by workers in each Census tract. Figures 4.1 and 4.2 represent some of these mobility and quality of service indicators that can be developed by using CTPP data only or open data only.

It is expected that Census tracts with higher number of workers driving would have higher congestion and more limited mobility. As noted in Figure 4.1, the highest number of driving commuter trips comes from the very core of the city of Chicago, the Loop. In addition, broader ring of Census tracts surrounding the city's center form an area where travel time to work (WTT) seems to be lower than in other areas of the city. This implies that the areas with the highest share of driving trips are also the areas with the shortest commute time to work. Further, this could indicate that for those transportation users living in the city center, their origin–destination (O-D) distances are shorter when compared to the outside of the Loop area, and most of their daily transport needs can be met within a close proximity to their residences, which is due to a good mix of land uses in the city center. Based on Figure 4.1, higher congestion levels are present in the downtown area, and this issue was not completely resolved by multilevel transportation infrastructure solutions present in Chicago. In addition, the major congestion generator in the city of Chicago—the Loop—is an area characterized by a very extensive multimodal network, which could provide a viable alternative if driving limitations (e.g., congestion pricing ring) are implemented in the city core. A simple visualization based on CTPP data indicates workers mode share spatial distribution and Census tracts with the highest share of long-distance trips; however, it is challenging to assess citywide mobility in a more detailed manner using CTPP data only.
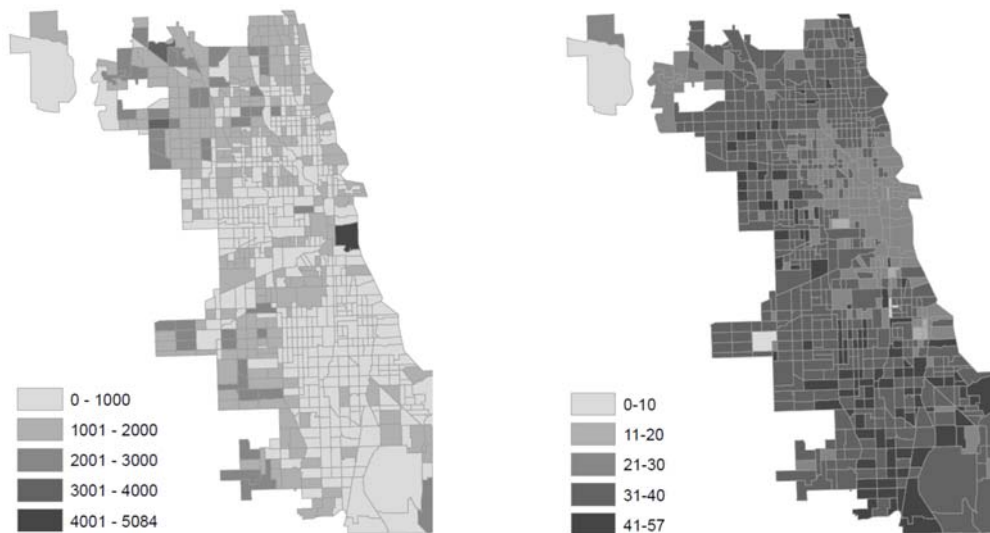
**FIGURE 4.1 CTPP-based mobility metrics:
number of workers driving (left) and average WTT (right).**
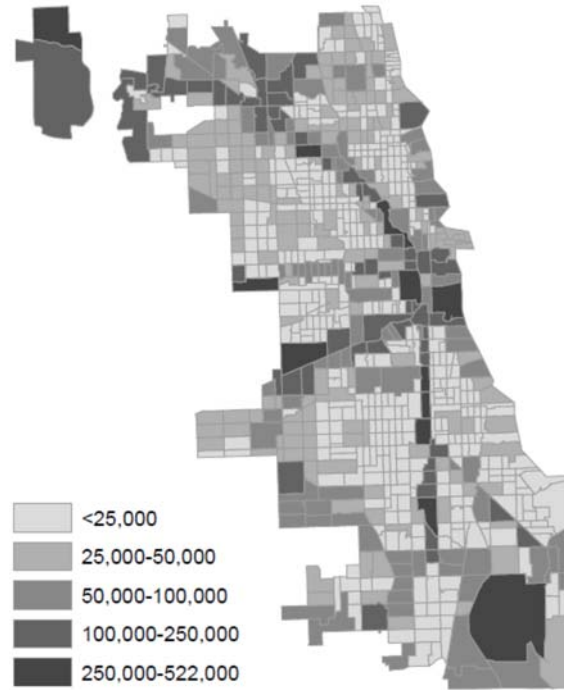
FIGURE 4.2 City open data-based mobility metrics: DVMT.

The open data from the city of Chicago can form their own indicators of mobility. For example CMAP and city's open data can be used to calculate DVMT, as shown in Figure 4.2.

The DVMT is calculated on the Census tract level, for each Census tract, by adding up the products of the average annual daily traffic volume (data available from CMAP) and their corresponding road segment lengths computed in ArcGIS:

$$\text{DVMT}_i = \sum_{ij} \text{AADT}_{ij} \times L_{ij}$$

where

DVMT_i = the total daily vehicle miles traveled (VMT) in Census tract i;
AADT_{ij} = the estimated annual average daily traffic (AADT) on road segment j within Census tract i; and
L_{ij} = the length of road segment j within Census tract i in miles.

The DVMT is a slightly better indicator of congestion that can be calculated from the available city data and CMAP data. Figure 4.2 shows how Census tracts with the highest congestion levels are those near the major freeway routes, including I-90, I-290, and I-55. These congested corridors are intersecting in the downtown area that shows the highest DVMT values in the city. The DVMT calculated on the Census tract level for the entire city of Chicago shows relatively balanced mobility services distribution throughout the entire city.

More comprehensive indicators of congestion–mobility can be derived if CTPP data are combined with alternative data sources. These combined metrics are derived from the traditional congestion measures such as travel time index and total delay, commonly used in Urban Mobility Report (*10*). These metrics refer primarily to working population within Census tracts, as CTPP data include mode share for work trips and average travel time for work trips. Travel time index is the ratio between the average peak hour travel time and the free-flow travel time in the observed roadway network (*10*). Here this index is adjusted to measure the commuter travel time index (CTTI) as the ratio between the WTT by a specific mode (e.g., drive or transit) and the total WTT in each Census tract:

$$\text{CTTI}_i = \frac{\text{TTI}_{ij}}{\text{WTT}_i}$$

where:

TTI_{ij} = the average travel time to work for Census tract i and mode j; and
WTT_i = total average travel time to work in Census tract i.

Figure 4.3 shows the results for the calculated CTTI on the Census tract level, for "drive alone" mode in Chicago. It should be noted that the fields valued as "zero" are the Census tracts where travel time data are currently unavailable for "drive alone" mode. The downtown area appears to have less competitive travel times by private vehicles when compared to other modes such as public transit, indicating that transportation users could be more likely to select other modes over private vehicle.

Legend:
0.00
0.00 – 1.25
1.26 – 1.50
1.51 – 1.75
> 1.75
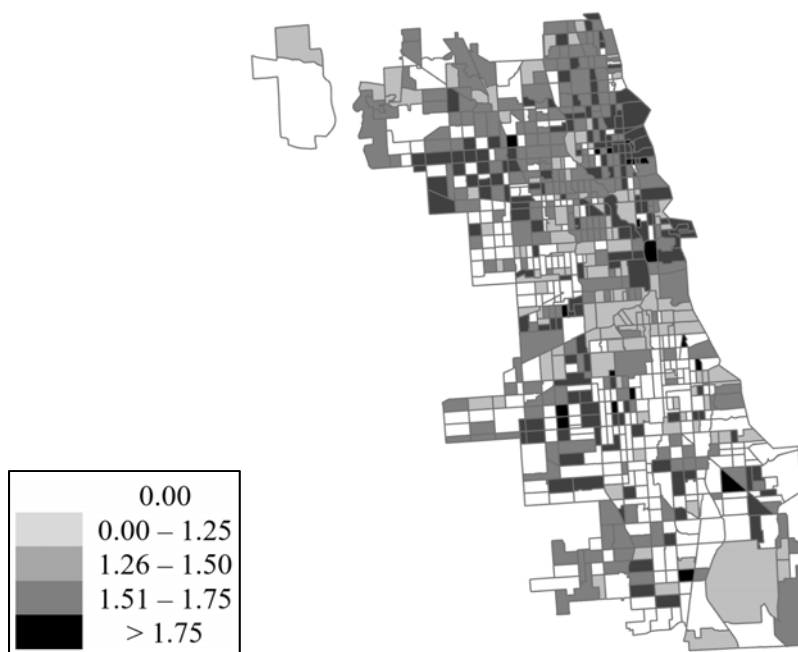
FIGURE 4.3 Travel time by car ("drive alone" mode) relative to the WTT.

Figure 4.4 shows the results for the calculated CTTI on the Census tract level, for "public transit" mode in Chicago. It should be noted that the fields valued as "zero" are the Census tracts where travel time data are currently unavailable for public transit mode. When compared to the average travel time for those Census tracts where transit travel time data are available there is a significant number of areas where traveling by transit almost doubles the commute time to work. The metrics presented in Figures 4.3 and 4.4 show how competitive public transit is as a mode choice, when compared to private vehicles and all transportation modes together. The downtown area of Chicago appears to have the most efficient public transit services with the most competitive travel times when compared to other modes of travel. The city's open data on multimodal infrastructure can be used to calculate the percentage of multimodal street network as an indicator of quality of transit service, related to the accessibility metrics presented in the following section.

Applying Census Data for Accessibility Evaluation

Accessibility is a relatively new addition to the current transportation performance measurement efforts. It describes the ability to reach desired destinations within the given spatial and temporal constraints (*11*). While mobility as a transportation performance metric relates to users need to reduce travel time to desired destination by ensuring that at least one option of travel is available; accessibility relates to reaching as many destinations as possible while using all available modes of travel.

Mobility is prioritized in areas where land use and transportation are highly disintegrated, with residential areas very distant from opportunities such as jobs, schools, hospitals, and shopping centers. Accessibility is prioritized in mixed land use areas with multimodal transportation infrastructure, where trip origins are in relative proximity to trip destinations and transportation users have diverse travel options with the opportunity to meet a broad range of travel needs within relatively short amount of time.
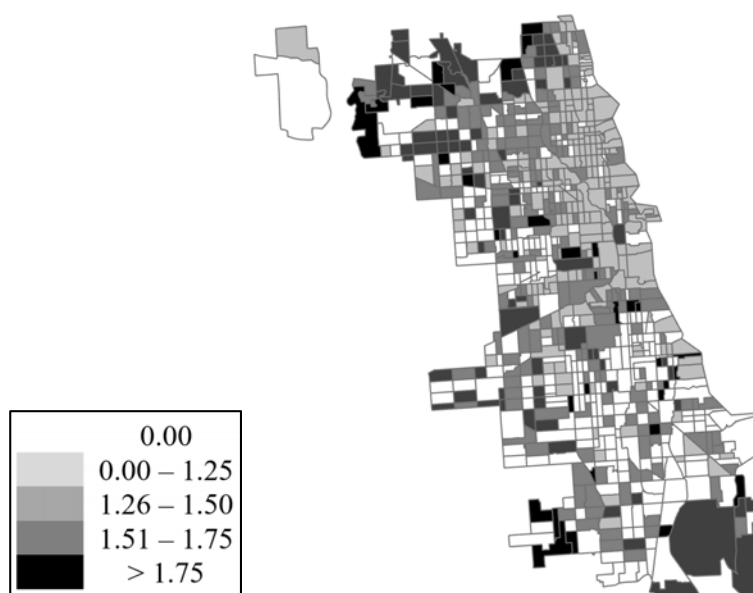
FIGURE 4.4 Travel time by public transit relative to the WTT.

Accessibility is considered when more-sustainable transportation solutions are incorporated in long-range transportation plans (*12*). Unlike mobility, which is essential for determining the capacity of the planned transportation network, accessibility is a measure crucial for spatiotemporal allocation of transportation resources while ensuring that freeways, transit lines, bike lanes, and sidewalks are layered in a manner that effectively connects transportation users to their trip destinations.

The first step towards evaluating accessibility using Chicago as a case study was to determine which transportation options are available and accessible on various parts of the entire transportation network. A simple network completeness analysis can provide this information by showing which network segments allow movements for all transportation user types (which segments can be considered "complete"). Figure 4.5 shows the percentage of roadway network that provides mobility opportunities for all four modes of transportation in Chicago (e.g., driving, transit, walking, and biking). The majority of the inner city area has more than 25% of street network that can be considered as "complete," while the very core of the city and some regions near lake Michigan have significantly higher presence of multimodal network when compared to the outer areas of the city. This further supports the findings related to mobility evaluation (Figure 4.1 from the previous section), which implies that the city center is the main car trip generator but with a high concentration of short car trips due to better land use mix that also contributes to better presence of the alternative and more sustainable modes of transportation. Based on the results from Figure 4.5, complete streets presence is higher along the major public transit (rail) corridors, indicating that areas around L-train stations facilitate access for pedestrians and bicyclists. As the distance from the city center increases, the presence of complete streets that provide access for all users decreases, as does accessibility to opportunities, which will further be discussed in this section.
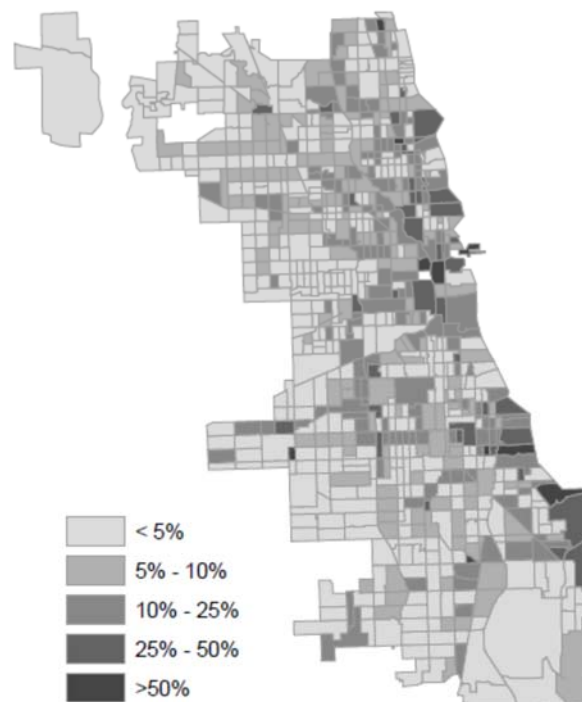
FIGURE 4.5 Network completeness.

Deriving accessibility measures based on CTPP data only would be challenging due to lack of information on multimodal infrastructure. For this particular metric, using CTPP as a standalone data source would not be a feasible solution. The combination of CTPP data and open data however, may result in more-comprehensive indicators of accessibility. A more-comprehensive review of different categories of accessibility measures may be found in the literature (*11*). For demonstration purposes in this study, CTPP data with the addition of open data are used to compute cumulative accessibility measures that indicate the total number of opportunities that may be reached by a specific mode of travel within the given timeframe.

The first step in cumulative accessibility analysis is identification of the potential origins and destinations. For this purpose, CMAP land use parcel-based inventory was used to define residential parcels as origins, and all other parcels as destinations (including the mixed land use parcels). In this manner, all Census tracts were split into purely residential parcels that represent the origins and other parcels that may be potential destinations that were coupled with CTPP data on socioeconomic characteristics. An example of an area from Chicago split into parcels is provided in Figure 4.6, showing how parcel-based separation increases the accuracy about the information on land use type and spatial coverage. The main limitation in this process is that while the information about the spatial allocation of opportunities is fully available (e.g., CTPP information on where jobs are located), the total number of opportunities within each parcel is not counted, and should be a subject of future research efforts.

The second step in accessibility calculation was to connect the defined origins and destinations by the existing transportation network links, using the information on roadway infrastructure, sidewalks, bike lanes, and transit lines. Multimodal infrastructure is overlaid on top of the defined origins and destinations to determine whether a feasible path by a specific mode exists between each origin and destination. If a feasible connection can be found the following third step is to compute travel time between each O-D pair. Using ArcGIS Network Analyst, travel time was computed for each mode of travel. In the case of public transit, only walking travel time was computed to the stations no further than 15-min walk from the defined origins. The final step of the analysis is a simple count of accessible

**FIGURE 4.6 Polygon-based (center) versus parcel-based (right)
land use inventory in Chicago (CMAP).**

destinations from each origin within the defined time budget (e.g., 5, 10,…120 min), and the summation of accessible destinations on the Census tract level.

The following general framework may be used to calculate cumulative accessibility measure for each travel mode of interest:

$$A_{ik} = \sum_i \sum_j \{d_{ij} \in N | T_{ij} \leq T\}$$

where

A_{ik} = total number of destinations accessible from origin i within time T, using model j;
 d_{ij} = destination j accessible from origin i within time T_{ij};
 N = total number of available destinations;
 T_{ij} = time needed to reach destination j from origin i;
  T = available time budget (5, 10, 15…120 min).

Figure 4.7 can serve as a simple example of cumulative accessibility calculations for public transit mode. If we assume that average transit speeds are available for each link in the example network, and that the link length is known, then travel time calculation for each link is computed simply as "link length/distance." Further, if we assume that node "1" is origin, while all other nodes are destinations, we can then compute the cumulative number of destinations reachable from node "1" within the defined travel time budget. Based on the information given in the example in Figure 4.7, the total number of destinations accessible from node "1" within the 30-min time budget is 5:

$$A_{1,transit} = d_{12} + d_{13} + d_{14} + d_{15} + d_{17}$$
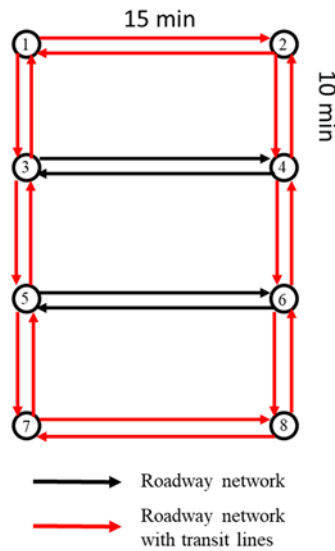
**FIGURE 4.7 An example of a simple roadway network for
cumulative accessibility calculations.**

Using the principle shown in the Figure 4.7, cumulative accessibility can be calculated on a large scale, for each Census tract in Chicago. A sample of this cumulative accessibility measure for destinations accessible by walking or transit within the 30-min travel time budget in Chicago is provided in Figure 4.8. Based on the results presented in Figure 4.8, the nature of the selected travel mode has a major influence on the overall destination accessibility. The other factors influencing accessibility include the availability of transportation infrastructure and its proper integration within the land use context. As the number of origins and destinations increases, the computational complexity related to travel time calculation also increases, and the resulting accessibility metrics may become more challenging to calculate. It is however important to provide indicators of accessibility whenever possible, particularly in long-range transportation planning, as they influence both land use and transportation policies.

Summary of Findings and Implications for Research and Practice

This paper focused on demonstrating how CTPP data can be used to develop and advance TPM, and what challenges may arise when developing TPM based on CTPP data only versus the possibilities that result from fusing CTPP data with other available transportation data sources. Three groups of measures—safety, mobility, and accessibility—were developed using CTPP data combined with alternative data sources with city of Chicago as the case study. These measures were presented for a broad range of transportation users, including private vehicle, public transit, pedestrians, and bicyclist users.

The results of the developed safety metrics show how the SPFs based on more-comprehensive datasets that combine CTPP data with alternative data sources outperform the SPFs that are based on CTPP data only. This is the case for all three examined crash types, showing the promising potential of harnessing data from multiple sources and platforms to
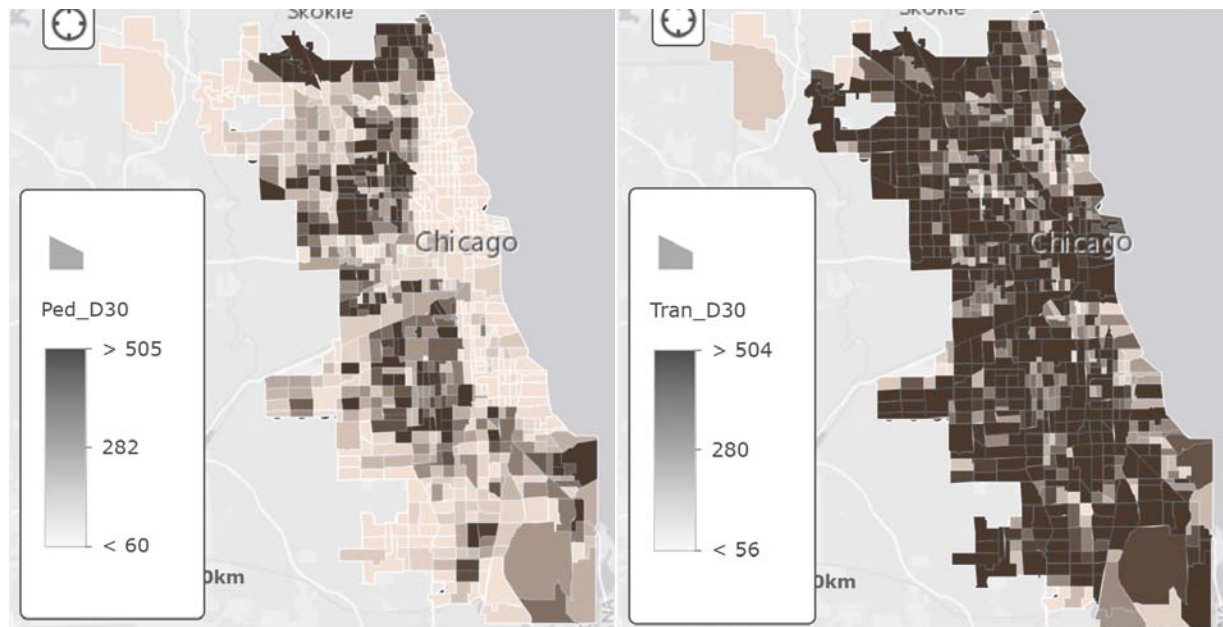
**FIGURE 4.8 Cumulative accessibility within 30 min by walk (left)
and public transit (right).**

improve crash predictions, and further investing transportation data infrastructure and multiagency collaborations.

Both simple mobility metrics (Figure 4.1) and combined mobility metrics (Figures 4.2–4.4) can be used to derive conclusions about transportation service efficiency. These metrics may serve as indicators of spatial allocation of mobility services for various modes of transportation, and reveal the hotspots where there could be a potential need to invest in operational improvements in order to achieve desired mode share. Combining CTPP and open data may serve to compute DVMT and travel time indices on the Census tract level. Data sources other than CTPP data currently provide better indications of mobility, but this may improve as CTPP flow data become available at a finer level.

The results of accessibility evaluation show how the exploration of accessibility measures can be combined with mobility metrics to inform practitioners about the overall availability of transportation service for various modes on the citywide level. These metrics are also calculated by fusing CTPP and open data, or data from transportation agencies. Accessibility evaluation clearly shows the distinction between modes of travel in terms of the ability to reach destinations of travel. Accessibility metrics can be expanded further beyond cumulative accessibility, to incorporate the weighting factors for destination attractiveness and spatiotemporal variations for different modes.

Table 4.5 provides a summary of developed performance metrics for each group of measures, including the supporting data used to develop these metrics, and whether these data can or cannot be obtained from the CTPP database. Based on the conducted analysis, the three groups of metrics developed here cannot rely on CTPP data only. In the case of safety performance measurement, the access to crash data is required in addition to CTPP data to develop the simplest form of SPFs. For mobility performance measurement, CTPP data can be used to derive conclusions about citywide mobility, however the metrics based on data combined from different sources are more informative when it comes to comparing different modes of transportation. The accessibility performance measurement requires multimodal network infrastructure data, and basic information about speeds for different modes, to enable the computation of accessible opportunities available from CTPP datasets. While the exposure information based on CTPP and used to develop safety metrics can be replaced by using exposure data from transportation agencies, the mobility and accessibility metrics rely on CTPP data-based information more strongly. The general role of CTPP data integration in all three groups of developed metrics remains significant, due to the fact that everyone can access and use CTPP data, which facilitates transferability of these metrics. Even though the city that served as a case study here (Chicago) is unique in terms of the broad range of transportation data sources, the metrics developed here, if relying on CTPP database, can be developed as long as information on crash data and transportation infrastructure is available to local and regional transportation agencies. This demonstration opens new possibilities for TPM development in regions with limited transportation data availability, particularly for the purpose of decision making related to long-range transportation planning for infrastructure investments.

Based on the summary given in Table 4.5, and the performance analysis conducted on the Census tract level, following conclusions and recommendations can be derived to guide researchers and practitioners attempting to use CTPP data for transportation performance measurement purposes:

**TABLE 4.5 Summary of Data for Performance Metric Development
from CTPP and Alternative Data Sources**

Performance Measurements	Required Data Input	Available from CTPP Database	Available from Other Data Sources
Safety	Crash data	No	Yes
	Exposure data	Yes	Yes
Mobility	Travel time by mode	Yes	Yes
	Travel demand by mode	Yes	Yes
Accessibility	Multimodal infrastructure	No	Yes
	Trip origins and destinations	Yes	Yes

1. Census tracts can be adequate units of analysis for the purpose of "big picture" performance management and analysis, as they provide compatibility between CTPP and other data sources, while the metrics developed on this macroscopic level are suitable for the purpose of long-range transportation planning.

2. The independent use of CTPP data for TPM development is not feasible. However, if crash data and transportation infrastructure data are available, all local and regional agencies can rely on CTPP database to develop Census-based transportation performance metrics, including safety, mobility, and accessibility evaluation for multimodal user types.

3. In the case of all three groups of performance measures (safety, mobility, and accessibility), combining CTPP data with alternative data sources is recommended whenever possible to advance the decision making based on the developed performance metrics.

The TPM process is improving as better data and methods become available to transportation practitioners. The CTPP data-based transportation performance measures are particularly important for transportation agencies in cities and regions where alternative data sources are still scarce or unreliable. The integration of CTPP data with the constantly improving transportation data sources and platforms has a promising potential to improve the efficiency and the quality of the decision making related to the investments in transportation infrastructure on all scales and in different environmental contexts.

References

1. Johnson, G., H. Scher, and T. Wittmann. Designing Shuttle Connections to Commuter Rail with Census Origin and Destination Data. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2534*, 2015, pp. 84–91. https://doi.org/10.3141/2534-11.
2. Marshall, N., and B. Grady. Sketch Transit Modeling Based on 2000 Census Data. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1986*, 2006, pp. 182–189. https://doi.org/10.3141/1986-24.
3. Kontokosta, C. E., and N. Johnson. Urban Phenology: Toward a Real-Time Census of the City Using Wi-Fi Data. *Computers, Environment and Urban Systems*, Vol. 64, pp. 144–153, 2017.
4. Jeon, C. M., A. A. Amekudzi, and R. L. Guensler. Sustainability Assessment at the Transportation Planning Level: Performance Measures and Indexes. *Transport Policy*, Vol. 25, pp. 10–21, 2013.
5. Naganathan, H., and W. K. Chong. Evaluation of State Sustainable Transportation Performances Using Sustainable Indicators. *Sustainable Cities and Society*, 2017. https://doi.org/10.1016/j.scs.2017.06.011.

6.   Transportation Performance Management. U.S. Department of Transportation, Federal Highway
     Administration. https://www.fhwa.dot.gov/tpm/. Accessed June 2017.
7.   Thornton, S. How Open Data Is Transforming Chicago. *Government Technology Magazine*, October
     2013. http://www.govtech.com/data/How-Open-Data-is-Transforming-Chicago.html.
8.   Hall, P. Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in
     Nonparametric Problems. *Journal of Multivariate Analysis*, Vol. 32, No. 2, pp. 177–203, 1990.
     https://doi.org/10.1016/0047-259X(90)90080-2.
9.   Hilbe, J. M. *Negative Binomial Regression*, Cambridge University Press, March 17, 2011.
10. Schrank, D., B. Eisele, T. Lomax, and J. Bak. 2015 Urban Mobility Scorecard. Texas AandM
     Transportation Institute and INRIX, August 2015.
11. Tasic, I., X. Zhou, and M. Zlatkovic. Use of Spatiotemporal Constraints to Quantify Transit
     Accessibility. *Transportation Research Record: Journal of the Transportation Research Board, No.
     2417*, 2014, pp. 130–138. https://doi.org/10.3141/2417-14.
12. Handy, S. Accessibility vs. Mobility—Enhancing Strategies for addressing Automobile Dependence
     in the U.S. Institute of Transportation Studies, University of California, Davis, 2002.

Facilitated Discussion

*Innovative Approaches: How are others using CTPP for Transportation Performance
Management? Which performance areas and measures? Are practitioners combining other data
with CTPP? How? What new innovative data sources are being used?*

Discussion around other datasets focused on using the National Performance Management
Research Dataset (NPMRDS), the Highway Performance Monitoring System (HPMS) in
combination with CTPP data. Incorporating pavement condition to the safety measure would
provide more information. To create meaningful measures for transportation agencies, which can
also be used/communicated at the public and policy level, the audience responses generally
pointed towards the fact that measures were inherently complex and thus hard to explain to those
outside the agency.

　　　When the audience was asked if they could replicate this methodology to produce these
measures in their agencies, no one said they could. Getting other data to combine with the CTPP
is an issue nationwide. Most areas do not have access to a dataset like the CMAP.

　　　When the audience was asked if anyone was using Census data in performance
management, no one said they were. Jim Hubble, MARC, shared how they use ACS for these
performance measures:

- Mode share (percent of people not driving alone);
- Average WTT if driving alone; and
- WTT when using transit.

　　　When the audience was asked if anyone is combining data sets, a representative from a
research center said his organization is looking at bike share data and demographics, which is
especially useful if the bike share stations are next to transit stations.

　　　When the audience was asked if anyone is looking at third-party data, a representative
from an MPO said that her organization is looking at Street Light data, but they don't have the
mega computer needed to crunch and analyze it.

Data Concerns: What limits use of CTPP/Census data for performance measures? What are the pitfalls or limitations to combining CTPP with other data sources? What impediments exist to implementing demonstrated measures immediately?

Individual audience members said that there were impediments to highly technical approaches to mobility measures resulting in decision makers not being able to make sense out of the process and outcomes. There are skillset challenges and a need for better software. It was pointed out by one audience member that Chicago has a very progressive open data program while other cities, especially small cities, don't have adequate data. Several audience members thought that FHWA might be able to provide appropriate open data for calculating performance measures.

Some MPOs have used AirSage data, but their staff needs specialized knowledge to work with these datasets. Several audience members mentioned issues with attempting to use private sector data sources including the lack of MOE and metadata on how the data was collected or manipulated.

User Community Needs: What tools, trainings, or other resources can be developed by the CTPP Program to facilitate use of CTPP data for performance measures? What additional research would help advance the use of CTPP data for performance measures?

When asked about data concerns, audience members provided the following responses:

- CTPP data not updated annually;
- The use of ACS overlapping data sets was supposed to work and it doesn't; and
- Pitfall of combining CTPP with other data is temporal mismatching and spatial coverage.

When the audience members were asked about method of training, they thought that webinars, YouTube, workshops, in-person training sessions work and would want to look at incorporating the discussion of CTPP data in already existing webinars (like quarterly National Performance Management Research Data Set webinars, TRB data committees, etc.). The upcoming release of the e-Learning module for the CTPP applications will provide ideas for new uses. Users could benefit from using communication products such as Slack, webinars, and YouTube tutorials.

Audience Suggestions for the CTPP Oversight Board

- Explore what datasets are available to transportation agencies (not just at national level but at regional level). People are afraid of big data, asking: can it be trusted? Can my agency handle it (in terms of technical analysis skills and computational capacity)?
- Explore combining NPRMRDS, HPMS, and other national data sets with CTPP. There is a research project soon to be completed comparing the ACS data to the NPMRDS.
- Explore training modules on combining data and on how to use CTPP data with performance measures.
- Explore the use of Slack to train/communicate; it's an interactive user group forum tool.

CHAPTER 5

Poster Session

YUE KE
KONSTANTINA GKRITZA
Purdue University

DANIEL RODRIGUEZ ROMAN
University of Puerto Rico, Mayaguez

SARA HERNANDEZ
University of Arkansas

MARIO SCOTT
MEGAN BROCK
Steer Davies Gleave

MATTHEW AIROLA
JIM GREEN
Westat Inc

RANJANI PRABHAKAR
Independent Consultant

STEPHANIE DOCK
District Department of Transportation

GREGORY NEWMARK
Kansas State University

LIYANG FENG
SAIMA MASUD
Southeast Michigan Council of Governments

MARKETA VAVROVA
University of Texas at El Paso

MICHAEL MEDINA
El Paso Metropolitan Planning Organization

MELISSA GROSS
CLAUDIA PASKAUSKAS
InNovo Partners

YOHAN CHANG
PROVEEN EDARA
University of Missouri, Columbia

LAURA SCHEWEL
StreetLight Data

PHIL LASLEY
Texas A&M Transportation Institute

MARK FOLEN
North Central Texas Council of Governments

PETER HAAS
Center for Neighborhood Technology Organization

CONSPICUOUS CONSUMPTION: GEOSPATIAL TRENDS IN VEHICLE TYPE CHOICE AND TRAVEL BEHAVIOR
Yue Ke and Konstantina Gkrtiza

In consumer economics, conspicuous consumption refers to the practice of purchasing expensive items to demonstrate wealth or power rather than to cover the real needs of a consumer. The theory states that consumers behave in such a manner so as to maintain or gain social status, which in turn causes others to emulate their behavior in order to maintain their respective social statuses. Colloquially known as "keeping up with the Joneses," conspicuous consumption is frequently associated with goods such as luxury yachts and imported sports cars.

This research applies the theory of conspicuous consumption to examine household vehicle type ownership and VMT demand. In this context, vehicle ownership type refers to both vehicle chassis configuration (e.g., sedan, pickup, SUV, etc.) as well as fuel type used (i.e., gasoline, diesel, hybrid, electric). Using data from the Census Bureau's 2015 5-year ACS and the Oregon Department of Environmental Quality, exploratory analysis using Moran's I revealed a significant spatial clustering of both VMT demand and vehicle type ownership. Spatial econometric models were then developed to demonstrate the extent to which a consumer's neighbor's vehicle type ownership and travel behaviors can influence the consumer's respective decisions in Portland, Oregon, and its surrounding urban areas. Findings include evidence of a positive and significant spatial lag in VMT demand, indicating that an increase in a household's neighbors' VMT induces the household to drive more. Further, the research indicates a similar trend in hybrid and electric vehicle ownership, which suggests that households may be more willing to own these types of vehicles if they see that their neighbors have them. The results of this study are useful to planners interested in understanding the adoption rates of new vehicle technologies (e.g., autonomous vehicles) as well as help guide policy makers in regulating the use of such emerging technologies.

ESTIMATING PARATRANSIT DEMAND MODELS USING ACS DISABILITY AND INCOME DATA
Daniel Rodriguez Roman and Sarah Hernandez

Travel demand models are useful tools for paratransit system planning in the face of changing demographics. Unfortunately, developing travel demand models that account for demographic information can be challenging for transit agencies that do not have the resources to obtain the travel behavior data needed to estimate these types of models. In response to this problem, a method is presented to fit paratransit demand models using disability and income data from publicly available data sources, such as the ACS, in addition to ridership data collected by transit agencies and general travel behavior information available from technical publications or derived from a transit agency's in-house knowledge. The latter is used as prior information that anchors the value of the model parameters during the fitting procedure (Figure 5.1).

The model fitting process is affected by the uncertainty associated with the available input data. Fortunately, the ACS and other U.S. Census data products report the margin of errors associated with their population estimates, and this information can be incorporated in model fitting procedures. To this end, a sample average approximation (SAA) approach was proposed that accounts for input data uncertainty. In the SAA approach a series of demographic data scenarios are generated based on the given uncertainty information. Then, a regularized least squares problem (RLS) is solved which minimizes an average error measure computed on the basis of the scenarios.

The demand models fitted with the proposed methodology can be used to forecast paratransit ridership given population projections or expected changes in system attributes. In this work, data from the Ozark Regional Transit (ORT) paratransit service and the population in its service area (Washington and Benton counties in Arkansas) were used to illustrate the application of the proposed procedures. Base-year data of the population with disabilities and their average income was obtained from the ACS. This base-year data was projected into year

A regularized least squares approach is proposed to fit paratransit demand models. Regularization offers a formal approach for incorporating prior information on the values of model parameters into the model fitting process. The prior information can originate, for example, from the professional experience of an agency's analysts or from the studies and models of other regions. In the proposed approach, it is assumed that there is a series of I paratransit demand observations ($\{y_1, y_2, \ldots, y_I\}$). The demand model produces ridership predictions ($\hat{y}_i$) given a set of parameters ($\boldsymbol{\theta}$) and explanatory variables ($\boldsymbol{X}$). For example, the prediction model can be specified as $\hat{y}_i = \alpha_i \sum_{n=1}^{N} \theta_n X_{in}$, where X_{in} is the population of group n and α_i is a factor that scales the ridership prediction to the desired time period of analysis. Representing the prior parameter information by $\tilde{\boldsymbol{\theta}}$, the model parameters are estimated by solving the following constrained, regularized least squares (RLS) minimization problem:

$$\min_{\boldsymbol{\theta}} L = \frac{1}{I} \sum_{i=1}^{I} \left(y_i - \hat{y}_i(\boldsymbol{\theta})\right)^2 + \sum_{n=1}^{N} \lambda_n \left(\tilde{\theta}_n - \theta_n\right)^2 \tag{1}$$

$$subject\ to$$

$$\theta_{n,min} \leq \theta_n \leq \theta_{n,max} \quad \forall\, n \tag{1.1}$$

FIGURE 5.1 Equations.

2030 using information obtained from county-level projections. The RLS and SAA–RLS models were fitted using monthly ridership information reported by ORT as demand data. Both models suggest that by 2030 the ridership for ORT's paratransit service will increase by around 36%, relative to 2016 values. This illustrative application shows how paratransit systems operators, using data from the ACS, can develop inexpensive quantitative models that can be used to guide their planning activities.

TRAVEL MODEL VALIDATION USING CTPP, HOUSEHOLD SURVEY, AND BIG DATA

Liyang Feng and Saima Masud

Two data sources are commonly used for travel model estimation, calibration and validation: HH travel survey and CTPP JTW data. In Southeast Michigan Council of Governments (SEMCOG), the household travel survey data has been used as primary sources for model development while the CTPP data is used as verification. Recently, SEMCOG started to explore big data options as another source of model verification. SEMCOG model verification uses a top-down approach. The approach looks regional patterns first, county level second, along with corridor level calibrations. This presentation focuses on high-level travel flow pattern and it provides a most important big picture of a travel model performance. Two HH surveys were conducted in years 2005 and 2015, and these data was used to develop SEMCOG's two sets travel models, E6 in the past and E7 for current, accordingly. For the CTPP, the county to county flow table was based on 2006–2010 ACS 5-year data.

To conduct a meaningful comparison, a set of tests is designed to compare travel model output, HH survey travel pattern and CTPP flows. Due to the nature of CTPP, only home-based work (HBW) travels were involved in tests. Percentage root mean square error and its related measures were used. The verification test found that the home to work travel patterns modeled in SEMCOG model were generally in line with the revealed flow patterns from both CTPP and HH survey, and the differences among any two of these three sources (survey, CTPP, and model)

were almost the same. Although the comparisons were conducted only for HBW trips, this encouraging result from the tests provided an extra confidence for other trip purposes from the model, as same approaches were used in model estimation process for other trip purposes. To enhance this study, the upcoming 2012–2016 ACS-based CTPP data has been considered to join the analysis, if time permits.

A CASE STUDY MEASURING THE EFFECT OF THE MARGIN OF ERROR IN CTPP DATA ON TRANSIT BUSINESS PLANNING
Mario Scott and Megan Brock

Cities across the United States are working to expand their transit offerings to better serve their commuters and visitors alike. Census data on commutation patterns can be extraordinarily valuable in assessing new transit service, and is at the base of many travel demand forecasting models. Steer Davies Gleave (SDG) frequently uses JTW data from the CTPP to estimate ridership for transit and intercity rail projects. In our experience using Census data for transportation projects, MOE are generally ignored in travel demand forecasting. In the Seattle region, SDG prepared passenger ferry ridership forecasts as part of a team developing a passenger ferry implementation and business plans for Kitsap Transit. JTW data from the 2006–2010 CTPP played a key role in forecasting demand and ridership. The flow data used was both at the county and tract level.

As is typical in prudent business planning, the ridership forecasts which drive the service's revenue are only assumed to realize a portion of the forecasted revenue. This is done to produce a viable business plan while acknowledging that the ridership models may have issues including: growth assumptions that do not materialize, erroneous input data, or faulty model assumptions. However, it is not clear that this would sufficiently insulate the business plan against the full MOE in the CTPP data and their effect on the ridership and revenue forecasts. A case study tests the effects of the margin of error on the business plan outcomes. The effects are tested by running the ridership models with varying levels of base JTW demand. The results are summarized by identifying the amount the base data must vary—positively or negatively—to result in different outcomes in the business plan for the planned transit services.

UTILIZING LEHD DATA IN JOB ACCESSIBILITY ESTIMATION
Ryan Westrom and Stephanie Dock

District of Columbia DOT (DDOT) is pursuing multiple explorations of accessibility within the District of Columbia. This research supports efforts to better estimate urban trip generation, expected parking utilization, and overall District mobility. In developing an accessibility metric that measures accessibility to jobs via various modes of transportation from any point in the District, choices in regards to the employment data sources needed to be made. DDOT explored use of multiple sources for such data, including: InfoUSA, the Metropolitan Washington Council of Governments (MWCOG), and the LEHD database from the U.S. Census Bureau. Ultimately, none of these sources were deemed perfectly accurate, and each had their downsides. For instance, the LEHD dataset underrepresents some federal or military jobs, which uniquely affects its accuracy in the District. Based on this, and seeking a more-accurate employment dataset, DDOT developed a unique methodology to create their own custom employment database at the

block level. This methodology was developed utilizing both the MWCOG and LEHD datasets for a more accurate picture of employment. This effort illustrates the importance of robust underlying data in accessibility measurement. As other cities and regions pursue this work, an understanding of the reliability of their employment data is vita. With use of the custom dataset, DDOT has been able to more accurately assess accessibility and create custom models and tools that more closely reflect reality.

PREDICTING VMT FROM PUMA DATA
Gregory Newmark and Peter Haas

Historically, the course and aggregate reporting of Census data has limited the direct application of nuanced travel models derived from disaggregated HH survey data. The emergence of PUMS data substantially alters this picture, because the PUMS data set preserves individual HH records. This research demonstrates the analytical and policy-making potential enabled by PUMS by applying a VMT model estimated on the California HH Travel Survey to HHs in metropolitan PUMAs throughout the state. This research explores transportation planning questions previously considered beyond the pale of Census data to generate new tools for modeling travel behavior. At the same time, this research reveals challenges of using PUMS data as well as possible approaches to overcoming these challenges.

UTILIZING CENSUS DATA FOR ACTIVE TRANSPORTATION
Marketa Vavrova and Michael Medina

The El Paso MPO is currently in the process of implementing the Active Transportation System, following the recent designation of seven regionally significant segments connecting the metropolitan planning area, which consists of El Paso County, Texas; southern Dona Ana County, New Mexico; and a small portion of Otero County, New Mexico. As a part of the Alternative Transportation Systems planning process, the research illustrates the use of Census datasets including LEHD Origin–Destination Employment Statistics (LODES), 2011–2015 ACS, and the 2006–2010 CTPP. These datasets are used to provide insight about social and demographic characteristics along those segments, such as population and employment densities, worker flows within the region, as well as transportation disadvantaged populations, by connecting people to opportunities.

FROM TRAFFIC COUNTS TO EQUITY: THE POWER OF INTEGRATED BIG DATA AND THE CENSUS
Laura Schewel

Ensuring transportation equity and EJ in today's communities is a major challenge for many planners. Unfortunately, empirically determining the impact of plans on different demographic groups is out of reach for many government agencies due to the difficulty and expense of collecting enough empirical data to measure it. Integrating Census data sets with big data— defined as the location records created by mobile devices—opens up new possibilities for planners including how planners can use analytics that combine big data sources with Census

data to understand the travel behavior of different demographic groups. There are two different types of big data that are most valuable for transportation planners: navigation–GPS, and location-based services data from smartphones. At the same time, there are biases that transportation professionals should be aware of for each type of data, and how these data sets can be combined with Census data to create comprehensive, empirical, equity analyses.

VISION FOR APPLYING MACHINE LEARNING TO CENSUS AND TRANSPORTATION PLANNING DATA
Melissa Gross and Claudia Paskauskas

Our society is more connected than ever and the volume of data being generated is rapidly expanding creating big data. Big data can be utilized to reveal new insights and provide a deeper level of analysis for safety, overall performance, or system reliability. The Transportation Systems Management and Operations (TSM&O) Program utilizes these big data sets to actively manage the safety and efficiency of the vehicular traveling public, but can also be applied to the multimodal network in support of improved mobility. By harnessing the power of the "big data" available, the TSM&O program can improve the safety and efficiency of the multimodal network through strategic planning for and the active monitoring of the system. With technology advancements, besides the traditional mechanisms of data collection, programmatic data gathering can access shareable data pools allowing for further data extraction from APIs. Examples of dynamic data sets that can be gathered from APIs include Bluetooth, social media, sensors, and third-party proprietary data such as WAZE, HERE.com, INRIX, and crowdsourcing, just to name a few.

Accessing historic and live data provides information to identify recurring congestion patterns, such as rush-hour traffic, and support planning for nonrecurring congestions. The relationship to determine traffic flow and performance can assist engineers and planners to manage the system in a proactive real-time situation, provide better data to calculate the return on investment for future improvement projects, prioritize funding allocation, and identify operational changes needed to improve traffic volume and flow on roadways. Census data can be utilized as an input to enable predictive technologies under the umbrella of artificial intelligence and machine learning, providing insights to a variety of performance measurements for multimodal system planning. The TSM&O program will utilize this insight to support decision making for the transportation network and infrastructure in preparation for connected and automated vehicles. For example, land use data, and historical traffic data, when combined with Census data, could be used to predict future traffic patterns, thus support efficient decision making for the transportation infrastructure. This research introduces a vision for predictive technologies that could utilize Census data applied to transportation.

HYBRID ORIGIN-AND-DESTINATION TRIP MATRICES
ESTIMATION MODEL USING MACHINE LEARNING TECHNIQUES
Yohan Chang and Praveen Edara

The O-D trip matrix is an essential ingredient in a variety of transportation planning and analysis studies. The traditional O-D matrix estimation models used license plate surveys, home interviews, roadside surveys, etc., but these methods have a disadvantage in the view of cost-effectiveness. The O-D matrix estimation method from observed link counts has interested from many researches and agencies, and various methods have been proposed to directly obtain O-D matrices based on link counts.

This research proposes a hybrid O-D estimation approach , combining both a mathematical based model and a group of machine learning models, such as random forests (RF), neural networks, and deep neural networks. The St. Louis area is used as a case study site. Open street map data was used for the seed network to capture a capacity and a geometric information, the Census data was used for seed O-D matrix, and the AADT value was utilized for a calibration of the mathematical model. A set of generated link counts from the mathematical model transformed a set of training data to feed the machine learning models. The proposed hybrid approach showed that RF outperformed other models for three classes: 20% seed O-D matrix changes, 20% seed network changes, and both.

ROAD SEGMENT SAMPLING USAGE AND EVALUATION OF THE
CENSUS BUREAU'S TIGER
Matthew Airola and Jim Green

Westat and other research organizations that conduct observational transportation studies (e.g., on seat belt use, motorcycle helmet use, traffic speeds) have often used the U.S. Census Bureau's TIGER files as the source of a road segment sampling frame. TIGER is used as a sampling frame, meeting the three most important characteristics of a good sampling frame criteria. Findings from a number of studies with respect to these sampling frame characteristics are presented, in addition to procedures for processing TIGER for use in Westat's statistical software. A number of improvements upon the use of TIGER as a road segment sampling frame are provided, including suggestions for future research, improvements, and approaches.

WHY PEOPLE CHOOSE TO LIVE WHERE THEY DO, TRANSPORTATION'S
ROLE IN THAT DECISION, AND HOW DATA CAN INFORM POLICY
Phil Laskley

A new study by the Texas A&M Transportation Institute (TTI) examined why people choose to live where they do in Texas and how important transportation is to that housing location decision. Understanding how these decisions are made could enable stronger policy decisions that address traffic and mobility issues through nontransportation planning means, providing a synergistic benefit to both transportation and other community and planning issues. This research examines the methods and results of the Texas Realtors Survey, and how these results, access to Census data, and other data sources could be combined to identify potential areas where targeted nontransportation tweaks could improve mobility and access. These improvements may be in the

form of using underutilized infrastructure more effectively, providing new mode options, or attracting residents to areas for improved transit service efficiency. By looking at the transportation paradigm holistically, policy makers and planners will be empowered to improve mobility by not directly addressing transportation.

HIGH-RESOLUTION DEMOGRAPHIC FORECASTING: THE CONVERGENCE OF SOCIOECONOMIC AND REMOTE SENSING DATA FOR SMALL-AREA FORECASTING
Mark Folden

Considerable work has been done at using a variety of methods to predict the spatial organization of urban areas in the future. Gravity models, Markov chains, cellular automata, real estate theory, microsimulation, and other techniques have all been used with varying degrees of success. Additional work has been done using remote sensing data to detect and measure change in urban land cover. Until now, relatively little work has been done to couple remote sensing data to county-level totals of forecasted population and employment to generate small area demographic forecasts at zone sizes suitable for travel demand modelling.

North Central Texas Council of Governments (NCTCOG) has assembled a temporally dense time series of 30- x 30-m LANDSAT imagery along with cadastral-based land use data to detect urban land use–land cover change and mathematically couple it to county forecasts of population and employment. This serves as a basis for predicting the quantity of future urban areal expansion based on external control totals for each county. Small-area socioeconomic data, terrain and land cover descriptor data, and proximity variables are then used to create a large dataset facilitating choice modelling of 30- x 30-m cells that urbanized during the observed change period. The validated choice model is used along with locally adopted future land use plans to predict future urban expansion based on "building forward" from the most recent observation.

Dasymetric mapping allocates household and employment from polygon-based sources to the same 30- x 30-m grid system at a known point in time. This allocation can be spatially interpolated across unurbanized area; or be used with multiple variables to create another large dataset allowing for modeling of HH, HH population, and employment density surfaces for each land use type. Adjusting density surfaces by a factor of uncertainty, subject to locally adopted density restrictions, ensures output totals match exogenous control totals at the county level. 30- x 30-m outputs of households, employment, and land use by category facilitates zonal tabulations at any geography needed by forecast users.

Topics discussed include: data sources, data preparation methods, modeling techniques, software tools used, previews of data generated from preliminary model runs, how the method presented meets the forecasting needs of NCTCOG, and potential for creating a feedback loop with a travel demand model to create a true Integrated Transportation and Land Use Model.

Census Bureau Potpourri Part 1

PHILLIP SALOPEK
U.S. Census Bureau (retired), presiding

BRIAN MCKENZIE
VINCE OSIER
U.S. Census Bureau

T he first of two sessions focused on the various Census Bureau programs and divisions, this session highlighted the Social, Economic, and Housing Statistics Division, and the Geography Division.

COMMUTING PROGRAMS AND PRODUCTS FROM THE CENSUS BUREAU
Brian McKenzie

It is important to distinguish between the Decennial Census and the ACS. The Census is a count of everyone in the United States while the ACS uses approximately 3.54 million addresses per year. The Census is considered a U.S. population count with core demographic characteristics. The ACS provides estimates of the population and housing characteristics, including demographic, social, and economic statistics.

Commuting data is collected through the ACS, beginning in 2005. The ACS questions related to commuting are the same as those on the 2000 Census Long Form and both only pertain to work trips. The transportation information collected in the ACS program includes means of transportation; occupants per vehicle; time leaving home for work; WTT; place of work; and vehicles per HH. For example, Figure 6.1 displays how people traveled to work in 2016 by mode.

The ACS survey data is collected continuously and is available for small areas (block groups). It has maintained comparability across years and geographies and includes MOE for quality checks. It contains a rich set of demographic characteristics and has several ways to access the data. The ACS release schedule is as follows: ACS 2016 1-year estimates for geographies of at least 65,000 populations (September 14, 2017); Supplemental Tables for selected geographies of at least 20,000 populations, with 58 tables containing basic demographics (October 19, 2018); the 2016 PUMS file (October 19, 2018); and the ACS 2012–2016 5- to 7-year estimates (December 7, 2018). The data is available on American FactFinder (factfinder,Census.gov).

Metropolitan and Micropolitan Areas were updated with the 2010 Census definitions. These updates were included in the 2013 ACS estimates that were released in 2014. Metro areas are aggregations of counties. The 2006–2010 county-level community flow files served as the inputs for the metro–micro area delineation process. The next update of the metro and micro areas will be based on the 2011–2015 ACS. It is assumed that the central counties won't have many changes, but commuting patterns could change for the outlying counties included in a metro based on the following qualifying criteria: at least 25% of workers living in the county work in the central county
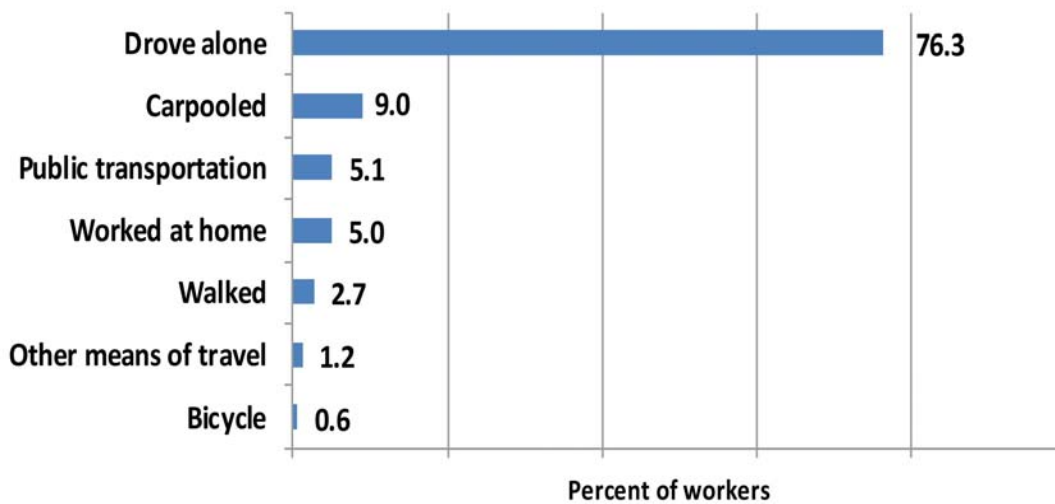
FIGURE 6.1 How people travel to work: 2016.

(Source: U.S. Census Bureau, 2016 ACS, Table S0801.)

or counties of the core-based statistical area (CBSA) and at least 25% of the employment in the county is accounted for by workers who reside in the central county or counties of the CBSA.

PUMS is a sample of population and housing unit records from the ACS. These data allow users to create custom tables that are not available through pretabulated ACS products. The smallest geographic summary level available is the PUMAs.

The ACS conducted two content tests to improve data collection and data use. The first was an update and clarification of the commute mode (Figure 6.2) and the second was a rephrasing of the time of departure question to address privacy concerns (Figure 6.3). The results of the two tests did not affect the overall response distribution, item missing data rate, or response reliability.

The LEHD is a program that uses administrative records information extracted from employers from across states. Key information about workers and firms includes worker counts; age; employment location; industry; and firm size. The LEHD also includes O-D worker flows at the block level as LODES. The data is available on a user-friendly website that offers several data extraction options.

Another data set that has transportation related information, but is often overlooked as a source for researchers and transportation planners, is the American Housing Survey (AHS). It is a longitudinal survey of housing units that is conducted every two years. The survey collects information on housing and neighborhood characteristics, including transportation and mobility. The tables are published for the nation, 15 metro areas, and 10 rotating metro areas with a sample size of 120,000 HHs.

New public transportation data is available for 2013 AHS, and new 2017 transportation data is forthcoming, with a release in 2018. Information is available at www.Census.gov /programs-surveys/ahs/. The 2017 AHS transportation-related questions include:

FIGURE 6.2 Time of departure questions test.

FIGURE 6.3 Commute mode question test.

- Public transportation use (specific modes);
- The frequency of public transportation use for work–school;
- Distance to closest public transportation stop;
- Access to amenities via public transportation;
- Biking and walking to work and amenities;
- Sidewalk availability and sidewalk lighting;
- Availability of bike lanes; and
- Costs associated with community.

Additional AHS new transportation-related questions will include:

- Number of commuting days each week;
- Number of days drives self all the way to work;
- Company car use;
- Drives self a portion of the way to work;
- Use of carpool;
- Roundtrip miles driven for commute;
- Cost of parking and tolls; and
- Use of public transportation for commute.

GEOGRAPHY DIVISION
Vince Osier

Statistical geographic areas are defined solely for data collection, tabulations, dissemination, and analysis, representing geographic concepts, such as urban, rural, and metropolitan and communities, localities, and other recognizable areas that do not have legally defined boundaries or are surrogates for legal entities. They provide a finer spatial resolution that is consistent and comparable over time for longitudinal analysis. Entities are designed to ensure statistical reliability and to protect confidentiality of the data. These designated statistical geographies are critical for analysis at lower levels of geography (i.e., place, neighborhood). They are used for a wide variety of federal programs including: Community Development Block Grants; Small Business Administration programs: rural development; rural health programs; and place-based planning and programs. Statistical geographies are often used by transportation and urban planners and policy makers.

Additional statistical geographies are those that lack legally defined boundaries including: zip code tabulation areas (ZCTAs); Census-designated places (CDPs or unincorporated places); tribal-designated statistical areas; and state-designated tribal statistical areas. These areas are defined specifically for data presentation and analysis for Census tracts, block groups, Census county divisions (CCDs), and PUMAs.

The process of obtaining delineation of boundaries includes publishing criteria, generally in the *Federal Register*. Boundaries often follow visible features and statistical areas may be aggregations of other geographic entities. The Geography Division partners with groups to develop and deploy new strategies to integrate the address list review program, street centerline update program, and boundary reporting programs, now existing as separate programs. They assist with improvements for accuracy, currency, and coverage of the Master Address File (MAF)–TIGER system. Partnerships include Geographic Support System Program; Boundary and Annexation Survey; LUCA; new construction; Participant Statistical Areas Program (PSAP); school districts' and Redistricting Data Program.

The PSAP is a decennial program that allows local participants, following Census Bureau criteria and guidelines, to review and suggest modifications to the inventory, boundaries, and names of statistical geographic areas. The 2020 PSAP geographies include Census tracts; block groups; CDPs; CCDs; and tribal statistical geographic areas. Figure 6.4 lists the general characteristics of PSAP entities. At this time, there are no changes to concepts or criteria from 2010.

CDPs began in 1940 as a supplementary report for unincorporated places, with a requirement of a population of at least 500. In 1950, unincorporated places were defined only outside of urbanized areas and were required to have a population of at least 1,000. From 1960 through 1990, CDPs were defined inside urbanized areas, with a minimum population threshold declining from 10,000 to 5,000, then to 2,500.

The outside urbanized areas had a threshold of at least 1,000. From 2000 forward, there is no minimum population threshold. Figure 6.5 displays an example of CDPs and Incorporated Places.

Type of Entity	Primary Purpose	Nationwide wall-to-wall coverage	Nest within
Census Tract	• Reliable small area data • Data comparability	Yes	County
Block Group	• Spatial resolution • Demographic characteristics	Yes	Census Tract
Census Designated Place (CDP)	• Demographic and economic data for unincorporated places	No-CDPs represent distinct communities	None
Census County Division (CCD)	• Subcounty units that have stable boundaries	21 states (nationwide coverage with MCDs)	County

FIGURE 6.4 General characteristics of PSAP entities.
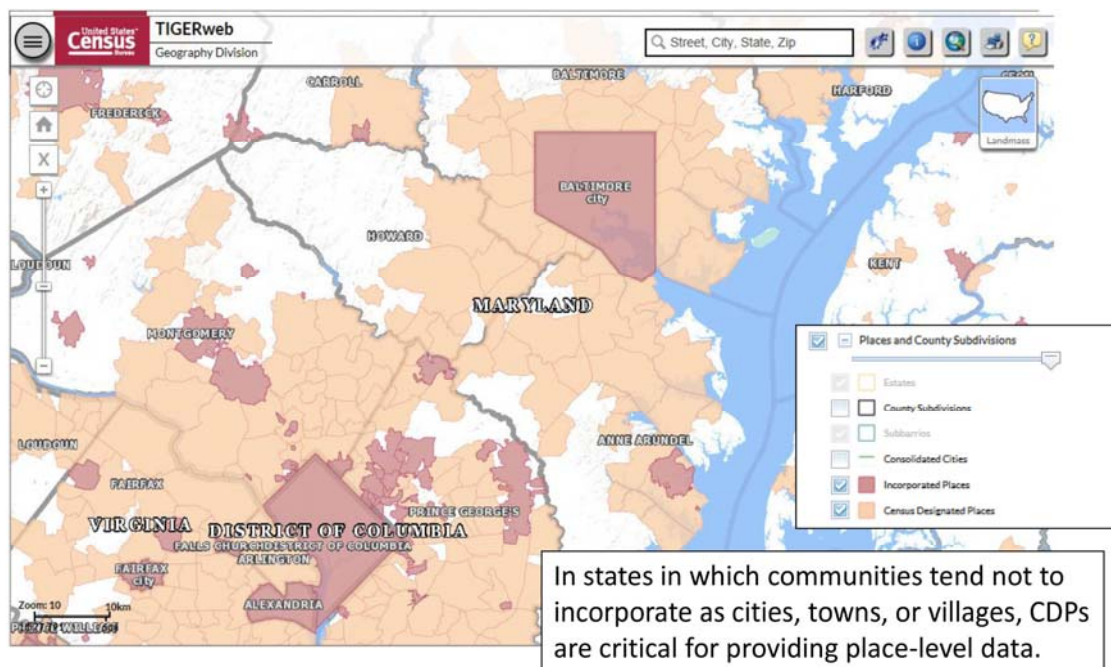
FIGURE 6.5 CDPs and incorporated places.

CHAPTER 7

Demographics, Equity, and Access

MICHAEL FRISCH
University of Missouri, presiding

BEN ETTELMAN
MAARIT MORAN
Texas A&M Transportation Institute

KIMBERLY KOREJKO
SHOSHANA AKINS
BENJAMIN GRUSWITZ
Delaware Valley Regional Planning Commission

REZA SARDARI
SHIMA HAMIDI
The University of Texas at Arlington

Census data is used in many ways. This session captured three of the more-specialized uses facing states and regional planning agencies throughout the country. These applications have broad applicability and the techniques can be transferred to other areas.

IDENTIFYING THE TRANSPORTATION NEEDS OF AGING TEXANS
Ben Ettelman and Maarit Moran

TTI conducted a policy research project for the Texas State Legislature to identify transportation solutions that promote healthy aging for the state's population. State demographers predict that the proportion of the Texas population that is age 65 years and older will continue to increase over the next 30 years. Given a similar pattern of in-migration that Texas experienced between 2000 and 2010, the proportion of the population in Texas that is age 65 and older is projected to increase from 11.5% in 2010 to 21% in 2050. In addition to developing a range of recommendations on innovative methods and best practices for meeting the transportation needs of the aging population in Texas, researchers used Census data to identify where within the state the aging population has the greatest mobility need. In order to accomplish this, researchers developed a mobility need index (MONI). The MONI utilized key demographic Census data that indicate high mobility needs, such as population aged 65 or older with a disability; population aged 65 or older with no household vehicles; and population 65 or older living in poverty. Researchers used these and other key data to develop indices, that when combined, provided a MONI for each Census tract and county throughout the state.

Researchers found that that the greatest absolute mobility need exists within all of the urban cores throughout the state of Texas. However, the suburban and rural geographies either

within or adjacent to Texas' metro areas have significant mobility need as well. Researchers highlight two key areas in their findings:

- In absolute numbers, Texas' urban cores exhibit the greatest mobility need for the aging population and considerations must be made to continue to find ways to provide accessible transportation options for Texans as their mobility needs increase.
- Suburban and rural areas, especially near urban cores, exhibit significant mobility need as well. In fact, these geographies have a larger proportion of mobility need as compared to their urban counterparts. Moreover, the aging population in these geographies are more at risk of suffering from a lack of mobility due to isolation and a lack of available transportation services.

This demographic analysis provides legislators with a better understanding of where mobility need exists for older adults in the state of Texas and could inform investment in future transportation services will be most effective and efficient.

LEVERAGING CENSUS DATA FOR MPO EQUITY ANALYSES

Kimberly Korejko, Shoshana Akins and Benjamin Gruswitz,

Title VI of the Civil Rights Act and the Executive Order on Environmental Justice (No. 12898) task agencies that receive federal funding to evaluate EJ and equity issues but do not provide specific guidance on how to complete this important task within a region's transportation planning process. Therefore, MPOs must devise their own methods for ensuring that EJ and equity issues are investigated and evaluated in transportation decision making. In 2001, Delaware Valley Regional Planning Commission (DVRPC) developed an EJ technical assessment to identify direct and disparate impacts of its plans, programs, and planning process on defined population groups in the Delaware Valley region. This assessment, the Indicators of Potential Disadvantage (IPD), formerly called the Degrees of Disadvantage Methodology, is used in a variety of DVRPC plans and programs.

DVRPC currently assesses the following population groups, defined by the U.S. Census Bureau: Non-Hispanic Minority, Carless Households, Households in Poverty, Female Head of Household with Child, Elderly (75 years and over), Hispanic, Limited English Proficiency, and Persons with a Physical Disability. Using ACS 5-year estimates, the demographic groups listed above's shares relative to their respective universes are calculated at the regional and the tract level. The regional share provides a threshold by which to evaluate the tract level shares. Any Census tract that meets or exceeds the threshold, is considered an EJ-sensitive tract for that IPD category. While using the IPD dataset over the past several years and as well as the ACS, DVRPC is striving for a data set that

- Better represents communities of concern identified in Civil Rights and EJ statutes;
- Employs a methodology that more clearly identifies those communities; and
- Responsibly communicates the reliability of the data sample.

TRANSIT ACCESSIBILITY AND THE SPATIAL MISMATCH BETWEEN JOBS AND LOW-INCOME RESDIENTS: EMPIRICAL FINDINGS IN THE DALLAS AREA

Reza Sardari and Shima Hamidi

Accessibility to public transit plays an important role in connecting residents to jobs and other opportunities, essential services, educational facilities, and recreational centers. The availability of transportation and housing choices can increase access to such areas and enhance economic growth. In a healthy society, residents in poverty should have access to jobs that help them improve their lives. Although the Dallas Area Rapid Transit (DART) light rail system is the longest light rail transit system in the nation, covering 90 mi with more than 60 stations across North Texas, transit ridership in Dallas is relatively low. The DART ranks 23 out of 39 large- and medium-sized transit agencies in the United States in terms of two transit ridership indicators: passenger miles per capita and passenger trips per capita (APTA, 2014).

Despite the vast literature on public transit and accessibility, practical research on the spatial mismatch hypothesis and job accessibility for low-income residents is missing from previous studies. Therefore, the primary objective is to analyze the patterns of low-wage job growth and the spatial clustering of residents below the poverty level. Using the General Transit Feed Specification (GTFS) combined with GIS spatial statistic tools, such as spatial autocorrelation and hot spot analysis, the equity and efficiency of public transit service in the Dallas area was investigated. Findings indicate a spatial mismatch between low-income groups and the location of low-wage jobs. According to the LEHD and Census data, poverty rates in Dallas city are considerably higher than in the suburbs while low-wage job growth has shifted from the city center to the suburbs without corresponding transit accessibility. LEHD annual job growth from 2002 to 2014,indicated low-wage jobs are growing outside the DART service area while low-income groups are concentrated in southern Dallas.

The findings of this research can help decision makers, transit agencies, and future transportation research by providing a comprehensive understanding of the geographic gap between low-income residents and the growth of low-wage jobs in the region. Finally, this research proposes a method to measure transit equity and compare public transit accessibility in urban areas using the LEHD, the GTFS feed, and ACS datasets.

CHAPTER 8

Census Bureau Potpourri Part 2

Phillip Salopek
U.S. Census Bureau (retired)

Matthew Graham
Ally Burleson-Gibson
U.S. Census Bureau

This was the second of two sessions focused on the various Census Bureau programs and division, including the Center for Economic Studies, the Center for Enterprise Dissemination Services and Consumer Innovation (CEDSCI), and the Decennial Communications Coordination Office.

CENTER FOR ECONOMIC STUDIES: LEHD PROGRAM
Matthew Graham

The LEHD uses the Local Employment Dynamics (LED) Partnership with States → State Unemployment Insurance (UI) and State Quarterly Census of Employment and Wages (QCEW) programs. The methodology uses a linked Employer–Employee Database. At the core of the process is administrative data. The program began in the late 1990s. The output is public-use data products on the workforce and the labor market. The LEHD Data Infrastructure concept is displayed in Figure 8.1.

Data products used in the process include the Quarterly Workforce Indicators data set. It has 32 indicators on employment, hiring, separations, and earnings. National indicators are in beta testing. The LODES provide an annual O-D dataset by firm and worker characteristics and detailed geography. The Job-to-Job Flows (J2J) provides quarterly statistics on transitions between jobs as well as transitions into and out of nonemployment.

Recent advances in these data sets include J2J, with 40 different measures of worker reallocation, data back to 2000 for some states, and firm and worker characteristics for national, state, and MSA tabulations. J2J recently has been expanded in scope. The J2J Explorer is a web-based analysis tool that enables comprehensive access to the J2J data in a dashboard interface. The beta version was released in June 2017.

New Pilots are being added through partnerships that include the University of Texas and the Colorado Department of Higher Education. The data is expected to fill major gaps in statistical infrastructure, particularly longitudinal earnings outcomes across state borders. The program will produce public-use statistics in 2018. The data products will include national earnings tabs by major and institution; flows from major institution to region industry; and employment rates by major and institution.

Research is being conducted on the design of a comparison between LODES and ACS Commuting (see https://ideas.repec.org/p/cen/wpaper/14-38.html). The detailed microdata comparison is available at https://ideas.repec.org/p/cen/wpaper/17-34.html. Main issues for the
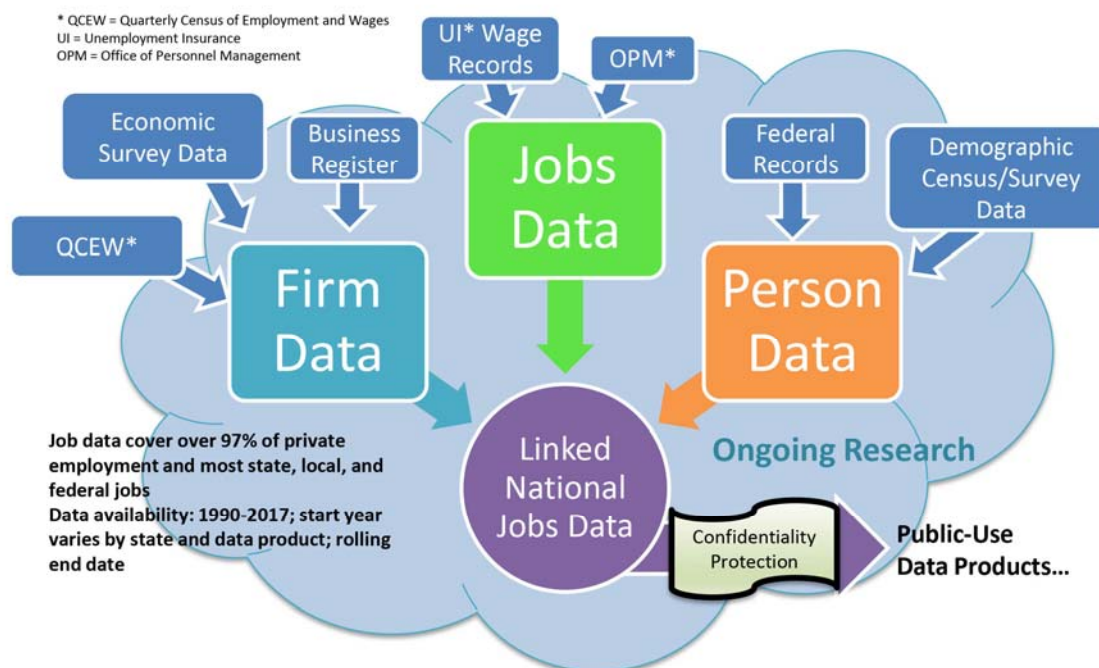
FIGURE 8.1 LEHD data infrastructure.

comparison research is the disagreement in workplace location between ACS and LODES and the missing data on establishments from administrative sources. A special project comparing micro was conducted by Green. The next step for the program is to add data on establishments (National Center for Education Statistics school districts) and the redesign unit-to-worker imputation. There are challenges with these advancements and staff resources will be critical. The goal across the board is to improve the accuracy of the data sources. In addition, a redesign of the OnTheMap application is planned for next year.

THE FUTURE OF CENSUS BUREAU DATA DISSEMINATION
Ally Burleson-Gibson

The mission of the Census Bureau is to serve as the leading course of quality data about the nation's people, places, and economy. The Census Bureau honors the privacy, protects confidentiality, shares their expertise globally, and conducts their work openly. The Census Bureau collects and disseminates data on a variety of topics, through a vast array of data tools and apps. At the same time, customers express frustration with finding and using Census Bureau content online. The decision has been made to move to an enterprise dissemination approach that will centralize and standardize the metadata, data, and software. The goal is to create a customer-oriented platform for easy access to Census Bureau data. The plan is to move dissemination from many tools to a single, streamlined, efficient search (Figure 8.2).
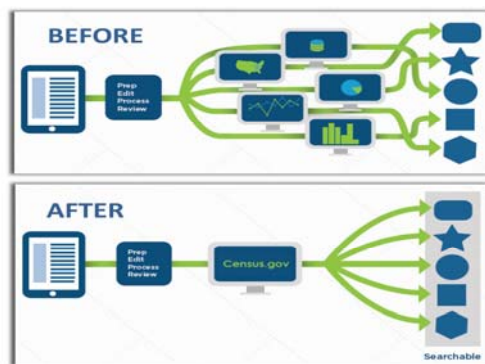
FIGURE 8.2 Before and after enterprise dissemination.

The data.census.gov development timeline includes the plan for the design, development and release of new features approximately every 2 months. Figure 8.3 provides the conceptual formation of CEDSCI. There will be an effort to provide a continuous integration of stakeholder and customer feedback, using an agile development methodology. The new interface is expected to make access to the ACS easier and will have a number of modern processing aspects now available with web-based applications. Features to explore Census data in data.census.gov platform illustrate the ease and transparency of the future user interface.

Figure 8.4 provides a graphic representation of the underlying concepts for the new platform. It integrates data services, metadata services, and geospatial services into a single application programming interface (API). The API supports requests for content, apps, and documentation.
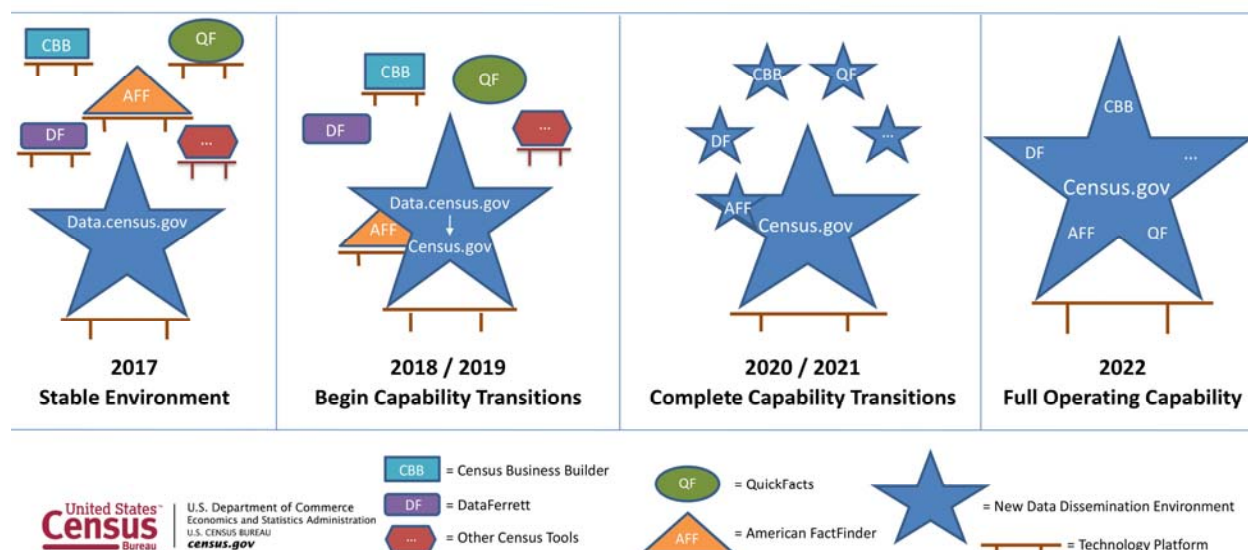
FIGURE 8.3 CEDSCI vision and scope.
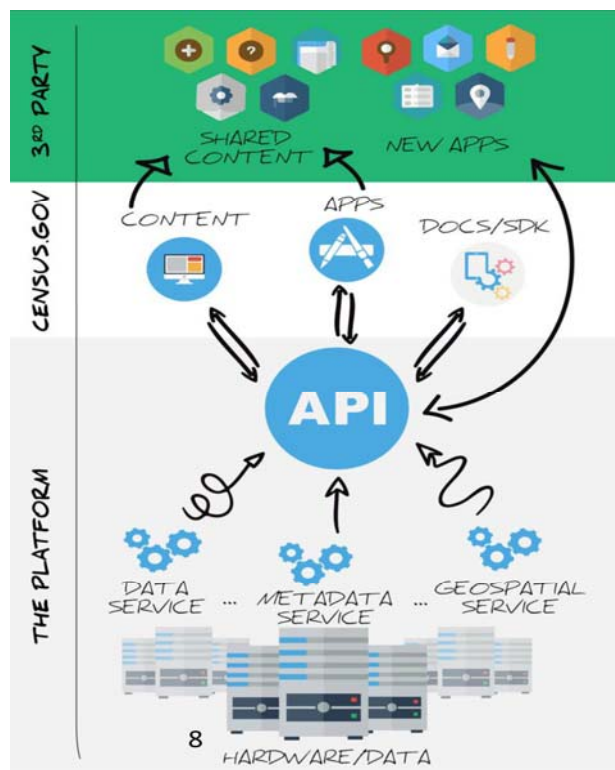
FIGURE 8.4 The new dissemination platform: how it all works.

Advanced Data Analysis

JOSEPH HAUSMAN
Federal Highway Administration, presiding

CEMAL AYVALIK
KIMON PROUSSALOGLOU
Cambridge Systematics

ARASH MIRZAEI
LIANG ZHOU
North Central Texas Council of Governments

JIANZHU LI
TOM KRENZKE
Westat

Working with MOE, evaluating data reasonableness and then applying the data for activities such as market segmentation are key challenges faced by the data analyst. This session addressed how these issues are being approached.

A FRAMEWORK FOR EVALUATING REASONSABLENESS OF TRAVEL TIME ESTIMATES AND MARGIN OF ERROR
Cemal Ayvalik and Kimon Proussaloglou

Acknowledging and incorporating the concept of MOE in the analysis of CTPP data is still "a work in progress." Although most practitioners possess a good understanding of ACS methods, the traditional absolute belief in these estimates is quite strong. Users are often cautioned to consider MOEs and some use those as a measure of data quality. MOE represent the uncertainty associated with sampling error and do not necessarily indicate the accuracy of the estimate; an estimate with a large MOE may be accurate when compared to another reliable data source. This research develops a framework for evaluating accuracy of travel time estimates and for contrasting actual errors to MOE by comparing mean travel times and categorical travel time distributions provided in the CTPP to those from another data source. Real-time travel time estimates were collected from Google Maps for a select group of Census tract pairs within the Detroit metropolitan area.

A set of tract pairs were selected that vary in size, proximity to each other, and employment density. In order to compare tract-level estimates from CTPP to point-level data from Google Maps API, at least twice the number of sampled housing units for the ACS and sampled at least 10 different workplace locations at the destination tract were sampled. Disaggregate HH and employment data were used as size variables to identify the sample O-D

pairs and incorporate a two-step probability-proportional-to-size sampling approach to identify origins and destinations.

Key hypotheses for this research include: accuracy of estimates is independent of magnitudes of MOE; increasing the sampling rate beyond ACS levels does not improve the accuracy of estimates, and accuracy of estimates is independent of the size of the Census tracts, the distance between place of residence and workplace, or employment density at the place of work.

Recognizing the temporal and structural differences in CTPP estimates and observed data from Google Maps, the proposed framework can be expanded into a more comprehensive effort. In addition, a similar API can be used as a means of quality assurance, and as a means of supplementing CTPP data with information on travel distances in developing future data sets.

USING CTPP DATA FOR MARKET SEGMENTATION OF HOUSEHOLDS AND EMPLOYMENT IN NORTH CENTRAL TEXAS REGIONAL TRAVEL MODEL
Arash Mirzaei and Liang Zhou

This research demonstrated the use of combination of CTPP and ACS data in the development of household and employment segments for the regional travel demand model of Dallas–Fort Worth (DFW) area. Using NHTS 2009 data, NCTCOG modelers identified that the best market segmentation for HBW trips is breaking down the households by number of workers by number of vehicles in small geographies. For the purpose of trip distribution, mode choice, and traffic assignment, the HBW market segmentation needs to have a breakdown of households by income also. To implement this, NCTCOG modelers used a combination of CTPP and ACS data. A similar process is utilized to break down employment into desired segments.

In the NCTCOG region, there are 243 TADs, 1,333 Census tracts, and 4,182 block groups. An iterative proportional fitting process (IPF) was used to connect the sources of the data into a desirable breakdown of the households by number of workers by number of vehicles by income at the block group level. The ACS data provided the distribution of households by number of workers by number of vehicles at the Census tract level. Each block group inherited this distribution from the Census tract. The ACS data also provided the distribution of households by income groups at the block group level. CTPP provides a three-dimensional breakdown of households by number of workers by number of vehicles by income at both the TAZ and TAD level. For purposes of stability and the reduction of sampling error, NCTCOG used the TAD level. The seed from CTPP at the TAD level was used for each block group within each TAD.

The IPF process started from the seed in each block group and distributed the households to match the control totals in each of the block groups. A similar two-dimension IPF process was used to break employment into 12 segments (income by industry) at the traffic analysis zone level. The results not only expanded our understanding of the distribution of households and employment in the DFW area, but also provide the foundation to recommend additional tables to be added to the CTPP data product.

USE OF PUBLISHED MARGINS OF ERROR FOR
AGGREGATING CTPP TABLES AND SENSITIVITY ANALYSIS

Jianzhu Li and Tom Krenzke

The CTPP comprises a set of special tabulations that are produced to meet the needs of transportation planners in understanding local JTW patterns. The tables relate worker and HH characteristics to travel mode based on the worker's residence, workplace, and travel from residence to workplace and involve cell estimates, and estimated MOEs for various geographic units such as Census tracts, TADs, and TAZs. TAZs are roughly the size of Census blockgroups. The 2006–2010 CTPP are based on 5 years of ACS data. One challenge of the CTPP is that the MOEs can be unstable, especially in small sample geographic areas.

Another challenge is to estimate the MOE when aggregating geographic areas that result in an area that is not published, or more generally, when aggregating any table cell estimates that result in an estimate that is not published in the set of tables. The naïve estimator which assumes zero covariance between cell estimates can overestimate the MOE and seriously break down when aggregating a medium to large number of cell estimates. As an alternative, the use of generalized variance functions (GVFs) for the purpose of stabilizing the variances and to address the issue of estimating MOEs for aggregated estimates was evaluted. Adjustments were proposed to improve the performance of the MOEs computed from GVF. A toolkit was produced to facilitate the estimation of MOEs for aggregated estimates using different approaches as well as the comparison between subgroups. Additionally, the researchers developed a replicated tables approach which can be used by transportation researchers as a diagnostic tool to assess the impact of the sampling and perturbation variance components in the CTPP tables on the transportation analysis models.

CHAPTER 10

TAZs
How Do We Move Forward?

KEVIN TIERNEY
Bird's Hill Research, presiding

JENNIFER MURRAY
Wisconsin Department of Transportation, recording

HUIMIN ZHAO
Independent Consultant, author

TAZs have been a part of the CTPP–UTPP data product since the inception of the tabulation. However, over time TAZs have become very costly to produce, redundant with other geographies and confusing in their structure. This commissioned paper looked at the CTPP TAZs while providing an assessment of the issues surround the continued production of TAZs for Census data.

TRAFFIC ANALYSIS ZONES: HOW DO WE MOVE FORWARD?
Huimin Zhao and Yong Zhou

TAZs have been a part of CTPP–UTPP tabulation geography for decades. However, over time defining and producing TAZ level tabulations have become costly. This paper reviews Census standard geographies and assesses the issues surrounding the continued production of TAZs for CTPP data. The issues include efforts for TAZ delineation, their usefulness, and data quality. The paper also presents the results of a CTPP data user survey that sought experts' opinions on the usefulness and utility of TAZ geography. The paper concludes that TAZ geography is essential for CTPP data and offers alternatives to address data quality issues.

Introduction

TAZs have been a part of the CTPP data product since the inception of the Census commuting special tabulations, dating back to 1980. TAZ is the most commonly used geography unit in travel demand models for transportation planning process. However, in many areas TAZs in CTPP (referred to as Census TAZs hereafter in the paper) are not the same as TAZs used for regional travel demand models (referred to as Model TAZs hereafter), which can be confusing and may lead to problems when referencing the data. In addition, the small geographic units such as Census TAZs impose confidentiality and privacy protection challenges, and data precision might be an issue due to the limited sample sizes of the ACS.

This paper assesses the issues surrounding the continued production of Census TAZs. The assessment is based on literature review, a series of interviews of experts in the field, and an online survey targeting professional staff at MPOs and state DOTs. The assessment is primarily

from a user perspective. Therefore, the survey aims to understand local agencies' efforts in Census TAZ delineation process, how CTPP data have been used for transportation planning, and how crucial the Census TAZ data structure is with comparison to other geographic units.

The next section describes various geography units for Census and CTPP, followed by a discussion on issues surrounding CTPP data and its TAZ geography unit. The user survey results are presented in the following section. The paper concludes with a discussion and recommendations of how to move forward with future TAZ requests.

Census Geography

Standard Census Geographic Entities

At the U.S. Census Bureau, virtually all Census data are geographically referenced. Currently, the standard hierarchy of Census geographic entities include Census blocks, block groups, Census tracts, Counties, States, Divisions, Regions, and Nation, with some variations for the island areas and American Indian, Alaska Native, and Native Hawaiian areas. Beyond the standard geographic hierarchy, the Census Bureau uses several other geographic entities including TAZs that help support specific data uses and user groups. The Census geographic hierarchy is illustrated in Figure 10.1.

The smallest geographic area for which the Census Bureau collects and tabulates decennial Census data is Census blocks. Block groups are the next level in the geographic hierarchy, which is generally defined to contain between 600 and 3,000 people. The block group consists of clusters of blocks and is the smallest geographic entity for which the decennial Census tabulated and published sample data when the long form was used and for which ACS presents data. The next level in the Census geographic hierarchy is Census tract, which is designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions. Census tracts are small and relatively permanent statistical subdivisions of a county with an average of 4,000 inhabitants. Although Census tracts are designed to be relatively permanent over time, they are updated every 10 years. Since the 1960 Census, the Census Bureau has assumed a greater role in promoting and coordinating the delineation, review, and update of Census tracts with local involvement.

Model TAZ and Census TAZ

The Census TAZ geographic delineation is not included in the standard hierarchy of Census geographic entities. Historically, Census TAZs were created specifically to support CTPP data, with the anticipation that these Census TAZs would be closely associated with Model TAZs in travel demand model and transportation planning process.

The Model TAZ is the unit of geography in conventional four-step travel demand models. In general, Model TAZs are designed to be relatively homogeneous, and the size of Model TAZs varies, with smaller zones in central business district and larger zones in the outer skirt area due to household and employment densities. Model TAZ's socioeconomic data, including population, households, and employment, is an input for travel demand models. Usually there is no minimum threshold requirement for Model TAZ population and employment. The total number of Model TAZs in a metropolitan planning area is determined to provide enough level of detail for models
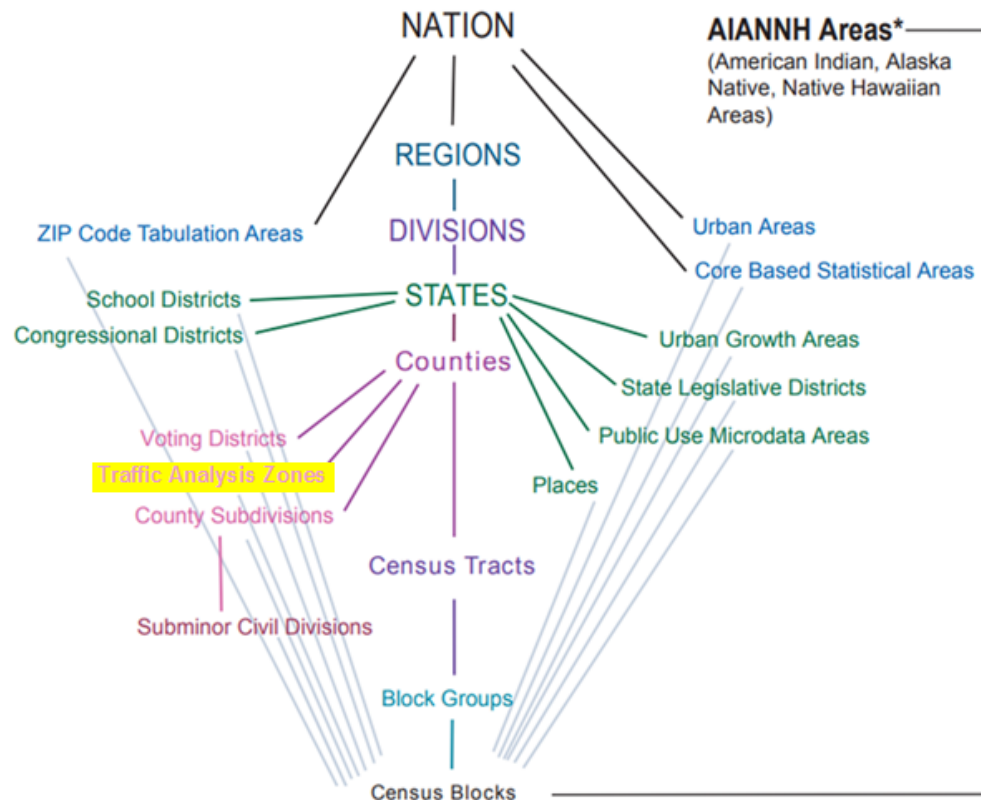
FIGURE 10.1 Standard hierarchy of Census geographic entities.
(Source: U.S. Census Bureau)

that support the regional or statewide transportation planning process. The complexity of the model is another factor that impacts Model TAZs in size. Prior to year 2000, most of travel demand models in the country were conventional four-step model and Model TAZs were similar to Census block groups in size with populations between 600 and 3,000. Since then, a new generation of travel demand models and land use models emerged that brought about the ability to provide traffic forecast information in detail and with these advances Model TAZs have tended to get smaller in size.

Census TAZs are not the same as Model TAZs. Census TAZ is a geography unit delineated by state or local transportation organizations for tabulating transportation-related data (i.e., CTPP data), especially JTW and place-of-work statistics (U.S. Census Bureau, 2002). The creation of Census TAZ as a geographic delineation was aimed to present the data in a way that is more convenient for data users to access and tabulate data. The U.S. Census Bureau requires Census TAZs to follow Census-designated boundaries (TIGER line boundaries). To ensure data quality, there are other minimum population or employment requirements for Census TAZs. These requirements are the main reason that Census TAZs differ from Model TAZs.

The Census Bureau first provided data for TAZs in conjunction with the 1980 Census, when it identified them as "traffic zones" (U.S. Census Bureau, 2002). For the 1990 Census, Census TAZs were defined as part of CTPP. For 2000 CTPP, the FHWA distributed the TAZ-

UP software to MPOs and state DOTs to delineate TAZs. The participation in the TAZ delineation program was not mandatory. MPOs and states who did not participate in the TAZ delineation program were given the option of requesting CTPP 2000 data at either the Census tract or block group level of detail.

Different from previous TAZ delineation process, the 2010 TAZ delineation included two geographic structures: Census TAZ and Census TAD. The Geographic Division of the Census Bureau, FHWA, and AASHTO developed the delineation business rules. It was required that all TAZs nest within a county and within a TAD, respectively. However, TADs were not required to nest within a county. TADs needed only to nest within the delineation coverage assigned to the MPO–state DOT.

The 2010 delineation business rules also provided guidelines that were suggested but were not required. For example, the Census Bureau recommended that the minimum resident worker population and workers by place of work level should be approximately 600 persons, which corresponds to the minimum threshold allowable for 2010 Census blockGroups. It was also recommended that Census TADs have an estimated population lower limit of 20,000 residents. Although these rules are recommended but not required, our user survey (see detailed descriptions in user survey section) indicates that many delineation program participants modified their TAZs to meet these suggested requirements.

CTPP and Census TAZ History and Issues

As Census TAZs were created and updated for tabulating CTPP data, it is meaningless to look at the future direction of Census TAZ without understanding the information presented in the CTPP data and how transportation planning professionals use the data, as well as some administrative issues surrounding CTPP.

History of Organizational Cooperation and Cost for CTPP

The CTPP is a historical example of organizational cooperation between the agencies and entities that rely on it (Christopher, 2002). The CTPP data are a set of special tabulations designed by transportation planners using large sample surveys conducted by the Census Bureau. The transportation community assumed the ownership of the program by demonstrating its willingness to pay for a set of special tabulations at the Census TAZ level to meet the transportation planning data needs.

The 1970 Census was the first decennial Census to offer the cost-reimbursable UTPP. There were 112 purchasers, most of which were MPOs. For the 1980 UTPP, there were 152 purchasers. The 1990 CTPP was the first pooled-fund program administrated by AASHTO that allows all the states and MPOs access to the data. The 2000 CTPP was also an AASHTO pooled-fund program. The approximate direct costs for CTPP–UTPP are shown in Table 10.1 (Christopher, 2002).

In addition to the direct charges to the states and MPOs for CTPP tabulation, there have been other costs and contributions from the transportation community to define the local tabulation geography, such as the Census TAZ delineation process. For the development of the 2000 CTPP, state DOT agencies invested three-quarters of million dollars for technical support, coordination and software for Census TAZ delineation (Christopher, 2002). The staff time to develop the Census TAZs was provided by local agencies for an unknown additional cost. Many agencies hired consultants to help with the TAZ delineation process.

Despite the extra cost of defining Census TAZs for CTPP tabulation, the transportation community has endorsed the Census TAZ delineation process probably because TAZs are an essential geography unit that is associated with transportation planning. For the 2000 CTPP, 282 of the 340 MPOs defined their own TAZs (Christopher, 2002).

On the other hand, the Census Bureau has approached CTPP tabulation as a cost-reimbursable product that was beyond the scope of the Bureau. There is no doubt that the transportation community is the main user of the data. Due to the organizationally cooperative nature of the CTPP program, it is important to have close communication between the transportation community and the Census Bureau in terms of data collection, processing, and tabulation to improve data quality and fulfill transportation planning data needs. The future of the Census TAZ should be guided by this broad conversation on data collection, processing, tabulation, and data usage.

Data Quality

From 1970 to 2000, CTPP and its predecessor, UTPP, used data from the decennial Census long form. Now the decennial Census long form has been replaced with the continuous ACS, and CTPP uses the ACS sample for the special tabulation. It is worth noting that past decennial Census long forms were mailed to one in six households (17% sample size) while the ACS samples the equivalent of 2.5% housing units annually (FHWA–FTA, 2007). The smaller sample size in ACS leads to larger sample errors. The estimated sample error is about 1.33 times of that of the 2000 Census (FHWA–FTA, 2007) and therefore, data precision becomes an issue especially for small geography area. Data precision improves for the CTPP data as they are tabulated for larger geography units.

To make sure that users understand that sample errors vary among places and variables, the CTPP tabulations report all the estimates with MOEs, and the Census Bureau strongly recommends that users incorporate this uncertainty in their analysis. In our user survey, we asked a question on how MOEs impact transportation planning professionals' decision on data usage–data analysis. The survey found that many transportation professionals did take into account of MOEs when making decisions for data usage–data analysis. At the same time, others found MOEs confusing.

The replacement of the decennial Census long form with ACS probably increased the Census Bureau's workload of CTPP tabulation significantly. This is because ACS is conducted on a continuous basis with much smaller sample size. A hierarchical geography system (such as 2010 delineated Census TAZ–TAD) may help to present the data with better quality and still serve for transportation planning data needs.

TABLE 10.1 Direct Cost for Transportation Planning Packages

	1970	1980	1990	2000
Buyers and users	112	152	All states, MPOs	
Cost	$0.6 M	$2.0 M	$2.5 M	$3.0 M
Tables	43	82	120	203

Data Contents

The CTPP tabulations include three geographies:

1. Residence-based tabulations summarizing worker and household characteristics;
2. Workplace-based tabulations summarizing worker characteristics; and
3. Worker flows between home and work, include travel mode.

The residence-based tabulations are much like regular Census products except that they have more two-way, three-way, and even four-way tables that depict population and household characteristics. This is tailored specifically for use by MPO travel demand models.

The workplace-based tabulations are the only Census product that contains summary data on workers at their place of work. It is also known that workplace geocoding has been a problem since 1970s due to reporting errors. The problem is especially troublesome when tabulating workplace data at a small geography unit such as Census TAZ. It directly leads to the loss of data when the workplace address contains errors and cannot be assigned to the corresponding TAZ. Tabulating the workplace data at standard Census geographic level such as block groups does not solve the problem because Census TAZs and Census block groups are comparable in size. Tabulating the workplace data at a larger geographic unit such as Census TADs or Census tracts, or even county level, can prevent the loss of survey records. However, doing so cannot prevent the loss of information as the worker flow information (such as trip length and travel time) between home and work relies on the accuracy of workplace location.

We believe the geocoding issue of workplace location can be tackled much more efficiently in the data-collection process rather than by altering the geographic unit in the data tabulating process. Collecting and geoprocessing location data used to be a hassle. But with advances in sensor technology and the widespread use of Bluetooth devices in transportation data collection, location information can be collected more easily and accurately. Furthermore, the digital form of the location data makes geoprocessing more straightforward.

Data collection requires careful planning and is beyond the scope of this paper. Due to the organizational cooperation nature of CTPP program, close communication between agencies is a first step to improve data quality.

Data Usage

Transportation planning professionals use CTPP data to

1. Evaluate the existing conditions;
2. Develop or update travel demand models; and
3. Analyze demographic and travel trends.

When evaluating the existing conditions or analyzing demographic and travel trends, we usually refer to a larger scale geography such as a planning corridor, a city, a county, or at a regional level. It is rare to see a travel trend analysis tailored specifically to one TAZ. In this sense, data tabulated at a larger geographic unit such as TAD may be sufficient if the trade-off must be made.

To use the CTPP data for travel demand model development and update, it is essential to have the data at TAZ level. The residence- and workplace-based demographic tabulations can be used as demographic inputs for a base year model. The workflow data between home and work can be used for the calibration and validation of trip distribution models for HBW trips. The calibration of trip distribution models is to reproduce the trip length distribution (not simply to replicate the current flows between zones). A valid trip length distribution of workflow relies on accurate home and work locations, presented at TAZ level.

Census TAZ Online Survey

Overview

As part of this study, an online survey was conducted to gauge state DOTs, MPOs, rural planning organizations (RPOs), and other planning organizations' experience and preferences of Census TAZ data. The results of this survey will help us to understand the current usage of Census TAZ data and provide some insights for future improvements.

Survey Development and Methodology

With inputs from the AASHTO oversight board, the research team utilized Google Form to develop a 15-question online survey regarding Census TAZ data use and preferences. An e-mail with the survey link was sent out to everyone on the CTPP TAZ Delineation contact list from the AASHTO website (http://ctpp.transportation.org/Documents/CTPP_TAZ_Delineation_Contact _List_Database_Master_to_Census_Bureau_March42011.xls).

After removing duplicated records, about 400 e-mails were sent on July 11, 2017. About 100 e-mails failed to deliver due to personnel changes or other reasons. Also, about 20 automatic replies were received due to the recipients' out of office or vacation status. Follow-up e-mails with reminders were sent out on July 17, 21, and 28, respectively. A PDF copy of survey questionnaires was attached with reminder e-mails so that the potential survey participants could share the survey with colleagues within their agencies. The survey asked one entry per agency. The survey was closed on August 3, 2017.

We received a total of 99 survey responses, of which 96 were online, and three via marked-survey PDF files.

Summary of Survey Results

There were 15 questions in this survey, grouped into three categories. The first category was about the planning agency and its transportation planning data sources, including two questions. The second category is about AASHTO's TAZ delineation program, including five questions. And the third category is about CTPP TAZ data usage, including eight questions. Besides the 15 questions, there were three additional information-collection questions regarding extra comments, contact information for follow-up questions. The survey results are summarized below.

Among the 99 respondents, 17% are from state DOTs, and the remaining 83% are from various regional planning organizations, including 76% from MPOs and 7% from RPOs. Among the MPOs, 41% are small MPOs (population less than 500,000), 19% are medium-sized MPOs, and 16% are large MPOs with population greater than 1 million. With 40 participating agencies

as small MPOs, it seems that small MPOs are quite interested in CTPP data, perhaps indicating greater importance of CTPP data to them. CTPP data are probably their main or sole data sources for transportation planning.

Survey Question Category 1: Agency and its Transportation Planning Data Sources

Question 1	Which of the followings describe your agency	
Choices	Response Options	Count
a	MPO with a population greater than 1 million	16
b	MPO with a population between 500 K to 1 million	19
c	MPO with a population less than 500 K	40
d	State DOT	17
RPO	Rural Planning Organization	7
	Total	99

Which of the followings describe your agency
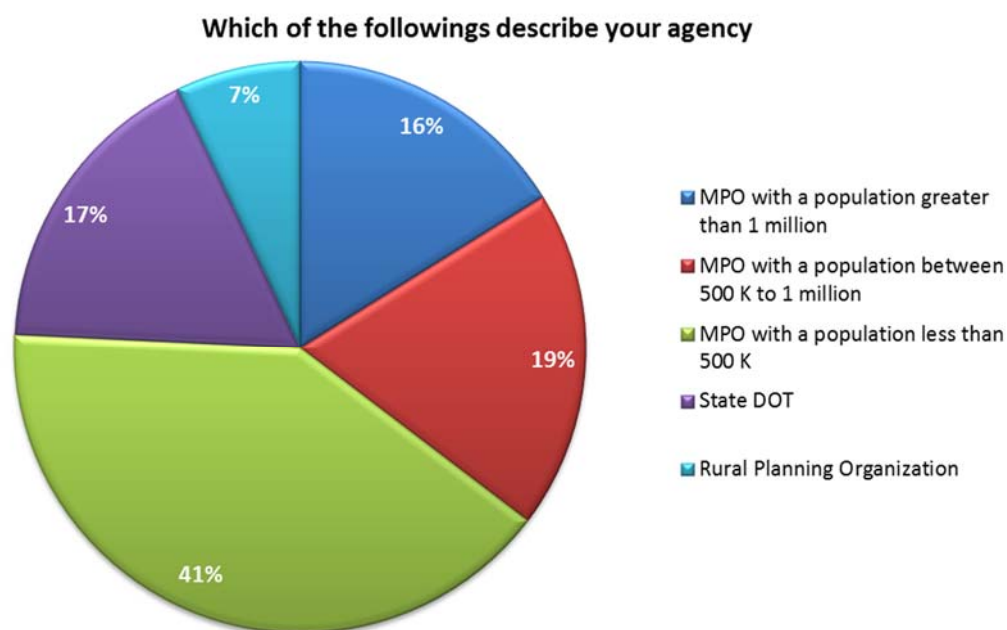
Figure 10.2 provides an overview of the geographic distribution of the survey respondent agencies. Please note that about five DOTs, 10 MPOs, and four RPOs did not provide the specific name of their agencies. So about 19 survey respondents are not marked in Figure 10.2.
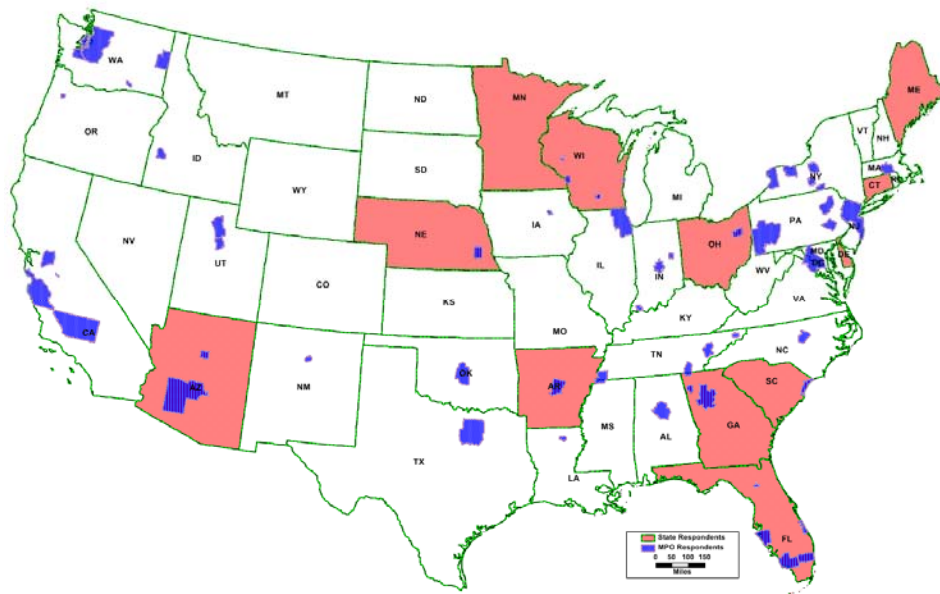
FIGURE 10.2 Census TAZ survey respondent distribution.

Survey Question Category 2: TAZ Delineation Program

Question 2 asks about the agency's general data acquisition practice and CTPP data usage. About 19% agencies solely rely on CTPP data package and state add-on data. About 45% conducted their own regional HH travel survey to supplement CTPP data. Another 16% agencies purchased additional data to supplement CTPP data. Only 18% of agencies mainly use HH survey data for planning purposes.

Question 2	Which of the following statements describe your agency's TAZ-related transportation planning data needs and data acquisition practice? (Please check all that apply)	
Choice	Response Options	Count
a	The agency solely relies on CTPP data package, State's add-on data when available, and other publicly available sources for its transportation planning data needs	19
b	The agency supplements regional household travel surveys and/or other locally collected demographic/employment data with CTPP data package/State's add-on data/other publicly available data sources for its transportation planning data needs	45
c	The agency purchases data to supplement CTPP data package/State's add-on data/other publicly available data sources for its transportation planning data needs	16
d	The agency mainly relies regional household travel surveys and/or locally collected demographic/employment data for its transportation planning needs and CTPP data package/State's add-on data/other publicly available data sources are used for reference	18
e	Other	5
Total		103

Question 3	Did your agency participate in the TAZ delineation program (2010/2011)?
Response Options	Count
Yes	82
No	7
Not Sure	10
Total	99

For most of the survey respondents (83%), their agencies participated in the TAZ delineation program in 2010–2011. This confirms the transportation community's support for Census TAZ geography.

Question 4	If No in Q3, what was the reason not to participate the TAZ delineation program? (Please check all that apply)	
Choice	Response Options	Count
a	Agency staff shortage/budget constraint	4
Other	State DOT did on our behalf	3
	Total	7

Among the seven agencies who did not directly participate in the TAZ delineation program, four did not participate due to staff availability while for three agencies, the state DOT participated on their behalf.

Question 5	If Yes in Q3, does the TAZ system for your regional travel model is the same as the CTPP TAZ geography for your region?
Response Options	Count
Yes	29
No	43
Not Sure	10
Total	82

For the 82 agencies who participated in the CTPP TAZ delineation program, 35% of them do have their regional travel demand model's TAZ system identical to the CTPP TAZ geography.

Question 6	What was the reason to use different TAZ systems for regional model and CTPP data?	
Choice	Response Options	Count
a	TAZs for regional model are small and can't fulfill the recommended minimum	25
b	There are TAZs in regional model that cross County lines	1
c	TAZs for regional model do not nest perfectly with census tract/block group	17
Others	Others	9
	Total	52

For the 52 agencies who participated in the CTPP TAZ delineation program but did not use the CTPP TAZ system for regional model, about 48% of them attributed the reason to the

fact that TAZs for their regional models are small and cannot fulfill the recommended minimum population and employment thresholds; about 33% of the agencies said that it was because the TAZs for regional model do not nest perfectly with Census tract and block group.

Question 7	How different are the TAZs for regional model and CTPP-TAZ geography?	
Choice	Response Options	Count
a	Less than 5% of the TAZs are in different shape for the regional model and CTPP	8
b	5% - 25% of the TAZs are in different shape for the regional model and CTPP	11
c	25%-50% of the TAZs are in different shape for the regional model and CTPP	11
d	More than 50% of the TAZs are in different shape for the regional model and CTPP	7
e	The CTPP-TAZ is completely different from the regional model TAZs	6
	Total	43

When the 52 agencies above were asked how different the TAZs for the regional model and CTPP TAZ geography were, 43 of them provided responses. The difference is spread across the response options: 19% for 5% or less difference; 25% for 5% to 25% difference; 26% for 25% to 50% difference; 16% for more than 50% difference; and 14% are completely different.

Survey Question Category 3: CTPP TAZ Data Usage

Most (92%) survey respondents have used CTPP tables or are familiar with CTPP tables.

Question 8	Have you used CTPP tables or are you familiar with CTPP tables?
Response Options	Count
Yes	91
No	8
Not Sure	0
Total	99

Question 9 is about the usage of CTPP tables. Among 99 survey respondents, 83 use the data for travel demand modeling, far more than any other transportation analysis listed. Besides these transportation analysis, survey respondents also used CTPP tables for population forecasting, O-D analysis, commuting flows for long-range planning, general travel pattern analysis, community impact analysis, and many other ad hoc requests.

Question 9	For which of the following transportation analysis do you use CTPP tables? (Please check all that apply)	
Choice	Response Options	Count
a	Travel demand modeling	83
b	Major corridor planning	37
c	Environmental justice analysis	45
d	Transit planning	31
e	Public involvement	25
Others	Others	10
	Total	231

Question 10 is about data contents. According to the survey respondents, the home-to-work flow tables are the most useful to serve their agencis' transportation data needs. This is probably because there are other data sources for population and employment data, but home-to-work flow tables are only available through CTPP.

Question 10	Which of the following CTPP tables are most useful to serve for your agency's transportation data needs?	
Choice	Response Options	Count
a	Residence-based tables	28
b	Workplace-based tables	16
c	Home-to-work flows tables	54
	Total	98

On whether it is crucial to have the CTPP data at TAZ geographic level for their use, 49% of survey respondents think that the CTPP data can be at Census track or block group level, but it's more convenient to have data at CTPP TAZ level. While 24% think the data must be at CTPP TAZ level, 18% think it is more convenient at Census tract and block group level. There are four responses in "others" group. Among the responses, one thinks that the smallest level of geography is the most beneficial; and another respondent stated that they normally use CTPP data at Census tract and block group level, but for model development, they have used the TAZ level data. In summary, the responses to this question confirmed the transportation community's preference of CTPP data at the TAZ level.

Question 11	How crucial is the CTPP data at TAZ geographic level for your use?	
Choice	Response Options	Count
a	The data must be at CTPP-TAZ level	23
b	The data can be at census tract/block group level, but it's more convenient to have data at CTPP-TAZ level	48
c	It is more convenient to have data at census tract/block group level	18
d	I did not use CTPP data at TAZ geographic level	5
Others	Others	4
	Total	98

In terms of data quality measures, about half of the respondents have used MOE information provided by CTPP while the other half have not.

Question 12	Have you ever used margin of error information provided by CTPP data?
Response Options	Count
Yes	48
No	50
Not Sure	0
Total	98

For the view on MOE, 54% of survey respondents think the MOE fields provide some insights on data quality but do not influence the way the data were used; 12% completely ignored the information. About 38% think MOE is significant enough to alter the way the data were used. About 9% think this information causes confusion.

Question 13	Which of the following statements closely describe your view on margin of error field? (please check all that apply)	
Choice	Response Options	Count
a	Margin of error information is completely ignored while using the data	12
b	Margin of error provides some insights on data quality but it does not influence the way the data was used	53
c	Margin of error provides insights on data quality and it is significant enough to alter the way the data was used	38
d	Margin of error provides information that leads to confusion in data usage	9
Others	Others	6
Total		118

For residence-based CTPP tables, 61% of survey respondents think it is the most important to provide the tables at TAZ geographic level with current household/person demographic variables. About 37% think it is more important to have a multi-dimensional joint distribution of household/person demographic variables and the data does not need to be presented at TAZ level.

Question 14	On residence-based CTPP tables, which of the following statements closely describe your view? (Please check all that apply)	
Choice	Response Options	Count
a	It is the most important to provide CTPP tables at TAZ geographic level with current household/person demographic variables	60
b	It is more important to provide multi-dimensional joint distribution of household/person demographic variables with a certain level of accuracy. The data does not need to be presented at TAZ geographic level	37
Others	Others	7
Total		104

Question 15 is a multiple choice question with a list of statements. It was intended to understand how transportation agencies use workplace-based and flow-based CTPP tables and what is their preferred geography unit for data presentation.

For workplace-based CTPP tables, more agencies do not use CTPP tables to develop employment demographics than those who do (35 versus 21). It is also noted that close to 40% of the survey respondents, (37 and 38 responses, respectively), indicated that the workplace demographic information, as well as home-to-work mode of transportation and travel time information, need to be presented at TAZ geographic level for their use. On the other hand, about 15% of the survey respondents indicated that workplace demographics and home-to-work flow information do not need to be presented at TAZ level for them to use.

Question 15	On workplace-based and flow-based CTPP tables, which of the following statements closely describe your view and/or data acquisition practice at your agency? (Please check all that apply)	
Choice	Response Options	Count
a	We do not use CTPP workplace-based tables to develop employment demographics for the region	35
b	We heavily rely on CTPP workplace-based tables to develop employment demographics for the region	21
c	The workplace demographic information does not need to be presented at TAZ geographic level for our use	16
d	The workplace demographic information need to be presented at TAZ geographic level for our use	37
e	The workplace mode of transportation and travel time data does not need to be presented at TAZ geographic level for our use	14
f	The workplace mode of transportation and travel time data need to be presented at TAZ geographic level for our use	38
Others	Others	11
	Total	172

Besides answering survey questions, survey participants also provided valuable comments on the CTPP data. There are two comments especially worth mentioning. One is related to the expansion of MPO coverage area. Several survey participants indicated that their agencies do not use CTPP tabulation because of the recent expansion of their MPO coverage area. As MPOs are required to update their long-range transportation plan every 5 years, the update cycle of TAZs may be something to consider in the future.

Another participant indicated that the base year for his agency's current travel demand model is 2015. The CTPP tabulations based on 2006–2010 ACS are outdated for their travel model update. Processing data takes time and effort. The next CTPP data set will be based on the 2012–2016 ACS, to be available in late 2018 or early 2019. By then many MPOs will still be updating their travel demand model with a base year between 2012 and 2016.

Moving Forward with TAZ

TAZ is an essential geographic unit in travel demand models for transportation planning. The CTPP tabulations at TAZ level have widespread support within the transportation community, as confirmed by our online TAZ user survey. In addition, small-sized MPOs rely more on CTPP data to fulfill their transportation data needs due to their limited resources.

Though tabulating CTPP data at TAZ level is the transportation community's preferred platform for data presentation, there are issues surrounding CTPP data tabulated at TAZ level. First, the CTPP tabulations are a cost-reimbursable product of the Census Bureau. TAZ geography is not a part of standard hierarchy of Census geography, so the Census Bureau is less supportive of this delineation. Second, CTPP tabulations are based on ACS. The small sample size of ACS leads to data precision concerns when the data are tabulated for small geographic areas such as TAZs. Additionally, the geocoding of workplace location has been a problem in the past due to reporting error. The workplace addresses with reporting errors cannot be assigned to a TAZ geography and will lead to loss of survey records and raise data quality concerns. Finally, Census TAZ definitions differ from Model TAZ definitions in many areas because of the recommended minimum population–employment requirement, the need to maintain linkages

with Census tract and block group definitions, and changes in MPO area definitions. The alternative TAZ definitions may cause confusion and limit the value of Census TAZ tabulations.

Our online TAZ user survey shows that more than 70% of the survey respondents prefer CTPP tabulations at TAZ level. About a quarter of survey respondents think that CTPP data at TAZ geography is a must for their use. As many small- and medium-sized MPOs with limited resources rely heavily or completely on CTPP to fulfill their transportation data needs, it is essential to maintain the TAZ geography for CTPP tabulation. Tabulating CTPP data at standard Census geography unit will not necessarily solve some of the data quality issues. For example, TAZs are comparable with Census block groups in size. If CTPP data were tabulated at block group level, all the geocoding and MOE issues would have remained with block group geography.

It is a practical solution to have a hierarchical geography system based on TAZs for CTPP's geocoding and data precision problems. The 2010 TAZ delineation included a new geography TAD that is an aggregation of TAZs. The data tabulated at TAD level will help to mitigate some of the data quality problems. This is a less-than-ideal scenario especially for home-to-work flow data, as trip distribution models aim to replicate trip length distribution. With larger geography, the trip length distribution estimates become less accurate.

A few immediate CTPP TAZ issues deserve our attention. One is about the data cycle. Unlike previous CTPP tabulations based on Census long form in a 10-year cycle, CTPP now is based on ACS and is released every 3 to 5 years. As MPOs are required to update their long-range transportation plan every 5 years, the more frequent CTPP data release is very helpful to fulfill the transportation planning data needs. However, in the process of updating the long-range transportation plan, MPOs may have expanded their coverage area due to growth. They may update their Model TAZ system. Therefore, the question is whether Census TAZ delineation process should be conducted in a more frequent basis, say, updated every 5 years, to keep up with the updates at local agencies. We believe it will be beneficial for the transportation community to have TAZ delineated every 5 years or paired with CTPP new release. On the other hand, more frequent delineation processes require staff work hours and administrative coordination. We are also unclear about the funding resources.

Our long-term goal is to have CTPP tabulations at TAZ level with improved accuracy. Past research has shown that the most-effective way to improve data quality is in the data-collection process. For example, if ACS data were collected digitally, workplace location can be instantly verified to a point on the map. The point layer then can be tagged easily to any geography entities, which will significantly reduce reporting errors and reduce the workload for geocoding and geoprocessing. On the other hand, CTPP is a product of cooperation among multiple agencies. The survey and data processing are administrated by the Census Bureau. It is not clear how progressive the Census Bureau is in adopting technologies in data collection process. Nonetheless, it is a key to keep close communication with the Census Bureau on issues of data collection, data processing, and data usage to achieve the long-term goal.

Acknowledgments

timely responses. The views presented in this paper are the authors. They do not necessarily reflect the official view of AASHTO.

References

Bower, K. Looking Back and Ahead: A History of Cartography at the Census Bureau and What the Future Holds, 2010. Available at https://www.Census.gov/history/pdf/cartographyatCensus.pdf.

Tierney, K., and Cambridge Systematics, Inc. Assessing the Utility of the 2006–2010 CTPP Five-Year Data: Summary Report, 2016.

Christopher, E. The CTPP: Historical Perspective, 2002. Available at http://www.trbCensus.com /articles/ctpphistory.pdf.

Using ACS Data in Transportation Planning Applications: Peer Exchange Report. FHWA–FTA, 2007. Available at http://trbCensus.com/SCOP/docs/acs_peer_exchange_may2007.pdf.

2000 Census of Population and Housing, Summary Population and Housing Characteristics. PHC-1-11. Florida and U.S. Census Bureau, Washington, D.C., 2002.

TAZ Delineation Business Rules. U.S. Census Bureau, FHWA, AASHTO, 2010. Available at https://www.fhwa.dot.gov/planning/Census_issues/ctpp/data_products/tazddbrules.cfm.

Geographic Area Reference Manual. U.S. Census Bureau, 2013. Available at https://www2.Census .gov/geo/pdfs/reference/GARM/.

Facilitated Discussion

Usefulness/Utility of Having the TAZ Geographic Delineation: What analyses are you and other transportation planners doing at the Census TAZ level of geography? What would be lost in these analyses if only larger geographic delineations were available?

Discussion around the usefulness and utility of Census-created TAZs supported the findings of the commissioned paper, with the recognition that it's the only information available. Overall, the audience recognized that there were differences between Census TAZs and modeling TAZs. It was noted that a number of MPOs are getting their employment from the CTPP and that the TAZs are based from residential population. One suggestion was for TAZs to be constructed with both worker TAZs and housing TAZs. One member of audience suggested imputing down to the TAZ level or using a percentage of the geography. It was noted that even though transportation professionals create their own TAZ's, it is useful to validate their data with Census-created TAZs. They also have confidentiality information requirements and need to account for confidentiality when working with geographic data.

The audience engaged in a robust discussion on how the Census built the previous TAZs. It was explained that TAZs were built from Census blocks, using an equivalency process. The importance of the PSAP Program was emphasized by a number of audience members. For example, one member stated that when you start to look at the creation of TAZs, you look to the PSAP processes first. There's room in that process to make the block groups your TAZ's or at least TAZs can be constructed to nest within block groups. Another audience member strongly recommended that MPOs will be involved with PSAP.

Census Bureau staff shared details of the TAZ creation process, describing how the blocks are the one of the last things produced. Blocks are recreated every 10 years. Once the block groups are set, the Census develops their blocks from them. Audience members were encouraged to be proactive and get involved in PSAP now. If they do, they can help set their own geographies, but they need to realize that participation is critical now.

The discussion continued with respect to the option to revisiting blocks now. First, Census works within a limited extent because the blocks are built from the limited geography. States would need to have someone designated as their local contact. This effort impacts a number of different local organizations doing outreach and getting some contact lists. It was also noted that there will be some outreach directly associated with a PSAP website that the Census is working on and that they can post information. The preliminary draft notice on the block group threshold will come out spring 2018. If they get a response through the *Federal Register*, the Census Bureau will respond.

A number of individual audience members expressed enthusiasm by the notion that if more people knew that they could define block groups, the transportation community might embrace a block group approach and the TAZ could then be eliminated. At the same time, Census is bound by particular requirements. For example, there are block group minimums—the current proposal is 600 people (recognizing that the block is based on decennial data).

Challenges and Limitations of the TAZ Geographic Delineation: How have your and other transportation planners' analyses at the Census TAZ level of geography been affected by larger MOE? How have your and other transportation planners' analyses at the Census TAZ level of geography been affected by the need for small area data perturbation? How have workplace geocoding issues affected your and other transportation planners' analyses at the Census TAZ level of geography and other geographic levels?

With respect to geographies, one audience member mentioned that MPOs use higher-level geography to calibrate their models. Specifically, worker levels at the TAZ level are perhaps the most valuable input, with the knowledge that the MOE will be lower than at the tract level. One audience member spoke on behalf of rural area needs. Using block groups as existing Census delineations, TAZs can be created and coded so the delineations do not violate any Census boundaries, and then these geographies can be used to capture the data. However, only housing occupied or vacant units can be captured due to data suppression of other variables.

Costs of the TAZ Geographic Delineation: In your experience, what are the marginal costs and resource requirements of defining TAZs, tabulating ACS results, and reporting data at the TAZ level?

A number of audience members thought that it is very hard to determine what the cost structure is for the creation of TAZs as it is spread over a number of sources. One member expressed concern regarding whether it is worth it to invest, while at the same time, acknowledged that having CTPP data at TAZ geographies gives great insight on travel patterns. Since the actual costs are unknown, it is hard to evaluate costs versus benefits.

It was noted that some of the costs in the past have been packaged in FHWA contracts. One audience member indicated that TAZs could be delineated later. Therefore, if an entity wants to pay for the delineation, they pay for it, and then everyone would have the opportunity to cut their block groups—it would not necessarily need to be a cost born by the MPOs—but the Oversight Board would need to consider this approach. Several members of the audience agreed to bring these suggestions to the Oversight Board to consider as they prepare their timelines. It was pointed out that there are other sources of data that the CTPP could attach to TAZ, or block group data, that does not come from the ACS.

Suggestions for the CTPP Oversight Board: What guidance do you want to provide to the CTPP Oversight Board as they consider the need for data at the TAZ-level geography?

Several audience members indicated that there is no commitment to the TAZ geography at this point. Further, they [Census Bureau] will only process if the CTPP calls for it as a special tabulation after the Census is completed. The Oversight Board could ask for block groups and the Census would like it because it saves time and money.

Audience Suggestions for the CTPP Oversight Board

- Encourage the Oversight Board to promote participation by all levels of government in the upcoming PSAP program.
- Stay in close contact with the Census Bureau to make sure all possible opportunities to influence the decision for data dissemination of the 2020 data and beyond meet the needs of transportation professionals.
- Encourage the Oversight Board to continue robust discussions on the need for, and use of, TAZ-level geographies to ensure the best outcomes for the future. If the efforts to use block groups as the optimal geography are successful, it may be possible to make user-defined boundaries for TAZs through CTPP software or other external opportunity. This could reduce the burden for the Census Bureau, and perhaps, lower costs for processing.

CHAPTER 11

We Like Our PUMS and We Use It

JENNIFER MURRAY
Wisconsin Department of Transportation, presiding

KEVIN TIERNEY
Bird's Hill Research

JONATHAN SCHROEDER
Minnesota Population Center

CHARLES PURVIS
Metropolitan Transportation Commission (retired)

The PUMS can be one of the most powerful resources in the data analyst's arsenal. This session provided insights into the uses of PUMS data.

USE OF PUMS BY STATE DOTS AND MPOS: A SYNTHESIS
Kevin Tierney

Transportation planners in many regions are using PUMS data as essential inputs to mission-critical analyses, but planners at many agencies are largely unfamiliar with these data or their benefits. To provide a broader understanding of PUMS, a synthesis of practice was sought to describe how transportation planners use PUMS data to help address their data needs.

PUMS data are unusual for Census data in that they are not tabulations of data summarized at a specified geographic area, but rather are a sample of the actual data records collected in the ACS. The PUMS records are subjected to data disclosure avoidance techniques to protect respondents' confidentiality, including limiting the most precise geographic reporting to PUMAs that are areas of at least 100,000 residents. The synthesis of practice was conducted in three tasks:

- Review of published and unpublished documentation on PUMS usage by transportation planners;
- In-depth interviews with transportation planners that use PUMS data; and
- Web-based survey scan of Census data users within transportation agencies.

Usage of the Census PUMS data is less prevalent than usage of other Census data products. Slightly more than one-third of the state DOT representatives contacted for the synthesis were regular or occasional users of PUMS data. About two-thirds of the large MPOs that participated in the synthesis use PUMS data. Only a few small or medium MPOs use the PUMS data. To some extent, the lower usage of PUMS reflects the fact that the

PUMS data set is a rather specialized niche data product. However, the most common reason that nonusers gave for not using the PUMS data was their lack of familiarity with these data.

In contrast, those that do use the PUMS data generally rated their importance highly and rated their level of satisfaction relatively highly as well. Because the PUMS data include full records with the full range of Census household and person data items, data users are able to cross-tabulate and explore relationships between different variable combinations than the Census Bureau can provide in its standard products or that are provided in the AASHTO CTPP data tables.

Transportation planners and researchers have found PUMS to be especially useful for the following types of analyses:

- Cross-tabulations of variables not readily available from CTPP;
- Cross-tabulations of variables in CTPP but with more currency;
- Disaggregate statistical analyses;
- Comparisons of different regions;
- Comparisons over time; and
- Validation of other data sources.

These tabulations and analyses frequently support focused studies, such as analyses of the commuting characteristics of specific population subgroups or the demographic characteristics of commuters by mode.

The PUMS data are also used to support travel surveys and travel demand models. PUMS data are used in the planning, design, expansion, and validation of HH travel surveys, and also provide input data for the development and validation of state-of-practice travel demand model subcomponents. In recent years, PUMS data have been used to support the development of HH composition and auto availability submodels, trip generation models, and external trip models. PUMS data also provide base year information for travel model validation and checking. Finally, as the developers of advanced travel demand models and integrated transportation land use models are relying on microsimulation techniques to a greater extent, the usage and importance of PUMS data have increased. All of the activity-based travel demand models that have been developed or are being developed in the United States rely on PUMS data as a key input into the population synthesis module of their model systems.

Over time, as more agencies increase the range of transportation planning analyses and travel demand modeling capabilities, it is likely more transportation planners will need to develop expertise in PUMS.

ENRICHED CENSUS DATA FROM IPUMS: MICRODATA, TIME SERIES, AND GIS DATA

Jonathan Schroeder

Integrated PUMS (IPUMS) is made available through a web interface (www.ipums.org). The website includes IPUMS microdata for the U.S. Decennial Censuses (1850–2010); ACS (2000–2016); samples from Puerto Rico (1910–2016); and complete count datasets for 1790–1840 households and 1850, 1880, 1910–1940 individuals and households. Still to be added are data from 1860, 1870, and 1900. The ACS microdata samples include: the 1-year 1% samples since 2005 (although 2000–2004 are smaller and limited to 1-year samples) and the

5-year 5% samples since 2005–2009. Suppression for confidentiality includes names, and addresses. Income variables are treated with top coding (the top value published for a variable) and there are geographic limitations.

The geography in the PUMS includes regions, divisions, states, and PUMAs. PUMAs must have at least 100,000 residents (the 2010 average was 131,000 with a maximum of 269,000). PUMAs have been used since 1970 (previously referred to as "county groups" in both 1970 and 1980). The 1970 units have more than 250,000 residents. IPUMS has also defined 1960 PUMAs and "mini-PUMAs" that have more than 50,000 residents. There are several recognized problems with PUMAs. They have limited spatial precision and are not consistent with counties, cities, or metro areas. The boundaries are revised after each Census and there is a change in the ACS PUMAs between 2011 and 2012. This change has resulted in inconsistencies within the 5-year sample.

The IPUMS–USA geographic resources include supplementary variables, based on PUMAs for counties, cities, metro areas, and metro status. "ConsPUMAs" are sets of PUMAs with consistent extents over time. There are GIS shapefiles and online maps available for PUMAs, Migrations, Place of Work PUMAs, and ConsPUMAs. Detailed documentation and composition files are available for users.

There are recognizes problems with place-of-work PUMAs as well. There is limited spatial precision and boundaries are not consistent with counties, cities, metro areas, etc. Boundaries are revised after each Census and there was a change in the ACS PUMAs between 2011 and 2012. This is similar to the problem with PUMAs, however, it is worse for place-of-work PUMAs based on reductions. For the 2000 to 2011 period, there were 2,071 PUMAs and 1,238 person-weight (PW) PUMAs, but for 2012 to 2016, there were 2,351 PUMAs and only 980 PW PUMAs.

There are also issues with geographically standardized time series. For example, there is data for 1990, 2000, and 2010 for 2010 units. There are 10 geographic levels including states, counties, tracts, block groups, county subdivisions, places, congressional districts, CBSAs, urban areas, and ZCTAs. There are approximately 1,600 time series in 109 tables. The "short form" counts are available only for race, ethnicity, age, sex, HH size and relationships, housing occupancy and tenure, but not for income, education, employment, and other variables.

Nominally, integrated time series are available for approximately 5,700 times series in 271 tables. Eight geographic levels available include nation, regions, divisions, states, counties, tracts, county subdivisions, and places. The time span covers 1970 through 2010 with "total population" back to 1790 and "persons by sex" back to 1820. The "long-form" tables use 2008 through 2012 ACS.

The National Historical Geographic Information System (NHGIS) has some unique features including historical Census tables and GIS files, time series tables, block data, universal data filtering and selection (e.g., all years, levels and data types accessible at one time), nationwide extent available for all levels, and agriculture, businesses, religious bodies, and more. NHGIS is available at no cost.

YEAR-TO-YEAR CHANGES IN COUNTY-TO-COUNTY COMMUTE PATTERNS: LESSONS FROM THE AMERICAN COMMUNITY SURVEY PUBLIC USE MICRODATA SAMPLE

Charles Purvis

The single-year products of the ACS provide important data on large area, county-to-county (or county groups) commute patterns. The standard tables from American FactFinder provide data on worker characteristics: by county-of-residence, by county-of-work, and intracounty workers. Data from the PUMS is available for PUMA-of-residence to the county (or county groups) of work, essentially a PUMA-to-county commuter flow that can be reduced to a county-to-county level.

PUMS data from the 2006– 2015 ACS was analyzed for California, with focus on year-to-year patterns within and between the San Francisco Bay Area and the rest of California. Earlier ACS PUMS files (through 2011) were tabulated using Census 2000-based PUMAs. In California, there were 233 PUMAs-of-residence and 71 PUMAs-of-work (Census 2000-based). There are 58 counties in California. Recent ACS PUMS files (2012 through the present) were tabulated using Census 2010-based PUMAs. In California, there are 265 PUMAs-of-residence and 41 PUMAs-of-work.

Details on commuters can be derived from PUMS data, including income, earnings, auto ownership, means of transportation to work, commute duration, worker industry and occupation, and race or ethnicity. Analyzing year-to-year changes in commute patterns is even more vital given the significant, year-to-year changes in the economy over the 2006–2015 decade. Commute patterns based on the 5-year CTPP cannot be used to tease out the detailed changes occurring throughout the decade. Use of replicate weights are used in terms of describing relevant, significant changes in patterns. This approach may be of great use in other very large metropolitan areas (New York and Washington, D.C.), but may be of lesser value for very large metropolitan areas with mega-counties (Los Angeles and Chicago).

To work with PUMS, start with the full 1-year ACS data from the American FactFinder in the following tables:

- Table B08007: county-of-residence, intra-county, intra-state, total;
- Resident workers;
- Table B08501: county-of-work (i.e., the workplace county);
- Table B08008: place-of-residence, intra-place, total resident workers; and
- Table B08501: place-of-work (i.e., workplace city/CDP).

The annual ACS sample size is between 1.45% and 1.78% of the population. From the 5-year ACS, add the following tables:

- 2009–2013 5-year ACS county-to-county commuting flows and
- 2006–2010 5-year ACS county-to-county commuting flows.

From the general Census, use Census 2000, Census 1990, Census 1980 (from your agency's UTPP), and Census 1970 (from your agency's UTPP).

More information and guidance is available at https://www.Census.gov/topics/employment /commuting/guidance/ or search "Guidance for Commuting Data Users: Commuting Flows."

Next, you will need the 1% Annual PUMS including:

- PUMA = residence PUMA (defined areas of 100,000 + population);
- POWPUMA = place-of-work PUMA;
- ACS PUMS 2005–2011 = Census 2000-based 5% PUMAs; and
- ACS PUMS 2012–2016 = Census 2010-based PUMAs.

You will need to concatenate State + POWPUMA codes.

For California PUMAs and POWPUMAS, there are 58 counties, 24 counties in seven multi-county PUMAs and 34 counties with one-or-more PUMAs. The goals is to produce a 41-to-41 matrix of counties-to-counties. Figure 11.1 compares California PUMS and POWPUMAs from 2000 to 2010 Census. In the San Francisco area, the Metropolitan Transportation Commission, the Association of Bay Area Governments, and local planners designed the Bay Area PUMAs with encouragement from the California State Data Center.

For the 2000 Census, there were 13 POWPUMAs in Los Angeles County: Lancaster City, Palmdale City, Santa Clarita City, El Monte City, Pomona City, East Los Angeles CDP, Inglewood City, Torrance City, Long Beach City, West Covina City, Downey City, Norwalk City, and "balance of Los Angeles County" (i.e., Los Angeles City and other unincorporated and incorporated places). However, in the 2010 Census, there was only one POWPUMA for Los Angeles County. Looking more closely at the Bay Area (Figure 11.2), data is available for the following:

- Intraregional (9-by-9) commuting: 1970–2016.
- Interregional (18-by-18) commuting: 1980–2016:
  – Interregional county-to-county commuting first available in the 1980 UTPP and
  – Interregional tract-to-tract commuting first available in the 1990 CTPP.
- Bay Area Counties: San Francisco, San Mateo, Santa Clara, Alameda, Contra Costa, Solano, Napa, Sonoma, and Marin.
  - Bay Area Neighbor Counties:
    – Mendocino + Lake, Yolo, Placer, Sacramento, San Joaquin, and Stanislaus and
    – Merced, Monterey + San Benito, and Santa Cruz.

Which Census?	ACS Years	PUMAs-of-Residence	POWPUMAs-of-Work
Census 2000	2005-2011	233	71
Census 2010	2012-2016	265	41

FIGURE 11.1 California PUMAs and POWPUMAs.
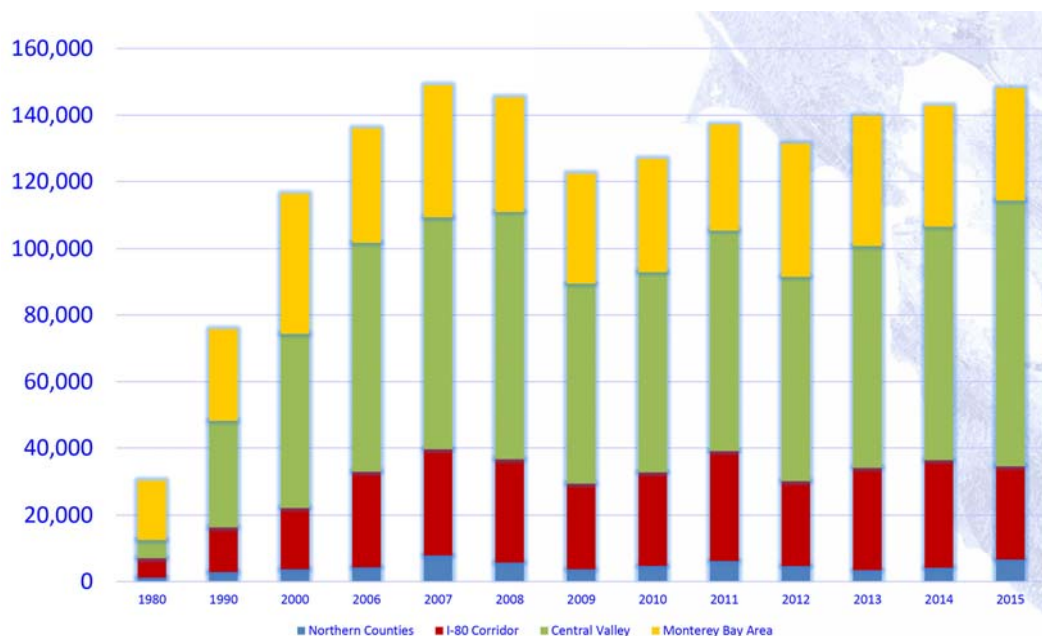
FIGURE 11.2 Total workers in-commuting to the San Francisco Bay Area, 1980–2015.

To deal with the replicate weights in PUMS to estimate standard error (SE) and coefficient of variation (CV), use:

- PWGT = person weight in PUMS.
- PWGT1 through PWGT80 = replicate weights in PUMS.
- Previous PUMS did not have replicate weights.
- Sum up the PWGT, PWGT1–PWGT80 in standard stat package.
- Calculate other variables in spreadsheets.

The key statistics to keep include estimate (e.g., total workers); sample size; average weight (estimate/sample size); sum of squared differences (PWGT less PWGT<n>); variance (previous calculation × 4, then divided by 80); SE (square root of variance); CV (SE divided by estimate); and MOE (90% or 95%).

It is important to pay attention to small sample sizes. In addition, if the CV is high (e.g., > .05), then flag for conditional formatting and add the footnote: "values are based on a very small sample sizes, analysis with caution." If the CV is too high, consider collapsing the data by grouping counties into corridors (recalculate estimates, SE, and CV). Be careful when describing issues to strictly identify the use a very small sample that will be accurate, but less precise, and avoid terms such as unreliable, inaccurate, bad, or warning not to use the analysis.

For more information see https://Censusmaven.wordpress.com, the Commuting to Silicon Valley (blog post); and https://Censusmaven.wordpress.com/2017/09/07/commuting-to-silicon -valley-part-2/.

CHAPTER 12

Transportation Modeling

JIM HUBBELL
MARC, presiding

JILAN CHEN
LIYANG FENG
Southeast Michigan Council of Governments

WILLIAM WOODFORD
Resource Systems Group

JAMES RYAN
Federal Transit Administration

SAM GRANATO
Ohio Department of Transportation

Census data products have always been integral parts of transportation models. Even as the models evolve and address new issues and problems, Census data can still be found at their core.

SYNTHESIZED TRAVEL MODEL INPUT AND ACS DATA CONSISTENCY CHECK: SEMCOG'S PRACTICE AND EXPERIENCE
Jilan Chen and Liyang Feng

The SEMCOG modeling program uses UrbanSim synthesized HH and population as a part of social-economic input data in its 2015 travel model development. The synthesizer used 2011–2015 5-year ACS data as a starting point, and then adjusted to the 2015 single-year HHs and population numbers for marginal controls. Other data, including Regional Economic Models, Inc., labor participation rate and regional unemployment rate, were also contributing factors in adjusting synthesized data for final travel model input. In addition, Census data was used to develop worker–income relationship to improve workers' employment location choices.

For travel model calibration, SEMCOG uses a top-down approach. The approach looks regional patterns first, county-level second, along with corridor-level calibrations. Naturally, this input data validation study applied the same hierarchy. Since the travel model is mainly used for forecasting purpose, short- and longer-term trends of different variables were also reviewed.

The research first explored HHs, population, and average HH size at the regional level using synthesized data, and then reviewed inconsistencies in the ACS. These inconsistencies were mainly caused by independently estimated HH and population data sets. Single-year ACS data sets from 2005–2015 and SEMCOG historical trend were referenced for reasonableness check. Simple trendline regression analysis is conducted to check data variations. In additional to aggregated average HH size, the distribution of HH size at both regional and county levels was

explored. Similar verification could be conducted on other variables, such as distribution on number of autos per HH, total number of children, etc., since these variables have significant impacts on travel model calibration and forecasting results. This approach enhanced staff understanding on both ACS variables and travel modeling, and advantages and shortcomings of the population synthesizer. Finally, the study outcome was adopted to enhance the synthesized data, and, in turn, an improvement on travel forecast capability is expected.

ROLE OF CENSUS DATA IN FTA'S SIMPLIFIED TRIPS-ON-PROJECT SOFTWARE
William Woodford and James Ryan

The Simplified Trips-on-Project Software (STOPS) is a key element of efforts by the FTA to streamline its Capital Investment Grant program. Transit agencies seek these grants to help fund fixed-guideway transit projects. Before STOPS, project sponsors were required to develop elaborate ridership forecasting models to quantify the mobility benefits of their proposed projects. These models often required years of data collection, development, and validation before they were ready to support the FTA's project evaluation process. Starting in 2011, FTA developed an alternative forecasting approach that relies on readily available national and local data to predict project ridership and mobility benefits. The centerpiece of STOPS is the development of matrices of O-D travel flows for all trips in a metropolitan area—separately for automobile, transit, and nonmotorized travel modes and work and nonwork trip purposes. Traditional forecasting models develop these matrices using behavioral choice models that weigh the attractiveness of alternative destinations and modes for each traveler.

While these models conform to theory on the distribution of travel throughout a region, they often generate travel patterns somewhat different from actual observations. Instead of synthesizing these patterns from scratch, STOPS builds off of JTW travel flow information from the CTPP. The relationship between the JTW and transit trips is strongest for commute-related travel but calibration against local transit survey information helped to establish in STOPS the relationship between these flows and nonwork travel as well. STOPS then refines these estimated transit flows with local transit count data to improve the fit. STOPS is now in use by transit agencies across the country to generate plausible ridership forecasts for nearly 100 projects in weeks rather than years. The outcome is made possible by the CTPP's large sample of JTW travel that effectively represents travel flows for all travelers at a high level of geographic precision. FTA provides aggressive technical support to STOPS users and to upgrade its capabilities as new methods and additional data sources become available.

The challenge to generate realistic projections of transit ridership requires reliable information on the number of trips between different origin and destination locations and the time and other impedances required to use each mode for each combination of origins and destinations. The solution is to start with each agency's computerized representation of its electronic schedule (GTFS), the same data used by online apps to suggest transit routings and travel times. The next step is to build O-D paths to identify the individual routes and stations involved, similar to the online apps.

The CTPP is crucial to understanding travel demand and exists throughout the United States with no new local data collection required. The tabulations provide a large sample data source for real O-D travel patterns for automobile, nonmotorized, and transit travel. It is usable

as a direct data input rather than a basis for model calibration that preserves the complexity of real travel patterns rather than creating an idealized or simplified model outputs.

CTPP data can be translated into travel demand. Starting with home-to-work travel, it can be used to represent ~50% of a transit market. CTPP Part 3 represents a solid foundation of total (all modes) trip making and transit travel. It is possible to generate a simple-trips-per-worker conversion factors. Other home-based travel represents 40% of transit market and can be scaled from home-to-work travel and other similar economic factors that affect other travel (except "special markets").

Nonhome-based travel represents ~10% of transit market. It can be scaled from home-to-work travel and from the number of workers arriving at each location. Nonhome-based trip patterns by arriving workers similar to trip patterns of residents. STOPS was originally developed using 2000 CTPP. It was later extended to 2006–2010 5-year ACS. The ACS results appear to be as good,or better than CTPP 2000 as they are more recent and have more detailed TAZ definitions.

To inform the model with these new data sources, it is possible to implement initial model with generic parameters from NCHRP and National Highway System (NHS) trip rates and transit routing parameters that general realistic paths. This represents conventional forecasting practice. The model can be tuned to match transit usage information from survey in six cities. In addition, the model performance was confirmed with information from twelve other cities. This resulted in STOPS understanding of observed traveler reactions to new fixed guideways. Figure 12.1 lists rider survey data by location.

Automatic adjustments of STOPS predictions of current ridership patterns were matched to actual transit usage in any particular city. Based on rider count data for individual routes, rail stations, and bus stops. This step is crucial for establishing model credibility for local decision-making (Figure 12.2).

Metro area	Comm. rail	Heavy rail	Light rail	Streetcar	BRT	Total
Atlanta		1				1
Charlotte *			1			1
Denver *			1			1
Phoenix			1			1
San Diego	1		2			3
Salt Lake City*	1		1		1	3
Subtotal	2	1	6	0	1	10

FIGURE 12.1 Systems with transit rider survey data. (*Indicates survey data on ridership both before and after recent project openings.)
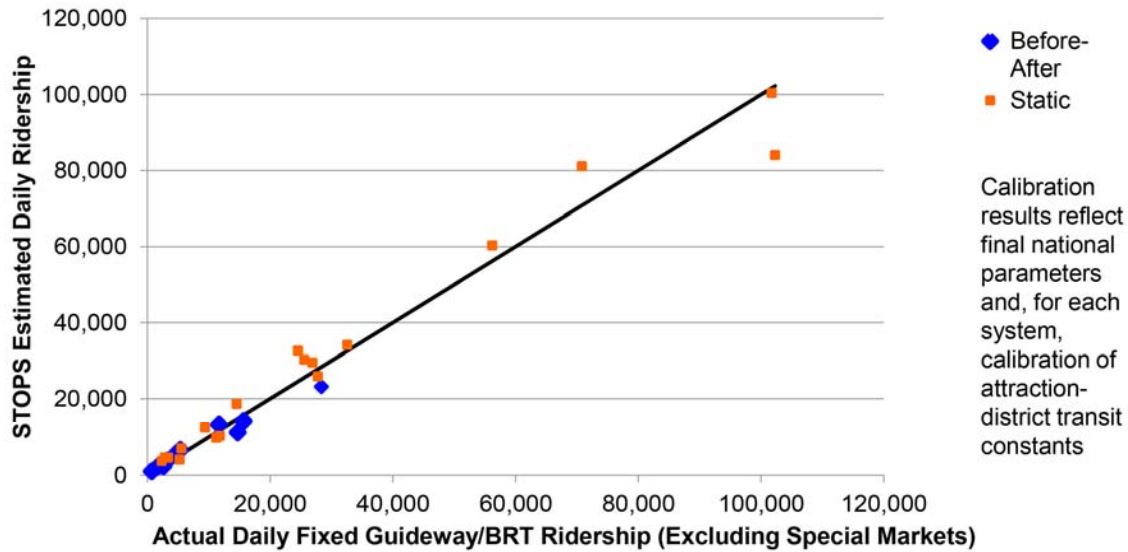
FIGURE 12.2 National calibration results.

STOPS has proved to be a big success as a new model can be implements in less than 2 weeks (compared to 6 months to years) and the results of the models are almost always plausible. STOPS forecasts for projects that have already been built are well within the FTA expectations for an indicator of project mobility benefit. The FTA is using these forecasts to make project-funding recommendations and the market has responded. Over 100 projects have adopted STOPS for forecasts and the FTA continues to provide technical support and training of the STOPS user community. STOPS project is only possible because of the CTPP due to its characteristics as a large sample size data set, the geographic location of worker residence and employment sites, the indication of transit reliance from auto ownership data, and the indication of transit usage from mode usage data. As a result, FTA is relying on CTPP to evaluate project mobility benefits for its capital investment program using STOPS.

USE OF TIME OF ARRIVAL AT WORK DATA FOR DYNAMIC TRAFFIC ASSIGNMENT (AND OTHER SUB-DAILY) TRAVEL MODELS

Sam Granato

For several decades, the CTTP has provided information on time of arrival at work, disaggregated geographically as far as the TAZ level, and derived from the long form–ACS questions on time of departure to work and WTT. This has allowed for its usage in travel demand models that seek to depict differences by area (or other characteristics) in the travel pattern by time of day in order to more accurately depict current travel flow and congestion patterns.

This research examines how such data was used in developing travel demand models that use Dynamic Traffic Assignment (DTA) for MPOs where the time of day pattern for home-to-work travel was varied by zone or district. This includes:

- Grouping of TAZs to focus on absolute differences from the areawide travel pattern by critical hours of day (where the amount of data available from the Census was extensive enough).
- Why alternate means of focusing on time-of-day differences (such as self-reported industry of worker) were considered but not utilized (at least yet).
- Adjustments made to conform to the travel patterns apparent from hourly traffic counts.

The existence in the ACS questionnaire of this question of time leaving home for work has been considered politically sensitive and scrutinized and in 2014 was included in a formal review of questions for either revision or elimination. It is shown here that there is at least some "constituency" for this type of question within the travel modeling profession, and it is anticipated that this question will be retained in future ACS surveys—albeit revised to ask directly about arrival at work instead. (Considerable differences were found between the responses to this question and local traffic counts during the a.m. peak period, suggesting some degree of carelessness or "resistance" to answering these questions is present locally as well.)

DTA-based travel modeling is similar to the "4-step process," except that a full day is broken into time intervals with variable trip start times within intervals. Travel paths can change mid-trip with spillover across intervals. The approach accounts for time-dependent network or traffic management and "traveler attributes." It incorporates saturation flows dynamic and set to lower initial values in off-peak periods based on driver and purpose characteristics. It can include deterministic (*Highway Capacity Manual* operational) intersection controls. This type of model uses trip assignment to time interval link-by-link based on the point in time that the link's "A-node" is reached. To understand the use of the data, it is necessary to view the format that is provided. For areawide use, a default set of trip percentages by hour of the day (by direction) has been available from the National Cooperative Freight Research Program (NCFRP) report of transferable parameters for travel models.

Initial adjustments can be made for an areawide sample MPO. The time-of-arrival figures from the Census provides the initial local area update of the NCHRP-based table (Home-Based Work—From Home), and then local traffic count data (both areawide and location specific) can be used to provide final adjustments. In fact, both the CTPP's figures and the local traffic counts can be used to make zone-specific adjustments as well as regionwide average values. A further adjustment can be made to zone-specific hour-of-day rates from CTPP can be developed proportionally. And then adjustments can be made for site-specific traffic counts.

How are the differences in time of arrival to work at traffic zone level being used? Time of arrival at work by zone in (mostly) 15- and 60-min time intervals can be illustrated on maps. These distributions are used to place zones into different "groups" for hourly rates for work-related trip purposes. These distributions re based on the number of workers arriving in that zone in the areawide peak hour compared to the average rate. The application of these calculations are estimations for planning purposes how the duration as well as the extent of congestion within the region could change in the future, as land use changes or capacity or operations-level projects may (or may not) get implemented to manage this. In addition, PUMS can be used to estimate the percent of responses that are "imputed"—in Ohio in 2000, 14% of time leaving for work, 11% for WTT, and 15½% either one or the other of these questions.

Keeping the "Census Data" Relevant

GREG ERHARDT
Department of Civil Engineering, University of Kentucky, author

STACEY BRICKA
MacroSys Research and Technology, presiding

PENELOPE WEINBERGER
AASHTO, recording

The data landscape is changing in terms of both data availability and the demands for new and more types of data. New data sources such as mobile devices, GPS, social media, and crowdsourcing expand the possibilities of data collection and analysis. The paper explored how Census data (CTPP, ACS, and LEHD) relate to these emerging and evolving data sets. Will Census data stay relevant? Can Census data be combines or integrated with these private data sets? Can Census data answer the policy questions of tomorrow?

UNDERSTANDING THE ROLE AND RELEVANCE OF THE CENSUS IN A CHANGING TRANSPORTATION DATA LANDSCAPE
Gregory D. Erhardt and Adam Dennett

The data landscape is changing in terms of both data availability and the demands for new and more types of data. New data sources such as mobile devices, GPS, and sensor data expand the possibilities of data collection and analysis. Using a review of recent literature as a starting point, this paper explores how Census data relate to these emerging and evolving data sets for transportation planning applications. It identifies areas where one or the other is used more commonly, and areas where they are complimentary, and finds that the Census data remain relevant, especially for the demographic and socioeconomic context they provide and for their universal availability.

The paper goes on to consider the prospects for keeping the Census data relevant to transportation planning, in the face of challenges such as the changing nature of mobility and of work, as well as opportunities to expand the role and relevance of Census data. It considers the results of a recent evaluation of the future of the United Kingdom Census and the overlap of the issues faced by the U.S. Census. The paper considers strategies to be considered for keeping the Census relevant, which are offered as a range of visions that the Census could take. The authors suggestion against the "give up and go home" strategy, and urging the Census Bureau, transportation planning organizations, and universities to continue their historic role of providing data as a public resource.

Introduction

The U.S. Census has long been an important data source for transportation planning and forecasting. The population and housing data provide the basis for populating TAZs; demographic and socioeconomic data are used to understand the effects of transportation projects on different populations; JTW data provide insight into commute patterns, mode shares, and the demand for transportation; and LEHD data provide consistent estimates of employment throughout the United States.

Transportation planning also has a long history of leveraging other data as a complement to the Census, including household travel surveys, traffic counts, transit ridership counts, state employment records and local land use data. More recently, a new generation of data have come online, and transportation planners have started developing methods to capture and use these so-called "Big Data."

Big Data include a range of sources that are typically passively collected, meaning that they emanate from sensors, transactions, or administrative records without the need for an active response on the part of the participant. In transportation, these include data such as transit automated vehicle location and automated passenger count data; transit farecard transactions; electronic toll transponder transactions; GPS traces from commercial vehicle movements; and trip tables derived from mobile phone data. These data offer several advantages over traditional travel surveys and Census data, including potentially much larger sample sizes, potential cost savings, and the ability to better measure changes due to their continuous nature. Big Data, however, brings its own set of challenges and limitations. Of note are the fact that the biases inherent in the data are often unknown, and that the data often excludes contextual information, such as demographics and socioeconomics, that can be included in an active data collection scheme. For these reasons, and due to the relative immaturity of the Big Data field, Smith (2013) argues for a hybrid approach that draws from the best aspects of each, while Johnson and Smith (2017) suggest that Big Data is viewed best as a supplement to, not a substitute for, traditional surveys.

This paper examines the relationship between Census data and emerging Big Data sources in the context of transportation planning, and considers the ways in which they serve as substitutes versus complements. It does this through a semistructured literature review that identifies recent transportation planning papers and articles that reference either the Census or Big Data. The search reveals both overlapping and nonoverlapping topic areas, indicating some potential for competition versus complementarity in those topic areas. A subset of the literature is reviewed in more detail to better understand the uses and limitations of each type of data.

The U.S. Census is not unique in facing the emergence of new data and technology—other nations are faced with similar issues and opportunities. This paper reports the recommendations of a recent effort to modernize the U.K. Census, and considers the relevance of those recommendations to the United States.

The paper goes on to consider some key policy questions of the future, and how the existing Census data structure fits or does not fit with those questions.

Given this three-tiered foundation, a menu of options is offered for keeping the Census relevant to transportation planning. These options are segmented into a competition track and a complementarity track. With a single exception, the authors refrain from recommending a path forward, and instead offer the options with the hope of stimulating a debate about the future of the Census.

Emerging Data Sources and Their Relationship to the Census

To identify areas of overlap and nonoverlap between the uses of Census data and Big Data, we conducted a semistructured review to identify relevant literature.

The Transportation Research International Database (TRID) database was used as the search engine. TRID combines the records from TRB's Transportation Research Information Services Database and the Organisation for Economic Co-operation and Development's Joint Transport Research Centre's International Transport Research Documentation Database, providing an extensive database focused specifically on transportation research. The search was limited to articles and papers, published in English, within the planning and forecasting subject area. The date range was from 2008 through August 2017. Papers focusing on research conducted outside the United States are included in an effort to learn from the international experience.

Two separate searches were conducted, one for the keyword "Census", and one for the key word "Big Data." The Census search returned 513 articles and the Big Data search returned 232 articles. A third search, for "Census" and "Big Data" returned only five articles, constituting a subset of both. While it would be possible to expand the results by searching for specific types of data—such as "mobile phone" or "GPS"—the 232 articles retrieved provides a sufficient basis for identifying the themes discussed in this paper.

Keyword Analysis

To get a sense of the topic areas that are prominent in the research, the key words from each of the 740 (513 + 232 – 5) articles returned from either search were tabulated. Supplemental Table 1 shows the frequency of each keyword in the Census search and in the Big Data search. Only the 253 keywords (out of 1,727 total keywords) used by more than five articles are shown. Each keyword is categorized as high frequency or low frequency for each search, with high frequency defined as being used by more than five papers in that set of search results. This grouping allows us to identify which keywords have a high frequency in both searches, in just the Census search, in just the Big Data search, or in neither search. Those without a high frequency in either search are of little interest and are not examined further.

Table 13.1 shows the number of articles returned for each year in the searches. It is clear that Big Data is a recent trend, with few articles published prior to 2014, but the numbers growing to be on par with the number of Census articles by 2015. The number of Census articles also grows during this period, indicating that Big Data research is not necessarily detracting from research that uses Census data.

Table 13.2 shows the keywords that occur with high frequency in both searches, sorted by the total frequency. The keywords show a number of terms indicating a range of applications relevant to transportation planning, travel forecasting and travel behavior analysis, traffic and transit. These are areas where there is potential for Big Data to serve as a substitute for Census data, although the mere presence of the terms in both searches does not necessarily indicate that it is a substitute. It could also be that each is used for different specific applications, or each is used in a complementary way.

TABLE 13.1 Articles by Year for "Census" and "Big Data" Search Terms

Year	"Census"	"Big Data"
2017	35	33
2016	68	57
2015	57	59
2014	60	35
2013	60	11
2012	55	11
2011	52	4
2010	42	9
2009	46	8
2008	38	5
Total	513	232

Table 13.3 shows the Census dominant keywords. These are keywords that occur frequently in articles within the Census search, but infrequently in articles within the Big Data search. For parsimony, only the top 40 are shown. The top keyword in this group is "traffic counts". An inspection of the papers using this keyword reveals that they are traffic-related, but not obviously Census related. It appears that either there is an anomaly in the coding, or that these articles use the term in a different context.

The remaining keywords in the Census dominant group are all more logical, and correspond to obvious applications of Census data. "Commuting," "work trips," and "commuters" all refer to analysis using the JTW data. "Demographics," "socioeconomic factors," and "equity (justice)" all use data that are available in the ACS or the Census long form. This is important because a characteristic of Big Data is that while they often provide detailed trajectory information, they usually lack characteristics of the individual or the household. Therefore, the Census remains the best source of this information. A number of terms that also show up relate to land use and the built environment ("land use," "neighborhoods," "land use planning," and "residential location"), highlighting another area where the Census shines. A fourth theme that can be observed is several terms relate to nonmotorized travel ("bicycling," "bicycles," "walking," and "nonmotorized transportation"). This may be due to the limitations of Big Data in capturing nonmotorized travel; people do not (yet!) have sensors built into their bodies that allow them to be directly tracked, and mode inference from GPS traces remains difficult, although inroads are starting to be made in this area (Bolbol et al., 2012), and more recently by technology start-ups such as TravalAi in the United Kingdom.

TABLE 13.2 Keywords with a High Frequency in Both Searches

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Travel demand	84	21	105	High	High
2	Origin and destination	74	19	93	High	High
3	Data collection	46	39	85	High	High
4	Travel behavior	62	19	81	High	High
5	Public transit	57	19	76	High	High
6	Travel surveys	55	10	65	High	High
7	Mode choice	50	9	59	High	High
8	Case studies	34	22	56	High	High
9	Urban areas	44	11	55	High	High
10	Transportation planning	34	17	51	High	High
11	Travel time	29	16	45	High	High
12	Data analysis	15	30	45	High	High
13	Traffic data	24	20	44	High	High
14	Mobility	25	18	43	High	High
15	Geographic information systems	36	7	43	High	High
16	Travel patterns	26	16	42	High	High
17	Planning	36	6	42	High	High
18	Traffic flow	28	12	40	High	High
19	Traffic models	20	13	33	High	High
20	Traffic volume	27	6	33	High	High
21	Traffic congestion	16	14	30	High	High
22	Forecasting	22	8	30	High	High
23	Algorithms	13	15	28	High	High
24	Traffic forecasting	18	8	26	High	High
25	Global positioning system	10	13	23	High	High
26	Choice models	15	8	23	High	High
27	Freight transportation	14	9	23	High	High
28	Vehicle sharing	15	7	22	High	High
29	Simulation	13	9	22	High	High
30	Ridership	12	9	21	High	High
31	Optimization	10	10	20	High	High
32	Decision making	8	12	20	High	High
33	Sustainable development	12	7	19	High	High
34	Infrastructure	11	6	17	High	High
35	Traffic simulation	9	7	16	High	High
36	Route choice	8	8	16	High	High
37	New York (New York)	8	6	14	High	High
38	Urban transportation	8	6	14	High	High
39	Sustainable transportation	8	6	14	High	High

TABLE 13.3 Census-Dominant Keywords (Top 40)

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Traffic counts	147	0	147	High	Low
2	Commuting	52	1	53	High	Low
3	Demographics	49	2	51	High	Low
4	Socioeconomic factors	47	2	49	High	Low
5	Spatial analysis	41	5	46	High	Low
6	Accessibility	36	4	40	High	Low
7	Land use	39	1	40	High	Low
8	Households	33	3	36	High	Low
9	Work trips	33	1	34	High	Low
10	Mathematical models	30	3	33	High	Low
11	Bicycling	27	4	31	High	Low
12	Traffic estimation	25	4	29	High	Low
13	Census	29	0	29	High	Low
14	Neighborhoods	27	1	28	High	Low
15	Commuters	23	4	27	High	Low
16	Automobile ownership	24	3	27	High	Low
17	United States	22	4	26	High	Low
18	City planning	20	5	25	High	Low
19	Walking	23	2	25	High	Low
20	Surveys	19	4	23	High	Low
21	Modal split	20	3	23	High	Low
22	Microsimulation	18	4	22	High	Low
23	Trip generation	20	2	22	High	Low
24	Canada	20	2	22	High	Low
25	Land use planning	19	2	21	High	Low
26	Nonmotorized transportation	21	0	21	High	Low
27	Activity choices	15	4	19	High	Low
28	Metropolitan areas	17	2	19	High	Low
29	Annual average daily traffic	18	1	19	High	Low
30	Demand	13	5	18	High	Low
31	Trip matrices	14	4	18	High	Low
32	Estimation theory	16	2	18	High	Low
33	Residential location	16	2	18	High	Low
34	Equity (Justice)	17	1	18	High	Low
35	Location	17	1	18	High	Low
36	Regression analysis	17	1	18	High	Low
37	Methodology	12	5	17	High	Low
38	Bicycles	15	2	17	High	Low
39	Statistical analysis	13	3	16	High	Low
40	Networks	10	5	15	High	Low

TABLE 13.4 Big Data-Dominant Keywords

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Big data	2	42	44	Low	High
2	Intelligent transportation systems	2	26	28	Low	High
3	Data mining	5	14	19	Low	High
4	China	2	15	17	Low	High
5	Logistics	4	11	15	Low	High
6	Real time information	3	11	14	Low	High
7	Cellular telephones	5	8	13	Low	High
8	Information processing	5	6	11	Low	High
9	Smartphones	3	8	11	Low	High
10	Smart cards	3	7	10	Low	High
11	High speed rail	2	7	9	Low	High
12	Technological innovations	2	7	9	Low	High
13	Netherlands	2	6	8	Low	High
14	Supply chain management	0	6	6	Low	High

Table 13.4 shows the Big Data dominant keywords. There are a more limited number of these, and several are general terms ("Big Data," "data mining," "information processing," technological innovations"). "Cellular telephones," "smartphones," and "smart cards" refer to specific types of data that are increasingly common. The applications in this group ("intelligent transportation systems," "real-time information," "logistics," and "supply chain management") are distinct from the other groups and are more operational or logistical in nature.

Finally, it is interesting to note that "China" and the "Netherlands" are in the Big Data-dominant group, whereas the "United States" and "Canada" are in the Census-dominant group. This may reflect clusters of research, but it also may relate to the quality and availability of Census data in those countries.

While the keywords provide an overview of the themes in each category, they provide little depth. To better understand the applications and uses of Census data and Big Data, the most frequent keywords within each group were examined in more detail. "Data collection" and "Big Data" were excluded from this exercise as not meaningful in this context, and "socioeconomic factors" was excluded because it is similar to "demographics," which was already included. For each keyword considered, a subsearch was conducted for articles using that keyword. The titles and abstracts of articles in the subsearch were examined, and a single paper was selected to illustrate a theme from that subsearch. Each of those papers is reviewed here in further detail.

Overlapping Topic Areas

Table 13.5 shows a summary of the articles reviewed for the top keywords with a high frequency in both the Census and Big Data searches. Ten articles are included—one for Census and one for Big Data with each keyword considered. The table shows the search terms, the author and year, the title, the full set of keywords used by that paper, the types of data used, and some brief notes.

TABLE 13.5 Summary of Selected Papers for Top Keywords with High Frequency in Both Searches

Search Terms	Author/ Year	Title	Keywords	Data Used	Notes
Travel demand and Census	Yasmin, Morency, and Roorda, 2017	Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model	Activity based models, activity choices, Montreal (Canada), O-D, travel demand, validation	O-D survey, Canadian Census	Transfers TASHA from Toronto to Montreal. O-D and Census provide validation data.
Travel demand and Big Data	Huntsinger, 2017	The Lure of Big Data: Evaluating the Efficacy of Mobile Phone Data for Travel Model Validation	Big Data, cost effectiveness, data analysis, data collection, data quality, households, mobile telephones, travel demand, travel surveys, validation	Mobile phone data (Airsage), HH travel survey	Airsage only available at district-level, but good for district-to-district flows. Proprietary nature makes it hard to evaluate.
O-D and Census	Çolak, Alexander, Alvim, Mehndiratta, et al., 2015	Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations and Opportunities	Boston (Massachusetts), cellular telephones, O-D, Rio de Janeiro, Brazil, traffic data, travel behavior, trip purpose	Mobile phone data (raw), Census, HH survey, O-D survey	Mobile phone data processed into O-D matrices and expanded to Census, validated against surveys. Worked reasonably well.
O-D and Big Data	Allos et al., 2014	New Data Sources and Data Fusion	Bluetooth technology, data files, data fusion, GPS, O-D, smartphones, trip matrices	GPS data (Traffic Master), mobile phone data (Telefonica)	Passive data lacks segmentation and potentially biased, but big/ complete sample size.
Travel behavior and Census	Jacques and El-Geneidy, 2014	Does Travel Behavior Matter in Defining Urban Form? A Quantitative Analysis Characterizing Distinct Areas Within a Region	Census tracts, characterization, factor-cluster analysis, travel behavior, urban form	Canadian Census, GIS land-use, O-D survey, satellite images	Census provides housing and HH measures.

Continued on next page.

**TABLE 13.5 (*continued*) Summary of Selected Papers for Top Keywords with
High Frequency in Both Searches**

Search Terms	Author/ Year	Title	Keywords	Data Used	Notes
Travel behavior and Big Data	Chen et al., 2016	The Promises of Big Data and Small Data for Travel Behavior (AKA Human Mobility) Analysis	Big Data, cooperation, data files, disciplines, mobility, transportation planning, travel behavior	Mobile phone data (raw), Big Data in general	Scaling factors needed. Imputing modes is hard. Not clear what to validate against. Representativeness unclear. Longitudinal nature is an advantage.
Public transit and Census	Wang, Lu, and Reddy, 2013	Maintaining Key Services While Retaining Core Values: NYC Transit's Environmental Justice Strategies	CTTP, costs, EJ, factor analysis, impacts, LOS, New York City Transit Authority, public transit, routes, service changes, social values, transportation operations	2000 Census JTW, Census racial and income counts, trip planner (route schedules)	Evaluate equity of proposed service cuts.
Public transit and Big Data	Oort and Cats, 2015	Improving Public Transport Decision Making, Planning and Operations by Using Big Data: Cases from Sweden and the Netherlands	Case studies, data sources, decision making, Netherlands, planning, public transit, smart cards, Sweden, transit vehicle operations, vehicle positioning systems	Transit smartcard data, automated vehicle location data, automated passenger count data	Illustrates range of applications: planning, operations, ridership prediction, real-time information. Promise in combining data sources.
Travel surveys and Census	Clark et al. 2014	Life Events and Travel Behavior	Aged, bicycling, commuting, travel behavior, United Kingdom, urban areas	U.K. HH Longitudinal Study, U.K. Census	Longitudinal data overcomes many estimation limitations.
Travel surveys and Big Data	Vij and Shankari, 2015	When Is Big Data Big Enough? Implications of Using GPS-Based Surveys for Travel Demand Analysis	Data files, data quality, errors, GPS, San Francisco (California), statistical inference, travel demand, travel diaries, travel surveys	HH travel survey, GPS-based travel survey	Higher volume of GPS data is often offset by lower quality due to limits of inferring mode, purpose, etc.

Within the subsearch on travel demand and Census, a number of the papers focus on travel demand model validation, followed by population synthesis, cycling, and O-D matrix estimation. Yasmin, Morency, and Roorda (2017) transfer the TASHA activity-based travel demand model from Toronto to Montreal, and use a combination of O-D survey data and Canadian Census data to validate the transferred model. This aligns with our own experience using U.S. Census JTW data and auto ownership data to validate travel models.

The subsearch on travel demand and Big Data includes substantial topical overlap with the travel demand and Census search. Core topics include using Big Data to validate travel models and for O-D matrix estimation, as well as one paper demonstrating the use of Big Data to estimate travel models. Huntsinger (2017) evaluates the effectiveness of Airsage mobile phone data for validating travel models. She compares the data to a HH travel survey for the same region. The comparison is necessary because the proprietary (black box) nature of Airsage makes it difficult to evaluate otherwise. The data comes in the form of district-to-district trip tables. It lacks the detailed travel characteristics and demographics of the survey, but due to the large sample size excels in the role of providing district-to-district flows.

Allos et al. (2014) examine the process of creating O-D matrices from GPS traces and mobile phone data in the United Kingdom. They report that the passive data provides a big/complete sample size, but lacks segmentation by purpose or income and is potentially biased. The potential for bias is important, with other research showing a transit smartcard data set to be biased against low-income and minority travelers, which can be problematic from an equity standpoint (Erhardt, 2016b).

The travel behavior articles are more diverse. Within the Census subsearch, urban form and transit-oriented development are a common theme. Carsharing and activity patterns also come up repeatedly. The Big Data and travel behavior subsearch includes several conceptual papers on how Big Data can be used and some on tracing travel patterns. Jacques and El-Geneidy (2014) study the effects of different urban forms, using the Canadian Census, among other sources. Chen et al. (2016) offer a review of Big Data applications, arguing for stronger collaboration between traditional transportation planners and computer scientists and physicists doing Big Data research. Their review highlights several advantages and limitations of Big Data, noting that imputing modes is difficult, the representativeness of the data is unclear, and it is not clear what to validate against. On the other hand, the longitudinal nature of Big Data offers a clear advantage that is often not available in traditional data.

For the public transit and Census subsearch, commuting, accessibility, and EJ emerged as core themes. Wang, Lu, and Reddy (2013) demonstrate a method of evaluating the equity of proposed service cuts using transit schedule data in combination with the Census JTW.

The public transit and Big Data papers were split between the use of smartcard data, conceptual papers on the value of Big Data, and approaches to imputing modes and walk distances. Oort and Cats (2015) illustrate a range of applications using smartcard data, automated vehicle location data and automated passenger count data. They note that the greatest promises of Big Data lie in combining multiple data sources.

The travel surveys and Census subsearch largely includes methodology papers for how to conduct travel surveys and analysis of travel survey data. An interesting application considers the effect of life events on travel behavior, using the U.K. Census and the U.K. Household Longitudinal Study (Clark et al. 2014). This paper demonstrates how longitudinal data can be used to overcome some of the limitations of cross-sectional data, such as self-selection bias and colinearity among certain variables.

The travel surveys and Big Data subsearch includes papers that discuss the strengths and limitations of Big Data and their value for travel model validation. Vij and Shankari (2015) examine GPS-only HH travel surveys where mode, purpose and other attributes are imputed from the GPS traces, in comparison to travel surveys that ask for those attributes explicitly. They find that

"In many cases, gains in the volume of data that can potentially be retrieved using GPS devices are found to be offset by the loss in quality caused by inaccuracies in inference. This study makes the argument that passively collected GPS-based surveys may never entirely replace surveys that require active interaction with study participants."

Census Dominant Topic Areas

Table 13.6 shows the papers reviewed within the Census-dominant topic areas. In this table, only papers from the Census subsearch are included.

Traffic counts was the most frequent keyword in the Census search. The papers within the traffic counts subsearch include a number of bicycle-related papers, as well as some about estimated AADT and others about O-D matrix estimation. The relevance to the Census is not immediately obvious for many, indicating either a possible anomaly in the keyword coding or an alternative use of the term Census. For example, the reviewed paper relates short-term bicycle counts to continuous bicycle counts for the purpose of estimating annual average daily bicycle traffic. While it does not use Census data, it is relevant with respect to the expansion of Census bicycle commute mode shares to annual totals.

Commuting is a common application of Census data, split between an analysis of mode shares and commuter patterns. Wang (2017) presents an interesting example that considers cohort changes in commute mode shares using IPUMS. The research demonstrates that it is valuable to be able to match data sets across time in a consistent format and with consistent data fields. Likewise, several national-level studies show up in this subsearch, highlighting that it is important to have consistent data across cities. This was also a theme to emerge from a recent workshop on the future of travel forecasting (Walker, 2017): that in order to advance our knowledge as a field we need data and models that are developed across multiple cities.

Both demographics and socioeconomic factors are common keywords within the Census dominant group. The demographics keyword includes papers on aging populations, spatial distributions, equity and car sharing. Tyndall (2017) illustrates several of these by studying the equity of carsharing, with respect to the demographics of the neighborhoods where the cars are located. They use a Big Data source from the carshare company to identify the car locations, but rely on Census data to understand the neighborhood demographics.

A number of papers also use Census data to study spatial effects. Often this applies to electric vehicles, urban form, or neighborhood characteristics. Liu, Roberts, and Sioshansi (2017) consider spatial effects on the adoption of hybrid electric vehicles, using a combination of Census, ACS, and state vehicle registration data.

Accessibility is becoming an increasingly important performance metric. Accessibility measures the ease of access to destinations, as opposed mobility, which measures the ease of movement. Owen and Levinson (2017) develop a comprehensive transit accessibility database. They use the GTFS for transit schedules, and the LEHD as a spatially detailed measure of employment.

TABLE 13.6 Summary of Selected Papers for Census Dominant Keywords

Search Terms	Author / Year	Title	Keywords	Data Used	Notes
Traffic counts and Census	El Esawey, 2016	Toward a Better Estimation of Annual Average Daily Bicycle Traffic	Adjustment factors, bicycle traffic, bicycles, traffic counts, traffic estimation	Automated bicycle counters (inductive loops)	Does not use Census data. Relevant to expansion of JTW bike mode shares.
Commuting and Census	X. Wang, 2017	Peak Car in the Car Capital? Double-Cohort Analysis for Commute Mode Choice in Los Angeles County, California, Using Census and ACS Microdata	ACS, carpools, Census, cohort analysis, commuting, demographics, forecasting, Los Angeles County (California), microdata, mode choice, PUMS, single-occupant vehicles	Integrated PUMS from 2000 Census and 2009–2011 ACS	Demographic data is important, as is the ability to match across multiple data sets for trend and cohort analysis.
Demographics and Census	Tyndall, 2017	Where No Cars Go: Free-Floating Carshare and Inequality of Access	Demographics, equity (justice), free-floating carsharing, location, mobility, mode choice, urban areas, vehicle sharing	Carshare location data (Car2Go), ACS	Big Data tells half the story, and is referenced to ACS demographics to understand equality considerations.
Spatial analysis and Census	Liu, Roberts, and Sioshansi, 2017	Spatial Effects on Hybrid Electric Vehicle Adoption	Adoption models, demographics, hybrid vehicles, neighborhoods, peer groups, spatial analysis, spatial effects	Census, ACS, Ohio vehicle registration data	Spatial distribution of demographic and socioeconomic factors is important.
Accessibility and Census	Owen and Levinson, 2017	Developing a Comprehensive U.S. Transit Accessibility Database	Accessibility, Alachua County (Florida), GIS, methodology, transportation disadvantaged persons	GTFS, LEHD	Accessibility is an increasingly important performance measure. Value in national consistency and availability of LEHD.

Big Data Dominant Topic Areas

Table 13.7 shows a summary of the papers reviewed within the Big Data-dominant topic areas.

The most common keyword among the Big Data-dominant topics is intelligent transportation systems. The papers in this area are focused on operational applications and on methods development. Xiao, Liu, and Wang (2015) develop a platform that combines a range of freeway-related data for performance management and operational analysis.

TABLE 13.7 Summary of Selected Papers for Big Data-Dominant Keywords

Search Terms	Author / Year	Title	Keywords	Data Used	Notes
Intelligent transportation systems and Big Data	Xiao, Liu, and Wang, 2015	Data-Driven Geospatial-Enabled Transportation Platform for Freeway Performance Analysis	Data analysis, data sharing, freeways, geospatial analysis, performance measurement, statistical analysis	Roadway geometric data, loop detector data, Bluetooth data, INRIX speed data, incident data, weather data, freeway travel time	Largely operational applications, and for performance management.
Data mining and Big Data	Zhang, Zhan, and Yu, 2017	Car Sales Analysis Based on the Application of Big Data	Automobile industry, automobile ownership, Big Data, data analysis, information processing, manufacturing, sales	Scraped car sale data and reviews	Aimed at providing insight to car makers.
China and Big Data	Hao, Zhu, and Zhong, 2015	The Rise of Big Data on Urban Studies and Planning Practices in China: Review and Open Research Issues	Big Data, China, review, urban planning, urban studies	GPS, mobile phone data, smartcard data, points of interests, volunteered geographic information, search engine data, digital land use data, parcel data, road networks	Chinese language papers more likely to focus on plan making and management applications than English language papers.
Logistics and Big Data	Coyle, Ruamsook, and Symon, 2016	Weatherproofing Supply Chains: Enable Intelligent Preparedness with Data Analytics	Data analysis, logistics, supply chain management, weather conditions, weatherproofing	50-year weather database, daily retail sales data by store	Ensure products are on shelves when storm hits. Applications from DOT or emergency management perspective are reasonable.
Real-time information and Big Data	Fusco, Colombaroni, and Isaenko, 2016	Short-Term Speed Predictions Exploiting Big Data on Large Urban Road Networks	Bayes' theorem, floating car data, mathematical prediction, networks, neural networks, Rome (Italy), speed prediction models, time series analysis, traffic models, urban highways	Floating car data (GPS), network	Short-term operational focus.

Data mining shows up frequently as well and the papers are often focused on mining a specific data set. One example uses scraped car sales data to provide insight to car makers (Zhang, Zhan, and Yu, 2017).

China is among the top keywords in the Big Data-dominant search, with the papers showing a range of applications including for transit, traffic, high-speed rail, and methods, as well as a wide range of data sets. Hao, Zhu, and Zhong (2015) provide an extensive review of Big Data applications in planning practice in China. It is recommended reading for anyone who wants a good overview of the range of applications of Big Data to planning. They note that Chinese language papers are more likely to focus on plan making and plan management than English language papers. It is interesting to consider why that may be: it could be a different research focus, that China lacks the same availability of other data sets, or that there are institutional differences in the planning structure that make Big Data more relevant.

Papers with the logistics keyword generally focused on supply chains, freight transportation or railroads. Coyle, Ruamsook, and Symon (2016), for example, considers the issue of delivering adequate supplies to stores prior to a coming storm.

Papers with the real-time information keyword are generally about traffic flow, speed predictions, or methodological developments. Fusco, Colombaroni, and Isaenko (2016) use GPS floating car data for short-term traffic predictions.

Common Themes and Observations

Several themes and observations emerge from the above review:

- There is substantial overlap between the use of Census data and the use of Big Data. The greatest overlap occurs in areas related to transportation modeling and public transit. Often, Census data and Big Data are used in combination, with the Census serving as a basis for expansion, or providing demographic and socioeconomic information.
- Census data remain the dominant source of demographic and socioeconomic information, as well as a widely available and widely used source of commute data.
- Big Data dominant topics tend to focus on shorter-term operational, traffic, and logistics issues.
- Due to their large sample sizes, Big Data also excel as the basis for generating O-D matrices.
- Big Data tend to be much less rich than survey data or Census data in terms of information content per observation. They generally lack information on demographics, household composition, trip purpose, mode, etc.
- The methods for inferring mode, purpose and other attributes from GPS or mobile phone traces remains weak, and the errors can offset the value of the additional observations.
- The quality of Big Data and the biases inherent in those data are often unknown and difficult to assess. This is especially true when commercial data are purchased, since the methods used in processing those data are often proprietary. This makes it especially important to have some external data source that they can be expanded to or validated against.
- Longitudinal data can overcome important limitations of cross-sectional data sets and open up new applications.
- The availability of Big Data remains sporadic, and even as they become more widely available, there is a risk that "data monopolies" will result in high prices (Erhardt, Batty, and

Arcaute, 2018). In contrast, the Census remains a widely available public resource, and the consistency across cities is important to allowing larger-scale analyses.

Beyond 2011: The Future of the U.K. Census

In the United Kingdom, the decennial Census (which is actually comprised of three separate Censuses with some country-specific questions asked in England and Wales, Scotland, and Northern Ireland and separate statistical authorities governing the collection and dissemination of the data) has captured information on the residential and workplace addresses of respondents since 1921 (Office for National Statistics 2012). From this locational information, estimates of the JTW have been derived and are available to access in digital form as O-D matrices dating back to 1981 (U.K. Data Service, 2017).

Information on the JTW is derived from the home address of Census respondent and then, historically, a question relating to their place of usual work. In 2011, JTW statistics were joined by "journey-to-learn" statistics relating to students and their location of educational establishment.

One of the major advantages of the Census travel-to-work data over any other measurement of commuting (apart from it being free to use and open) is its coverage. It is a legal requirement to complete a Census return in the United Kingdom and in 2011 a national 94% response rate was achieved (Office for National Statistics, 2017c), meaning that even before estimation and imputation, nearly all geographic and demographic dimensions of the population were covered. This is clearly a significant benefit to anyone using the data for travel-to-work analysis, as volumes and close to the full range of O-Ds are well represented. Taking advantage of this feature of the data, for a number of decades now, travel-to-work areas have been defined using these flow data for the purpose of local labor market analysis and statistical reporting leading to policy decisions made by the Department for Work and Pensions in relation to out-of-work benefits.

Clearly, however, Census travel-to-work data is not without its issues. Aside from the well-established issues such as errors in recording peripatetic working/other irregular travel to work patterns and timeliness, O-D data contain no routing information or detail on modal shifts and reveal little about other important travel activities not associated with work (such as shopping, school runs, and leisure). All of this means that researchers are starting to explore the potential of other datasets in conjunction with Census data to enhance our understanding of travel patterns.

Work is underway to determine whether detailed route and mode data captured continuously from a mobile application can be used to validate modeled detailed journal estimates using Census O-D data (Innovate UK, 2017). The smartphone application TravelAi (http://www.travelai.info/) provides recommended routing across travel modes, but also monitors the location of the user to provide that data to transportation agencies. The Office for National Statistics in England and Wales are also actively looking at the potential of other mobile telephone related data for mobility–transportation research (Office for National Statistics, 2017b). They propose to evaluate the comparability of flows derived from mobile telephone data and those estimated from the Census, however this is no indication of whether any headway has been made with this as yet.

The last Census in 2011 cost the U.K. government around £480 million to run (Office for National Statistics, 2017a), which despite being a very low cost per capita over the 10-year life

span of the data, contributed the opening of a conversation on whether the Census is still value for money or even necessary in a world where alternative population data exist amongst the myriad of administrative, commercial and survey datasets now in existence. There is no constitutional requirement for a Census to take place in the United Kingdom and the Beyond 2011 program explored the potential for replacing all of part of the Census using these data sources, as well as other options such as short-form and rolling Censuses.

After an extensive research and consultation period, the National Statistician recommended that the 2021 Census would be a full Census, however the data collection methods would be entirely online (Office for National Statistics, 2014). This approach eliminates the need to post paper forms out to households, the feature of previous Censuses that had created the most cost. The National Statistician also recommended that the 2021 Census feature an increased use of administrative data and surveys to enhance statistics from the Census and improve statistics between Censuses. The report recommended against an approach that eliminated the Census and instead used only administrative data to construct population statistics. While other countries successfully use such an approach, those countries have a population register, which the United Kingdom does not. A population register is a centralized data system for recording, and keeping current, vital statistics for all residents of a country (United Nations Statistical Office, 2014). Such registers are common in northern Europe, with the vital statistics recorded typically including births, deaths, marriages, name changes, and other changes of interest. Assuming it is accurate, a population register would make the Census function of counting people unnecessary because the register contains that count, although address and other attributes may or may not be recorded. The administrative data approach was viewed as a risky endeavor without a population register. The government accepted the recommendation, but expressed interest in moving towards an administrative approach in the future (Maude, 2014).

The Policy Questions of Tomorrow

When planning on 10-year Census timeframes, it is valuable to consider not just competing and complementary data sources but also how the relevant policy questions may change over those timeframes. This section discusses policy areas that should be on Census planners' minds. It does not suggest that these issues are resolved, or will definitively come to be—just that they are questions worth grappling with.

The Future of Mobility

The past several years have seen both the rise of new shared-mobility modes, and massive investment in developing the technology of self-driving cars.

Over the past decade, advances in payment and smartphone technology have enabled new uses for old transportation modes. The literature review above has already identified carsharing as a mode of interest (Tyndall, 2017), but bike sharing systems have proliferated as well (Shaheen et al., 2012). The option to share vehicles has the potential to reshape decisions about owning a vehicle and the demand for parking (Martin, Shaheen, and Lidicker, 2010).

Transportation Network Companies (TNCs), such as Uber and Lyft, also represent a re-invention of an old mode. TNCs allow a user to book and pay for a ride with a smartphone app, with the ride delivered by an independent driver in their personal vehicle. At current rates, the cost to the user is generally much lower than a taxi, and some drivers prefer the convenience of

the app and payment system. They did not exist a decade ago (Uber was founded in 2009), but they are no longer a niche mode, at least in major cities. In San Francisco, for example, TNCs make over 170,000 vehicle trips within the city, which is approximately 12 times the number of taxi trips, and 15% of all intra-San Francisco vehicle trips (San Francisco Count Transportation Authority, 2017).

TRB Special Report 319 identifies, but does not resolve, many of the policy questions related to shared mobility and technology-enabled transportation services (Transportation Research Board, 2016b). Among these are questions of regulation, safety, and security, the impact on congestion and transit ridership, equity of access, and the effects on the labor market.

In the future, drivers themselves may become unnecessary. Both technology companies and traditional automakers are investing billions of dollars in developing self-driving cars or autonomous vehicles. The prospects and timeframe for broad adoption of the technology remain uncertain (Litman, 2014; Bansal and Kockelman, 2017; Rohr et al., 2016), but the implications for the transportation system and transportation policy are profound (Fagnant and Kockelman, 2015; Anderson et al., 2014). The effects depend in part on how they are used. Will households replace their personal vehicles with self-driving cars? Will they be used as fleet vehicles by TNCs? Perhaps they will first become common for freight transportation, as opposed to personal travel? These are important questions that transportation planners must grapple with, and as the technology emerges, it is important to have the data to understand these trends.

The Future of Work

The future of mobility highlights issues related to the future of work that extend beyond transportation.

Arguably, TNCs biggest innovations have happened not in transportation, per se, but in the labor market. Special Report 319 (TRB, 2016b) considers these employment and labor issues. Drivers are not treated as employees, but as independent contractors who own and maintain their own vehicles, pay for their own health insurance and manage their own payroll/self-employment taxes. This represents an important shift from a traditional employer–employee relationship, with looser ties between the two. There are implications not only on the levels of net compensation, but also brings potential for less regularity of working hours, lower stability of employment, a higher share of part-time works, an increased ability to engage in multiple jobs, and a decreased stability of employment. It is easy to see how these trends may extend beyond transportation to a wide range of jobs, and it is sometimes referred to as the "gig economy". From a transportation perspective, such a situation is very different than commuting to regular shift work.

Self-driving cars and trucks may have an even bigger impact on labor markets. According to the Bureau of Labor Statistics, the United States has 1.8 million heavy truck drivers, 1.3 million delivery truck drivers, 665,000 bus drivers, and 233,000 taxi and chauffer drivers. As self-driving vehicles emerge, it is logical to expect that these workers will be displaced, that the cost to consumers of delivering goods is reduced, and that the firms that own the vehicles see their profit margins increase. These trends are likely to increase income and wealth inequality in the United States. As drivers are pushed out of regular employment, they may also engage in the gig economy, accentuating the trends discussed above.

It easy to dismiss such concerns as speculation, and future employment is indeed difficult to forecast, but Vardi (2017) argues that the future is already here. He notes the combination of

high manufacturing output with low manufacturing employment and stagnant wages over the past several decades. While it is difficult to pinpoint the exact reasons for such trends, increasing automation is likely a contributing factor.

While Vardi uses an example of the shift from horse-powered transportation to automobiles a century ago, a better analogy may be the rise of containerization 40 years ago. Containerization dramatically reduced the labor involved in shipping, greatly reducing its cost. Beyond the direct labor market implications, this contributed to the rise of global trade, a major shift in the nature of our nation's ports, and the repurposing of waterfront areas and entire neighborhoods in many cities.

The shifting nature of work and shifting mobility options may also contribute to regional disparities in several dimensions. While TNCs and bike sharing systems are popular in large cities, they are most effective when combined with a certain level of density. It is easy to envision fleets of autonomous cars shuttling people around Pittsburg (as Uber is doing today) or San Francisco, but their market may be more limited in the smaller cities in Kentucky where auto ownership is higher and the distances are greater. Changes in labor markets and employment are likely to be geographically uneven, and there is evidence that people are less likely to move to follow jobs than in the past (Cooke, 2013; Molloy, Smith, and Wozniak, 2017).

Long-Distance Travel: A Policy Question of Today

Rather than a policy question of tomorrow, accommodating the demand for long-distance travel is a commonly overlooked policy question of today. In the United States, personal vehicle trips longer than 50 mi account for 2% of total trips, but 23% of VMT. While a precise estimate of resources is not available, it is clear that long-distance travel commands far less than 23% of the effort involved in transportation planning, data and forecasting. In spite of the fact that there appears to be a renewed call for spending billions of dollars on intercity high-speed rail every few years, and the huge portion of our roadway system dedicated to intercity travel, the data and resources available for long-distance planning are woefully inadequate, as illustrated by the reliance of recent long-distance models on either the 2001–2002 NHTS, or the 1995 American Travel Survey (Moeckel, Fussell, and Donnelly, 2015; Outwater et al., 2015).

The TRB Executive Committee recognized this deficiency and commissioned *Special Report 320: Interregional Travel: A New Perspective for Policy Making* (Transportation Research Board, 2016a). Two of the reports key findings are especially noteworthy here:

> "Because of outdated travel behavior survey data, long-distance travel is not nearly as well understood as local travel."

> "To encourage the development of urban transportation systems that are integrated and function well across a metropolitan region, the federal government has long required state and local authorities to coordinate their urban highway and transit investments. The goal of this coordination, which is often challenging to implement, is to guide transportation investments from a multimodal and multijurisdictional perspective that is informed by sound data and objective analysis. Because interregional travel corridors often span multiple states, many lack the coordinated planning and funding structures needed to ensure that investments in transportation capacity are made from a corridor-level perspective."

In other words, there is both a lack of reliable data on the topic, and a challenge in overcoming the institutional and jurisdictional coordination problems associated with investing in new data.

Options for Keeping the Census Relevant

In this final section, we consider several options for keeping the Census data relevant to transportation planning. These are grouped into two general tracks: the competition track considers strategies where the Census data is directly competing against other data sources, while the complementary track considers strategies associated with identifying a unique niche for the Census to fill. The first three strategies constitute the competition track, while strategies four through seven are on the complementary track.

Strategy 1: Give Up and Go Home

Strategy 1 is based on the premise that emerging Big Data sources are becoming so good and so cheap that they are making the Census obsolete. This represents a vision of the future (or the present) where technology is so omnipresent that our every movement is recorded in a database, where it is linked to every credit card purchase we have ever made, every social media comment we have ever posted, and a facial-recognition database of every photo we have ever been in. This is a vision of total knowledge, where it is unnecessary to ask about travel behavior because we already know the answers.

In such a future, the Census may very well become obsolete. Even in a world that only partially approximates this vision, it may seem a reasonable strategy to decide that the Census is irrelevant and to no longer use it. There are two problems with this strategy.

First, it is clear from the literature review above that, regardless of grand visions for where the world may be heading, we are not nearly at the point where Big Data can be considered "all knowing." The Big Data studies identified above are limited in scope to specific applications and specific geographies. They often have limitations and biases that arise from the way the data are collected, such as the tendency of transit smart card data to underrepresent minority and low-income travelers (Erhardt, 2016b). Those biases and limitations can be difficult to detect and evaluate, especially when the methods are not fully transparent (Huntsinger 2017), and those data limitations can easily offset the value of a larger sample size (Vij and Shankari 2015). For these reasons, it is common for Big Data to be used in combination with Census data or other actively collected data, as illustrated by many of the studies cited above.

Second, to the extent that such a vision of the future is viable, it is much closer to reality in the private sector than it is for transportation planners shaping public infrastructure, services and policy. Technology companies are in a position to invest heavily in acquiring data resources and the computing infrastructure necessary to support them, and to hire talented engineers and computer scientists. They also operate with a different set of political and legal constraints than the public sector—what may be viewed as inappropriate government intrusion in Washington might be perfectly acceptable in Silicon Valley. Will the role of transport planners be that of a customer purchasing these data? Or will it be to work with the companies providing transportation (via self-driving vehicles) to develop optimization strategies to regulate traffic volumes along routes so that congestion is avoided and efficiency maximized?

While these are reasonable and appropriate roles for transportation planners to play, there are risks in limiting the planning role in this way. One such risk is the danger of a "data monopolies" (Erhardt, Batty, and Arcaute, 2018). A data monopoly can occur when a single company has exclusive rights to all the data of a certain type or on a certain topic. In such cases, that company can exert control over the price, at the expense of those purchasing the data.

A second issue is that private-sector interests may or may not align with the interest of serving the public good. Shuldiner and Shuldiner (2013) consider how the public interest can be best served when the transportation data of greatest value is collected by private entities, and how the current situation differs from the historic development of transportation planning models based on public data. If the data show a picture of the real world that is inconsistent with a company's public image or corporate strategy, what incentive do they have to share those data? The Freedom of Information Act would not apply, so a company would be within its rights to filter the data that it releases.

The current experience with TNCs illustrates the types of issues that can arise. Uber has been in conflict with multiple cities over regulatory issues, most recently resulting in a Transport for London's (TfL's) decision ending its ability to operate in London (Rao and Isaac 2017). From its own operations, Uber has extensive data about travel in London that may be useful to planners at TfL, but it is not realistic to expect Uber to provide those data to planners at TfL while it appeals TfL's decision in court, nor is it realistic to expect planners at TfL to trust those data should they be made available.

Going beyond appropriate restrictions to protect privacy, which all good data stewards have an obligation to uphold, do we really want to put ourselves in a position where private interests can control and filter the data that shapes our understanding of the world? It is precisely to avoid this situation that there will always be a role for data as a public resource, and the authors urge the Census Bureau to continue its historic role providing this resource.

Strategy 2: Keep Calm and Carry On

The second strategy considered is for the Census to continue its transportation data program in its current state, a strategy we label "keep calm and carry on." The rationale for this strategy is that the review of research studies show that the Census is clearly continuing to play a role in transportation planning. In particular, it has an important role in providing context with respect to household, socioeconomic, and demographic characteristics, and it is often use in combination with Big Data as a basis for expanding or supplementing those data. The fact that it is universally available as a public data resource ensures that a wide variety of actors can each conduct independent analyses using these data, contributing to a diversity of ideas and viewpoints, and a rich environment for innovation.

This strategy may be combined with some minor adjustments to the existing approach. For example, the U.K. Census includes questions on the journey-to-school in addition to the JTW, and the American Census could benefit from the same. This would provide planners with more complete travel information, particularly in locations where colleges or universities are major attractors, such as Arizona State University, which contributes a substantial portion of ridership to Phoenix's light rail line (FTA, 2013).

It may also be beneficial to add questions designed to provide consistency with external data sources. For example, the Census asks about usual place of work and usual mode to work, whereas most travel surveys record the destination and mode of work commutes for a designated

travel day. This makes it difficult to compare the data between the two, and can be particularly important when reflecting the variability of travel, particularly for something like a bicycle commute, which can be affected by the weather (Nosal et al., 2015).

If such questions are added, they should be supplemental to, not in place of, the existing JTW questions. Consistency with past Census and ACS data is important to ensure that trends can be monitored cleanly.

Strategy 3: If You Can't Beat 'Em, Buy 'Em

The third strategy considered on the competition track is labeled, "if you can't beat 'em, buy 'em." The goal here is to use the relative advantages of similar data sources to get a more complete picture of the JTW.

Currently, the main advantage of mobile phone data is that the large sample size provides a strong basis for creating trip tables at a reasonable level of geographic detail. In contrast, the ACS JTW data becomes noisy for more-detailed geographies simply because there are a limited number of observations. The ACS data, however, provide more information than mobile phone trip tables, such as the usual mode to work and characteristics of the workers.

This strategy would involve purchasing mobile phone data for a region, specifically focusing on work commutes, which are expected to be the most reliable purpose that can be extracted due to their regularity. These data would be compared in detail to the Census JTW data, and the expansion factors would be adjusted for each to create a unified, best-estimate trip table. It is expected that this approach would be most effective if the adjustments could be made on disaggregate data, and then released as aggregate trip tables to protect privacy. Such an approach would require an appropriate licensing arrangement with the mobile phone data vendor, and if Census restricted data were to be used, it would need to be conducted in an established secure data center.

There are a few possible paths toward making this happen. One is for the Census bureau to do the analysis and expansion on their end, and then release it as part of the JTW data products. Alternatively, an arrangement can be made where the data vendors better incorporate the Census data into their own products. A third option would be to do the analysis as a post-processing step, starting from both sources. This third option could be done as a pilot test for a single region.

A more sophisticated approach would be to manage the integration as part of the data collection process, rather than after-the-fact. For example, when the ACS surveys a household, the questionnaire could ask for permission to access the mobile phone records for individuals in that household. Those data would be combined with the survey to improve the data quality. Alternatively, the mobile phone records could serve as a sampling frame for the survey, allowing for integration in that direction. The privacy elements of such an approach would need to be carefully managed.

Strategy 4: Administrative Integration

The "administrative integration" strategy draws from the future envisioned for the U.K. Census to integrate appropriate administrative data sets for the purpose of improving the Census. Already, the Census is doing this through its LEHD data product that integrates unemployment

insurance records, tax records, and other data to create spatially detailed estimates of employment, workers, and commute flows.

The types of data integrated could be expanded in several directions. Feeney et al. (2015) provide a useful overview of the types of administrative data that researchers have used in the past, and could potentially be integrated with Census transportation data. Some promising options include:

- Birth and death records for monitoring population changes.
- School enrollment data, both for primary schools and colleges and universities. Most institutions can be expected to have address lists of their students, which would be useful for developing journey to school matrices.
- School districting data, which may be valuable for restricting the journey to school matrices based on district boundaries.
- Incarceration records, representing a portion of the group quarters population that does not travel.
- Social Security Administration data, which can be used both as a means of merging data across multiple sources and as a means of linking age, income and retirement status.
- State vehicle registration data, as a means of linking auto ownership information.
- Utility records, particularly power usage data that potentially could be used to identify when a unit is occupied either seasonally or by time of day.
- Parcel data from county assessors' offices, which are already public, as a means of integrating land use.
- Transaction data from toll transponders, transit farecards, and similar transportation transactions.
- Credit report information, which is widely available, and could be used to infer information about income, housing tenure and vehicle ownership.
- Credit card transaction data, which may provide information on the location and type of purchases.

For some of these data, the value contributed may be outweighed by privacy concerns or by the trouble of compiling the data. It is, however, worth being deliberate in assessing that trade-off.

For the administrative integration strategy, the role of the Census Bureau (or another agency that took on the task), would be that of a data aggregator. As envisioned, it would

1. Gather disaggregate data from multiple jurisdictions.
2. Code the data to be as consistent as possible across jurisdictions, and merge them into a unified data set.
3. Link those data across types. For example, vehicle registration data could be linked to utility usage data, and parcel records to improve the estimates of car ownership currently included in the ACS.
4. Clean and check the unified data.
5. Aggregate them in such a way as to protect the privacy of individual records.

The data aggregator is able to add value both by working with disaggregate records, but keeping those records hidden behind a firewall, and by ensuring consistency across regions allowing for larger-scale analyses.

Strategy 5: Capture the Future

The "capture the future" strategy is aimed at adjustments to the Census JTW data collection to better reflect current, and possible future trends in mobility and work.

The key change needed for capturing the future of mobility is simply to expand the list of modes included in the JTW questionnaire. Already, the change in travel modes is prominent, at least in certain cities. For example, between 2005 and 2015, the ACS show that in San Francisco, the share of work commutes by taxi, bike, and other modes more than doubled, from 3.4% to 6.9% (Erhardt, 2016a). This represents a combination of what has been called a "bicycle renaissance" and (Pucher, Buehler, and Seinen, 2011) and the emergence of TNCs, which as of 2016 composed 15% of intra-San Francisco vehicle trips (San Francisco Count Transportation Authority, 2017). While still small shares relative to other modes, these are important trends in their own right. It would be valuable to split the other category to consider TNCs, or at least to clarify that they are included in the taxi category. Moving to the future, it would be valuable to break out autonomous modes, both in the commute mode choice questions and in vehicle ownership. This would be most effective ahead of the trend, such that the annual ACS data be used to monitor trends in those modes.

For the future of work, the key issue is how to account for informal and irregular work. Options here include the option to collect more than one workplace, with a usual mode associated with each, as well as further clarifying the definition of work. It may be that respondents have different understandings of whether a "gig" should be reported as work, leading to ambiguity in the responses. However it is counted, there is value in consistency.

Strategy 6: Go Long (Distance)

The "go long (distance)" strategy deviates from the focus on work commutes and considers an important, but neglected, travel market—long-distance travel. Because it spans state and municipal boundaries, it is important that long-distance travel data be collected at a national level. Extending the Census transportation data offerings could be a natural way to accomplish this.

Such a survey approach would likely be a retrospective question asking respondents to list long-distance trips made by members of their household in the last month. While the definition offered for long-distance trips can vary—often 50+ mi, sometimes 100+ mi—defining the question based on overnight trips would define a clean breakpoint for respondents in terms of identifying and remembering those trips. The information collected could be very simple and would include:

- Destination, recorded with city/county, state, and zip code;
- Mode of travel;
- Departure and return dates; and
- Purpose: business versus leisure.

Such a data set, collected across the country for a reasonably large sample size, would be a tremendous resource for this important component of travel demand.

Strategy 7: Go Long(itudinal)

The general lack of longitudinal data has been recognized as a limitation of transportation research for nearly 30 years (Kitamura, 1990). This is a problem because cross-sectional correlations among different variables can make it difficult to detect the effects of certain policy interventions or other changes. For example, a time-of-day model might wish to consider the effect of congestion on changes in the temporal distribution of trips. A model estimated from cross-sectional data would likely find that congestion is higher in the peak period, and people prefer to travel in the peak period, so more congestion would lead to a higher likelihood of traveling in the peak period. Of course, the directionality of this assessment is wrong, but the model estimation cannot distinguish that. Conversely, if longitudinal data were available where the same households were observed in subsequent years, the data and resulting models would correctly show that an increase in congestion between those two years would make the travelers in that household less likely to travel in the peak period.

There are a range of other examples that can be used to illustrate this effect, but the issue is that our interest as transportation planners extends beyond describing the state of the system as it is today. Our interest in transportation data is also in understanding the factors that cause the system to change, and applying that understanding to predict how the system will change in response to our interventions. For this purpose, cross-sectional data that does not observe change is inherently limited.

As discussed above, Big Data do offer some advantage in this area. Because they tend to be continuously collected, they provide an opportunity to measure change, which can be leveraged to measure the impacts of transportation projects (Erhardt, 2016a).

The ACS could evolve into a panel survey, where a portion of the households are resurveyed in subsequent years. The German Mobility Panel has taken this approach since 1994 (Weiss et al., 2017). In Germany, this approach has enabled a range of applications and analyses that otherwise would be difficult or impossible, such as assessing the individual-level stability in commute patterns (Hilgert et al., 2016) and studying the effect of life changes on travel behavior (Scheiner, Chatterjee, and Heinen, 2016). Together these provide a means of understanding the levers that can be used most effectively to induce changes in travel behavior.

Next Steps

This paper has found that in spite of the emergence of a variety of Big Data sources, the Census remains relevant to transportation planning. The paper considered the types of applications where one or another is more commonly applied, and found a large area of overlap where the two are used together as complementary data sources, even in studies that are labeled as "Big Data" studies. In spite of this relevance, the Census faces challenges in maximizing its relevance and value for transportation applications going forward, and these challenges are not unique to the American context. They include a natural desire for cost effectiveness, and the evolving nature of mobility and work. There are also opportunities, such as the dearth of long-distance and longitudinal data where the Census is in a position where it could step up to provide important resources.

Seven strategies are considered for keeping the Census data relevant to transportation planning. Three consider the Census' role in direct competition with Big Data, and four consider the ways in which it could be more complementary.

- Strategy 1: Give up and go home.
- Strategy 2: Keep calm and carry on.
- Strategy 3: If you can't beat 'em, buy 'em.
- Strategy 4: Administrative integration.
- Strategy 5: Capture the future.
- Strategy 6: Go long (distance).
- Strategy 7: Go long(itudinal).

Among the strategies considered, we advise against the "give up and go home" strategy, and we urge the Census Bureau, transportation planning organizations, and universities to continue their historic role of providing data as a public resource.

The remaining strategies are intended to provide a menu of options, which are not necessarily mutually exclusive. They will serve as a starting point for discussion at the TRB Conference on Applying Census Data for Transportation in Kansas City, Missouri, in November 2017. The authors hope that that discussion will continue in the broader community as we renew our effort to keep the Census relevant and valuable for transportation planning purposes.

Acknowledgments

References

Allos, A., A. Merrall, R. Smithies, R. Fishburn, J. Bates, and R. Himlin. 2014. New Data Sources and Data Fusion. European Transport Conference, Association for European Transport, 2014.

Anderson, J. M., N. Kalra, K. D. Stanley, P. Sorensen, C. Samaras, and O. A. Oluwatola. *Autonomous Vehicle Technology: A Guide for Policymakers*, 2014.

Bansal, P., and K. M. Kockelman. Forecasting Americans' Long-Term Adoption of Connected and Autonomous Vehicle Technologies. *Transportation Research Part A, Policy and Practice*, Vol. 95, 2017, pp. 49–63. https://doi.org/10.1016/j.tra.2016.10.013.

Bolbol, A., T. Cheng, I. Tsapakis, and J. Haworth. Inferring Hybrid Transportation Modes from Sparse GPS Data Using a Moving Window SVM Classification. *Computers, Environment and Urban Systems: Special Issue: Advances in Geocomputation*, Vol. 36, No. 6, 2012, pp. 526–537. https://doi.org/10.1016/j.compenvurbsys.2012.06.001.

Chen, C., J. Ma, Y. Susilo, Y. Liu, and M. Wang. The Promises of Big Data and Small Data for Travel Behavior (AKA Human Mobility) Analysis. *Transportation Research Part C, Emerging Technologies*, Vol. 68, 2016, pp. 285–299. https://doi.org/10.1016/j.trc.2016.04.005.

Clark, B., K. Chatterjee, S. Melia, G. Knies, and H. Laurie. Life Events and Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board, No. 2413*, 2014, pp. 54–64. doi:10.3141/2413-06.

Çolak, S., L. P. Alexander, B. G. Alvim, S. R. Mehndiratta, and M. C. González. Analyzing Cell Phone
    Location Data for Urban Travel. *Transportation Research Record: Journal of the Transportation
    Research Board*, No. 2526, 2015, pp. 126–135. https://doi.org/10.3141/2526-14.
Çolak, S., L. P. Alexander, B. G. Alvim, S. R. Mehndiretta, and M. C. Gonzalez. Analyzing Cell Phone
    Location Data for Urban Travel: Current Methods, Limitations and Opportunities. Presented at 94th
    Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
Cooke, T. J. Internal Migration in Decline. *The Professional Geographer*, September 2013.
    https://doi.org/10.1080/00330124.2012.724343.
Coyle, J. J., K. Ruamsook, and E. J. Symon. Weatherproofing Supply Chains: Enable Intelligent
    Preparedness with Data Analytics. *Transportation Journal*, Vol. 55, No. 2, 2016, pp. 190–207.
    https://doi.org/10.5325/transportationj.55.2.0190.
El Esawey, M. Toward a Better Estimation of Annual Average Daily Bicycle Traffic. *Transportation
    Research Record: Journal of the Transportation Research Board*, No. 2593, 2016, pp. 28–36.
    https://doi.org/10.3141/2593-04.
Erhardt, G. D. Fusion of Large Continuously Collected Data Sources: Understanding Travel Demand
    Trends and Measuring Transport Project Impacts. PhD thesis. University College London, 2016a.
Erhardt, G. D. How Smart Is Your Smart Card? Evaluating Transit Smart Card Data with Privacy
    Restrictions and Limited Penetration Rates. *Transportation Research Record: Journal of the
    Transportation Research Board*, No. 2544, 2016b, pp. 81–89. https://doi.org/10.3141/2544-10.
Erhardt, G. D., M. Batty, and E. Arcaute. Recommendations for Big Data Programs at Transportation
    Agencies. *Big Data for Regional Science* (L. A. Schintler and Z. Chen, eds.), Routledge, New York,
    2018.
Fagnant, D. J., and K. Kockelman. Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers
    and Policy Recommendations. *Transportation Research Part A, Policy and Practice*, Vol. 77, 2015,
    pp. 167–181. https://doi.org/10.1016/j.tra.2015.04.003.
Federal Transit Administration. Before-and-After Studies of New Starts Projects: Report to Congress,
    2013.
Feeney, L., J. Bauman, J. Chabrier, G. Mehra, K. Murphy, and M. Woodford. Catalog of Administrative
    Data Sets. J-PAL North America, 2015.
    https://www.povertyactionlab.org/sites/default/files/documents/AdminDataCatalog.pdf.
Fusco, G., C. Colombaroni, and N. Isaenko. Short-Term Speed Predictions Exploiting Big Data on Large
    Urban Road Networks. *Transportation Research Part C, Emerging Technologies*, Vol. 73, 2016, pp.
    183–201. https://doi.org/10.1016/j.trc.2016.10.019.
Hao, J., J. Zhu, and R. Zhong. The Rise of Big Data on Urban Studies and Planning Practices in China:
    Review and Open Research Issues. *Journal of Urban Management*, 2015.
    https://doi.org/10.1016/j.jum.2015.11.002.
Hilgert, T., C. Weiss, M. Kagerbauer, B. Chlond, and P. Vortisch. Stability and Flexibility in Commuting
    Behavior: Analyses of Mode Choice Patterns in Germany. Presented at 95th Annual Meeting of the
    Transportation Research Board, Washington, D.C., 2016.
Huntsinger, L. F. The Lure of Big Data: Evaluating the Efficacy of Mobile Phone Data for Travel Model
    Validation. Presented at 96th Annual Meeting of the Transportation Research Board, Washington,
    D.C., 2017.
Innovate UK. Local Authority Solutions for Integrated Transport—Innovateuk. Accessed August 21,
    2017. https://connect.innovateuk.org/web/local-authority-solutions-for-integrated-transport.
Jacques, C., and A. El-Geneidy. Does Travel Behavior Matter in Defining Urban Form? A Quantitative
    Analysis Characterizing Distinct Areas within a Region. *Journal of Transport and Land Use*, Vol. 7,
    No. 1, 2014, pp. 1–14. https://doi.org/10.5198/jtlu.v0i0.377.
Johnson, T. P., and T. W. Smith. Big Data and Survey Research: Supplement or Substitute? In *Seeing
    Cities Through Big Data*, 2017. https://doi.org/10.1007/978-3-319-40902-3_7.
Kitamura, R. Panel Analysis in Transportation Planning: An Overview. *Transportation Research Part A,
    General*, Vol. 24, No. 6, 1990, pp. 401–415. https://doi.org/10.1016/0191-2607(90)90032-2.

Litman, T. Autonomous Vehicle Implementation Predictions. *Victoria Transport Policy Institute*, Vol. 28, 2014. http://leempo.com/wp-content/uploads/2017/03/M09.pdf.

Liu, X., M. C. Roberts, and R. Sioshansi. Spatial Effects on Hybrid Electric Vehicle Adoption. *Transportation Research Part D: Transport and Environment*, 2017. https://doi.org/10.1016/j.trd .2017.02.014.

Martin, E., S. Shaheen, and J. Lidicker. Impact of Carsharing on Household Vehicle Holdings: Results from North American Shared-Use Vehicle Survey. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2143, 2010, pp. 150–158. https://doi.org/10.3141/2143-19.

Maude, F. Government's Response to the National Statistician's Recommendation, July 18, 2014. http://webarchive.nationalarchives.gov.uk/20160204145156/https://www.statisticsauthority.gov.uk/w p-content/uploads/2015/12/letterfromrthonfrancismaudemptosirandrewdilnot18071_tcm97- 43946.pdf.

Moeckel, R., R. Fussell, and R. Donnelly. Mode Choice Modeling for Long-Distance Travel. *Transportation Letters*, Vol. 7, No. 1, 2015, pp. 35–46. https://doi.org/10.1179/1942787514Y .0000000031.

Molloy, R., C. L. Smith, and A. Wozniak. Job Changing and the Decline in Long-Distance Migration in the United States. *Demography*, Vol. 54, No. 2, 2017, pp. 631–653. https://doi.org/10.1007/s13524- 017-0551-9.

Nosal, T., L. F. Miranda-Moreno, Z. Krstulic, and T. Götschi. Accounting for Weather Conditions When Comparing Multiple Years of Bicycle Demand Data. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

Office for National Statistics. Census 1911-2001. Accessed May 18, 2012. http://webarchive .nationalarchives.gov.uk/20160111030756/http://www.ons.gov.uk/ons/guide-method/Census/2011 /how-our-Census-works/about-Censuses/Census-history/200-years-of-the-Census/Census-1911- 2001/index.html.

Office for National Statistics. Frequently Asked Questions. 2017a. Accessed August 21. https://www.ons .gov.uk/Census/2011Census/2011Censusdata/2011Censususerguide/frequentlyaskedquestions.

Office for National Statistics. ONS Methodology Working Paper Series No. 8: Statistical Uses for Mobile Phone Data: Literature Review. 2017b. Accessed August 21, 2017. https://www.ons.gov.uk/methodology /methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpapers eriesno8statisticalusesformobilephonedataliteraturereview.

Office for National Statistics. Response Rates. 2017c. Accessed August 21, 2017. https://www.ons.gov.uk /Census/2001Censusandearlier/dataandproducts/qualityoftheCensusdata/responserates.

Office for National Statistics. The Census and Future Provision of Population Statistics in England and Wales: Recommendation from the National Statistician and Chief Executive of the UK Statistics Authority. United Kingdom, 2014. www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and- projects/beyond-2011/beyond-2011-report-on-autumn-2013-consultation--and-recommendations /national-statisticians-recommendation.pdf.

Oort, N. v., and O. Cats. Improving Public Transport Decision Making, Planning and Operations by Using Big Data: Cases from Sweden and the Netherlands. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp. 19–24. doi:10.1109/ITSC.2015.11.

Outwater, M., M. Bradley, N. Ferdous, C. Bhat, R. Pendyala, S. Hess, A. Daly, and J. LaMondia. Tour- Based National Model System to Forecast Long-Distance Passenger Travel in the United States. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

Owen, A., and D. M. Levinson. Developing a Comprehensive U.S. Transit Accessibility Database. *Seeing Cities Through Big Data*, 2017. https://doi.org/10.1007/978-3-319-40902-3_16.

Pucher, J., R. Buehler, and M. Seinen. Bicycling Renaissance in North America? An Update and Re- Appraisal of Cycling Trends and Policies. *Transportation Research Part A, Policy and Practice*, Vol. 45, No. 6, 2011, pp. 451–475. https://doi.org/10.1016/j.tra.2011.03.001.

Rao, P. S., and M. Isaac. Uber Loses License to Operate in London. *The New York Times*, September 22, 2017. https://www.nytimes.com/2017/09/22/business/uber-london.html.

Rohr, C., L. Ecola, J. Zmud, F. Dunkerley, J. Black, and E. Baker. Travel in Britain in 2035: Future Scenarios and Their Implications for Technology Innovation. RAND Corporation, 2016. http://www.rand.org/pubs/research_reports/RR1377.html. https://doi.org/10.7249/RR1377.

San Francisco Count Transportation Authority. TNCs Today: A Profile of San Francisco Transportation Network Company Activity, 2017. http://www.sfcta.org/sites/default/files/content/Planning /TNCs/TNCs_Today_061317.pdf.

Scheiner, J., K. Chatterjee, and E. Heinen. Key Events and Multimodality: A Life Course Approach. *Transportation Research Part A, Policy and Practice*, Vol. 91, 2016.

Shaheen, S. A., E. W. Martin, A. P. Cohen, and R. S. Finson. Public Bikesharing in North America: Early Operator and User Understanding. Mineta Transportation Institute, San José State University, 2012. http://transweb.sjsu.edu/PDFs/research/1029-public-bikesharing-understanding-early-operators-users.pdf.

Shuldiner, A. T., and P. W. Shuldiner. The Measure of All Things: Reflections on Changing Conceptions of the Individual in Travel Demand Modeling. *Transportation*, Vol. 40, No. 6, 2013, pp. 1117–1131. https://doi.org/10.1007/s11116-013-9490-5.

Smith, T. W. Survey-Research Paradigms Old and New. *International Journal of Public Opinion Research*, Vol. 25, No. 2, 2013, pp. 218–229. https://doi.org/10.1093/ijpor/eds040.

Special Report 320: Interregional Travel: A New Perspective for Policy Making, 2016a. doi:10.17226 /21887.

Special Report 319: Between Public and Private Mobility: Examining the Rise of Technology-Enabled Transportation Services, 2016b. doi:10.17226/21875.

Tyndall, J. Where No Cars Go: Free-Floating Carshare and Inequality of Access. *International Journal of Sustainable Transportation*, Vol. 11, No. 6, 2017, pp. 433–442. https://doi.org/10.1080/15568318 .2016.1266425.

U.K. Data Service. WICID: About the Data Sets. Accessed August 21, 2017. http://wicid.ukdataservice .ac.uk/cider/about/data_int.php?type=2.

United Nations Statistical Office. *Principles and Recommendations for a Vital Statistics System*. United Nations, New York, 2014.

Vardi, M. Y. Humans, Machines and the Future of Work. Presented at the College of Engineering Distinguished Lecture, University of Kentucky, Lexington, April 13, 2017.

Vij, A., and K. Shankari. When Is Big Data Big Enough? Implications of Using GPS-Based Surveys for Travel Demand Analysis. *Transportation Research Part C, Emerging Technologies*, Vol. 56, 2015, pp. 446–462. https://doi.org/10.1016/j.trc.2015.04.025.

Walker, J. Advancing the Science of Transportation Demand Modeling. National Science Foundation, Washington, D.C., 2017. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1648930.

Wang, T., A. Lu, and A. Reddy. Maintaining Key Services While Retaining Core Values: NYC Transit's Environmental Justice Strategies. *Journal of Public Transportation*, Vol. 16, No. 1, 2013, pp. 123–152. https://trid.trb.org/View/1248230?ajax=1. https://doi.org/10.5038/2375-0901.16.1.7.

Wang, X. Peak Car in the Car Capital? Double-Cohort Analysis for Commute Mode Choice in Los Angeles County, California, Using Census and ACS Microdata, 2017. https://trid.trb.org/View /1439710?ajax=1.

Weiss, C., B. Chlond, C. Minster, C. Jödden, and P. Vortisch. Assessing Effects of Mixed-Mode Design in a Longitudinal Household Travel Survey. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017

Xiao, S., X. C. Liu, and Y. Wang. Data-Driven Geospatial-Enabled Transportation Platform for Freeway Performance Analysis. *IEEE Intelligent Transportation Systems Magazine*, Vol. 7, No. 2, 2015, pp. 10–21. https://doi.org/10.1109/MITS.2014.2388367.

Yasmin, F., C. Morency, and M. J. Roorda. Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model. *Transportmetrica A: Transport Science*, Vol. 13, No. 3, 2017, pp. 222–249. https://doi.org/10.1080/23249935.2016.1249437.

Zhang, Q., H. Zhan, and J. Yu. Car Sales Analysis Based on the Application of Big Data. *Procedia Computer Science*, 2017. https://doi.org/10.1016/j.procs.2017.03.137.

Appendix

SUPPLEMENTAL TABLE 1: All Keywords with 6 or More Appearances in Search Results

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
1	Traffic counts	147	0	147	High	Low
2	Travel demand	84	21	105	High	High
3	Origin and destination	74	19	93	High	High
4	Data collection	46	39	85	High	High
5	Travel behavior	62	19	81	High	High
6	Public transit	57	19	76	High	High
7	Travel surveys	55	10	65	High	High
8	Mode choice	50	9	59	High	High
9	Case studies	34	22	56	High	High
10	Urban areas	44	11	55	High	High
11	Commuting	52	1	53	High	Low
12	Demographics	49	2	51	High	Low
13	Transportation planning	34	17	51	High	High
14	Socioeconomic factors	47	2	49	High	Low
15	Spatial analysis	41	5	46	High	Low
16	Travel time	29	16	45	High	High
17	Data analysis	15	30	45	High	High
18	Traffic data	24	20	44	High	High
19	Big data	2	42	44	Low	High
20	Geographic information systems	36	7	43	High	High
21	Mobility	25	18	43	High	High
22	Planning	36	6	42	High	High
23	Travel patterns	26	16	42	High	High
24	Land use	39	1	40	High	Low
25	Accessibility	36	4	40	High	Low
26	Traffic flow	28	12	40	High	High
27	Households	33	3	36	High	Low
28	Work trips	33	1	34	High	Low
29	Mathematical models	30	3	33	High	Low
30	Traffic volume	27	6	33	High	High
31	Traffic models	20	13	33	High	High
32	Bicycling	27	4	31	High	Low
33	Forecasting	22	8	30	High	High
34	Traffic congestion	16	14	30	High	High
35	Census	29	0	29	High	Low
36	Traffic estimation	25	4	29	High	Low
37	Neighborhoods	27	1	28	High	Low
38	Algorithms	13	15	28	High	High
39	Intelligent transportation systems	2	26	28	Low	High
40	Automobile ownership	24	3	27	High	Low
41	Commuters	23	4	27	High	Low

Continued on next page.

SUPPLEMENTAL TABLE 1 (*continued*): All Keywords with 6 or More Appearances in Search Results

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
42	United States	22	4	26	High	Low
43	Traffic forecasting	18	8	26	High	High
44	Walking	23	2	25	High	Low
45	City planning	20	5	25	High	Low
46	Modal split	20	3	23	High	Low
47	Surveys	19	4	23	High	Low
48	Choice models	15	8	23	High	High
49	Freight transportation	14	9	23	High	High
50	Global Positioning System	10	13	23	High	High
51	Trip generation	20	2	22	High	Low
52	Canada	20	2	22	High	Low
53	Microsimulation	18	4	22	High	Low
54	Vehicle sharing	15	7	22	High	High
55	Simulation	13	9	22	High	High
56	Nonmotorized transportation	21	0	21	High	Low
57	Land use planning	19	2	21	High	Low
58	Ridership	12	9	21	High	High
59	Optimization	10	10	20	High	High
60	Decision making	8	12	20	High	High
61	Annual average daily traffic	18	1	19	High	Low
62	Metropolitan areas	17	2	19	High	Low
63	Activity choices	15	4	19	High	Low
64	Sustainable development	12	7	19	High	High
65	Data mining	5	14	19	Low	High
66	Equity (Justice)	17	1	18	High	Low
67	Location	17	1	18	High	Low
68	Regression analysis	17	1	18	High	Low
69	Estimation theory	16	2	18	High	Low
70	Residential location	16	2	18	High	Low
71	Trip matrices	14	4	18	High	Low
72	Demand	13	5	18	High	Low
73	Bicycles	15	2	17	High	Low
74	Methodology	12	5	17	High	Low
75	Infrastructure	11	6	17	High	High
76	China	2	15	17	Low	High
77	Statistical analysis	13	3	16	High	Low
78	Traffic simulation	9	7	16	High	High
79	Route choice	8	8	16	High	High
80	Bicycle facilities	14	1	15	High	Low
81	Social factors	13	2	15	High	Low
82	Data quality	13	2	15	High	Low
83	Pedestrians	12	3	15	High	Low
84	United Kingdom	11	4	15	High	Low
85	Behavior	11	4	15	High	Low
86	Cluster analysis	11	4	15	High	Low

Continued on next page.

SUPPLEMENTAL TABLE 1 (*continued*): All Keywords with 6 or More Appearances in Search Results

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
87	Traffic assignment	11	4	15	High	Low
88	Trip length	11	4	15	High	Low
89	Networks	10	5	15	High	Low
90	Logistics	4	11	15	Low	High
91	Housing	14	0	14	High	Low
92	Conferences	13	1	14	High	Low
93	Automobile travel	12	2	14	High	Low
94	Built environment	12	2	14	High	Low
95	New York (New York)	8	6	14	High	High
96	Urban transportation	8	6	14	High	High
97	Sustainable transportation	8	6	14	High	High
98	Real time information	3	11	14	Low	High
99	Low income groups	13	0	13	High	Low
100	Vehicle miles of travel	12	1	13	High	Low
101	Multinomial logits	10	3	13	High	Low
102	Policy	9	4	13	High	Low
103	Cellular telephones	5	8	13	Low	High
104	Traffic count	12	0	12	High	Low
105	Rural areas	11	1	12	High	Low
106	California	11	1	12	High	Low
107	Transportation	11	1	12	High	Low
108	Transportation modes	10	2	12	High	Low
109	Traffic surveillance	9	3	12	High	Low
110	Validation	9	3	12	High	Low
111	Cyclists	9	3	12	High	Low
112	Logistic regression analysis	9	3	12	High	Low
113	Level of service	9	3	12	High	Low
114	Disaggregate analysis	10	1	11	High	Low
115	Aged	10	1	11	High	Low
116	Transit oriented development	10	1	11	High	Low
117	Conference	10	1	11	High	Low
118	Population	9	2	11	High	Low
119	Travel	8	3	11	High	Low
120	Network analysis (Planning)	7	4	11	High	Low
121	Performance measurement	6	5	11	High	Low
122	Multimodal transportation	6	5	11	High	Low
123	Logits	6	5	11	High	Low
124	Information processing	5	6	11	Low	High
125	Smartphones	3	8	11	Low	High
126	Montreal (Canada)	10	0	10	High	Low
127	Commodity flow	10	0	10	High	Low
128	Bicycle commuting	10	0	10	High	Low
129	Population density	10	0	10	High	Low
130	Employment	9	1	10	High	Low
131	Australia	9	1	10	High	Low

SUPPLEMENTAL TABLE 1 (*continued*): All Keywords with 6 or More Appearances in Search Results

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
132	Gender	9	1	10	High	Low
133	Policy analysis	8	2	10	High	Low
134	Modal shift	7	3	10	High	Low
135	Data fusion	6	4	10	High	Low
136	Trend (Statistics)	6	4	10	High	Low
137	Trip purpose	6	4	10	High	Low
138	Railroad transportation	5	5	10	Low	Low
139	Neural networks	5	5	10	Low	Low
140	Environmental impacts	5	5	10	Low	Low
141	Smart cards	3	7	10	Low	High
142	Jobs	9	0	9	High	Low
143	Minneapolis (Minnesota)	9	0	9	High	Low
144	Activity based modeling	9	0	9	High	Low
145	Monte Carlo method	9	0	9	High	Low
146	Land use models	8	1	9	High	Low
147	Accuracy	8	1	9	High	Low
148	Pedestrian safety	7	2	9	High	Low
149	Toronto (Canada)	7	2	9	High	Low
150	Pollutants	6	3	9	High	Low
151	Traffic distribution	6	3	9	High	Low
152	Calibration	6	3	9	High	Low
153	Costs	6	3	9	High	Low
154	Regional planning	5	4	9	Low	Low
155	Bus transit	5	4	9	Low	Low
156	Mathematical prediction	5	4	9	Low	Low
157	Stochastic processes	5	4	9	Low	Low
158	Strategic planning	5	4	9	Low	Low
159	Rail transit	4	5	9	Low	Low
160	Data files	4	5	9	Low	Low
161	Transit operating agencies	4	5	9	Low	Low
162	High speed rail	2	7	9	Low	High
163	Technological innovations	2	7	9	Low	High
164	Transportation disadvantaged persons	8	0	8	High	Low
165	Dublin (Ireland)	8	0	8	High	Low
166	Population forecasting	8	0	8	High	Low
167	Chicago (Illinois)	8	0	8	High	Low
168	Least squares method	8	0	8	High	Low
169	Immigrants	8	0	8	High	Low
170	Road networks	7	1	8	High	Low
171	Errors	7	1	8	High	Low
172	Central business districts	7	1	8	High	Low
173	Bicycle travel	7	1	8	High	Low
174	Freight traffic	6	2	8	High	Low
175	Traffic safety	6	2	8	High	Low

**SUPPLEMENTAL TABLE 1 (*continued*): All Keywords with 6 or
More Appearances in Search Results**

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
176	Highway traffic control	6	2	8	High	Low
177	Residential areas	6	2	8	High	Low
178	Arterial highways	6	2	8	High	Low
179	Systems analysis	6	2	8	High	Low
180	Beijing (China)	4	4	8	Low	Low
181	Mobile telephones	4	4	8	Low	Low
182	Sensors	4	4	8	Low	Low
183	Policy making	4	4	8	Low	Low
184	Netherlands	2	6	8	Low	High
185	Multivariate analysis	7	0	7	High	Low
186	Walkability	7	0	7	High	Low
187	Economic factors	7	0	7	High	Low
188	Synthetic populations	7	0	7	High	Low
189	Data acquisition	7	0	7	High	Low
190	Urban area	7	0	7	High	Low
191	Bayes' theorem	6	1	7	High	Low
192	Probits	6	1	7	High	Low
193	England	6	1	7	High	Low
194	Transport planning	6	1	7	High	Low
195	Energy consumption	6	1	7	High	Low
196	Databases	5	2	7	Low	Low
197	Pedestrian-vehicle crashes	5	2	7	Low	Low
198	Planning and design	5	2	7	Low	Low
199	Electric vehicles	4	3	7	Low	Low
200	Evacuation	4	3	7	Low	Low
201	France	4	3	7	Low	Low
202	Stated preferences	4	3	7	Low	Low
203	London (England)	3	4	7	Low	Low
204	Quality of service	3	4	7	Low	Low
205	Cities	3	4	7	Low	Low
206	Alternatives analysis	3	4	7	Low	Low
207	Hybrid vehicles	6	0	6	High	Low
208	Seasons	6	0	6	High	Low
209	School trips	6	0	6	High	Low
210	Ireland	6	0	6	High	Low
211	Freight Analysis Framework	6	0	6	High	Low
212	Carpools	6	0	6	High	Low
213	Peak hour traffic	6	0	6	High	Low
214	Multi-agent systems	6	0	6	High	Low
215	Traffic counting	6	0	6	High	Low
216	Links (Networks)	6	0	6	High	Low
217	Population synthesis	6	0	6	High	Low
218	Texas	6	0	6	High	Low
219	Loop detectors	6	0	6	High	Low
220	Hamilton (Canada)	6	0	6	High	Low

Continued on next page.

**SUPPLEMENTAL TABLE 1 (*continued*): All Keywords with 6 or
More Appearances in Search Results**

Rank	Keyword	Census Count	Big Data Count	Total Count	Census Category	Big Data Category
221	Peak periods	6	0	6	High	Low
222	Urban transportation policy	6	0	6	High	Low
223	Agent based models	6	0	6	High	Low
224	Geography	6	0	6	High	Low
225	Urban development	5	1	6	Low	Low
226	Light rail transit	5	1	6	Low	Low
227	Traffic analysis zones	5	1	6	Low	Low
228	Days	5	1	6	Low	Low
229	Markov chains	5	1	6	Low	Low
230	Signalized intersections	5	1	6	Low	Low
231	Railroad commuter service	5	1	6	Low	Low
232	Intersections	5	1	6	Low	Low
233	Crashes	5	1	6	Low	Low
234	Estimating	5	1	6	Low	Low
235	Revealed preferences	4	2	6	Low	Low
236	Suburbs	4	2	6	Low	Low
237	Paratransit services	4	2	6	Low	Low
238	Income	4	2	6	Low	Low
239	Parking	3	3	6	Low	Low
240	Greenhouse gases	3	3	6	Low	Low
241	Data banks	3	3	6	Low	Low
242	Rapid transit	3	3	6	Low	Low
243	India	3	3	6	Low	Low
244	Routes	3	3	6	Low	Low
245	Developing countries	3	3	6	Low	Low
246	Medium sized cities	3	3	6	Low	Low
247	Bluetooth technology	2	4	6	Low	Low
248	Floating car data	2	4	6	Low	Low
249	Urban highways	2	4	6	Low	Low
250	Routing	1	5	6	Low	Low
251	Traveler information and communication systems	1	5	6	Low	Low
252	Special events	1	5	6	Low	Low
253	Supply chain management	0	6	6	Low	High

Facilitated Discussion

What has been your experience with integrating Census Data with other data sources? How do these complement/supplement what Greg and Adam have found?

In general, the audience had little hands-on experience with integrating Census data with other data sources. However, it was mentioned that in the maritime industry, onboard spatial data using automatic identification systems (AIS) can be appended to information on the commodities on the vessel (based on tagging to the associated paperwork) and other characteristics of the operations, to produce a rich integrated data source. This approach will be able to inform future

freight behavioral models, transportation policies on port operations, as well as contribute to the overall understanding of commodity movements across modes.

Several audience members mentioned attempts to use AirSage data to understand visitor travel behavior in Wilmington, North Carolina. There was a mention of integrating with transit and taxi data, and some efforts to supplement HH travel surveys with Census data. Others recalled states using INRIX or Streetlight mobile data for statewide modeling. In this case, the Streetlight data was used for modeling and the Census data was used to weight the model results.

Strategies: Anything Missing?

A number of audience members thought that an eighth strategy should be included—a hybrid approach that would take into consideration several of the seven strategies put forward in the research, but considering "gluing" data together from many sources in a myriad of ways. In addition, a number of audience members thought that the biggest problems with surveys are response rates, cost, and expectations (e.g., public perception). It was suggested that these problems might actually kill surveys before Big Data replaces them.

It was noted that the connection between the desire for (specific) data and the need for it are not well established and not always strong. For example, the desire for data on long-distance travel is a case where we want the data, but the need is not well demonstrated. Also, the methodology for how to gather long-distance attributes is problematic, particularly with a survey instrument. Currently, BTS has aviation data available for analysis of long-distance trips, but there are no surface transportation data sources (e.g., no charter bus, passenger rail, or auto) and no good solutions have been identified to date. It was noted that in Colorado, they have a policy that was enacted in the absence of data, based on instinct on how Denverites travel.

Based on the overall discussion, the biggest overall data gaps today are urban freight and intercity passenger travel. An emerging gap would be TNC data (e.g., Uber, Lyft). It was also noted that MPOs could use income by industry by workplace and more frequent tables.

What role do these strategies discussed from our session today play in CTPP Board and Census Bureau decisions?

There was strong concern that what is currently at risk is the balance between continuity and adapting to change (re: changing questions on the ACS). Careful consideration needs to be made when making recommendations for changes and making sure there are methods for reconciling previously collected data (e.g., aggregations). In addition, more needs to be learned about projects where other data sets (e.g., INRIX, Streetlight) are being used for O-D studies.

Challenges continue with traditional data include response rates, costs and expectations. Even more challenging are those data sets that have been attempted, but not successfully collected. For example, long-distance attempts have been unsuccessful and require a completely new approach. While technology issues continue to plague everyone, they also could offer some yet to be used solutions.

What are the opportunities for data fusion/integration?

While the audience expressed interest in opportunities for data integration, the more pressing issues for immediate attention include:

- Continuing to improve how mode is asked in order to get more accurate understanding of travel (e.g., walk, bus, subway, walk);
- Developing a methodology for collecting intraurban freight and intercity passenger;
- The production of more frequent tables; and
- More widespread recognition of the relevance of CTPP for day to day operations and analysis.

Audience Suggestions for the CTPP Oversight Board

Continue to explore methodologies for long-distance travel by all modes and for the production of data sets that can be used for urban freight planning and analysis.

Explore opportunities (e.g., public–private partnerships) to acquire TNC data and produce PUMS for use by transportation professionals.

Identify opportunities to conduct a synthesis of current and up-coming data projects that use integration–fusion techniques, particularly with respect to Census products (e.g., CTPP, ACS).

CHAPTER 14

Using Census Data to Understand Alternative Modes

JIM HUBBELL
MARC, presiding

ROXANA J. JAVID
Savannah State University

RAMINA J. JAVID
Shahid Beheshti University

ZHUYUN GU
ANURAG KOMANDURI
Cambridge Systematics

MEGAN BROCK
MARIO SCOTT
PIERRE VILAIN
Steer Davies Gleave

Census data is not only about demographics. In the transportation community, it is used for modal analysis, especially when assessing new and emerging modes. Three examples were presented in this session.

INVESTIGATING THE FACTORS INFLUENCING ELECTRIC VEHICLE ADOPTION IN CALIFORNIA: A COUNTY-LEVEL DATA ANALYSIS
Roxana J. Javid and Ramina J. Javid

Plug-in electric vehicles (PEVs) are believed to be one of the means to improve the sustainability of the road transportation. To investigate which types of incentives are most effective at encouraging PEV adoption, a set of multiple regression models relating infrastructure, cost-related and sociodemographic variables to PEV adoption rates in 58 California counties were used. Multiple datasets were integrated including California Household Travel Survey, National Renewable Energy Laboratory, and ACS data to demonstrate how this model is able to quantify the impacts of these variables on PEV adoption rates at the county level, where decisions are typically made. The potential factors include charging station per capita (infrastructure); commute time and energy price (cost related); age and gender of the buyer; HH's maximum level of education; homeownership status; and average number of vehicles (sociodemographic). To test for multicollinearity, correlation coefficient matrix and variance inflation factor tests were employed. The model was applied to California county-level data, with the goal of quantifying how public charging station infrastructure and other potential factors contribute to PEV purchasing in each individual county.

Findings indicate that charging station per capita are effective in promoting PEV adoption, particularly among male buyers in households with less number of vehicles available. Sociodemographic factors such as gender and household's number of vehicles have significant influences on PEV adoption rate across individual counties in California. While sociodemographic factors cannot be controlled, they can be used to modify the effectiveness of changes in charging station availability in altering PEV purchasing for a given county, providing valuable input into regional decision making. It is clear that in California, people would choose a PEV if charging from a public station is a feasible option, regardless of their commute time. With sufficient data availability, this model could be used by regional and city-level policy makers and transportation planners to optimize their infrastructural investments by identifying counties where the response of drivers to added charging station would be maximized, implying that larger benefits can be achieved.

PREDICTIVE MODELS FOR BIKE-SHARE UTILIZATION USING OPEN-SOURCE AND CENSUS DATA
Zhuyun Gu and Anurag Komanduri

Cities across the country have open-sourced their bikeshare utilization data to allow planners and analysts to understand and quantify how the system is being utilized. The research team downloaded and synthesized bikeshare data from a variety of large, medium, and small cities including New York; Chicago; San Francisco; Los Angeles; Washington, D.C.; Austin; Minneapolis; Philadelphia; and Chattanooga to create a single, massive repository of bikeshare data. Activity is aggregated at a station level to support predictive modeling. Each row in the database represents activity at a station for a unique combination of date and time of day. To this station-level dataset, several additional data sources were appended to the station-level dataset. For example, ACS data at a block level and block group level were merged using buffer-based analyses. This allowed the researchers to quantify population and demographics that are within walking distance of the bike stations. Supply side variables such as bike-lane mileage and transit station access distance were also appended. Local land-form variables were also incorporated. Network connectivity variables such as access to other stations within time and distance bands were also captured. Finally, extraneous variables such as wind speeds, temperature, precipitation, and special events were also captured and included in the database.

The research team built two separate predictive demand models: one that models aggregate system-level utilization, and the second that models station-level activity. Since the models use information from a variety of cities over a large period of time, the models are robust enough to be used when studying either the development of an entirely new system in a new city, or in the expansion of existing systems. Without the availability of detailed Census data that captures the marketshed for possible riders, these predictive models would not have been possible.

USING CTPP DATA FOR PASSENGER FERRY DEMAND FORECASTING

Megan Brock, Mario Scott, and Pierre Vilain

Cities across the United States are working to expand their transit offer to better serve their commuters and visitors alike. For waterfront cities like Seattle and New York City, passenger ferry services are often a viable transit option. Census data on commutation patterns can be extraordinarily valuable in assessing new ferry service, and is at the base of many travel demand forecasting models. SDG has recently used JTW data from the CTPP to estimate ridership for new ferry services in the New York and Seattle regions. In both cases, a series of mode choice models using base demand from the CTPP data to determine the capture rate of new ferry services were estimated. In 2013, SDG was commissioned by New York City Development Corporation to complete a study on a citywide ferry service in New York City.

As a part of this research, SDG estimated ridership six potential route configurations, two of which are up and running as of May 1, 2017. The remaining are expected to rollout incrementally over the next year and a half. JTW data from the 2000 CTPP was used. Growth from the 2000 data was calculates using 2010 Census and ACS data. Following the citywide ferry study, SDG prepared passenger ferry ridership forecasts as a part of a team doing long-range planning for Kitsap Transit. This research used the 2006–2010 JTW data from the CTPP to estimate ridership for three passenger-only ferry services in the Seattle region.

CHAPTER 15

National Household Travel Survey
Building on 50 Years of Experience

DANNY JENKINS
Federal Highway Administration, presiding

ALAN PISARSKI
Consultant

STEVE POLZIN
University of South Florida

CEMAL AYVALIK
Cambridge Systematics

CLARA RESCHOVSKY
Bureau of Transportation Statistics

2017 NATIONAL HOUSEHOLD TRAVEL SURVEY
Danny Jenkins

The NHTS is a periodic national survey conducted by FHWA to provide travel and transportation patter data for transportation planners and policy makers in the United States. The survey has been conducted every 5 to 8 years since 1969, providing nearly 50 years of data. The most recent data collection effort is the NHTS for 2017, with data collected on trips taken by all members of participating households over a 24-h period. The data includes purpose of the trip, means of transportation, travel time of the trip, and time of day or day of week.

On August 8–9, 2018, the *Using NHTS Data Workshop* will be held in Washington, D.C. In spring of 2018, Summary of Travel Trends will be available and there will be website upgrades (see www.nhts.ornl.gov).

The data collection began in April of 2016 and ended in April of 2017, with data collected across 365 days. The final travel day assigned was April 30, 2017, with data collection efforts ending in early May. The respondents were encouraged to record all travel, even if "out of town" on their assigned travel day. According to the definitions, a "complete household" had 100% of all household members (5 years and older) responding. There were approximately 130,000 completed surveys and is planned for release in early 2018 (26,000 national samples and 130,112 add-on samples). The add-on agencies included nine state DOTs and four MPOs (Figure 15.1).

On August 8–9, 2018, the *Using NHTS Data Workshop* will be held in Washington, D.C. In spring of 2018, Summary of Travel Trends will be available and there will be website upgrades (see www.nhts.ornl.gov).
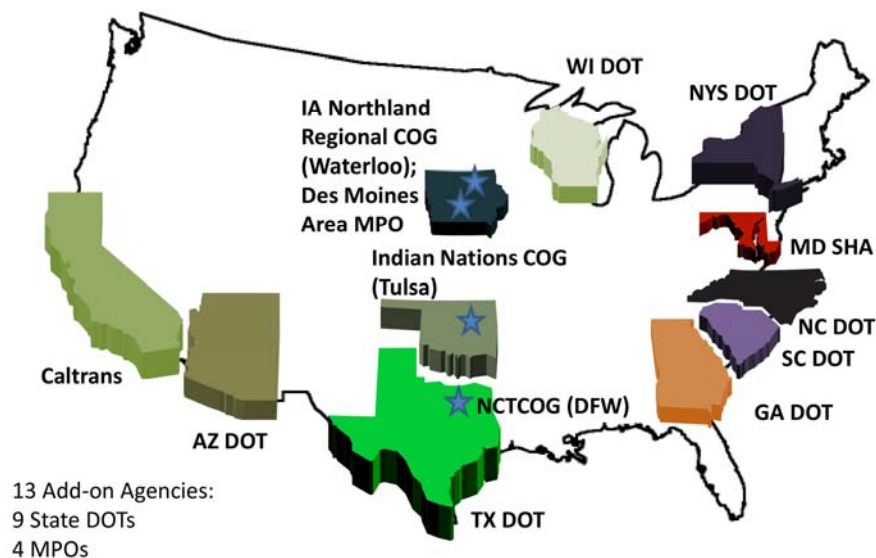
FIGURE 15.1 NHTS add-on agency locations.

NPTS–NHTS AND THE CENSUS JTW

Alan Pisarski

The linkage between the National Personal Travel Survey (NPTS), the NHTS, and JTW data began in an era when major metropolitan areas were conducting surveys. The Census JTW question was added to the Census in 1960, related to a 1961 travel survey at the Commerce Department. The NPTS was first conducted in 1969, then administered in 1977 and 1983 by the Census Bureau, but was conducted privately in 1990. At one time, the Decennial JTW was matched to the NPTS, and then with U.S. Housing and Urban Development AHS, providing key metropolitan updates. The original goal of the NHTS was to establish a fixed schedule for the NHTS in 2000 and 2005 so it could be more easily linked to Census years.

The great strengths of the NPTS, in the context of a Census JTW, include its ability to include multiple jobs, multimodes to work, work trips embedded in all trip making. It collects "usually used" instead of "yesterday" providing a great validity tes; and it has added seasonality, distance, and geographic specificity. NHTS has a strong demographic base through the travel activity of household workers.

It is unclear why the NPTS was changed to the NHTS. The original work to create the NPTS, which was deployed in 1969, preceded the establishment of the U.S. DOT. The DOT had been established in 1967 and the NPTS was in the field by 1969. The late 1960s was the era of big travel surveys, due to the first round of mandated planning processes as per the 1962 Federal-Aid Highway Act (having just finished New York and Washington, D.C.). At that time, a number of transportation surveys were developed including truck, taxi, hotel, and external screen lines by the Bureau of Public Roads (BPR) and FHWA. Attempts were made to sum these various surveys into a harmonized database, but unfortunately there were challenges as all of these surveys were given at varying times, with different definitions of variables and methods, and with different sample sizes. Under these circumstances, the Census (also located in the Department of Commerce) and the BPR, with their strong relationship, designed a mandated travel survey and provided training to the BPR staff. The 1961 Commerce Department survey of

travel characteristics was discussed at the second National Conference on Highways and Urban Development, held in Williamsburg, Virginia, in 1965. Census data was used to update the urban transportation studies in 1972. Today, the NHTS, the Decennial Census, and the ACS are the source of the data needed for meeting the continuous aspect of transportation planning.

NEW DATA, NEW RESEARCH
Steve Polzin

The NHTS remains the preeminent resource for understanding travel behavior, using a national sample, creating a longitudinal data set. It captures all (or most) travel by household members and provides a rich understanding of household characteristics. It provides a robust set of historical studies and can be fused with other data sets. Figure 15.2 compares 2001 per capita VMT with 2008 per capita VMT by age group, revealing a decline in VMT for young adults. When the NHTS data is made available in 2018, there are a number of demographic issues to be researched including:

- Are multi-adult untraditional households behaving differently?
- Are travel differences across race–ethnic groups changing?
- What is the travel behavior of the growing downtown residential populations?
- How pronounced are the travel differences by cultural geography (urban–rural, coastal–interior)?
- Do low-income households travel in different urban locations?
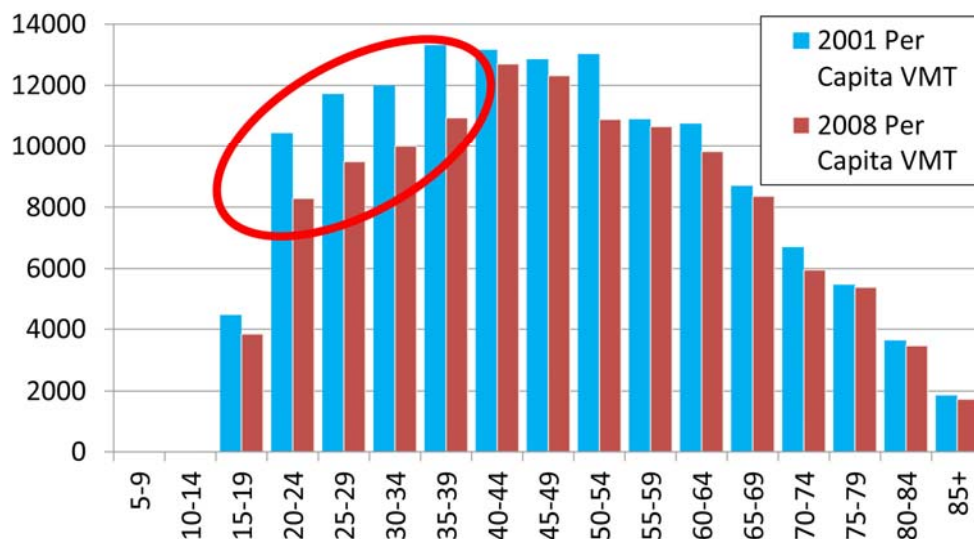
FIGURE 15.2 Millennial travel: PMT and VMT per capita by age.

Modal issues include:

- What can we learn about the bounce back in VMT?
- Insights on declining transit use?
- Who is carpooling?
- Who is working at home?
- Bike, pedestrian trends?
- Communication substitution (trip purpose trends)?

Emerging issues include:

- TNCs;
- Electric vehicle use; and
- Propensity to use Mobility as a Service or automated vehicles.

Figure 15.3 illustrates previous research findings with respect to vehicle availability over time and by household size. There are a number of behaviors related to the propensity to travel, including modes and business models:

- Dynamics of vehicle use in household;
- Temporal pattern of household travel;
- Pattern of vehicle use over vehicle life cycle;
- Travel group size;
- Trip chaining/tours;
- Mobility aids/child seats;
- Products, tools, materials, attachments/trailers; and
- Temporal trip distribution of all travel.

The NHTS has been used by a number of other sectors to conduct research in the following:

- Transportation and energy/environment;
- Transportation and health;
- Transportation and land use/built environment; and
- Transportation funding

New opportunities for exploration for the NHTS include:

- Better geocode data and
- More data to fuse/merge, integrate with NHTS.
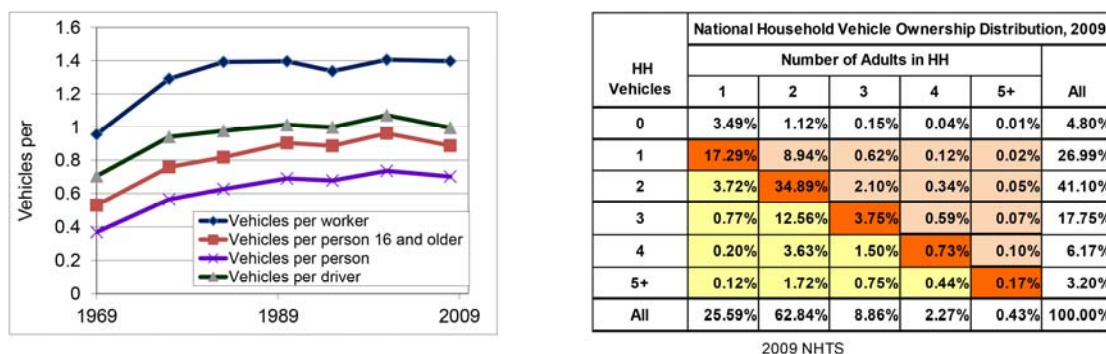
FIGURE 15.3 Technology related hot topics: vehicle availability.

UPDATING NHTS WITH ACS DATA
Cemal Ayvalik

The lack of current and available data to update NHTS is the motivation behind this research to use ACS data. The need for recent demographic, behavioral, and technological trends requires more-frequent data than NHTS cycles. This research develops a model based on the compilation of 2009 NHTS and 2009 ACS data segment and population estimation models using interpolation. Linear regression models predict number of household and person trips, and amount of person and vehicle travel. Multinomial logit models predict travel behavior for different portions of the population, with the entire population fully segmented. In addition, departure times are predicted. Independent variables with significant explanatory power included: HH size; vehicles in the HH; workers in the HH; HH income; gender; age (65+); education; employment status; retired HH member; licensed driver; population density; urban–rural; and the availability of heavy rail. Validation is based on backcasting to 2001 and using 2000 Census PUMS to predict travel behavior indicators.

Suggestions for moving forward with updating NHTS with ACS includes short- and long-term steps. In the short term, it will be necessary to remove outliers in the comparison NHTS dataset and testing models by inputting comparison NHTS demographics. This would need to be followed by a test of the revised models with 2016 NHTS data. In the longer term, it may be necessary to synthesize population for more accurate joint distributions. In addition, researchers should segment the analysis to explore and incorporate causal relationships between life cycle, lifestyle and travel.

LEVERAGING FEDERAL DATA: FOCUSING ON CTPP AND NHTS
Clara Reschovsky

Compared to NHTS, CTPP provides commuting trips only, is available for small geographies, and has a larger sample size, with source data that is collected continuously. The NHTS, on the other hand, provides all trip types, has a smaller sample size, is not available for small geographies, and requires the deployment of a surveying effort. The long-standing data challenges for both these data sources include timeliness of the data. The data is released at a minimum of approximately 2 years post collection, even longer for the multiyear datasets from

ACS. Issues with trip details at small geographies is hampered by sample sizes that preclude sample geographies or renders it unreliable. Averaging effect of larger geographic area reporting limits full understanding of travel behavior. In addition, there are privacy concerns for survey respondents puts the balance of data collection and data release in conflict. Most importantly, there is no additional funding available for a full, comprehensive data collection effort, particularly at the national level. Budgets are limited at all levels of government and it is hard to plan multiyear projects with budget uncertainties.

A possible way forward is to consider alternative data sets including cell phone data (e.g., INRIX or cell phone data directly from phone companies); app data (e.g., TNCs such as Uber or Lyft, car-/bikeshare data); social media (e.g., crowdsource data); or modeled data instead of observed data (e.g., "B" tables in CTPP, Local Area Transportation Characteristics for Households or Freight Analysis Framework-style data that is modeled from the Commodity Flow Survey).

It has been recognized that reported travel behavior provides "normal" travel behavior, but this can change over time, but is relatively stable. It is useful for habitual travel (e.g., JTW) and can help users validate modeled data results. At issue are concerns about model choices as technology changes. This would affect the use of car and bike share. In the future, it would include self-driving cars. Currently, bicycles with electrical assist are already available for purchase or for use in bike sharing operations. Questions remain regarding further trip replacement with online communication and shopping as household members increase the use of shipping goods to personal residences resulting in more truck traffic in neighborhoods and congestion. It is possible to use survey data with alternative data to model additional characteristics. This approach is not a new concept, but more doable with different types of data.

Another travel behavior of concern is long distance travel. Questions remain on how long distance should be defined. What minimum distance traveled should be considered long distance? What modal considerations need to be taken (e.g., airplane or intercity train)? What should be collected regarding long-distance trip purpose? Last mile of travel for freight impacts passenger travel behavior in terms of trip replacement and road congestion. Additionally, what are the issues with connectivity of transportation modes, particularly with respect to linked and unlinked trips? Also needed in additional knowledge of intermodal connectivity for use in modeling.

NHTS and the CTPP remains critical to travel behavior. These national data sets are necessary for understanding non-statistically stratified data sets. Geographic bias in nonnational data needs to be account for in analysis. In addition, private sources of data tend not to be open in their methodology, making it difficult to discern the inherent biases in their data.

The next steps for the data community are to keep using both NHTS and CTPP data for analysis with citations; participate in conferences and user groups to share information and learn more about data; and to document usage of national data in the development of data collection or post-data collection weighting efforts when implementing survey at the local level.

CHAPTER 16

Workplace Data
Achieving Its Potential

BRIAN MCKENZIE
U.S. Census Bureau, presiding

MICHAEL FRISCH
University of Missouri, recording

JUNG SEO
Southern California Association of Governments

One of the mainstays of the CTPP special tabulation is the abundance of data on workers at their work locations. However, there is concern moving forward that the workplace tables have been underused and may be reduced in the future tabulations. This commissioned paper explored the many potential uses of the data both from current applications and "what could the data be used for" perspective.

THE CTPP WORKPLACE DATA FOR TRANSPORTATION PLANNING: A SYSTEMATIC REVIEW
Jung H. Deo, Tom Vo, Shinhee Lee, Frank Wen, and Simon Choi

CTPP data has been a valuable resource for transportation planning community, providing information about where people live and work, their JTW commuting patterns, and their socioeconomic and travel characteristics. While the CTPP data has been widely utilized by transportation planning agencies and researchers as a key input for various transportation planning subject areas including, but not limited to, travel demand modeling, descriptive statistics, policy and planning strategies, environmental analyses, and survey and sampling methods, the CTPP Oversight Board believes that the CTPP workplace data is underutilized. To understand the potential enhancements to the CTPP workplace data for better utilization in the future, this paper provides an overview of the CTPP and other data products that have been widely utilized in transportation planning and research, such as the LEHD, LODES, and NHTS. It then discusses the strengths and limitations of the CTPP workplace data as compared to those two data products. In addition, this paper summarizes the previous and current utilization of the CTPP data by reviewing over 300 studies that cited the use of the CTPP data, and identifies the key subject areas and the emerging topics of those studies.

Introduction

The CTPP program is a Technical Service Program of AASHTO, funded by member state transportation agencies. The CTPP data is a set of special tabulations from ACS data, designed for transportation community. The CTPP data has been a valuable resource for transportation planners and researchers, and it has been utilized for various transportation planning subject areas including, but not limited to, travel demand modeling, descriptive statistics, policy and planning strategies, environmental analyses, and survey and sampling methods. The CTPP provides invaluable information about where people live and work, their JTW commuting patterns and their means of transportation to work. One of the unique features of the CTPP data product making it different from other Census data products is that it provides more workplace-based tables than the ACS data. CTPP workplace data, one of three components of the CTPP data product, provides detailed workplace based socioeconomic and travel characteristics information for workers, although the CTPP Oversight Board believes the CTPP workplace data has been underutilized.

The main goal of this paper is to assist the CTPP Oversight Board in the development of future workplace data with the purpose of encouraging transportation planners and researchers to better utilize the CTPP workplace data. The objectives of this paper are (1) to explore the multiple data products relevant for transportation planning, (2) to discuss about the strengths and limitations of the CTPP as compared to other products, (3) to summarize a variety of previous and current uses of the CTPP and its workplace data, and (4) to suggest potential enhancements to the CTPP workplace data for better utilization. To examine the strengths and limitations of the CTPP workplace data, this paper conducts a comparative analysis between the CTPP and other data products such as the LODES and NHTS. And then to better understand the utilization of the CTPP data, this paper conducts the literature review of the 305 studies that cite the use of the CTPP data and summarizes those studies by subject area.

Design Comparison of Workplace Data Products

Analyzing characteristics of workplace is crucial for understanding and mitigating traffic congestion, commuting patterns, EJ, and so forth. The analysis requires reasonable and accurate dataset. Transportation planners have utilized numerous national and local datasets, including but not limited to the CTPP, the LODES and the NHTS. Each workplace data product has their own strengths and limitations. It is important to use the appropriate data for certain types of analysis. This section explores the multiple data products that have been widely utilized in transportation planning and research, and then, discusses the strengths and limitations of the CTPP workplace data as compared to those products.

Overview of Workplace Datasets

ACS Data

Data is a mandatory component in both qualitative and quantitative analyses. An accurate and comprehensive dataset gives an advantage of unfolding many insights of a subject (i.e. means of transportation by household income in urban and rural areas, commute time by age compositions by minority status); thus, it will help to produce a high quality and empirical finding. The Census

Bureau produces many useful and publicly available tools and datasets that are used by various sectors such as governmental agencies, private companies, non-profit organizations, universities, etc. The Census Bureau is a federal government overseen by the Economics and Statistics Administration, which is a part of the Department of Commerce (*1*). The Census Bureau produces two major datasets with information about commuting: ACS and the LEHD. Within each of these datasets, it provides detailed information related to workplace and commuting characteristics. These datasets are crucial and useful for transportation planners.

Within each of the described datasets, they have special tabulations dedicated specifically to transportation planning (i.e., CTPP and LODES). It is important to know the background of the CTPP, which is a subset of the ACS. The Census has been conducting an annually continuous social and economic survey called ACS since 2005; this particular survey was created to provide information more frequently and eventually replaces the decennial long form in 2010 (*2*). The responses of ACS are combined and available at 1-year, 3-year, and 5-year period estimates. (The last ACS 3-year estimates were for 2011–2013 and have been discontinued since 2014.) It means that these estimated periods represent the social and economic characteristics over a specific data collection time frame (*3*). Of course, the decennial Census—during the period between 1960 and 2000—provided much more in-depth and diverse information because of its sample size, but its frequency was an issue. With the ACS, the Census can release yearly up-to-date information of social and economic data for communities within the United States. For example, the results from the ACS allows a city to examine the changes of commuting time for minority population every year for its EJ analysis. The data compilation and estimation within the ACS can provide an opportunity for data users to analyze trends and compare across geographical units (i.e., states, counties, cities, communities, Census tracts) and population groups.

The decennial Census mailed out survey questions of households nationally; for instance, about 17% of all U.S. households (about 19 million) were sampled with the long form in Census 1990 and 2000 (*2*, *5*, *6*). The ACS, as mentioned, was created to improve the decennial Census (*4*) because of its frequent data availability; however, the accuracy of the ACS data is not as high as the decennial Census due to its sample size. Per Census, the ACS 1-year surveyed roughly 3% of all U.S. households (about 3.5 million) and group quarters such as military barracks, nursing homes, and prisons. For the 5-year estimates, the sample size of the ACS is less than 10% of all households in the United States (about 11.5 million) The MAF was used to randomly select households during the ACS survey period; and, these households should not be selected again within five years (*2*). The MAF is a comprehensive database that contains the latest address information, location codes, source, and history data for U.S. residents (*5*).

The questionnaires in the ACS are similar to the traditional long form, which includes questions about sociodemographic, housing, economic, and JTW. The sampling periods of the ACS depict the availability of its geography (i.e., large, medium, and small). For 1-year estimates, the data is available only at large areas of 65,000 population and over. For 3-year estimates, the data is available at medium areas of 20,000 population and over. For 5-year estimates, the data is available in the smallest areas (e.g., Census tracts and Census block groups), which represent between 600 to 3,000 resident populations. The ACS 5-year estimates (2006–2010) have the smallest sampling errors comparing to 1- and 3-year estimates due to its sample size of roughly 11 million housing units; however, it still has a higher MOE than the 2000 decennial Census of 18 million sample size (*7*). It is obvious that larger sample size has better quality but it is expensive and time-consuming.

CTPP Data

The ACS estimates have produced a rich database for many special tabulations related to social, demographic, and economic characteristics, home and work locations and commuting flows; these tabulations that have been utilized by many transportation planning agencies as a key input to various transportation-related policies and planning efforts (e.g., corridor and project studies, environmental analyses, emergency operations management). Due to its usefulness and demand of such tabulations, the CTPP program was created through a pooled fund and collaborative effort between the Census Bureau, the DOTs, state DOTs, the AASHTO, and TRB committees. In addition, CTPP has been utilized for model validation and calibration purposes by MPOs and DOTs in their long-range transportation plans (*2*).

The CTPP predecessors were called UTPP and UTP in 1980 and 1970, respectively. In 1990, the UTPP changed its name to the CTPP, and 2006–2010 CTPP has been using ACS 5-year estimates to process the unique tabulations related to transportation (*8*). These packages used the decennial Census long form data to generate special tabulations. The 2000 Census was the last one to include the long form, and after that, all questions related to commuting were moved to ACS. Consequently, CTPP datasets produced after 2000 Census were based on ACS data, which is the only source for information on commuting and several other demographic characteristics. The latest CTPP data was generated using ACS 2006–2010 5-year estimates and was released in 2013. These special tabulations are available at TADs, TAZs, and Census tracts. TADs and TAZs are defined by states and MPOs. CTPP provides special tabulations for residence, workplace, and flows between home and work. The advantage of having these tabulations is the capability to analyze detailed information related to residence, workplace, and commute flows. For instance, the commuting flows of the workers can be customized to analyze the difference in average commuting times between low-income and high-income workers from location A to location B.

LEHD O-D Employment Statistics Data

Another major transportation planning dataset produced by the Census Bureau is the LODES, which is a collaborative effort between the Census Bureau and Departments of Labor in various states. LODES data is not available for all states due to data unavailability and data sharing limitation (*9*). Like the CTPP, the purpose of this program is to explore the LED by connecting residence and workplace with each other. The LEHD provides detailed information on the local labor market based on actual administrative records; the Census uses such information to improve its economic and demographic data programs (*9*). This dataset can be utilized to investigate various workplace-related topics such as firm size, earnings and commuting flows. The LEHD is known as another potential data source, besides the CTPP, that provides information regarding workplace characteristics and commuting flows.

One unique feature of the LEHD dataset is that it uses administrative data, which covers more than 95% of the total workforce in the United States (*3*). This data includes information from state QCEW and federal administrative records. The QCEW program publishes employment and wages information from employers within the United States at various geographical levels (i.e., county, MSA, state and national) by detailed industry. The QCEW primarily collects workplace characteristics from administrative records of all private sectors, and local and state governments covered under the UI programs. In addition, the Annual Refiling

Survey and the Multiple Worksite Report from the BLS are also used to supplement the missing data gap of QCEW microdata. Under the LEHD program, the mentioned data sources about firms and workers are combined to tabulate job-level quarterly earnings, workplace, and residence information, and firm characteristics (e.g., industries). The LEHD program, since 2012, has included federally mployed (not including military jobs) and self-employed workers. The employment data of federal employees is obtained from the U.S. Office of Personnel's Management and the self-employment data is collected from tax files. It is important to understand how the LODES data computes its job counts. A job is counted only if the employee is employed at the same place in both first quarter (previous) and second quarter (current). The LODES data files are state-based and organized into three types: O-D, residence area characteristics, and workplace area characteristics. The LODES data is available for most states for the years 2002–2014, and the latest LODES data was enumerated by 2010 Census block. The LODES data has been integrated into a web-based map application called OnTheMap, an online mapping and reporting application that shows where workers are employed and where they live.

National Household Travel Survey Data

Another major dataset in the transportation planning field is the NHTS. This dataset has been referenced and utilized by transportation planners to assist them in understanding the travel patterns and behaviors in the United States. According to the 2017 compendium of uses, there were 198 reports and papers that utilized the NHTS in 11 categories (i.e., bicycle and pedestrian studies, energy consumption, environment, health, policy and mobility, special population groups, survey, data synthesis, and other applications, traffic safety, transit planning, travel behavior, and trend analysis and market segmentation) (*10*). The major usage of the NHTS is to explore travel behavior, which is important for program initiatives, review programs, and policies, mobility issues, and long-range plans. The NHTS is not updated as frequently as other datasets (i.e., CTPP, LODES); a total of eight NHTS was conducted between 1969 and 2017. The NHTS was known as the NPTS prior to 2001.

The NHTS dataset collects daily travel information—that means the data is collected from trips within the 24-h timeframe. The questionnaires in the survey ask for trip purpose, modes, travel time, departure time, departure date, vehicle occupancy, driver characteristics, and vehicle characteristics. The 2009 NHTS is organized into four different data files, including HH file, person file, vehicle file, and travel day trip file. Every time the survey is conducted, it always introduces new emerging variables. The 2009 NHTS data includes unique information such as telecommuting, public perceptions of the transportation system, internet usage, and active transportation trips. Users of the NHTS have also identified additional variables needed for future collection; these extra variables are costs of travel, specific travel routes, travel of the sampled household changes over time, household and workplace location, traveler's reason for selecting a specific mode of travel over another mode (*11*). The latest 2016 NHTS recently completed its data collection phase in April 2017. There were slightly more than 129,000 households participating in the survey. The 2017 NHTS data was made publicly available in early 2018.

The Add-on Program is unique for the NHTS. This program provides an opportunity for states and MPOs to purchase additional samples of the household travel survey that are compiled into a geocoded database within their jurisdictions for more localized transportation-related planning and forecasting. The location file of the add-on deliverables provides latitude and

longitude of origin and destination address and is linked with the four main files by household ID, person ID and trip ID. In 2016 NHTS, nine state DOTs and four councils of governments (COGs) were add-on partners, including Arizona, California, Georgia, Maryland, New York, North Carolina, South Carolina, Texas, Wisconsin, Des Moines area MPO, Indian Nations COG, Iowa Northland Regional COG, and North Central Texas COG (*10*).

Strengths of CTPP and Other Datasets

The CTPP provides useful special tabulations for transportation planning purposes from using sample dataset (e.g., ACS 5-year estimates) to statistically represent all areas within the United States. This data is available at various geographical units such as county, place, tracts, and so forth (*3*). The CTPP is derived from the ACS, and therefore, it allows users to analyze workplace and travel patterns with more customized tabulations than the LODES. CTPP includes unique variables and cross-tabulations at small geography (i.e., TAZs or Census tracts) at three summary levels, which are residence geography, work place geography, and home-to-work flows (*2*). These tables are tabulated from the ACS dataset. The CTPP has undergone a tremendous improvement in its contents from 1990 to 2010, and consequently, the data has added more customized tables and enhanced statistical processes (*2*). Per Weinberger, in 2018, the tabulations in the CTPP will be reduced by roughly 30% of the current 2006–2010 CTPP version but it will still have more workplace information than the LODES. Another unique feature of the CTPP is the freedom for users to create customized reports based on interested geographical units (i.e., Census tracts) or demographic variables (i.e., low-income, minority, vehicle availability by household income). Additionally, as compared to the LODES, the CTPP includes several unique transportation-related variables such as mode choice and travel time (*12–14*). The CTPP application provides O-D flows for several special tabulations such as poverty status, minority status, travel time, age of worker, industry and more.

As compared to the CTPP, the LODES provides information on workplace and commuting flows at a finer geography (down to the Census block level), while the LODES provides less workplace characteristics than the CTPP. Spear has stated in his report of "NCHRP 08-36, Task 098 Improving Employment Data for Transportation Planning" that the CTPP 2000 and 2006–2008 datasets include more O-D flows than the LODES data. Spear also suggested combining the CTPP with the LODES "to smooth out the geographic distribution of home-to-work trips, and to develop more complete areawide O-D matrices for HBW trips that could be used in travel modeling applications" (*14*). In 2003, one study has evaluated feasibility of generating workplace data from LEHD program (*15*); the author has stated that the CTPP captures more internal trips (i.e., people who live and work in the same tract), which is an important variable for transportation planning. The study found that Census tracts with internal trips are higher in the CTPP as compared to the LEHD. Furthermore, the reason behind the difference in internal trips between these workplace datasets may be "attributable to the LED data capturing only those employers who pay unemployment insurance, missing self-employed worker" (*15*). Also, the result of the statistical linear model has shown better fitness with the CTPP than the LEHD. Overall, it is a major drawback that the LEHD does not have detailed information of residence and workplace (i.e. mode choice, travel time, self-employment)—comparing to the CTPP.

Compared to the CTPP and the LODES, the NHTS provides more detailed variables of households, persons, travel day trips, and vehicles and long-distance trips (*16*). This data also

provides specific information of travel behaviors of people on multiple trip purposes (e.g., shopping trip, recreational trip). The NHTS provides travel characteristics during weekends, which makes it unique (*17*). In addition, the NHTS committee has been actively collecting feedbacks and comments from data users to improve the next version of the NHTS. There are several improvements from the 2001 to 2009 NHTS based on the Summary of Travel Trends 2009 NHTS. Besides the general adjustments (i.e., data collection, odometer reading, eligible household members) to the 2009 NHTS, the survey has also incorporated emerging transportation-related questions about 1) safe routes to schools, 2) hybrid vehicles, 3) detailed work-related travel questions (i.e. whether the worker can set or alter their work schedule, whether the worker has the option of working from home, frequency of working at home, and self-employed status), and 4) online shopping and shipping. The survey has also improved its geocoding technique. Instead of post-processing location data, it uses the real-time interactive online tool during the interview to geocode locations. Similar to the CTPP dataset, the NHTS utilizes the Census population estimates for its final adjustment.

Limitations of CTPP and Other Datasets

The CTPP special tabulations are derived from a continuous survey called the ACS, which surveys roughly 3.5 million U.S. households annually. To produce CTPP tabulations at small geographies (e.g., Census tracts) with low MOE, ACS 5-year estimates are used because of their its sample size relative to the 1-year data. This makes it difficult to perform temporal analyses using the CTPP dataset (*2, 13*). The CTPP only accounts for workers of age 16 and older, primary jobs, and institutionalized group quarters. The responded workplace locations may not be accurate because some jobs require workers to travel to multiple places (i.e. construction workers or employees attending the conference). Due to confidentiality, some of the information is suppressed, which results in unsatisfactory statistical reliability (*3, 18*). This statistical issue also occurs in the LODES and the NHTS. The suppression factor is related to the geographic detail available in each data source. The more geographic detail, the greater the chance there is suppression and the more error created by suppression. The CTPP does not include nonwork trips such as shopping, school, recreational, and so forth as the NHTS does. It does not include trip chain information. For instance, an individual may drive to the park-and-ride, take the train to work, and Uber home.

　　Although CTPP provides detailed data on workplace and O-D flows at small geographic level, the LODES provides more geographically detailed data (i.e., Census block level) when performing small area analysis related to workplace and O-D flows (*12, 13*). Also, the commute distance is not reported in the CTPP dataset (*12*). The CTPP may not cover an entire range of workers because, if workers who were on vacation or sick leave during the survey timeframe, they will not be included in the survey. Not every response may be accurate due to misreporting of workplace geocoding. For workplace address, it sometimes cannot be geocoded correctly because of missing address information. For example, it is difficult to accurately assign a worker who works for Boeing in Seattle without a proper address because of many Boeing offices. The unidentified or un-geocoded workplace will be assigned to the county and place level (*3, 19*). The CTPP has roughly 9% to 10% percent of workplace records geocoded to county or place level, which may be difficult to be further allocated to TAZ or Census tract level.

　　It is difficult to perform quality control on the survey data because the survey respondents may respond inaccurately, which results in reporting errors (*2*). In September 2005, there was an

intense debate about the 2000–2004 ACS data used to process the CTPP, in replacement of the 2000 decennial Census. One of the issues that were raised in the debate was about the "errors in the annual ACS data for 2000–2004 are very large and the data cannot be used to make rational conclusions in transportation planning" (*20*). Though the errors have been improved over time (e.g., 2011–2015 ACS 5-year estimates), it is important to keep improving the data quality. These quality issues occur due to low sample size. CTPP data users have also raised the possibility of eliminating some of the smaller geographies such as TAZs, because these tables have the most impacts. Estimates for larger geographies are much more stable due to their larger sample size. Furthermore, the ACS uses population estimates as population controls for its weighting methodology, instead of actual Census counts (*3*, *20*). As compared to the CTPP, LODES collects actual administrative records and collaborates with states to get consistent socioeconomic counts.

Although CTPP provides more variables and covers all areas in the U.S. as compared with LODES (*21*), the CTPP commuting flows (i.e.. CTPP 2000, CTPP 2006–2008) do not include low-frequency O-D pairs (e.g. work trip using bicycle or trips between distant zones) because the CTPP is based on sampled data. Therefore, LODES delivers more realistic home-to-work flows than any other sample-based datasets like the CTPP. Because the CTPP datasets are developed from using sampled data, it may omit some low-frequency O-D pairs that are not captured by the surveys, which may not provide a clear picture on commuting patterns. In transportation analysis and modeling, the sample weighting methodology is used widely to expand survey data to estimate the universe of home-to-work trips (*14*). For instance, low frequency O-D flows may not get captured and will be assumed a zero probability of occurrence in the statistical model—this will assign more weight to other trips. Spear also explained that "O-D pairs with a low frequency of home-to-work trips that are sampled in the CTPP get weighted more heavily, while low frequency O-D pairs that are not sampled are assumed to have no home-to-work flows" (*14*). This is a downside of using sampled data because not every aspect of O-D flows can be captured.

Because the NHTS and the CTPP are produced from surveys, both use a statistical method to generalize the survey responses to represent all population characteristics in the U.S. There are two types of errors when conducting a survey, which are non-sampling error and sampling error. As explained in the 2011 Summary of Travel Trends 2009 NHTS and NHTS Task C: Sample Design in 2017, nonsampling error may have resulted in several discrepancies; these include "the inability to obtain information about all persons in the sample; differences in the interpretation of questions; inability or unwillingness of respondents to provide correct information; inability of respondents to recall information; errors made in collecting and processing the data; errors made in estimating values for missing data; and failure to represent all sample households and all persons within sample households (known as under-coverage)" (*22*). On the other hand, the sampling error refers to when the sampled group's estimates don't represent the true population values. Confidence interval or MOE is used to examine and control the quality of estimates.

On the other hand, LODES has several limitations related to employment coverage, data availability, data continuity, and geography. The definition of workplace may be misinterpreted for LODES; it means that "an address from administrative data may or may not be the actual location that a worker reports to most often" (*3*). One example of this is the employees within the construction industry. Their workplaces are varied depending on the projects. The LODES dataset does not cover a full range of employment; the employment groups that it does not cover

are self-employment, military employment, the U.S. Postal Service and informal employment. Limited workplace-related variables are also another limitation of LODES. As compared to the CTPP, LODES does not include as many variables, such as means of transportation, WTT, vehicle available and poverty status. Another limitation is data discontinuity. For certain variables, LODES dataset does not have a consistent set, which makes it hard to perform longitudinal analysis. For instance, it is impossible to track down the changes of ethnicity of employees for the last 10 years because this variable only became available from 2009. Finally, the geography of LODES is not available for the whole United States as the LEHD program is a voluntary program.

Compared to the CTPP and the LODES, the NHTS is not updated frequently. The NHTS survey is conducted roughly every 5 to 10 years. Moreover, workplace data is not a part of NHTS's main data files, although the location file of the NHTS Add-on deliverables provides the detailed location information of origin and destination address and users can link the location file with the four main files by household ID, person and trip ID. The NHTS does not include contain specific information on costs of travel, information about specific travel routes or types of roads used, or travel of the sampled household changes over time, and the traveler's reason for selecting a specific mode of travel over another mode. Table 16.1 summarizes the characteristics of the CTPP, the LODES and the NHTS.

Uses of the Census Transportation Planning Products Data

Literature Review

This paper discusses the myriad uses of the CTPP data and its workplace data in transportation planning and research. In order to review the research subject areas, methodologies and data sources of the literature and studies that utilized the CTPP data, this study searched journal articles, dissertation, reports, and conference presentations that cited the use of the CTPP data from academic libraries, journal websites such as the TRB's *Transportation Research Record: Journal of the Transportation Research Board* website and Journal of American Planning Association online access, various conference publication websites, and Google search engine results pertaining to CTPP data. The key word and search engine terms used were "Census Transportation Planning Products," "Census Transportation Planning Package," or "CTPP". The resulting literature and studies were examined to select those that are most relevant to this study. The literature and studies reviewed in this paper cover a diverse range of subjects in transportation planning including, but not limited to, modeling, policy, demographics, equity, survey, and general planning issues.

This paper reviewed 305 studies that cited the use of the CTPP data. The publication dates of those studies range from 1989 to 2017 and their publication types included journal articles, dissertation/thesis, books, reports, conference proceedings, and poster presentations. In this paper, those studies were grouped into 12 categories based on the primary subject area identified in their abstracts, although there is, of course, much overlap between these categories in many studies. Some studies were categorized into multiple subject areas as they discuss multiple subject areas and no single subject area was considered the primary category. For the category classification, this paper reviewed previous similar studies and reports on the uses of

TABLE 16.1 Characteristics of CTPP, LODES, and NHTS

Categories	CTPP (ACS)	LODES (LEHD)	NHTS
What is the main source of data?	Used ACS to create special tabulations on commuting characteristics including residence and workplace.	Used LEHD dataset which from administrative records.	Used customized survey to randomly survey households on travel behaviors.
What is the sample size?	2006–2010 5-year CTPP was derived from ACS 2006–2010 5-year estimates (roughly 10% of all U.S. HHs).	Collected administrative records from 50 states via UI program and Office of Personnel.	2016 NHTS surveyed roughly 129,000 households. Add-on program allows agencies to purchase additional data.
What is data coverage?	Provides special tabulations for residence, workplace, and flows between home and work for the whole U.S.	Provides O-D, residence area characteristics, and workplace area characteristics for most states.	Survey samples represent all areas within the U.S.
How frequently does it update?	2006–2010 5-year CTPP is based on 2006–2010 ACS. The next version of CTPP uses 2012–2016 ACS. Release roughly every 5 years.	Available annually since year 2002 with the exceptions of some states.	Release roughly every 5–10 years. The 2016 NHTS Public Use Data will be released in early 2018.
What workplace information does it have?	Have 115 workplace-based tables for over 200,000 geographies. Standard tables include workplace location, commute mode, departure time from home, arrival time to work, travel time (minutes), sex, age, race, ethnicity, citizenship status, language spoken, earnings, poverty status, occupation, industry, class of worker, hours worked each week, weeks worked in the past 12 months, earnings, number of vehicles available, household size, number of workers in household.	Provides workplace characteristics (i.e., firm size, firm age, NAICS industry sector, work location) and worker characteristics (i.e., primary–secondary job, earnings, education, age, gender, ethnicity, house location).	The NHTS add-on deliverables provides the detailed location information of origin and destination address, which can be linked with main data files. The main data files includes characteristics for each household, person, worker, vehicle, and daily travel data. For each worker, NHTS provides information on full/part-time, number of jobs, job types, workplace location, usual mode, distance, and arrival time to work, drive alone/carpool, and flexibility in work arrival time.
What is the smallest geographic unit available?	TAZs	Census blocks	Latitude and longitude of trip ends (for add-ons only)
Who is included in the survey?	Collects employment characteristics from workers of 16 years and over including telework and noninstitutional group quarters (i.e., college dormitories and military barracks). On the other hand, the data does not capture secondary job and excludes workers living in institutionalized group quarters such as prisoners and nursing homes.	Includes all ages of workers. It includes all jobs under state UI law, which is 95% of private sector wage and salary employment. Also, it covers most of civilian federal employment using records from the Office of Personnel. Does not cover self-employment, military employment, the U.S. Postal Service, and informal employment.	Includes civilian, non-institutionalized population of the U.S. of 5 years and older. It excludes institutionalized group quarters (i.e., motels, hotels, nursing homes, prisons, barracks, convents or monasteries and any living quarters with 10 or more unrelated roommates.
How does it geocode residential–employment?	92% of worker records are successfully geocoded to place level. The leftover cases are allocated to a workplace location for geographies down to the place level.	Geocode using detailed addresses within the administrative records, which is 95% of private-sector wage and salary employment.	Uses online interactive tool to real-time geocode during the interview process.

the CTPP and NHTS (*1*, *23*) and then classified 12 categories based on the review of subject areas and keywords of the 305 studies. Table 16.2 summarizes the list of subject areas used in this paper and their keywords. Appendix A contains a listing of the 305 studies examined in this paper, including their titles, authors, and subject area categories.

Summary of Uses and Applications of the CTPP Data

This section summarizes the various uses and applications of the CTPP data by subject area, based on the review of 305 studies that cited the use of the CTPP data. Among the CTPP's three component tables—Part 1 residence based tables, Part 2 workplace based tables and Part 3 home-to-work flow tables, it is observed that Part 2 workplace based tables was most frequently used, followed by Part 3 home-to-work flow tables, which indicates the CTPP workplace data is a critical component of the CTPP. Among the 305 studies, Part 2 workplace based tables were used in 179 studies (59%) and Part 3 home-to-work flow tables were used in 170 studies (56%) while Part 1 residence based tables were used in 127 studies (42%). It is observed that 126 studies (41%) used Part 1 residence based tables and/or Part 3 home-to-work flow tables only. The majority (73%) of those 126 studies that didn't use the CTPP workplace data utilized Part 3 home-to-work flow tables.

Figure 16.1 summarizes the uses of the CTPP data by subject area, comparing between studies that used the CTPP workplace data and studies that didn't use the CTPP workplace data. Among the 12 subject area categories, the most common uses of the CTPP data are commuting patterns and job-housing mismatch and travel demand modeling and forecasting, followed by transit planning, policy analysis and travel behavior analysis. It is observed that 66 of the entire 305 studies (22%) utilized the CTPP data for the subject of commuting patterns and job-housing mismatch, 61 studies (20%) for the subject of travel demand modeling and forecasting, 37 studies (12%) for the subject of transit planning, 37 studies (12%) for the subject of policy analysis, and 36 studies (12%) for the subject of travel behavior analysis. Of the 179 studies that cited the use of Part 2 workplace-based tables, the five most common uses are commuting patterns and job-housing mismatch (38 studies, 21%), travel demand modeling and forecasting (29 studies, 16%), built environment and accessibility study (26 studies, 15%), trend analysis and market research (24 studies, 13%), and policy analysis (22 studies, 12%). Of the 126 studies that didn't cite the use of the CTPP workplace data, the five most common uses are travel demand modeling and forecasting (32 studies, 25%), commuting patterns and job-housing mismatch (28 studies, 22%), travel behavior analysis (21 studies, 17%), transit planning (20 studies, 16%), and policy analysis (15 studies, 12%). The results indicate that the CTPP workplace data are useful especially for the subjects of trend analysis and market research, built environment and accessibility study, policy analysis, and commuting patterns and job-housing mismatch.

TABLE 16.2 Subject Areas and Relevant Keywords

Subject Area	Relevant Keywords
Bicycle and Pedestrian Studies	Bicycle commuting, bikeway, off-road trail system, pedestrian, physical activity, walking
Built Environment and Accessibility Study	Accessibility, built environment, decentralization of residence and employment, job accessibility, job opportunities, job proximity, land use intensity, polycentric city, spatial concentration, spatial inequality, spatial mismatch, sprawl, street connectivity, transportation infrastructure, urban spatial structure
Commuting Patterns and Job-Housing Mismatch	Commute distance and time, commute flow, commute pattern, job-housing balance, JTW trips, spatial relationship between residence and workplace, travel patterns
Demographics Study	Baby boomers, demographic, gender, household attribute, immigrant population, income, millennials, neighborhood type, poor job seekers, race/ethnicity, socioeconomic characteristics, wage
EJ and Title VI	Accessibility, education attainment, EJ, gender, impact equity analysis, immigrants, limited English proficiency, low income, low wage workers, minority, national origin, poverty, social equity, social impact, spatial inequality, Title VI, transportation cost and needs
Health, Safety and Environmental Issues	Asthma, cholesterol, crime, electric power plants, electric vehicle charging, energy analysis, greenhouse gas reductions, environmental analyses, health impact, heat, plug-in hybrid electric vehicles, obesity, ozone, vehicle emission
Policy Analysis	Congestion management, congestion relief strategies, disaster relief strategies, enterprise zone policy, gasoline tax revenue, highway congestion pricing, park-and-ride, parking requirements, regulations, ridesharing, transit subsidies, transportation pricing strategies, urban containment policy, urban growth control
Survey, Data Synthesis and Research Methods	Cellular data, data fusion, data matching, data synthesis, fuzzy clustering method, indicator development, interview, IPF, methodology, model-based synthesis, sampling, synthetic data techniques, transportation indicators, travel survey
Transit Planning	Bus rapid transit, bus transit system, commuter rail system, interurban rail trip, light rail, multimodal transportation, new transit services, public transit study, transit access, transit demand analysis, transit dependent populations, transit feasibility analysis, transit mode share, transit planning, transit propensity index, transit ridership, transit subsidies
Travel Behavior Analysis	Behavior uncertainty, commuting behavior, driving alone, household travel, immigrants, individual characteristics, minority travel patterns, mode choice, segregation, social interaction, socioeconomic characteristics, travel behavior, travel pattern, travel-related characteristics, vehicle ownership, vehicle transit behavior
Travel Demand Modeling and Forecasting	Activity based model, discrete choice model, freight model, gravity model, mode and destination choice model, model calibration and validation, multinomial logit, regional transportation plan, socioeconomic forecasting, surface model, travel demand model, travel forecasting, travel simulation, trip attraction model, trip distribution, trip generation, VMT
Trend Analysis and Market Research	Central business district, changing patterns, economic centers, economic activity centers, economic structure, edge cities, edgeless cities, employment centers, housing price, interurban movements, location quotient, market analysis, population distribution pattern, spatial trend, sprawl, temporal dynamic, trend analysis, typology of land use patterns

NOTE: Subject areas and relevant keywords are sorted in alphabetical order.
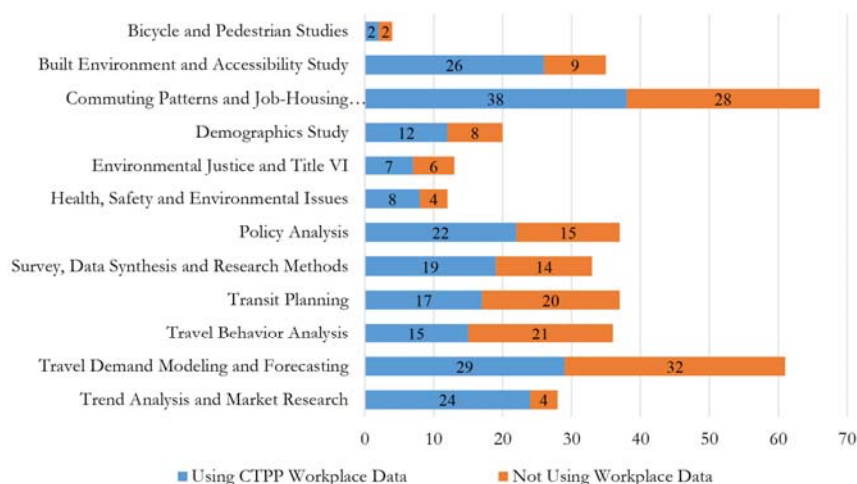
FIGURE 16.1 Uses of the CTPP data by subject area. (Note: Some studies were categorized into multiple subject areas as they encompass multiple subject areas and no one subject area was considered the primary category.)

Figure 16.2 summarizes the uses of the CTPP data by publication year. The publication dates of the 305 studies range from 1989 to 2017—27 studies before year 2000, 129 studies from year 2000 to year 2009, and 149 studies since year 2010. Of the 179 studies that cited the use of Part 2 workplace-based tables, 16 studies were published before year 2000, 83 studies were published from year 2000 to year 2009, and 80 studies were published since year 2010. As shown in Figure 2, it is observed that the CTPP data has been increasingly utilized since year 2005. Among the entire 305 studies, 247 studies (81%) were published since year 2005, and among the 179 studies that used the CTPP workplace data, 126 studies (79%) were published since year 2005.

Figures 16.3 and 16.4 summarize the uses of the CTPP data and its workplace data by subject area and publication year. During review periods, two subject areas—commuting patterns and job–housing mismatch and travel demand modeling and forecasting—have been constantly popular uses of the CTPP data. The subject of commuting patterns and job-housing mismatch accounts for 4 of 27 (15%) studies published before year 2000, 29 of 129 (22%) studies published between year 2000 and year 2009, and 33 of 149 (22%) studies published since year 2010. The subject of travel demand modeling and forecasting accounts for 26%, 17%, and 21%, respectively. On the other hand, some subject areas such as bicycle and pedestrian studies, EJ and Title VI, and health, safety and environmental issues are newly analyzed since year 2000. While there were no studies that cited the use of the CTPP data for those three subject areas before year 2000, those three subject areas, taken together, account 9% of 129 studies published between year 2000 and year 2009, and 12% of 149 studies published since year 2010. Of the 179 studies that cited the use of Part 2 workplace based tables of the CTPP data, commuting patterns and job-housing mismatch, travel demand modeling and forecasting, and built environment and accessibility study have been constantly popular uses during review periods. The subject of commuting patterns and job-housing mismatch accounts for 3 of 16 (19%) studies published before year 2000, 17 of 83 (20%) studies published between year 2000 and year 2009, and 18 of
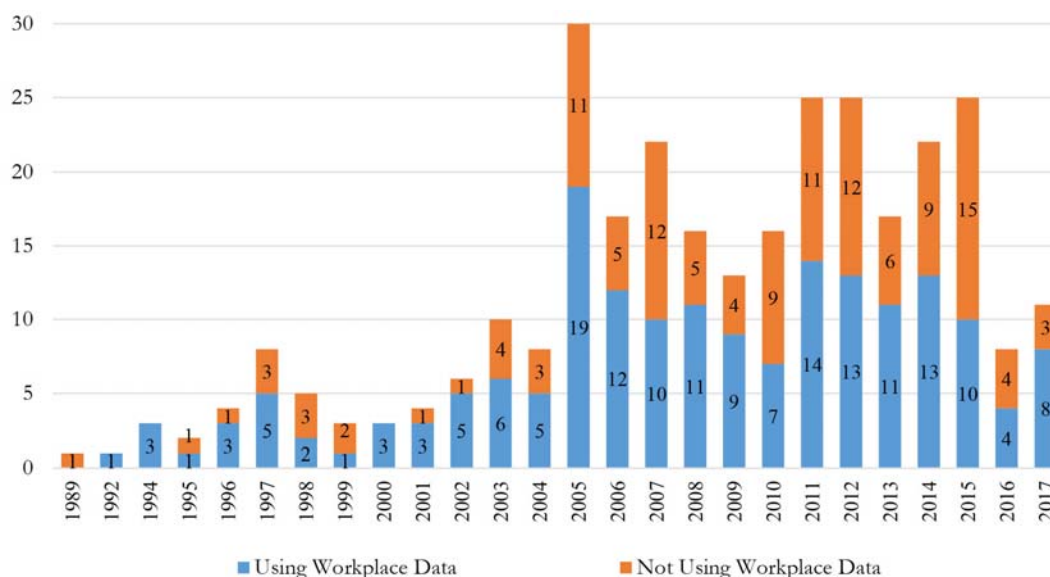
FIGURE 16.2 Uses of the CTPP data by publication year.
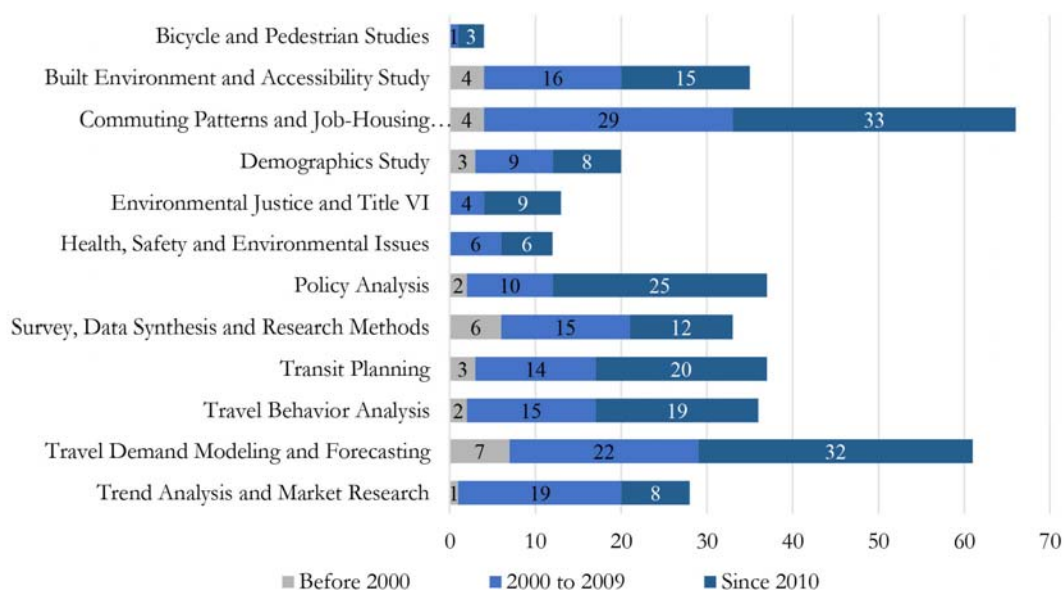

FIGURE 16.3 Uses of the CTPP data by subject area and publication year. (Note: Some studies were categorized into multiple subject areas as they encompass multiple subject areas and no one subject area was considered the primary category.)
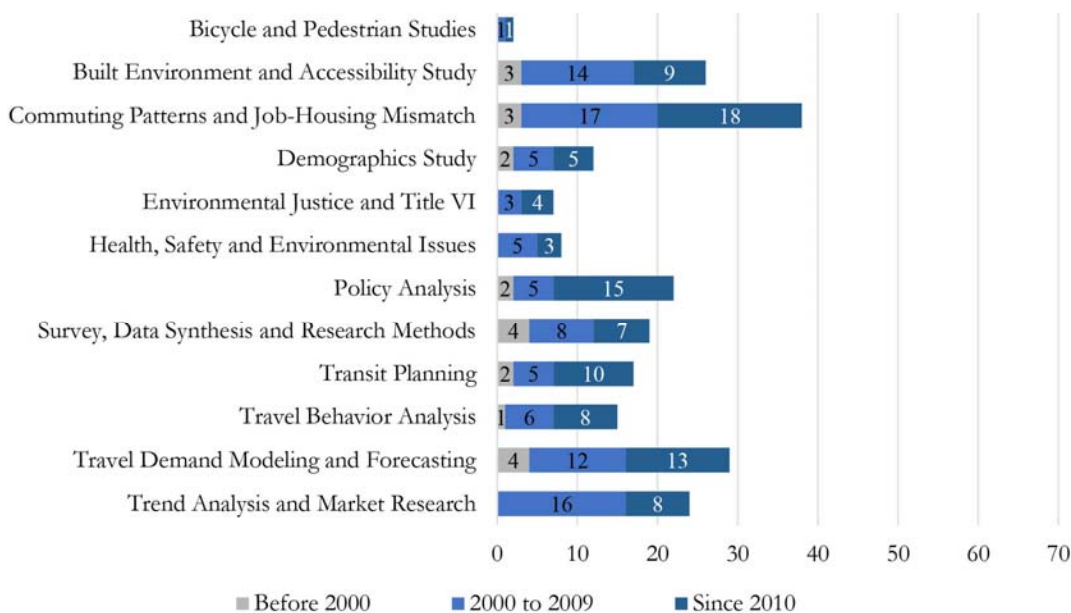
FIGURE 16.4 Uses of the CTPP workplace data by subject area and publication year. (Note: Some studies were categorized into multiple subject areas as they encompass multiple subject areas and no one subject area was considered the primary category.

80 (23%) studies published since year 2010. The subject of travel demand modeling and forecasting accounts for 25%, 14% and 16%, and built environment and accessibility study accounts for 19%, 17% and 11%, respectively. The results indicate that, during review period, the CTPP workplace data has been constantly utilized for a significant number of research on the subject of commuting patterns and job-housing mismatch, travel demand modeling and forecasting, and built environment and accessibility study. In addition, the CTPP workplace data has been utilized in research on newly emerging subjects since year 2000 such as trend analysis and market research, health, safety and environmental issues, EJ and Title VI, health, and bicycle and pedestrian studies.

Case Studies: Utilizing the CTPP Workplace Data in Transportation Planning and Research

This section introduces the case studies of how the CTPP workplace data is utilized in transportation planning and research. The purpose of this section is to explore some of the applications related to transportation planning and research that were performed using the CTPP workplace and to indicate how essential the CTPP workplace data was to the completion of the applications, including whether or not the data was essential; if the data was, what made them so; and, if the data was not essential, what information might have been substituted to complete the application.

Spatial and Socioeconomic Analysis of Commuting Patterns in Southern California: Using LODES, CTPP, and ACS PUMS

As a part of EJ analysis of the regional transportation plan, the Southern California Association of Governments (SCAG) examined commuting distance by income to better understand the relationship between commuting pattern and socioeconomic characteristics in Southern California region. Multiple workplace data were used in this study, including the LODES Version 7.1 data, the CTPP 5-Year 2006–2010 ACS data, and the 2009–2013 ACS 5-year PUMS (*24*).

Due to the differences in data structure, variable and geographic units among those three datasets, this study uses different methodologies to examine the relationship between commute distance and income level. Using the LODES data, this study examined the median commute distance, by wage group, for six counties in the region for the years 2002, 2008, and 2012. The commute distance measured is the Euclidean distance, straight-line distance, or distance measured "as the crow flies" between the centroid of origin block and destination block, and the commute distance is weighted by block-level commuter number. Given its minimum geographic unit is Census block, the LODES data allowed this study to conduct analysis in a more geographically detailed way than other two datasets. Using the CTPP data, this study examined the median commute distance by income group for six counties in the region. The commute distance measured is the Euclidean distance between the centroid of origin tract and destination tract and the commute distance is weighted by tract-level commuter number. As the CTPP data provides more detailed information of workplace compared to the LODES data, this study examined the median commute distance by additional CTPP variables, such as household income, poverty status and vehicles available. Using PUMS data, this study examined the median wages for intercounty and intracounty commuters to compare the median wages between workers residing in their destination–work counties and outside their destination-work-counties. The most detailed unit of geography contained in the PUMS dataset is the PUMA.

The results of this study showed the similar patterns in commuting distance by income group among LODES, CTPP, and PUMS datasets: (1) higher wage workers tend to commute longer distance than lower wage workers; (2) the commute distance is growing in all six counties between 2002 and 2012; and (3) the commute distance of workers in inland counties (Riverside and San Bernardino counties) is longer and grows more rapidly than in coastal counties (Los Angeles and Orange counties). However, it was also observed that the median commute distance from the LODES data is longer than those from the median commute distance from the CTPP data, possibly resulting from differences between two datasets in data input source, data coverage, geographic tabulation level, time period, and characteristics level.

Small Area Applications Using 1990 CTPP: Gainesville, Florida

This study presents a case study of the main CTPP applications, limitations or problems encountered with the CTPP data, and results of the applications for the Gainesville Urbanized Area in its long-range transportation planning efforts (*25*). This study demonstrates that the CTPP provided detailed information about socioeconomic and travel characteristics that was unavailable from other sources and the CTPP data were of value during several stages of development of the Gainesville Urbanized Area 2020 Transportation Plan. The study focuses on how the CTPP was used to validate the travel demand model in preparation for the development and evaluation of multimodal alternatives for the plan. The study notes that the CTPP workplace

data was the best source of employment data by TAZ. Several categories of employment by occupation were collapsed into the three required by TRANPLAN, the standard travel demand forecasting software used in Florida. The study also noted that some errors were observed during the validation data review process, e.g., misallocating employees of the University of Florida to a single TAZ located across the street from the campus.

The study underscored that the household travel survey for Gainesville was not up to date when preparing the plan and limited staff and financial resources required that the CTPP be used to identify key travel parameters to improve the accuracy of the forecasts. Despite some errors, the study highlights that the CTPP data was essential to the completion of the plan as it provided information unavailable from other sources. It also states that, without the CTPP data, the planning effort would have been less refined, would have had less public support and likely would have resulted in a different transportation plan than the one adopted.

Access to Growing Job Centers in the Twin Cities Metropolitan Area

The Twin Cities Metropolitan Area has experienced significant decentralization of population and jobs during recent decades (*26*). This study investigated job growth, job decentralization, and commuting patterns in the Twin Cities Metropolitan Area during the 1990s, focusing particularly on how these patterns affect the opportunity structures that is, the ease of access to growing job centers and adequate, affordable housing facing people of color and lower income households. The study utilized the workplace-based tables of the CTPP compiled by TAZ in 1990 and 2000 to identify small- and large-scale job clusters, to examine job growth by job center type, to examine commuting patterns to the job centers, and to show the racial breakdowns of the workers commuting to each center. This study used the 1990 and 2000 CTPP data compiled by TAZ to identify job centers that were defined as adjacent TAZs with greater-than-average numbers of jobs per square mile and total employment exceeding 1,000 jobs. The 1990 and 2000 CTPP data also used to analyze the racial breakdown of workers broken out—workers of Hispanic origin or other racial–ethnic backgrounds—by the type of job center they work in. Additionally, data for WTT of the CTPP were used for commuter-shed analysis, deriving the areas around each job center representing 20-, 30-, and 40-min commutes in 1990 and 2000.

The results of this study indicate that, if current patterns continue, the potential for transit in the Twin Cities Metropolitan Area would decline, and consequently, job opportunities available to workers who rely on transit—lower-income workers who are disproportionately people of color—will decline. Additionally, the study highlights serious shortfalls in affordable housing in fast-growing job centers and social equity implications for people working in declining job centers—limiting workers' future opportunities and lessening their potential for higher earnings in the future.

Conclusion

This paper explored the major data products that have been widely utilized in transportation planning and research—the CTPP, the LODES, and the NHTS; and then, it examined the strengths and limitations of the CTPP workplace data as compared to the LODES and the NHTS. It is important to have a full understanding of each data's characteristics before incorporating it into a project. The CTPP workplace data has been utilized by various organizations and agencies due to its unique and rich tabulations even at small geography like Census tract. Over the years, CTPP has

shown a tremendous improvement in its contents from 1990 to 2010 by introducing more customized tables. Also, the CTPP workplace data generated from the ACS 5-year estimates allows users to perform temporal and spatial analysis with relatively lower MOE than using the ACS 1- or 3-year estimates, although it still has a higher MOE than using the decennial Census.

As compared to the CTPP, the LODES provides users with workplace information in more geographically detailed manner, and therefore, it allows users to perform small area analysis related to workplace and O-D flows. On the other hand, the CTPP provides invaluable information for transportation planners and researchers that are not included in the LODES; and therefore, it allows users to analyze workplace and travel patterns with much more socioeconomic and travel characteristics, such as means of transportation, WTT, vehicle available and poverty status. Although the LODES provides the longitudinal employment statistics annually, the LODES data is not available prior to the year 2002 and it does not have a consistent information for certain variables. Also, the LODES data is not available for the whole United States. Those limitations make it hard to perform certain longitudinal analysis, especially when users need workplace information prior to year 2002, while the CTPP allows users to utilize the workplace data back to the year 1990.

The upcoming CTPP version uses the ACS 2012–2016 5-year estimates to generate its special tabulations. On an important note, the customized tables in this upcoming CTPP version will be reduced by about one-third, as compared to the 2006–2010 CTPP. Accuracy of geocoding workplace locations is also considered an important component in improving the CTPP workplace data. Incorporating real-time mapping application for respondents when responding to ACS may improve geocoding issues. The CTPP workplace data may be integrated with other major datasets such as LODES and NHTS to unlock more unique workplace tabulations. Additionally, developing user friendly applications to easily retrieve the customized tables from the big CTPP datasets, sharing the success stories through CTPP website and professional conferences, and collaborating with partner agencies, including MPOs and COGs, in the nation to provide a technical support to local jurisdictions and data users could encourage users to better utilize the CTPP workplace data in the future.

This paper also summarized the various uses and applications of the CTPP data product and its workplace data. Over 300 studies that cited the use of the CTPP data were reviewed in this paper and were grouped into 12 subject area categories based the review of the studies. According to the review results, a considerable number of research reports have been conducted on the subjects of commuting patterns and job-housing mismatch, and travel demand modeling and forecasting, and it is expected that they will be the key subject areas in the future. The results indicate that the CTPP workplace data is useful especially for transportation planning and research on the subjects of trend analysis and market research, built environment and accessibility study, policy analysis, and commuting patterns and job-housing mismatch. Also, given that research has increased since 2000 on the subjects of bicycle and pedestrian studies, EJ and Title VI, and health, safety and environmental issues, and trend analysis and market research, the CTPP workplace data can be more widely utilized in the future on those newly emerging subject areas. Additionally, demographics may also be the emerging topic area, given the trend of an aging population and the millennial generation and workforce not only in the nation.

Appendix A: Studies That Cite the Use of CTTP Data

Author/Year	Title	Subject Area(s)
Alexander et al., 2015	Assessing the Impact of Real-Time Ridesharing on Urban Traffic Using Mobile Phone Data	PO
Alexander et al., 2015	Origin–Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data	CJ
Anas, and Hiramatsu, 2012	The Effect of the Price of Gasoline on the Urban Economy: from Route Choice to General Equilibrium	PO
Antipova et al., 2011	Urban Land Uses, Sociodemographic Attributes and Commuting: A Multilevel Modeling Approach	TB, CJ
Appold, 2015	Airport Cities and Metropolitan Labor Markets: an Extension and Response to Cidell	BA
Atlanta Regional Commission, 2005	Comparison of 2000 JTW Census Data, Gravity Model Results, and SMARTRAQ Household Travel Survey Data, in the Trip Distribution Model at the ARC	MF
Baltimore Metropolitan Council, 2014	Web Application to Examine Commuting in Baltimore Region Baltimore Metropolitan	CJ
Barnes, 2005	The Importance of Trip Destination in Determining Transit Share	TP
Baum-Snow, 2010	Changes in Transportation Infrastructure and Commuting Patterns in U.S. Metropolitan Areas, 1960–2000	CJ, BA
Becker et al., 2011	A Tale of One City: Using Cellular Network Data for Urban Planning	SD
Bhat et al., 2013	A Household-Level Activity Pattern Generation Model with an Application for Southern California	MF, TB
Bohon et al., 2008	Transportation and Migrant Adjustment in Georgia	BA, TB
Boyce, and Bar-Gera, 2003	Validation of Multiclass Urban Travel Forecasting Models Combining Origin–Destination, Mode, and Route Choices	MF
Bricka, 2004	Variations in Hispanic Travel Based on Urban Area Size	TB, DM
Cambridge Systematics, 2017	Using Census Data to Develop Efficient Household Travel Survey Sampling Plans	SD
Cambridge Systematics, 2013	Counting Workers: Comparison of Employment Data for CPS, ACS and LODES	SD
Cambridge Systematics Inc., 2005	Use of CTPP Data in the Cook–DuPage Corridor Study	TB
Cambridge Systematics Inc., 2009	Analysis of Iterative Proportion Fitting in the Generation of Synthetic Populations	SD
Cambridge Systematics Inc., 2009	Model-Based Synthesis of Household Travel Survey Data	SD
Cambridge Systematics, Inc., 2009	Disclosure Avoidance Techniques to Improve ACS Data Availability for Transportation Planners	SD
Cambridge Systematics, Inc., 2004	CTPP Workers-at-Work Compared to Other Employment Estimates	SD
Cambridge Systematics, Inc., 2011	NCHRP 08-36, Task 98: Improving Employment Data for Transportation Planning	SD
Cambridge Systematics, Inc., 2011	Using 2006–2008 CTPP in Planning for San Juan Light Rail Transit Study	TP

Continued on next page.

Author/Year	Title	Subject Area(s)
Cambridge Systematics, Inc., 2014	FTA New Starts Project Using CTPP	TP
Cambridge Systematics, Inc., 2016	Research for the AASHTO Standing Committee on Planning. Task 127. Employment Data for Planning: A Resource Guide	SD
Capon, 2007	Health Impacts of Urban Development: Key Considerations	HS, BA
Case et al., 2008	Simulating the Economic Impacts of a Hypothetical Bio-Terrorist Attack: A Sports Stadium Case	PO
Catala, 2005	Florida Journey to Work GIS Web-Site	CJ
Catanzarite, 2012	Edge Cites Revisited: The Restless Suburban Landscape	TA
Center for Transportation Research, 2011	Understanding Emerging Commuting Trends in a Weekly Travel Decision Frame Implications for Mega Region Transportation Planning	CJ
Center for Urban and Regional Studies, 2012	Using CTPP 2000 Employment and Worker Flow Data to Build Integrated Land Use–Travel Demand Models of Small Communities and Rural Areas	MF
Center for Urban Transportation and University of South Florida, 2005	Online Web Application Using Journey-to-Work Data from CTPP 2000	CJ
Cervero and Kockelman, 1997	Travel Demand and the 3Ds: Density, Diversity, and Design	MF, TB, BA
Cervero and Landis, 1997	Twenty Years of the Bay Area Rapid Transit System: Land Use and Development Impacts	TP
Cervero and Wu, 1997	Polycentrism, commuting, and residential location in the San Francisco Bay area.	CJ
Cervero et al., 2002	Transportation as a Stimulus of Welfare-to-Work: Private versus Public Mobility	PO
Cervero, 1994	Use of Census Data for Transit, Multimodal, and Small-Area Analyses	TP
Cervero, 2001	Efficient Urbanisation: Economic Performance and the Shape of the Metropolis	BA
Chattanooga Transportation Planning Organization, 2015	Chattanooga–Hamilton County North Georgia Data Collection Phase II	MF
Chen and Suen, 2010	Richmond's Journey-to-Work Transit Trip-Making Analysis	TP
Chen et al., 2007	Role of the Built Environment on Mode Choice Decisions: Additional Evidence on the Impact of Density	TB
Chen et al., 2011	Development of Indicators of Opportunity-Based Accessibility	SD
Chirumamilla, 1998	Discrete-Continuous Model of Household Vehicle Ownership and Trip Generation	TB, MF
Choand Rodriguez, 2015	Location or Design? Associations Between Neighbourhood Location, Built Environment, and Walking	BA, BP
Chow et al., 2010	Subregional Transit Ridership Models Based on Geographically Weighted Regression	TP

Continued on next page.

Author/Year	Title	Subject Area(s)
Chu, 2012	Census/ACS/CTPP Data for Transit Planning	TP
Chung, 2003	Temporal Analysis of Land Use and Transportation Investments With Geographic Information System	BA
City of Madison, 2007	Downtown Madison Market Analysis	TA
Clifton et al., 2012	Household Travel Surveys in Context-Based Approach for Adjusting ITE Trip Generation Rates in Urban Contexts	SD
Coleman, 1999	Forecasting Interurban Rail Trips: an Overview of Two Scenarios	TP
Columbia University Graduate School of Architecture, Planning and Preservation, 2014	Promoting Bus Rapid Transit Options on the New Tappan Zee Bridge and I-287 Corridor	TP
COPAFS, 2012	A Preview of Small Area Transportation Data from the American Community Survey	SD
Cutsinger and Galster, 2006	There Is No Sprawl Syndrome: A New Typology of Metropolitan Land Use Patterns	TA
Cutsinger et al., 2005	Verifying the Multi-Dimensional Nature of Metropolitan Land Use: Advancing the Understanding and Measurement of Sprawl	TA
Cutsinger et al., 2005	Verifying the Multi-Dimensional Nature of Metropolitan Land Use Advancing the Understanding and Measurement of Sprawl	BA
Delaware Valley Regional Planning Commission, 2006	Development of Zonal Employment Data for Delaware Valley Region Based on Census 2000	TA
Delaware Valley Regional Planning Commission, 2006	Evaluation of Census Transportation Planning Package 2000 for the Delaware Valley Region	SD
Deloitte, 2015	Ridesharing: the Easiest (and Hardest) Approach to Congestion Reduction	PO
Denise, 2011	Comparing Methods for Estimation of Daytime Population in Downtown Indianapolis, Indiana	MF
Dentel-Post et al., 2017	Getting People Around After the Trains Stop Running : A Transit Propensity Index for Late-Night Service Planning	TP
Denver Regional Council of Governments, 2010	Using ACS/CTPP data in Activity-Based Model Calibration	MF
Department of Sociology–Anthropology Illinois Sate University, 2002	Use of CTPP files for Analysis of Metropolitan Area Multiple Nuclei	TA
Des Moines Area MPO, 2005	U.S. Census, CTPP, and NHTS Data Used in the Des Moines Area MPO's Travel Demand Model	MF
Diao, 2015	Are Inner-City Neighborhoods Underserved ? an Empirical Analysis of Food Markets in a U.S. Metropolitan Area	BA
Dolney, 2009	Using Simulation to Estimate Vehicle Emissions in Response to Urban Sprawl Within Geauga County, Ohio	HS
Eastgate MPO, 2006	Use of CTPP at the Eastgate MPO, Youngstown, Ohio	MF

Continued on next page.

Author/Year	Title	Subject Area(s)
Ed, 1996	Census Data Use in Illinois By Small Metropolitan Planning Organizations	SD
Eisman, 2012	Spatial Analysis of Urban Built Environments and Vehicle Transit Behavior	TB, BA
Employment and Training Institute and University of Wisconsin–Milwaukee, 2005	Neighborhoods at Work	TA
Evans, 2016	CTPP Tract-to-Tract Commute Visualization	CJ
Ewing et al., 2003	Urban Sprawl and Transportation	TA
Farber et al., 2015	Measuring Segregation Using Patterns of Daily Travel Behavior: A Social Interaction Based Model of Exposure	TB, CJ
Farhan and Murray, 2008	Siting Park-and-Ride Facilities Using A Multi-Objective Spatial Optimization Model	PO, MF
Fayyaz et al., 2017	Dynamic Transit Accessibility and Transit Gap Causality Analysis	TP
Federal Emergency Management Agency, 2008	HAZUS – MH: FEMA's Software Program for Estimating Potential Losses from Disasters	HS
FHWA and Cambridge Systematics Inc, 2005	Disclosure and Utility of Census Journey-to-Work Flow data from the American Community Survey: Is There a Right Balance?	SD
FHWA, 2007	Peak Spread of Journey-to-Work	CJ
FHWA, 2008	Using Census Data to Analyze Limited English Proficiency Populations for Transit Applications	TP, EJ
FHWA, 2009	Vehicle Availability and Mode to Work by Race and Hispanic Origin, 2007	TB, DM
FHWA, 2013	Commutation Flow: CTPP 2000, ACS and CTPP, and LEHD-OTM	CJ
FHWA, 2014	How Much Do We Spend on Housing and Transportation?	TB
FHWA, 2014	How Hard is it to Count Workers? Self-employment data in Nonemployer statistics and in American Community Survey	SD
FHWA, 2006	Use of CTPP 2000 in FTA New Starts Analysis	TP
FHWA, 2010	CTPP Data to Support Transit Ridership Forecasting	TP
FHWA, 2013	Census Data Application for Title VI Service Equity Analysis	EJ
Fredericksburg Area Metropolitan Planning Organization, 2013	Population and Employment Projection Dataset and Methodology	MF
Freedman et al., 2008	New Approaches to Creating Data for Economic Geographers	SD
Freedman, 1999	Comparing Stratified Cross-Classification and Logit-Based Trip Attraction Models	MF
Funderburg et al., 2010	New Highways and Land Use Change: Results from A Quasi-Experimental Research Design	MF, BA

Continued on next page.

Author/Year	Title	Subject Area(s)
Gabbe, 2017	Why Are Regulations Changed? A Parcel Analysis of Upzoning in Los Angeles	PO
Glaeser, 1996	Spatial Effects Upon Employment Outcomes: The Case of New Jersey Teenagers. Discussion	TB
Gottlieb and Lentnek, 2001	Spatial Mismatch Is Not Always A Central-City Problem: an Analysis of Commuting Behaviour in Cleveland, Ohio, and Its Suburbs	CJ
Greater Buffalo-Niagara Regional Transportation Council, 2003	2002 Regional Transportation Survey	SD
Greaves, 1989	Simulating Household Travel Survey Data in Metropolitan Areas	MF
Greenberg and Evans, 2015	Pay-to-Save Transportation Pricing Strategies and Comparative Greenhouse Gas Reductions: Responding to Final Federal Rule for Existing Electric Utility Generating Units	HS, PO
Gregor, 1998	Assessing Intercity Commuting Patterns in the Willamette Valley Using the Census Transportation Planning Package	CJ
Grengs, 2010	Job Accessibility and the Modal Mismatch in Detroit	BA, PO
Guldmann, 2013	Analytical Strategies for Estimating Suppressed and Missing Data in Large Regional and Local Employment, Population, and Transportation Databases	SD
Hampton Roads Planning District Commission, 2005	A Compendium of 2000 Census Commute Analyses for the Hampton Roads Region	CJ
Han and Zegras, 2016	Exploring Model and Behavior Uncertainty	TB
Henson, 2011	Travel Determinants and Multiscale Transferability of National Activity Patterns to Local Populations	TB
Herb and Herb, 2007	Racial Profiling and the Police : Utilizing the Census Transportation Planning Package to Benchmark Traffic Stops Made By the North Carolina State Highway Patrol	DM
Hirsch et al., 2017	Municipal Investment in Off-Road Trails and Changes in Bicycle Commuting in Minneapolis, Minnesota Over 10 Years: A Longitudinal Repeated Cross-Sectional Study	BP
Holleran and Duncan, 2012	Sketch-Level Feasibility Analysis of Commuter Rail Service Between Kannapolis and Charlotte, North Carolina	TP
Homer, 2004	Spatial Dimensions of Urban Commuting: A Review of Major Issues and Their Implications for Future Geographic Research	CJ, BA
Horner and Marion, 2009	A Spatial Dissimilarity-Based Index of the Jobs—Housing Balance: Conceptual Framework and Empirical Tests	TA
Horner and Mefford, 2005	Examining the Spatial and Social Variation in Employment Accessibility: A Case Study of Bus Transit in Austin, Texas	TP
Horner and Mefford, 2007	Investigating Urban Spatial Mismatch Using Job-Housing Indicators to Model Home–Work Separation	CJ
Horner and Murray, 2003	A Multiobjective Approach to Improving Regional Jobs-Housing Balance	PO

Continued on next page.

Author/Year	Title	Subject Area(s)
Horner, 2002	Extensions to the Concept of Excess Commuting	CJ
Horner, 2007	A Multi-Scale Analysis of Urban Form and Commuting Change in A Small Metropolitan Area (1990–2000)	CJ, TA
Horner, 2008	'Optimal' Accessibility Landscapes? Development of a New Methodology for Simulating and Assessing Jobs—Housing Relationships in Urban Regions	MF
Horner, 2010	How Does Ignoring Worker Class Affect Measuring the Jobs-Housing Balance? Exploratory Spatial Data Analysis	CJ
Houston-Galvaston Area council, 2005	How Census 2000 and CTPP 2000 Data Helped Us in the Use of Regional Travel Demand Forecast	MF
Hu and Wang, 2015	Decomposing Excess Commuting: A Monte Carlo Simulation Approach	CJ
Hu et al., 2017	Commuting Variability by Wage Groups in Baton Rouge, 1990–2010	CJ, EJ
Hu, 2013	Changing Job Access of the Poor: Effects of Spatial and Socioeconomic Transformations in Chicago, 1990–2010	BA, DM
Huntsinger, 2012	Temporal Stability of Trip Generation Models: an Investigation of the Role of Model Type and Life Cycle, Area Type, and Accessibility Variables	MF
Hwang and Thill, 2007	Using Fuzzy Clustering Methods for Delineating Urban Housing Submarkets	SD
Immergluck, 1998	Job Proximity and the Urban Employment Problem: Do Suitable Nearby Jobs Improve Neighbourhood Employment Rates?	BA, DM
Immergluck, 1998	Neighborhood Economic Development and Local Working: The Effect of Nearby Jobs on Where Residents Work	CJ, DM
Indian Nations Council of Governments, 2011	Using 2006-2008 CTPP and CTPP 2000 Data to Evaluate the Reliability of Travel Forecast Assumption	TB
Jang and Yao, 2011	Interpolating Spatial Interaction Data	SD
Jang and Yao, 2014	Tracking Ethnically Divided Commuting Patterns Over Time: A Case Study of Atlanta	CJ, DM
Jang et al., 2014	Spatial Analysis of the Baby Boomers' Jobs and Housing Patterns in a GIS Framework	TB, DM
Jeon et al., 2015	Application of CTPP Data for Validation of Regional Transportation Forecasting Models: MAG Experience	MF
Kawabata and Shen, 2007	Commuting Inequality between Cars and Public Transit: The Case of the San Francisco Bay Area, 1990-2000	TB
Kawabata, 2003	Spatial Distributions of Low-Skilled Workers and Jobs in U.S. Metropolitan Areas	CJ
Kawabata, 2002	Access to Jobs: Transportation Barriers Faced by Low-Skilled Autoless Workers in U.S. Metropolitan Areas	BA
Kawabata, 2009	Spatiotemporal Dimensions of Modal Accessibility Disparity in Boston and San Francisco	TB

Continued on next page.

Author/Year	Title	Subject Area(s)
Kentucky Transportation Cabinet, 2005	Use of Census Transportation Planning Package Data to Update the Kentucky Statewide Traffic Model	MF
Kentucky Transportation Center, 2010	Investigating Contextual Variability in Mode Choice in Chicago Using a Hierarchical Mixed Logit Model	TB, MF
Kim and Sang, 2006	Disaggregated Travel Forecasting	MF
Kim and Hewings, 2012	Integrating the Fragmented Regional and Subregional Socioeconomic Forecasting and Analysis: A Spatial Regional Econometric Input-Output Framework	MF
Kim and Hewings, 2013	Land Use Regulation and Intraregional Population-Employment Interaction	PO
Kim et al., 2012	Exploring Urban Commuting Imbalance By Jobs and Gender	CJ, BA, EJ
Kim et al., 2014	Exploring Job Centers By Accessibility Using Fuzzy Set Approach: The Case Study of the Columbus MSA	BA, DM
Kim, 2005	Trip Generation Model for Pedestrians Based on NHTS 2001	BP
King County Department of Transportation, 1999	Guidelines for Local Travel Demand Model Development	MF
Kirkpatrick, 1997	Conversion of GIS Databases for Modeling Rural Transportation Networks	MF
Kockelman, 1997	Effects of Location Elements on Home Purchase Prices and Rents in San Francisco Bay Area	TA
Krenzke and Hubble, 2009	Toward Quantifying Disclosure Risk for Area-Level Tables When Public Microdata Exists	SD
Kwon, 2015	The Effects of Urban Containment Policies on Commuting Patterns	PO
Lane, 2011	TAZ-Level Variation in Work Trip Mode Choice Between 1990 and 2000 and the Presence of Rail Transit	TB, TA
Lanton, 1996	Small-Area Applications Using 1990 Census Transportation Planning Package: Gainesville, Florida	MF
Larisa Ortiz Associates, 2014	Trenton Citywide Economic Market Study	TA
Layman and Horner, 2010	Comparing Methods for Measuring Excess Commuting and Jobs-Housing Balance Empirical Analysis of Land Use Changes	CJ
Lee et al., 2011	The Attributes of Residence/Workplace Areas and Transit Commuting	CJ, TP
Lee, 2005	A Spatial Analysis of Disaggregated Commuting Data: Implications for Excess Commuting, Jobs -Housing Balance, and Accessibility	CJ, BA
Lee, 2006	Urban Spatial Structure, Commuting, and Growth in United States Metropolitan Areas	BA
Lee, 2007	Edge or Edgeless Cities? Urban Spatial Structure in U.S. Metropolitan Areas, 1980 to 2000	TA

Continued on next page.

Author/Year	Title	Subject Area(s)
Levinson and Marion, 2010	The City Is Flatter: Changing Patterns of Job and Labor Access in Minneapolis–Saint Paul, 1995-2005	TA, TB
Limoges, 1996	Improvement of Decennial Census Small-Area Employment Data: New Methods to Allocate Ungeocodable Workers	SD
Lin and Long, 2006	What Neighborhood Are You in? Empirical Findings on Relationships Between Residential Location, Lifestyle, and Travel	TB
Lin and Long, 2008	What Neighborhood Are You in? Empirical Findings of Relationships Between Household Travel and Neighborhood Characteristics	TB, DM
Lindfors, 2012	Exploring the Commuting Interactions of Neighboring Metropolitan Areas	CJ
Linesch, 2012	Building a Statewide Traffic Count Database : A California Statewide Travel Demand Model Application	MF
Liu et al., 2009	Using GIS and CTPP Data for Transit Ridership Forecasting in Central Florida	TP
Long et al., 2014	Model-Based Synthesis of Household Travel Survey Data in Small and Midsize Metropolitan Areas	SD
Long, Liang and Lin, Jie, 2007	an Investigation in Household Mode Choice Variability across Metropolitan Statistical Areas for Urban Young Professionals	TB
Lu, 2015	Urban Mobility Evaluation Using Small-Area Geography and High-Resolution Population Data	CJ
Luce et al., 2006	Access to Growing Job Centers in the Twin Cities Metropolitan Area	TA
Madison Metropolitan Planning Area, 2006	Environmental Justice Analysis—Madison Area Transportation Regional Transportation Plan 2030	EJ
Maricopa Association of Governments, 2015	Use of GIS in the Validation of Travel Forecasting Models	MF
Maricopa Association of Governments (MAG), 2017	Application of ACS and CTPP Databases in Environmental Justice Assessment—Examples from MAG	EJ
Massachusetts Institute of Technology, 2009	The Effectiveness of Job–Housing Balance as a Congestion Relief Strategy	CJ, PO
Matsuo, 2013	Competition Over High-Income Workers: Job Growth and Access to Labor in Atlanta	DM
McCahill and Garrick, 2012	Automobile Use and Land Consumption: Empirical Evidence from 12 Cities	MF
McCahill, 2012	The Influence of Urban Transportation and Land Use Policies on the Built Environment and Travel Behavior	BA, TB
McCall et al., 2016	A County Level Methodology to Study the Impact on Emissions and Gasoline Tax Revenue of Plug-in Hybrid Electric Vehicles in New Jersey	PO, HS
McCall et al., 2016	Effect of Plug in Hybrid Electric Vehicle Adoption on Gas Tax Revenue, Local Pollution and Greenhouse Gas Emissions	HS

Continued on next page.

Author/Year	Title	Subject Area(s)
McGill University, 2010	The Spatial Patterns Affecting Home to Work Distances of Two-Worker Households	CJ
McNeely, 2007	Development of a Ridership Forecasting Tool for Small Public Transit Systems Using GIS	TP
Metro North Rail, 2015	Measuring Change in Transit Ridership for A New Mode Using ACS: The Case of Hudson Bergen Light Rail and Light Rail Overall	TP
Metropolitan Transit Authority and New York City Transit, 2004	Second Avenue Subway in the Borough of Manhattan, New York County, New York	TP
Metropolitan Transportation Commission, 2005	Environmental Justice for Long-Range Regional Transportation Plans: Using Census Data to Target Communities of Concern	EJ
Metropolitan Transportation Commission, Oakland, 2003	Commuting Patterns of Immigrants	CJ, DM
Mishra et al., 2011	A functional integrated land use-transportation model for analyzing transportation impacts in the Maryland–Washington, D.C., Region	PO, MF
Mississippi River Regional Planning Commission, 2017	Commuter Feasibility Study—Arcadia to La Crosse and Tomah to La Crosse	TP
Missoula Metropolitan Planning Organization, 2015	2016 Missoula Long-Range Transportation Plan	MF
Mix, 2005	Evaluating the Local Employment Dynamic Program as an Alternate Source of Place of Work Data for Use By Transportation Planners	SD
Mohan, 2004	Household Travel Survey Data Fusion Issues	SD
Moore and Campbell, 2014	The Correlates of Congestion: Investigating the Links Between Congestion and Urban Area Characteristics	PO
MTA New York City Transit, 2013	New York City Transit's Environmental Justice Strategies: Using CTPP Journey-to-Work Data to Perform Service Change Impact Analysis by Demographics	EJ, PO
Mulbrandon, 2007	An Agent-Based Model to Examine Housing Price, Household Location Choice, and Commuting Times in Knox County, Tennessee	MF
Murakami et al. , 2014	Workplace Geocoding Issues	SD
National Academic of Science, 2012	Smoothing the Borders of Labor Markets and Payment Areas: Use of the "Journey to Work" Data in Recommendations to Refine Medicare's Geographic Payment Adjusters	HS
National Research Council et al., 1994	Historic Uses of Census Data in Transportation Planning and Future Needs	SD

Continued on next page.

Author/Year	Title	Subject Area(s)
National Research Council et al., 1994	The Decennial Census and Transportation Planning: Planning for Large Metropolitan Areas	PO
Nelson et al., 2007	Transit in Washington, D.C.: Current Benefits and Optimal Level of Provision	TP, PO
New York State DOT, 2011	Commuting Flow: CTPP 2000, ACS and CTPP, and LEHD-OTM	CJ
New York University Wagner School of Public Service, 2010	The High-Speed Rail Development in the Northeast Megaregion of the United States: A Conceptual Analysis	TP
Newburger et al., 2011	The City in the Twenty-First Century: Neighborhood and Life Chances : How Place Matters in Modern America	TA
Newman and Bernardin, 2010	Hierarchical Ordering of Nests in A Joint Mode and Destination Choice Model	MF
North Central Texas of Governments, 2017	Using CTPP Data to Segment Households and Employment	MF
Nyerges and Orrell, 1992	Using Geographic Information Systems for Regional Transportation Planning in a Growth Management Context	PO
Ogura, 2010	Effects of Urban Growth Controls on Intercity Commuting	PO, CJ
O'Regan and Quigley,		BA, DM
Pan and Ma, 2006	Employment Subcenter Identification : A GIS-Based Method	TA
Pan et al., 2014	Effects of Rail Transit on Residential Property Values: Comparison Study on the Rail Transit Lines in Houston, Texas, and Shanghai, China	TP
Pan, 2003	Non-Survey Regional Freight Modeling System	MF
Pan, 2006	Freight Data Assembling and Modeling: Methodologies and Practice	MF
Parsons Brinckerhoff, 2006	Calculating/Analyzing Transit Dependent Populations Using 2000 Census Data and GIS	TP
Paschai et al., 2011	The Use of ACS and Decennial Census Data Products in the Demographic Forecasting Process at NCTCOG	MF
Principal of Schaller Consulting, 2007	Use of CTPP to assess transit access to the Manhattan CBD	TP
Public Policy Institute of California, 2004	Transportation Spending by Low-Income California Households: Lessons for the San Francisco Bay Area	TB, EJ
Rae, 2015	Mapping the American Commute: from Mega-Regions to Mega Commutes	CJ
Rahmani, 2013	Aggregate Relation Between Residence and Workplace Travel Time in Large Urban Areas	CJ
Rashidi and Mohammadian, 2011	Household Travel Attributes Transferability Analysis: Application of A Hierarchical Rule Based Approach	MF
Rashidi et al., 2012	A Behavioral Housing Search Model: Two-Stage Hazard-Based and Multinomial Logit Approach to Choice-Set Formation and Location Selection	MF

Continued on next page.

Author/Year	Title	Subject Area(s)
Regional Transportation Authority, 2006	Northeastern Illinois CTPP Journey to Work Flow Summaries	CJ
Regional Transportation Authority Chicago, Illinois, 2009	Interactive CTPP Analysis Using RTAMS for Northeastern Illinois : A Web-Based Analysis Tool (an Online Journey to Work Data Application)	TB
Rensselaer Polytechnic Institute, 2013	Conduct Urban Agglomeration with the Baton of Transportation: Effects of Jobs-Residence Balance on Commuting Pattern	CJ
Roanoke Valley Transportation, 2017	Vision 2040: Roanoke Valley Transportation	PO
Rothblatt and Colman, 1997	Comparative Study of Statewide Transportation Planning Under ISTEA	SD
Rudin Center for Transportation, 2012	The Emergence of the "Supercommuter"	CJ
Sabre Systems Inc., 2004	Allocation of Missing Place of Work Data in Decennial Censuses and CTPP 2000	SD
Sailor and Lu, 2004	A Top-Down Methodology for Developing Diurnal and Seasonal Anthropogenic Heating Profiles for Urban Areas	HS
Saint Mary's University of Minnesota, 2009	Geographic Information Systems and the Economic Structure of the Seven Rivers Region	TA
Salem-Keizer MPO, 2000	Use of CTPP for Transportation Planning and Modeling in the Salem-Keizer (Oregon) MPO	MF
San Diego Associations of Governments, 2005	Getting Around Rounding and Suppression Issues with CTPP	SD
Sandoval et al., 2011	The Transition from Welfare-to-Work: How Cars and Human Capital Facilitate Employment for Welfare Recipients	BA
Sang et al., 2011	Examining Commuting Patterns: Results from a Journey-to-Work Model Disaggregated by Gender and Occupation	CJ, DM
Sang, 2008	Examining Commuting Patterns and Spatial Mismatch By Occupation and Gender: Disaggregate Journey-to-Work Model	CJ, DM
Santa Barbara County Association of Governments, 2014	Santa Barbara County State of the Commute	CJ
Seattle Office of Housing, 2007	Gaining Clues to Seattle's Workforce Housing Needs	TA
Sen et al., 1995	Household Travel Survey Nonresponse Estimates : The Chicago Experience	SD
Serulle and Cirillo, 2016	Transportation Needs of Low Income Population: A Policy Analysis for the Washington D.C. Metropolitan Region	EJ, PO, TP
Severen, 2017	Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification	CJ
Sherman-denison Metropolitan Planning Organization, 2012	Sherman-Denison Metropolitan Planning Organization annual report	PO

Continued on next page.

Author/Year	Title	Subject Area(s)
Sivanandan et al., 2007	Method to Enhance Performance of Synthetic Origin–Destination Trip Table Estimation Models	SD
Smart, 2014	A Nationwide Look At the Immigrant Neighborhood Effect on Travel Mode Choice	TB
SEMCOG, 2014	Using CTPP Data to Visualize Commuting Patterns in Southeast Michigan	CJ
SCAG, 2015	Visualization of Origin–Destination Commuter Flow Using LEHD Origin–Destination Employment Statistics (LODES) Data	CJ
SCAG, 2016	Spatial and Socioeconomic Analysis of Commuting Patterns in Southern California: Using LEHD Origin–Destination Employment Statistics, Census Transportation Planning Products, and ACS Public Use Microdata Sample	CJ, DM, EJ
State of Maryland, 2013	The Maryland Statewide Transportation Model	MF
Sultana and Weber, 2007	Journey-to-Work Patterns in the Age of Sprawl: Evidence from Two Midsize Southern Metropolitan Areas	CJ, BA
Sultana, 2002	Job/Housing Imbalance and Commuting Time in the Atlanta Metropolitan Area: Exploration of Causes of Longer Commuting Time	CJ
Sultana, 2005	Racial Variations in Males' Commuting Times in Atlanta: What Does the Evidence Suggest?	CJ, DM
Sultana, 2005	Effects of Married-Couple Dual-Earner Households on Metropolitan Commuting: Evidence from the Atlanta Metropolitan Area	CJ
Sweet, 2013	Traffic Congestion's Economic Impacts: Evidence from US Metropolitan Regions	PO
Tal and Handy, 2010	Travel Behavior of Immigrants: an Analysis of the 2001 National Household Transportation Survey	TB, MF, EJ
TTI, 2015	Austin State Agency Congestion Footprint	PO, CJ
Thaithatkul et al., 2015	A Passengers Matching Problem in Ridesharing Systems By Considering User Preference	PO
The Association of American Geographers, 2007	GIS integration of daily commuting movement and population density surface model	MF
The Champaign County, 2015	The Champaign County Travel Demand Model	MF
The Florida DOT, 2016	Guidebook for Florida Stops Applications	MF
The University of Tennessee Center for Transportation Research, 2008	Minimum Travel Demand Model Calibration and Validation Guidelines for State of Tennessee	MF
Transport Foundry, 2015	Using Passive Data to Build an Agile Tour-Based Model - A Case Study in Asheville	MF

Continued on next page.

Author/Year	Title	Subject Area(s)
TRB and the Division on Engineering and Physical Sciences, 2008	Metropolitan Spatial Trends in Employment and Housing	TA
TRB, 2006	Commuting in America 2006	CJ, TA
TRB, 2011	Research for the AASHTO Standing Committee on Planning. Task 111. U.S. Commuting and Travel Patterns: Data Development and Analysis	CJ, TB
TranSystems Corporation, 2006	The Use of CTPP Data for Commuter Rail Demand Analysis in Danbury Connecticut	TP
Tri-County Regional Planning Commission et al., 1997	Socioeconomic Forecasting Model for the Tri-County Regional Planning Commission	MF
U.S. Census Bureau: American Community Survey, 2011	Commuting in the United States: 2009	CJ, TA
University of California Transportation Center, 1995	Job Accessibility as a Performance Indicator: An Analysis of Trends and Their Social Policy Implications in the San Francisco Bay Area	BA
University of Kentucky, Lexington, 2014	Intercounty Commuter Public Transit Services and Opportunities in the Central Bluegrass	TP
University of North Carolina at Chapel Hill, 2014	Using CTPP Data to Improve the Wichita Area Trip Distribution Model	MF
University of South Florida, 2007	Development of Alternative Measures of Transit Mode Share	TP
University of Southern California, 2006	The U.S. Context for Highway Congestion Pricing	PO
University of Texas at Austin, 2014	Understanding Transit Ridership Demand for the Multidestination, Multimodal Transit Network in Atlanta, Georgia: Lessons for Increasing Rail Transit Choice Ridership while Maintaining Transit Dependent Bus Ridership	TP, TB
University Transportation Center for Alabama, 2005	The Impact of Sprawl on Commuting in Alabama	PO, CJ
Upchurch and Kuby, 2014	Evaluating light rail sketch planning: Actual versus predicted station boardings in Phoenix	TP, MF
Urban Transportation Center, UIC, 2013	Analysis of Travel Behavior Using the ACS	TB, DM
Urbanomics, 2005	Trip Making, Induced Travel Demand, and Accessibility	MF
UrbanTrans Consultants Parsons Brinckerhoff, 2005	Portland Metro Rideshare Market Research and Implementation Plan	TP

Continued on next page.

Author/Year	Title	Subject Area(s)
VanLandegen Chen, Xuwei, 2012	Micro-Simulation of Large Scale Evacuations Utilizing Metrorail Transit	TP
Walker et al., 1997	Updating Existing Travel Simulation Models With Small-Sample Survey Data Using Parameter Scaling Methods	MF
Wall, 2001	Use of 1990 CTPP and NCHRP 365 Report to Build a Travel Demand Model for Las Cruces, New Mexico	MF
Wang and Monor, 2003	Where The Jobs Are: Employment Access and Crime Patterns in Cleveland	HS, BA
Wang and Wang, 2013	Modeling Population Settlement Patterns Using A Density Function Approach: New Orleans Before and After Hurricane Katrina	MF
Wang et al., 2011	Street Centrality and Land Use Intensity in Baton Rouge, Louisiana	BA
Wang et al., 2012	Incremental Integration of Land Use and Activity-Based Travel Modeling	MF
Wang et al., 2013	Incremental Integration of Land Use and Activity-based Travel Modeling: Using CTPP2000 for Model Validation and Calibration	MF
Wang, 2000	Modeling Commuting Patterns in Chicago in a GIS Environment: A Job Accessibility Perspective	CJ
Wang, 2000	Modeling Commuting Patterns in Chicago in a GIS Environment: A Job Accessibility Perspective	CJ, BA
Wang, 2001	Explaining Intraurban Variations of Commuting By Job Proximity and Workers' Characteristics	CJ, DM
Wang, 2003	Job Proximity and Accessibility for Workers of Various Wage Groups	TB, DM
Wang, 2005	Job Access and Homicide Patterns in Chicago: an Analysis At Multiple Geographic Levels Based on Scale-Space Theory	BA
Wang, 2005	Job Access and Homicide Patterns in Chicago: an Analysis At Multiple Geographic Levels Based on Scale-Space Theory	BA, HS
Wang, 2011	Job Density and Employment Subcenters in the Four U.S. Metropolitan Areas	TA
Wang, 2012	Modeling Population Patterns in New Orleans 2000–2010 : A Density Function Approach	TA
Weber and Sultana, 2008	Employment Sprawl, Race, and the Journey to Work in Birmingham, Alabama	CJ, BA
Weigel, 2012	Development of a Commercial Building/Site Evaluation Framework for Minimizing Energy Consumption and Greenhouse Gas Emissions of Transportation and Building Systems	HS
Weinberger, 2012	Death by a Thousand Curb-Cuts: Evidence on the Effect of Minimum Parking Requirements on the Choice to Drive	TB, PO
Welch et al., 2005	The Effects of Ozone Action Day Public Advisories on Train Ridership in Chicago	PO, TB, HS

Continued on next page.

Author/Year	Title	Subject Area(s)
Widener and Horner, 2011	A Hierarchical Approach to Modeling Hurricane Disaster Relief Goods Distribution	PO
Wilbur Smith Associates and Kentucky Transportation Cabinet, 2005	Using Census Data to Develop A New Kentucky Statewide Traffic Model	MF
Wilbur Smith Associates, 2007	Using CTPP 2000 Data for the Trans Texas 35 Corridor Model	MF
Wiosna, 2015	Changing Bike Mode Share Between Time Periods for Suffolk County, MA	BP
Woo and Guldmann, 2011	Impacts of Urban Containment Policies on the Spatial Structure of US Metropolitan Areas	PO
Woo and Guldmann, 2014	Urban Containment Policies and Urban Growth	PO
Woo et al., 2014	Impacts of the Low-Income Housing Tax Credit Program on Neighborhood Housing Turnover	PO, EJ
Xiao, 2015	Spatial Representation in the Social Interaction Potential Metric: an Analysis of Scale and Parameter Sensitivity	BA
Yang and Ferreira, 2008	Choices Versus Choice Sets: A Commuting Spectrum Method for Representing Job– Housing Possibilities	CJ
Yang and Ferreira, 2009	Informing the Public of Transportation–Land Use Connections	PO
Yang and Salling, 2002	Integrating GIS and DMBS to Deliver Computation Support on Job Accessibility	BA
Yang, 2005	Commuting Impacts of Spatial Decentralization: A Comparison of Atlanta and Boston	CJ
Yang, 2005	The Spatial and Temporal Dynamics of Commuting: Examining the Impacts of Urban Growth Patterns, 1980-2000	CJ, TA
Yang, 2008	Policy Implications of Excess Commuting: Examining the Impacts of Changes in US Metropolitan Spatial Structure	CJ, TA
Yao, 2007	Where Are Public Transit Needed—Examining Potential Demand for Public Transit for Commuting Trips	TP, MF
Yoon et al., 2012	Feasibility of Using Time–Space PRISM to Represent Available Opportunities and Choice Sets for Destination Choice Models in the Context of Dynamic Urban Environments	MF
Zhan and Chen, 2008	Intercity Commute Patterns in Central Texas	CJ
Zhang, 2008	Metropolitan Dynamics of Accessibility, Diversity, and Locations of Population and Activities	TA, BA
Zhang, 2015	Impacts of Enterprise Zone Policy on Industry Growth: New Evidence from the Louisville Program	PO

NOTE: Abbreviations for Subject Area Categories are as follows: BP = Bicycle and Pedestrian Studies; BA = Built Environment and Accessibility Study; CJ = Commuting Patterns and Job-Housing Mismatch; DM = Demographics Study; EJ = Environmental Justice and Title VI; HS = Health, Safety and Environmental Issues; PO = Policy Analysis; SD = Survey, Data Synthesis and Research Methods; TP = Transit Planning; TB = Travel Behavior Analysis; MF = Travel Demand Modeling and Forecasting; and TA = Trend Analysis and Market Research.

References

1. U.S. Census Bureau. Who We Are. https://www.Census.gov/about/who.html. Accessed Aug. 4, 2017.
2. Weinberger, P. Z. National Data Sets, How to Choose Them, How to Use Them. *European Transport Conference*, 2016.
3. Graham, M. R., M. J. Kutzbach, and B. McKenzie. Design Comparison of LODES and ACS Commuting Data Products. *Center for Economic Studies*, Vol. CES 14-38, 2014. https://doi.org /10.1017/CBO978 1107415324.004.
4. The Federal Highway Administration. American Community Survey. https://www.fhwa.dot.gov /Planning/Census_issues/american_community_survey/. Accessed Aug. 4, 2017.
5. U.S. Census Bureau. Master Address File Description. https://www.Census.gov/did/www /snacc/publications/MAF-Description.pdf. Accessed Aug. 4, 2017.
6. Census. Decennial Census (2010, 2000, 1990). *U.S. Census Bureau*. https://www.Census.gov /data/developers/data-sets/decennial-Census.2010.html. Accessed Nov. 9, 2017.
7. Long, L. Five-Year CTPP Data Product, 2013.
8. Census Data for Transportation Planning TRB Subcommittee on Census Data for Transportation Planning. http://www.trbCensus.com/. Accessed April 8, 2017.
9. U.S. Census Bureau. LEHD Origin–Destination Employment Statistics Dataset Structure Format Version 7.2.
10. U.S. Department of Transportation. National Household Travel Survey: Understanding How People Get from Place to Place. http://nhts.ornl.gov/index.shtml. Accessed May 8, 2017.
11. The Federal Highway Administration. 2009 NHTS User's Guide (Version 2), October, 2011.
12. Kermanshah, A., and S. Derrible. A Geographical and Multi-Criteria Vulnerability Assessment of Transportation Networks against Extreme Earthquakes. *Reliability Engineering & System Safety*, Vol. 153, 2016, pp. 39–49. https://doi.org/10.1016/j.ress.2016.04.007.
13. Hu, Y., F. Wang, and C. G. Wilmot. Commuting Variability by Wage Groups in Baton Rouge, 1990– 2010. *Papers in Applied Geography*, Vol. 3, No. 1, 2017, pp. 14–29. https://doi.org/10.1080 /23754931 .2016.1248577.
14. Spear, B. D. *Improving Employment Data for Transportation Planning*. NCHRP 08-36, Task 098, 2011.
15. Mix, W. A. Evaluating the Local Employment Dynamic Program as an Alternate Source of Place of Work Data for Use by Transportation Planners, pp. 1–28.
16. Long, L., and J. Lin. An Investigation in Household Mode Choice Variability Across Metropolitan Statistical Areas for Urban Young Professionals, 2007.
17. Pendyala, R. M., and A. Agarwal. Travel Characteristics on Weekends : Implications for Planning and Policy Making, 2004.
18. Chu, X. Census/ACS/CTPP Data for Transit Planning, 2012.
19. Murakami, E., C. Baber, and E. Christopher. *Workplace Geocoding Issues: CTPP Flow Data Place of Work Is Critical*, 2014.
20. Stopher, P. R., and S. P. Greaves. Household Travel Surveys: Where Are We Going? *Transportation Research Part A, Policy and Practice*, Vol. 41, No. 5, 2007, pp. 367–381. https://doi.org/10.1016 /j.tra.2006.09.005.
21. Grengs, J. Nonwork Accessibility as a Social Equity Indicator. *International Journal of Sustainable Transportation*, Vol. 9, No. 1, 2015, pp. 1–14. https://doi.org/10.1080/15568318.2012.719582.
22. Santos, A., N. McGuckin, H. Y. Nakamoto, D. Gray, and S. Liss. *Summary of Travel Trends: 2009 National Household Travel Survey*, 2011.
23. Ayvalik, C., P. Weinberger, and K. Tierney. *Assessing the Utility of 2006–2010 CTPP Data*, 2017.
24. Seo, J., T. Vo, F. Wen, and S. Choi. Spatial and Socioeconomic Analysis of Commuting Patterns in Southern California: Using LEHD Origin-Destination Employment Statistics. *Census Transportation Planning Products and ACS Public Use Microdata Sample*, 2017.

25. Blanton, W. Small-Area Applications Using 1990 Census Transportation Planning Package, Gainesville, Florida, 1996.

26. Luce, T., M. Orfield, and J. Mazullo. Access to Growing Job Centers in the Twin Cities Metropolitan Area. Center on Urban and Regional Affairs, University of Minnesota, Minneapolis, 2006.

Facilitated Discussion

Other than the traditional datasets, what other workplace data sources are commonly used, and for what applications? To what extent have data sources from private vendors supplemented large-scale survey data in your organization? What advantage does survey data maintain?

One member of the audience noted that SCAG supplements with local data, and other members of the audience mentioned purchasing InfoUSA for additional workplace data. However, whenever agencies rely on purchased private-sector data as their source of data, questions remain regarding with private data, specifically, how do you know what you are getting? It is likely that the private sector is benchmarking their products with Census data.

New data sets being produced by the private-sector vendors have the similar issues. Several members of the audience expressed their concern. For example, there is great uncertainty regarding Sidewalk Labs type products using sensor real-time data. It is unclear what these data really mean. SCAG also uses a local panel of experts to review their estimates.

When are CTPP workplace data not an option, and what data source is used instead? For what applications are CTPP data the only appropriate workplace data available? What policy questions are you answering with CTPP data?

In the discussion, audience members recalled the comparisons between LEHD to CTTP workplace data, noting that LEHD has 95% coverage. Even so, while LEHD covers 95%, it does not cover school districts, and in addition, there are significant employers in geographies with residential land uses. It is important to point out that surveying workers is different from surveying jobs. One observer noted that Bureau of Economic Analysis (BEA) employment totals are significantly higher. There are questions regarding employment numbers because of the Uber effect, seasonal workers, second jobs, etc.

It was noted that Census used to have a category for "no fixed" place of work. One observer noted that ACS had a continuous team doing data quality checks and that this might mean a reduction in error, yet changes in ACS funding may have reduced this data quality effort.

The question of whether or not we really knew the real MOE was raised. It was also noted that we do not use the ACS hours worked variable.

What level of geography is most commonly used? Why? Beyond basic counts, what variables are most used in flow tables? Among data users, which sectors are not represented, but could be?

In the discussion of geography, it was mentioned that some transportation professionals use TAZ and Census tracts even with MOE issues. Participants in the session wanted the data, even with high coefficients of variation. While the paper made a point of talking about how LODES goes down to the block level—yet the block level data in LODES is not useful—too much variation much of it inserted to maintain confidentiality.

Possible users not represented at the conference, or this session, include: Federal Reserve analysts; real estate industry; political consultants; and marketing analysts. One audience member recalled that emergency preparedness personnel previously relied upon CTPP data.

Given CTPP's variation in data quality across tables and geographies, would prerelease data quality parameters be welcomed?

After a lengthy discussion on aspects of place of work data, the conversation turned back to issues around TAZ and Census block groups. It was suggested that the conference attendees be involved in discussions of Census block group boundaries to enhance aspects of work place data.

Audience Suggestions for the CTPP Oversight Board

- Continue to expand opportunities to improve workplace data, including the use of private-sector vendor data. However, consideration needs to be made on how best to provide transparency and data quality to the extent possible for future data users.
- Consider strategies to renew connections with private-sector data users (e.g., marketing firms, real estate industry representatives) as they bring new ideas and approaches that further the usefulness of Census and related data products.
- Include the concerns of workplace data into the conversations with the construction of TAZ and Census blocks.

CHAPTER 17

The Future of Data for Transportation Planning

KRISHNAN VISWANATHAN
Cambridge Systematics, presiding

STACEY BRICKA
MacroSys Research and Technology

ANURAG KOMANDURI
Cambridge Systematics

BHARGAVA SANA
San Francisco County Transportation Authority

NANDA SRINIVASAN
Energy Information Administration

This session focused on the future of transportation planning data, how Census data relate to it and the experience of agencies and universities working in a future data context.

A PUBLIC AGENCY'S PERSPECTIVE
Bhargava Sana

San Francisco County Transportation Authority has recently made efforts to incorporate new forms of data into their data analysis. They are using INRIX speed–travel time data for congestion monitoring; Lyft or Uber vehicle GPS data (with Northeastern University) to estimate the number of trips; cellphone and GPS O-D data (from Google) to estimate freeway facility-specific O-D matrices. However, these sources still require conventional data such as travel diary surveys, Travel Demand Management data, and onboard surveys. The new sources of data have no demographic information, no trip details (e.g., purpose, occupancy), and need advanced data processing techniques to be useful at all.

In the near-term, conventional data sources will continue to be used with testing–validating new Big Data sources to determine usefulness. In the medium term, there are plans to use Big Data sources to augment our understandings while continuing to field travel diary and special purpose surveys. In the longer-term though, the potential to identify and incorporate emerging mobility options in models and support transparent and nonproprietary data sources will be explored.

A CONSULTANT'S PERSPECTIVE

Anurag Komanduri

It is a good time to be in transportation as transportation data with so many new forms of sensor data. These could be used for equity analysis to look at access for everyone, something the private sector may place less emphasis on than the public sector. There are concerns about how new forms of data are shared and used. More understanding of how data is being generated and need it to be transparent to use with the public needs to be a focus. There are pros and cons of these new forms of data that need to be taken into consideration.

These data can be classified into three different types. Vendor data is a product (e.g., cell phone data) with little transparency although INRIX does provide speed data, but without clarification on the underlying data. Semi-open sources can be accessed using a computing technique referred to as "scraping data" from an Internet site. For example, information on rental activities can be scraped from sites with rental offerings (e.g., Airbnb). Large amounts of data can be obtained using this technique on public-facing sites. Another approach is for agencies to establish a data-sharing agreement with private-sector vendors, making it possible for the agency to have access to data. LA Metro has been able to access on-demand responsive service data from a private-sector vendor. The private sector will build the service and the public sector will provide the funding and ask the private sector to share the data generated. Open data is now available on parking and bike activities without the need of vendor sales. The question is how to use these free data sources in more flexible platforms. There could be new models that can use these new forms of data.

Going forward, many new sensor data sources will be producing data in the connected vehicles and autonomous vehicles, but there are questions regarding who will have access to these vast data sources. Processing these data will require new skills. It is a great time to be involved in any data as these resources grow.

CONSULTANT–FHWA'S PERSPECTIVE

Stacey Bricka and Wenjing Pu

There is a great deal of interest in Big Data and there are concerns about it as a black box. It appears to give a rich volume of O-D flows and algorithms can provide trip purpose, but there are concerns. Travel surveys have been underway since the 1950s and have become simpler, but costs are increasing and funding is declining. They are a random probability small sample. Big Data provides the "what's going on" while the survey provides "who's doing it, what and why."

The Census is the bridge between the two sources. Big Data has been providing destination and origin data for agencies, with the home location inferred and attached to Census data. Travel surveys are used for weights and expansion and simulations. There are many caveats with a general trend to use Big Data, made stronger because travel surveys are at the core. Big Data makes it possible to get the big picture, while surveys can be used to validate and calibrate models. The three data sources provide the volume, with the "who and why" for policy. Details need to be unpacked and sorted out, with projects exploring options. There are questions we do not even know to ask. The conversation has the momentum to better understand our options.

FHWA has several exploratory projects underway and state DOTs and MPOs need to share the challenges they are facing. The first contract is for a new mobile data project with University of Maryland, with subcontractors AirSage and INRIX/Streetlight. It is to be a national O-D—county-

to-county flows fused with NHTS and Census to build a prototype dataset for mobility. Another Exploratory Advanced Research is the long-distance survey report that looks at the development of an instrument. There is also NEXTGen NHTS project. FHWA is transforming Big Data into smart data with the promise of studies to see how this data can answer questions. The answer lies in working together with an open discussion to foster solutions. There will be a pooled-fund study as a framework for data products that will use Census data.

A DATA EXPERT'S PERSPECTIVE
Nanda Srinivasan

During the 1960s, 1970s, and 1980s, there was an unfulfilled vision of CTPP being used integrate data for planning. This vision still exists, with CTPP data being integrated with other sources of mobility data. Four major themes that are driving data today include usefulness of Big Data; Google API and INRIX being used for known purposes; timeliness and repeatability of Big Data; and sufficiency of observations for hypothesis testing. People seem to see Big Data as the new shiny thing, a free data set that could solve problems for transportation. What about data quality? While new forms of mobility data create huge data sets, quantity does not equal quality. There can be processing errors and private companies might not even know that these errors are occurring. In traditional surveying, errors can occur (e.g., respondent errors, coding errors), and are successfully dealt with using statistical techniques and practices.

Big Data can be validated with NHTS, given knowledge of the process and how the survey data was collected. The documentation is available and can commented on it through TRB opportunities. On the other hand, when data is purchased (e.g., buying INRIX as proprietary data), the data is "as is." This issue is not new as transportation planners and researchers have previously purchased Dunn and Bradstreet and InfoUSA and have gained experience with purchased data. Without transparency, it is difficult to know if, or how, data may have been manipulated. When data is to be used for public good, and it must have governance so the public can trust it as independent source.

Data accessibility is a key concern. Even though the Google searches for data are "free", costs will be incurred in the process, along with potential legal issues. In addition, at any time, private-sector data providers can increase the cost of a primary data source. There are the issues of how to cite or quote private data. There are issues with repeatability, auditability, and archiving strategies for new data sources. Private sources can be different across MPOs, resulting in impacts on transferability. One of the best NHTS projects was the transferability project that made it possible for every Census tract to have new variables.

There are questions regarding methodologies for Big Data. Two national conferences have been held that addressed the change from the long form to ACS, with sessions on the role of government, the role of the private sector, and the role of academia. The role of the federal government was made even more apparent when the American Transportation Research Institute data was used for performance measure. To strengthen data resources, institutional linkages are necessary with the federal government to build sustainable and strong connections with local and state data sources for local use. Academia and consultants have a role here with federal government, and will require funding for their efforts. Data vendors' data analysts need to participate (e.g., Uber, Lyft, AirSage) as well. Research is needed on new technologies and how to integrate with the CTPP. It would be helpful to have an anthology of existing and emerging examples. In 1990, a book of Census uses was produced. Now, a book of visualizations and

projects using Big Data could be produced (e.g., an NCFRP Synthesis). All of these efforts will require collaboration. A pooled-fund strategy, over the next 5 years, may be the right mechanism to explore best practices and lessons learned with respect to combining different sources of data and accomplishing the vision of integrated data.

CHAPTER 18

Comparing Census Data Sets

MARA KAMINOWITZ
Baltimore Metropolitan Council, presiding

FRANCISCO TORRES
ARASH MIRZAEI
LIANG ZHOU
North Central Texas Council of Governments

SAM GRANATO
Ohio Department of Transportation

T he experienced data user knows there is no perfect data set. Each comes with nuances, methodological concerns, and issues. The savvy analyst knows to check any single source with other relevant sources before drawing any major conclusions.

COMPARISON OF TRAVEL TIME DISTRIBUTIONS FROM ACS 2015 AND NPMRDS
Francisco Torres

The ACS 2015 5-Year Estimates provide travel times distributions by 5-min intervals ranging from 0 to 90 min. These distributions are available at the county level and correspond to commuter trips. The NPMRDS provides very accurate travel time data that can be transferred to a travel model network, which when applied to a matrix of HBW trips can be used to estimate travel time distributions.

The purpose of this research is to compare the travel time distributions obtained from the NPMRDS with those of ACS for the same year at the county level for each one of the 12 counties that are part of the DFW metropolitan area. Since July 2013, NCTCOG, the MPO of the DFW metropolitan area, has been using and storing the travel time data made available by FHWA as part of the NPMRDS. The database from this source for the North Central Texas area contains millions records of travel times at 5-min intervals for the period between July 2013 and January 2017. These records are associated to a roadway network of 9,300 segments, called traffic message centers (TMCs), which cover a length of 9,126 mi; where 70% of the regional VMT are generated.

For each TMC, all the travel times for 2015 were aggregated and averaged at 30-min intervals for all typical weekdays (Tuesdays, Wednesdays, and Thursdays) in the whole year. The aggregated travel times for each TMC were transferred to the roadway network used in the calibration on the travel demand model. Since NPMRDS covers only the roads that are part of the NHS, all freeways and main arterials, the travel times on the remaining links of the roadway network were taken from the estimations of the travel model. With this travel times, an impedance matrix was calculated from all travel survey zones (TSZ) to all TSZs. As part of the development of the regional travel model a matrix of HBW trips has been calculated. The travel

time distributions are then estimated having this matrix and the matrix of NPMRDS travel times as basic inputs.

COMPARING THE USE OF CTPP AND LEHD TO CREATE AN EMPLOYMENT DISTRIBUTION IN THE NORTH CENTRAL TEXAS REGIONAL TRAVEL MODEL

Arash Mirzaei and Liang Zhou

During the development of a new model, NCTCOG needed to develop a distribution of employment by industry at the TAZ level. In development of the employment breakdown, employment by industry control totals for the calibration year of 2014 would be needed. In this research, two methods of creating this distribution of employment are tested and compared.

The first method of creating this employment distribution uses 2013 BEA data and 2010 CTPP data. For the first step, the CTPP employment by industry table is aggregated into industry categories of basic, retail, and service for each TAZ. Next, The CTPP TAZ data is processed to calculate each TAZ's share of the county employment of each industry category. Then, BEA county employment by industry data is aggregated into three industry categories: basic, retail, and service. Finally, the TAZ employment for each industry category is determined by multiplying the TAZ's share of the county employment by industry category by the BEA total county employment by industry category.

The second method of creating this employment distribution uses 2013 BEA data and 2010 LEHD modified by local knowledge. This method is similar to the first method, but uses LEHD instead of CTPP. To begin, the modified LEHD TAZ data is processed to calculate each TAZ's share of the county employment of each industry category. Then, BEA county employment by industry data is aggregated into industry categories of basic, retail, and service. Lastly, by multiplying the TAZ's share of the county employment by industry category determined from LEHD by the BEA total county employment by industry category, the TAZ employment for each industry category is found. A comparison of the employment distributions using these two methods will be presented. In addition, the results of the travel demand model for year 2014 using each of the two employment distribution methods will be compared against the ground counts and transit ridership. These comparisons will provide an indirect validation of the methods described.

COMPARING CTPP AND LEHD ON JOURNEY-TO-WORK TRIP LENGTH DISTRIBUTIONS STATEWIDE

Sam Granato

For several decades, both the PUMS and CTPP data sets have provided information on WTT, disaggregated geographically to the PUMA or traffic zone level. This has allowed for its usage in travel demand models that can depict differences by area (or other characteristics) in the JTW travel pattern.

More recently, a partnership between the Census Bureau and state Labor Market–UI agencies has enabled the creation of the LEHD which provides estimates (with some disclosure-proofing) of travel between home and work at the Census block level. Numerous differences in the data between these two sources have been previously noted by others that can lead to differences in conclusions reached. These include differences in the workforce being measured

(all wage and salary jobs versus "primary" jobs including proprietorships), self-identified workers via survey versus administrative records, and differences in breakdown by age, income and industry—along with differences between modeled versus reported travel times or distances.

Statewide—within Ohio at least—it has also been found that the LEHD data reports far more long-distance travel to work than can plausibly be explained by differences in source (beyond such things as being temporarily away from home, like college students, or recent moves rendering one's most recently used tax address different from the address that would have been reported on the Census ACS form). For example, 13% of the LEHD home-to-work trips exceed an estimate of 70 min, while less than 2% of such trips from the Census survey exceed that time threshold. Census data (predominantly the CTPP) is compared with LEHD data concerning the pattern of home-to-work travel in terms of modeled travel time. This includes not only the pattern statewide in Ohio, but also a fairly large university within an isolated small city (to determine how much of the differences seen statewide could be reduced to being temporarily away from home) and other locations where large employers in other industries can be identified.

From the work to date, it can be concluded that the large discrepancies in JTW between the two data sources are due to a variety of issues. Workers in certain industries or age groups can be more prone to erroneous reporting than others, but the larger issue could simply be employers either not consistently reporting multiple worksites, not assigning workers accurately to multiple sites, or the nature of the work frustrates easy assignment to a specific workplace. Depending on the application, the transportation planner has a variety of means of adjusting the JTW data from LEHD data to a distribution deemed more "reasonable."

CHAPTER 19

Closing Session

CLARA RESCHOVSKY
Bureau of Transportation Statistics, presiding

CATHERINE T. LAWSON
State University of New York, Albany

PENELOPE WEINBERGER
AASHTO

GUY ROUSSEAU
CTPP Oversight Board Vice Chair

ED CHRISTOPHER
Independent Transportation Planning Consultant

REPORT BACK FROM COMMISSIONED PAPER BREAKOUT DISCUSSIONS
Catherine T. Lawson

The first paper explored the use of CTPP in the realm of Performance Measures. The research objective was to demonstrate the application of CTPP for the purpose of advancing TPM. The three areas chosen were safety, mobility, and accessibility, with a case study set in Chicago using three data sources: CTPP, DOT crash data, and Chicago's Open Data. The research demonstrated that the CTPP could be used with local data to produce satisfactory performance measures for safety, but not for mobility or accessibility, due to missing data and complexity issues. Audience members mentioned the potential of conflating the CTPP with the NPMRDS and pavement condition data; using bike data and demographic; and other "big data" sources.

The second paper focused on TAZs: the TAZs that Census produces and then there are the TAZs you might produce in your own modeling group. Sometimes they are the same and sometimes they are not—that was a big discussion. The Census tries to use their nesting structure so that everything fits into everything from the Census block up to the whole nation. It's a very good strategy. However, not all TAZs like to fit into those blocks. The authors conducted an online survey and found people are using TAZs, even if they're not using them in their own models, they are using them for validation. Respondents who were working in the rural areas indicated they were using block groups as a concept. Some used the PSAP program and then there emerged a sense of consensus. Working together, successfully gathering information from local sources might be the best way forward.

The third paper looked at how CTPP is "mixing with" the emerging data environment of sensor data, and GPS data, and different ways that people want to bring data to the table. The Census has some dominant areas in this research including traffic counts and Census and commuting and demographics. Big Data focuses on new data sets and particular applications. The authors provided seven strategies. For example, giving up and going home. No one thought

this was what was needed (especially the people at this conference). Just keep calm and carry on. Incrementally, journey-to-school could be added. Some small changes would keep data consistent with other external data. Rather than trying to beat private data, agencies could purchase data from the private vendors and make the data available in a public release. Progress has been made with administrative data integration (e.g., LEHD). Questions could be made to be more fluid in nature, and be tied more to what is happening with our mobility options. Long-distance data could be pursued again (previous efforts were not satisfactory). Finally, consideration for using survey methodologies that incorporate longitudinal data collection (e.g., similar to deployments in Germany, the Netherlands, or Seattle).

In the audience discussion, it was mentioned that GPS is available on vessels as AIS. With AIS, linkages can be made between a vessel and the goods onboard. For surface transportation, private-sector data (e.g., INRIX or StreetLight data), can be used to establish trip O-Ds. But these data lack trip purpose. Questions remain on how best to use these data, and how they would be helpful for our understanding of travel behavior. The challenges include response rates, deployment costs, and expectations, particularly with regards to long-distance travel data. An eighth strategy was suggested to "go hybrid." How should all of these various data sources be glued together? Data is needed for such as intraurban freight and intercity passengers—two big gaps. More tables are designed that are produced more frequently. In addition, more needs to be known about the TNCs because they really are impacting our traffic on the margin or maybe in bigger ways.

The final paper focused on the CTPP workplace data. Looking at what is currently available using the CTPP, what are the competing ways to answer those same questions about workplaces? Sources include the LODES, the LEHD, and the NHTS. For the CTPP, 2006 through 2010 covers about 10% of the households that can extrapolated with 115 tables about workplace locations, commute patterns, and person characteristics. LODES, or the LEHD, is produced using administrative data. The NHTS provides national-level statistics from a nationwide random sample; however, for a state that purchased the add-ons, the latitude–longitude locations are available and add value. The CTPP has more socioeconomic variables than the LODES, but again, the LODES is giving you the spatial granularity that you might need. The confidentiality requirements in the small samples that happened with the ACS, of course, are giving us issues with missing data when we look at things too closely, when we want those tiny geographies. It was noted that Federal Reserve analysts, real estate industry members, political consultants, and marketing analysts were missing from the conversation. Emergency preparedness people used to be in our meetings. Close guidance going into the future could help capture what works and what doesn't work—a playbook of what to do and how to do it correctly.

THE FUTURE OF CTPP
Penelope Weinberger

In 2012, AASHTO established a Technical Service Program, funded by participating states to focus on the CTPP. The upcoming release of the 2012–2016 CTPP will have the same structure as the 2006–2010 data set, with Parts 1, 2, and 3—residence, workplace, and JTW data. There are no changes to the TAZ geographies in this release. The data itself will be coming in 2018, a including fewer tables (e.g., tables that had a really high failure rate for cell confidence, not being accessed by the public, or not considered valuable by the subcommittee reviewing the

tables). We held electronic town hall meetings through the CTPP listserv, and Travel Model Improvement Program and other listservs.

In the end, tables were reduced the current 343 to 176 to meet the request from the Census Bureau. Twenty-two tables will be created in the software. Seventy-six will be available for large areas only and 99 of them will be all areas. Collapsed tables will be generated in the upgraded software and all of the univariate tables that can be derived from other tables. Place-to-county flows, HH-size tables, poverty by mode, the workplace by class of worker by industry, time arrival, mean travel time by mode, and vehicles available by number of workers will be available.

CONFERENCE CLOSING REMARKS:
APPLYING CENSUS DATA FOR TRANSPORTATION
Guy Rousseau

At the conference, our "founding fathers" have shared their wisdom on the CTPP evolution. In the first commissioned paper, it was demonstrated that performance measures can be produced using CTPP data. In the second paper on TAZs, Census TAZs versus model TAZs were examined, including impacts on modeling analysis. The third paper explored the emergence of Big Data and how to improve and enhance our tools and our data sets looking at Big Data using data fusion and hybrid approaches to better leverage Big Data sources. The last paper focused on making better use of CTPP data to validate work location in our models and different types of analysis. The posters displayed a variety of applications with the CTPP data.

Notable sessions include the CTPP 101 session, a refresher is very helpful as a reminder of the tables, graphics, and data analytics. The Census Bureau Potpourri sessions described activities by the Census Bureau, LEHD, and other data sets. CTPP data can be used for equity analysis, aging, accessibility, and the importance of the CTPP for policy analysis. In the Advanced CTPP data analysis session, MOEs were explored along with what can be done with them, and how to account for them when in data analysis. Details were provided on PUMS data, PUMAs, and IPUMS, the public use of the microsamples and how to make better use of our analyses. The role of CTPP data in transportation modeling included DTA models and the FTA STOPS model. Further, Census data can be used for alternative modes including: ferries, electric vehicles, and all these different modes of transportation. NHTS is being used with CTPP data, with evidence of cross-pollination between these data sets. This combination of data is the foundation for CIA that brings all these data sets into this very useful report.

With respect to reflections on the outcome of this conference and next steps for the Oversight Board, rebranding needs to be considered (e.g., strategic marketing–visioning) and perhaps a Big Data purchase similar to what the FHWA's purchase of the NPMRDS. With all these performance measures requirements, State DOTs and MPOs really needed that data, a unified data set that is very helpful. Perhaps, the next CTPP might do something similar to that (or it may not be). There are many friends of the CTPP program, including FTA and their use of CTPP in their STOPS model to make big decisions as far as transit investments where to invest by regions of the country.

Coordination with the Census Bureau, particularly the issue of data relevance versus the privacy rights, a very delicate equilibrium is critical. We want more data, we want more relevant,

accurate data. Yet, at the same time, there is this pressure for privacy and disclosure reviews. The future of CTPP will require funding, perhaps a public and private partnership.

FINAL REMARKS
Ed Christopher

The Census Bureau offers a Fact of the Day that demonstrates how we can extend the value of data through fusing data sets together. It is an example of the future of our data sets. The CTPP Oversight Board has a big job ahead to deal with the topics addressed at this conference (e.g., TAZs and purchasing private sector data supplements). There is a concern with having so many topics on the table and no easy answers for any of them. In particular, there are issues surrounding enhancing the quality of Big Data because users need to look at the data, learn about its warts, and understand the data, before using it. Attention needs to be paid to quality, not just the quantity of our data resources.

CHAPTER 20

Conference Participants

Matt Airola
Westat

Shoshana Akins
Delaware Valley Regional
    Planning Commission

Sulabh Aryal
Richmond Regional Planning
    District Commission

Cemal Ayvalik
Cambridge Systematics

Charles Baber
Baltimore Metropolitan Council

Chris Bonyun
Beyond 2020

Dany Bouchard
CartoVista

Nick Bowden
Sidewalk Labs

Stacey Bricka
MacroSys Research

Megan Brock
Steer Davies Gleave

Clarlynn Burd
U.S. Census Bureau

Ally Burleson-Gibson
U.S. Census Bureau

Paul Bushore
Mid-America Regional Council

Braden Cale
Oklahoma Department of Transportation

Laura Chaney
Oklahoma Department of Transportation

Yohan Chang
University of Missouri

Jilan Chen
Southeast Michigan Council of Governments

Yu-Jen Chen
Tennessee Department of Transportation

Ed Christopher
Independent Consultant

Hayley Collins
Cambridge Systematics

Josh Coutts
U.S. Census Bureau

Greg Erhardt
University of Kentucky

Ben Ettelman
Texas Transportation Institute

Tom Faella
La Crosse Area Planning Committee

Dave Faucett
Ozarks Transportation
    Metropolitan Planning Organization

Li Feng
Southeast Michigan Council of Governments

Alison Fields
U.S. Census Bureau

Mark Folden
North Central Texas Council of Governments

Michael Frisch
University of Missouri

Mike Giangrande
Westat

Jacob Gonzalez
Benton-Franklin Council of Governments

Matthew Graham
U.S. Census Bureau

Sam Granato
Ohio Department of Transportation

Jo Anne Gray
Texarkana Metropolitan Planning Organization

Melissa Gross
InNovo Partners

Ben Gruswitz
Delaware Valley Regional
  Planning Commission

Joe Hausman
Federal Highway Administration

Laine Heltebrindle
Pennsylvania Department of Transportation

Sarah Hernandez
University of Arkansas

Thomas Hill
Florida Department of Transportation

Sara Hintze
Mid-America Regional Council

Jim Hubbell
Mid-America Regional Council

Shimon Israel
Metropolitan Transportation Commission

Danny Jenkins
Federal Highway Administration

Ken Joh
Metropolitan Washington
  Council of Governments

Jessie Jones
Arkansas Department of Transportation

Mara Kaminowitz
Baltimore Metropolitan Council

Yue Ke
Purdue University

Jessica Keldermans
Illinois Department of Transportation

Hoheila Khoii
California Department of Transportation

Jaehoon Kim
Tennessee Department of Transportation

Anurag Komanduri
Cambridge Systematics

Kim Korejko
Delaware Valley Regional
  Planning Commission

David Kruse
University of Texas San Antonio

Lisa Lam
Oklahoma Department of Transportation

Tracy Larkin
Nevada Department of Transportation

Phil Laskley
Texas Transportation Institute

Catherine T. Lawson
State University of New York, Albany

Mai Le
Transportation Research Board

Jane Li
Westat

Jon Lupton
Metroplan

Saima Masud
Southeast Michigan Council of Governments

Kim Maxey
New Bern Area Metropolitan Planning
  Organization

Brian McKenzie
Census

Michael Medina
El Paso Metropolitan Planning Organization

Phil Mescher
Iowa Department of Transportation

Jasmy Methipara
Federal Highway Administration

Bruce Millar
Transportation Research Board

Karen Miller
Missouri Department of Transportation

Arash Mirzaei
North Central Texas Council of Governments

Jen Murray
Wisconsin Department of Transportation

Tom Palmerlee
Transportation Research Board

Claudia Paskauskas
InNovo Partners

Jeff Pinkerton
Mid-America Regional Council

Alan Pisarski
Alan Pisarkski Consulting

Steve Polzin
Center for Urban Transportation

Ranjani Prabhakar
Self-employed

Chuck Purvis
Metropolitan Transportation Commission

Andrea Repinsky
Mid-America Regional Council

Clara Reschovsky
Bureau of Transportation Statistics, U.S.
  Department of Transportation

Guy Rousseau
Atlanta Regional Commission

Brandon Rudd
Oklahoma Regional Council of Governments

Phil Salopek
U.S. Census Bureau (retired)

Bhargava Sana
San Francisco Transit Authority

Rolf Schmitt
Research and Innovative Technology
  Administration, U.S. Department of
  Transportation

Jonathan Schroeder
University of Minnesota

Mario Scott
Steer Davies Gleave

Jung Seo
Southern California Association of Governments

Wendy Sheppard
Illinois Department of Transportation

Kim Smith
Greater Buffalo–Niagara Regional
  Transportation Council

Nanda Srinivasan
Department of Energy–
  Energy Information Administration

Deborah Stempowski
U.S. Census Bureau

Ivana Tasic
University of Utah

Shelby Templin
Oklahoma Department of Transportation

Gail Thomas
Central Oklahoma Economic
  Development District

Curt Thye
StreetLight Data

Kevin Tierney
Bird's Hill Research

Francisco Torres
North Central Texas Council of Governments

Catherine Tulley
Maryland Statewide Personnel System

Sharif Ullah
Lochmueller Group Inc.

Shawn Urbach
Maryland Area Regional Commuter

Marketa Vavrova
University of Texas at El Paso

Tori Velkoff
U.S. Census Bureau

Krishnan Viswanathan
Cambridge Systematics

Tom Vo
Southern California Association of Governments

MaryAnn Waldinger
Community Planning Association

Penelope Weinberger
American Association of State Highway and
  Transportation Officials

Bill Woodford
Resource Systems Group, Inc.

Eileen Yang
Maryland Area Regional Commuter

Kathy Yu
North Central Texas Council of Governments

J.J. Zang
Cambridge Systematics

Huimin Zhao
Independent Consultant

Caitlin Zibers
St. Joseph Area Transportation
  Study Organization

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

TRB

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