Using Census Data for Transportation Applications

DETAILS

0 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-43444-7 | DOI 10.17226/22600

AUTHORS

Turnbull, Katherine F.

CONFERENCE PROCEEDINGS ON THE WEB 10

Using Census Data for Transportation Applications

Summary of a Conference

Katherine F. Turnbull, *Rapporteur*
Texas A&M Transportation Institute
Texas A&M University System

October 25–27, 2011
Arnold and Mabel Beckman Center of the National Academies
Irvine, California

Sponsored by
Transportation Research Board
American Association of State Highway and Transportation Officials
Federal Highway Administration
Federal Transit Administration

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Washington, D.C.
2012
www.TRB.org

Transportation Research Board Conference Proceedings on the Web 10

Transportation Research Board publications are available by ordering individual publications directly from the TRB Business Office, through the Internet at www.TRB.org or national-academies.org/trb, or by annual subscription through organizational or individual affiliation with TRB. Affiliates and library subscribers are eligible for substantial discounts. For further information, contact the Transportation Research Board Business Office, 500 Fifth Street, NW, Washington, DC 20001 (telephone 202-334-3213; fax 202-334-2519; or e-mail TRBsales@nas.edu).

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the project were chosen for their special competencies and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to the procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This project was sponsored by the American Association of State Highway and Transportation Officials, the Federal Highway Administration, the Federal Transit Administration, and the Transportation Research Board.

Committee for a Conference on Using Census Data for Transportation Applications
Jonette Kreideweis, *Minnesota Department of Transportation (retired), Chair*
Nathan Erlbaum, *New York State Department of Transportation*
Thomas J. Kane, *Thomas J. Kane Consulting*
Keith Killough, *Arizona Department of Transportation*
Kathleen Lindquist, *Washington State Department of Transportation*
Alan E. Pisarski, *Consultant*
Steven E. Polzin, *University of South Florida*
Kristen Rohanna, *San Diego Association of Governments*
Guy Rousseau, *Atlanta Regional Commission*
Robert L. Santos, *The Urban Institute*

Liaisons
Ken Cervenka, *Federal Transit Administration*
Elaine Murakami, *Federal Highway Administration*
Penelope Weinberger, *American Association of State Highway and Transportation Officials*
Alison K. Fields, *U.S. Census Bureau*

TRB Staff
Thomas M. Palmerlee, Associate Division Director–Data and Information Services
Nanda Srinivasan, Senior Program Officer
Matthew A. Miller, Senior Program Associate
Jaclyn Hubersberger, Senior Program Associate

Publications Staff
Diane LeBlanc Solometo, Editor
Javy Awan, Production Editor
Jennifer J. Weeks, Editorial Services Specialist
Mary McLaughlin, Proofreader
Paul deBruijn, Typesetting and Design

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. 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 Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

Contents

Preface

A TRB conference, Using Census Data for Transportation, was conducted on October 25–27, 2011, at the Arnold and Mabel Beckman Center of the National Academies in Irvine, California. The conference was supported by AASHTO, Census Transportation Planning Products; FHWA, Office of Planning; FTA; and TRB.

TRB assembled a planning committee appointed by the National Research Council (NRC) to help organize and develop the conference program. The planning committee was chaired by Jonette Kreideweis, from the Minnesota Department of Transportation (retired). 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, identifying speakers, reviewing submitted abstracts, and developing topics for the breakout group discussions. Katherine Turnbull, from the Texas Transportation Institute, served as the conference rapporteur and prepared this document as a factual summary of what occurred at the conference.

The conference brought together approximately 115 individuals from across the transportation communities at the national, state, regional, and local levels, and from the public and private sectors and academia. The conference focused on the critical role of census data in a wide range of transportation planning applications.

The conference provided a forum for participants to share experiences with the use of census data in transportation planning and decision making. Participants also learned about recent and forthcoming census products. The conference further provided the chance 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 conference began with general sessions highlighting activities at the Census Bureau; data at the federal, state, and local levels were used. Breakout sessions included presentations on the use of census data and data from other sources for a wide range of transportation applications. Two breakout discussion group sessions allowed participants to share their ideas and experiences, discuss challenges and opportunities, and identify research and training needs. In addition, there were five discussion groups focused on different topics. The five topics were content specification; integration with other sources and private-sector data; data dissemination and data access, tools, and models; funding and institutional arrangements relating to the census and alternative sources; and research and professional development. The final general sessions included an update on research activities and a data power user's forum. The discussion group leaders highlighted key points and common themes in the closing session.

The views expressed in this summary are those of individual conference participants and do not necessarily represent the views of all conference participants, the planning committee, TRB, or the NRC.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC Report Review Committee. The purposes of this independent review are to provide candid and critical comments that will assist the institution in making its published summary as sound as possible and to ensure that the report meets institutional standards for clarity, objectivity, and responsiveness to the project charge. The review comments and draft manuscript remain confidential to protect the integrity of the process.

TRB thanks the following individuals for their review of this report: Stacey G. Bricka, Texas A&M Transportation Institute, Austin; Mell D. Henderson, Mid-America Regional Council, Kansas City, Missouri; Phillip J. Mescher, Iowa Department of Transportation, Ames; and Guy Rousseau, Atlanta Regional Commission, Atlanta, Georgia.

Although the reviewers listed above provided many constructive comments and suggestions, they did not see the final draft of the conference summary before its release. The review of this report was overseen by C. Michael Walton of the University of Texas at Austin. Appointed by the NRC, he was responsible for making certain that an independent examination of this summary was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this summary rests entirely with the author and the institution.

The committee thanks Katherine F. Turnbull for her work in preparing the conference summary report. The committee also extends a special thanks to AASHTO Census Transportation Planning Products, FHWA, and FTA for providing the vision and encouragement that made the conference a success.

DEDICATION

These proceedings are dedicated to **Nathan Erlbaum** of the New York State Department of Transportation and member of the Conference Planning Team, who passed away in 2012. Active in TRB, Nathan served on the Urban Transportation Data and Information Systems Committee, the National Transportation Data Requirements and Programs Committee, and numerous panels and task forces. He was known for his problem-solving focus and his ability to explain complex models and data to policy members. Nathan's work will continue to guide others throughout the transportation community.

Welcome and Opening Session

Jonette Kreideweis, *Minnesota Department of Transportation (retired), presiding*
Thomas L. Mesenbourg, *Deputy Director, U.S. Census Bureau*

The opening general session featured a welcome and overview of the conference from Jonette Kreideweis, the Chair of the Conference Planning Team. Thomas L. Mesenbourg from the U.S. Census Bureau provided an overview of current activities and future directions at the Census Bureau.

CENSUS DATA IN A DYNAMIC TIME FOR TRANSPORTATION PLANNING

Jonette Kreideweis

Let me share a little history with you. In participating in this conference, we celebrate a unique partnership that has evolved over the years between the U.S. Department of Transportation, the states, metropolitan planning organizations (MPOs), AASHTO, and the Census Bureau to collaborate, purchase, and develop data for transportation planning. This partnership has informed the standard demographic and economic data products produced by the Census Bureau.

More specifically, it provided the foundation for the Census Transportation Planning Products (CTPP) program. The CTPP program, with funding support from all the states and many MPOs, has become the standard for providing residence, workplace, journey-to-work, and flow data for the transportation community. In addition, the CTPP program has provided great training opportunities, rich research, and tremendous technical support.

Many of you have participated in previous census data conferences. In looking back, it seems the principal goal of these conferences was to bring data experts together to identify what was needed from the Census Bureau, the research community, and other stakeholders to support planning efforts. This list of needs then became the platform for what has been described as a sort of unilateral negotiation with the Census Bureau to sort out what would and could be provided.

The focus of this conference is a little different from conferences of the past. It is more than just a dialogue between all of us and the Census Bureau—it is about understanding the value of demographic, travel, flow, and economic data to the work we do in this time of unprecedented change and challenge.

The conference is about discussing how we can best take advantage of the American Community Survey (ACS) and overcome some of the challenges and technical issues that result in part from smaller sample sizes, higher margins of error, and increased data suppression. The conference is about learning what you are doing to supplement and integrate new public and private data sources. It is about hearing what you are doing to improve data access, analysis, modeling, and use of data to more effectively support decisions.

Last, the conference is about where we go from here as a transportation community to ensure that we have the data, tools, training, research, institutional arrangements, and partnerships in place to do the best job that we can to support the decision makers in our organizations.

From the preconference survey, we know that the variety of ways you are using census data is increasing. We learned that many of you are combining data from multiple sources in new ways. We also know that geographic information systems and mobile technologies are beginning to play a larger role than in the past, and for some, new partnerships are emerging to fill data gaps and needs. We know there is growing interest in transit, bike, and pedestrian travel.

In larger, growing metropolitan areas, there is an increasing need for much more granular data at smaller geographies to support modeling, livability, transit, and other transportation planning studies. Census data needs and uses are also changing in many of the smaller, slow or no growth areas of the country.

This conference is about hearing how these trends and changes are affecting current and future census data needs. It also focuses on how we can best ensure that we have the data we need to meet current and future needs.

The conference planning team worked hard to develop an interesting, informative, and great program. We begin with an excellent panel that will set the stage for understanding some of the political realities behind the census data we use. We then move into a series of sessions with speakers sharing their experiences, successes, and challenges in using census data. Another panel will discuss the results of current research to improve the utility and reduce the effects of data suppression and take advantage of some newer census data products. Tomorrow includes a power user forum that will give us the chance to hear how a few major players and vendors are enhancing and improving their applications, tools, and methods to better meet our needs.

Beginning late tomorrow and into the third day of the conference, we have scheduled facilitated breakout sessions. We will be asking you to share your insights, concerns, suggestions, and ideas for where the transportation community needs to go from here. This is your chance to tell us how census data products add value and how we can begin to address new opportunities and technical issues.

We will be recording all your thoughts and sharing them in conference proceedings. As a TRB function, we are not empowered to make recommendations; however, we will be providing a summary that documents individual observations and common themes as input to stakeholders. Katie Turnbull from the Texas Transportation Institute will be preparing the conference summary.

In addition, personnel from the Census Bureau and members of the AASHTO CTPP Advisory Committee are here to share their ideas and to listen to your comments. The committee will be meeting immediately following this conference. So now is the ideal opportunity to communicate your ideas and needs.

CENSUS BUREAU DIRECTIONS

Thomas L. Mesenbourg

My comments focus on the 2010 census and activities under way at the Census Bureau. I will begin by summarizing a few highlights from the 2010 census, including those related to transportation characteristics. I will discuss the Local Employment Dynamics (LED) program and the Commodity Flow Survey (CFS). I will conclude by describing the Census Bureau change initiatives, including the FY 2012 budget, Census Bureau reorganization, 2020 census, regional office restructuring, and ACS program review.

As you are aware, the 2010 census results indicate an increasingly diverse and rapidly growing U.S. population. The 2010 resident population was 308,745,538, which represents a 9.7 percent increase from the 2000 census. Although the population of the United States has continued its steady increase since 1910, the 9.7 percent growth rate from 2000 to 2010 was lower than the 13.2 percent increase from 1990 to 2000.

All states, except Michigan, experienced an increase in population between 2000 and 2010. Nevada experienced the largest percentage increase, with a 35.1 percent growth rate. Puerto Rico experienced a slight decline in population.

The 2010 census is used for the apportionment of congressional seats. Texas gained four seats, and Florida gained two. Arizona, Georgia, Nevada, South Carolina, Utah, and Washington all gained one seat. Both New York and Ohio lost two seats. Illinois, Iowa, Louisiana, Massachusetts, Michigan, Missouri, New Jersey, and Pennsylvania each lost one seat.

A number of 2010 census products have been released and are available for use. These products include Apportionment and Redistricting counts, demographic profiles, Island Area population counts, and Summary File 1. Other products include the ACS (1-year), Special Tabulation of Same-Sex Couple Households, Voting and Registration in the United States, Briefs on Race Groups, Housing Statistics from the 2010 Census, and Foreign Born Population in the United States: 2010. The ACS 3-Year was just released, and the 5-year estimates for 2006–2010 will be available in December. The Geographic Mobility: 2008–2009 Summary and Summary File 2—counts and characteristics for more than 300 detailed population groups—will be available during the next few months.

Let me take a few minutes to highlight some of the transportation characteristics from the 2006, 2007, 2008, 2009, and 2010 ACS and the 2010 census. The percentage of workers 16 years old and older who drove alone to work increased from 76.1 percent in 2009 to 76.6 percent in 2010. The percentage of workers who carpooled decreased from 10.0 percent in 2009 to 9.7 percent in 2010. The percentage of workers who took public transportation decreased from 5.0 percent to 4.9 percent.

In 2009, the foreign-born population carpooled at a rate of 16 percent, compared with 9.4 percent for the native-born population. Public transportation use among the foreign born was 10.8 percent in 2009, more than twice that of the native-born population, at 4.1 percent. 2010 saw a similar relative distribution of travel modes across these groups.

In 2010, most workers, 72.6 percent, were employed in the same county as their residence. The states with the highest rate of workers who worked outside their county of residence in 2010 were Virginia, 51.3 percent; Maryland, 47.0 percent; and New Jersey, 45.7 percent.

The average travel time for workers increased slightly from 25.1 in 2009 to 25.3 minutes in 2010. The 2010 average commute represents an increase from the 21.7 minutes recorded in the 1980 census, but a decrease from the 25.5 minutes recorded in the 2000 census.

Private vehicles continue to represent the major mode of transportation for work trips. These data can be examined for areas. For example, the mean travel time to work from 2006 to 2010 here in Irvine, California, can be examined. The results indicate little change during the 5-year period. These data can be displayed graphically by census tracts.

The New York–Northern New Jersey–Long Island and New York–New Jersey–Pennsylvania metropolitan areas, with 30.5 percent, were at the top of 15 U.S. metropolitan areas ranked by the number of workers who commuted by public transportation in 2009. Other metropolitan areas in the top five include Chicago, Illinois; Washington, D.C., and Los Angeles and the San Francisco Bay area in California.

The New York metropolitan area recorded the longest mean travel time to work in 2009, at 34.6 minutes. The Washington, D.C., metropolitan area was second, at 33.4 minutes. The Lewiston metropolitan area in Idaho and Washington recorded the shortest mean travel time to work, at 14.7 minutes.

The LED data can be used to analyze where workers are employed and where they live. Work flows in, out, and within a community, and the data can be obtained and analyzed. The changes in work flows over time can be examined.

The CFS is a partnership between the Census Bureau and the Bureau of Transportation Statistics. It is the primary source of information about freight movement in the United States. The CFS is the only source of nationwide data on the movement of goods from origin to destination by all modes of transportation. Data from the CFS are used by policy makers and transportation planners for assessing the demand for transportation facilities and services, energy use, safety risks, and environmental concerns.

Data included in the CFS are shipment value, tons (weight), ton-miles, and average miles per shipment. Data published by the CFS are mode, distance shipped, commodity, and hazardous shipment characteristics. Data are also available by export shipment characteristics, North American Industry Classification System (NAICS), and geography, including U.S., state, and selected CFS-defined metropolitan area.

Preliminary data from the 2007 CFS were released in December 2008, and the final data were available in December 2009. A number of improvements were made in the 2007 CFS. Precanvassing was introduced in 2006 to identify out-of-scope establishments before the 2007 CFS. Integrated modeled transportation network software was used to compute origin–destination mileages. A disclosure avoidance method was applied to 2007 CFS published data. The 2007 CFS was the first time estimates were released by NAICS.

Planning for the 2012 CFS has been under way for more than a year. Data collection is scheduled to begin at the end of 2011. The preliminary data are expected to be released in December 2013, and the final data should be available in December 2014. Improvements in the 2012 CFS include testing and revising questions and using integrated computer-assisted data entry forms processing, including optical character recognition. Another improvement is the use of electronic reporting, including the capability to download and upload spreadsheets with shipment data.

The Census Bureau change initiatives focus on the FY 2012 budget update, Census Bureau reorganization, 2020 census planning, restructuring of the regional offices, and ACS program review. Major elements of the Census Bureau FY 2012 budget include the flatlining of salaries and expenses. The request included $16 million in planned program terminations and reductions, plus the three new initiatives. The House markup does not fund the three new initiatives, and the Senate markup cuts $5 million more. The House markup cuts $156 million, or 21 percent of the periodic programs request, including funds for the FY 2011 ACS improvements and geographic initiative. It also includes the first funds for 2020 census planning. However, if the House markup prevails, it could threaten the FY 2012 Economic Census, a $124 million request, and the 2020 census, a $165 million request. The Senate committee markup for periodics is $63 million below the request, but is $94 million higher than the House markup. The total budget House markup is 16 percent below the request, and the Senate markup is 8 percent below the request.

The Census Bureau reorganization includes reestablishing the Research and Methodology Directorate, establishing a new 2020 Census Directorate including the American Community Survey Office, strengthening the CIO position, and establishing a Chief Technology Office. It also includes new senior technical positions, including a chief demographer, and establishes an Enterprise Risk Management Office.

Cost is a key challenge with the 2020 census. The cost per housing unit of the 1970 census—in 2010 dollars—was $14. In comparison, the estimated cost per housing unit of the 2010 census was $108, and the projected cost of the 2020 census is $181.

The 2020 census vision is to provide an efficient and quality census that counts people, once, only once, and in the right place. The 2020 census mission is to conduct a census of the population and housing and disseminate data to the President, the states, and the American people. The guiding principles for the 2020 census are to reduce cost, maintain quality, reduce the filed timeline, and tailor response modes. Other guiding principles are to leverage the ACS, support continual frame updating, support agile decision making, and reuse data. Additional guiding principles are to leverage systems and methods, leverage partnerships, support organizational solutions, and build a 21st-century workforce. The 2020 census goals are a complete and accurate census, embraced and valued results, efficient 2020 census, and well-managed 2020 census program.

There are four major cost drivers with the 2020 census. Cost Driver Number 1 is the increased population diversity and decreased willingness to cooperate with self-response and nonresponse follow-up. Cost Driver Number 2 is the limited 2000 to 2009 updating of the master address file and Topologically Integrated Geographic Encoding and Referencing System (TIGER) maps, which led to a design incorporating "last-minute" (2009) updates. Cost Driver Number 3 is the failure to link acquisitions, schedule, and budget. Cost Driver Number 4 is the demand for the Census Bureau to strive for improving accuracy over previous censuses.

The regional office restructuring includes a change in the number of offices and a change in the management of data collection. The restructuring plan was developed by the 12 regional directors. The possible designs were evaluated against goals and the regional boundaries, and city selections were data driven. The regional office restructuring reduced the number of offices from 12 to six. Regional offices in New York City, Philadelphia, Pennsylvania; Atlanta, Georgia; Chicago, Illinois; Denver, Colorado; and Los Angeles, California, remain in operation. Regional offices in Boston, Massachusetts; Charlotte, North Carolina; Detroit, Michigan; Kansas City, Missouri; Dallas, Texas; and Seattle, Washington, will close. The new regional office boundaries will be effective in January 2013.

The selected regional office design includes no change in the responsibilities of the almost 7,600 interviewers. Changes will be made in supervisory structures and responsibilities, with more supervisory staff working out of their homes.

Improved management information systems and other tools will maintain high data quality and support increased efficiency and lower costs. The reorganization is projected to save $15 to $18 million, net of transition costs, beginning in FY 2014. Of the approximately 8,200 field staff, 330 regional office staff in the six closing offices will be affected, with a net loss of between 115 and 130 positions nationwide. Assistance and support are being provided to the affected personnel.

The ACS is undergoing a comprehensive program review. The processes, methods, products, program management, and research are all being examined. The Committee on National Statistics panel was initiated in the fall of 2011, and a separate California State Association of Counties and Race and Ethnic Advisory Committee review is being conducted. Sample expansion began in June 2011, with increased sample size for small tracts and government units. There is computer-assisted personal interviewing follow-up of all nonmailed and nonresponding addresses in American Indian/Alaskan Native areas, Hawaiian homelands, and remote Alaska.

A positive Internet test was conducted in April 2011. A second test will be conducted in November 2011. A content reinterview survey is scheduled for January 2012, and the results will be available in the summer of 2013. Population controls for 2010 ACS products include county and subcounty based on intercensal population estimates. Population controls for 2011 ACS products include county and subcounty controls, which will be based on population estimates and vintage 2011 population estimates.

For more information about commuting and to access commuting in the United States, 2009, visit: http://www.census.gov/hhes/commuting/.

Data User and Provider Panel

Mary Lynn Tischer, *Federal Highway Administration*
Martin Tuttle, *California Department of Transportation*
Vincent Barabba, *Chairman, Market Insight Corporation*
Steven E. Polzin, *University of South Florida, presiding*

This general session focused on census data users in government agencies and the private sector. Mary Lynn Tischer described data needs at FHWA. Martin Tuttle of the California Department of Transportation provided a state and local perspective. Vincent Barabba of the Market Insight Corporation discussed census data reliability and data quality for small geographic areas.

TRANSPORTATION AGENCY NEEDS AND USES OF CENSUS DATA: A FEDERAL PERSPECTIVE

Mary Lynn Tischer

My comments focus on the needs and uses of census data at the federal level, especially at the U.S. Department of Transportation and FHWA.

The mission of the U.S. Department of Transportation is to "serve the U.S. by ensuring a fast, safe, efficient, accessible, and convenient transportation system that meets our vital national interests and enhances the quality of life of the American people, today and into the future."

To meet this mission, the scope and focus of the department and FHWA are evolving. FHWA is no longer focusing just on pure engineering design. Examples of the new and emerging issues being considered at FHWA include livability, sustainability, energy, the environment, and performance management.

The U.S. Department of Transportation and FHWA are also facing new constraints. Examples of these constraints are population growth, social and demographic changes, and scarce resources. The challenge for the department is to chart a course through these challenges that will fulfill our mission.

Data can help us understand new and emerging issues quantitatively. Data, along with analytic tools, can assist in testing various policy and program scenarios and foreseeing the outcomes without actual implementation.

There are numerous federal data needs. Data are needed for national level indicators for policy analysis, program evaluation, and regulation and legislation. Areas of policy interest include pricing, livability, and greenhouse gas (GHG) emissions. Quality trend data are needed for forecasting. Consistent and timely data are needed across national, state, and local levels for programmatic analysis and performance measurement

Data provide the robust inputs for national, state, and local modeling. Data are key inputs for travel flows for safety, operations, and infrastructure planning. More modeling inputs with the ability to make comparisons with the overall population are also needed.

Historically, major metropolitan areas conducted 5 percent survey samples for use in travel demand models. These surveys were replaced with smaller surveys of less than 1 percent combined with census data for activity-based models and microsimulation. Many state and metropolitan areas funded add-on samples to the National Household Travel Survey (NHTS). The NHTS is not large enough to provide small-area estimates directly. The American Community Survey (ACS) is of benefit in combining data sets.

Census data are important in transportation planning and other transportation analyses. The census provides consistent and comparable data across local, state, and national levels. It provides a robust link between sociodemographics, geography, and transportation. The census provides detailed home and work location data for geographic, demographic, and transportation data points and analyses.

The census uses a consistent, well-tested method and survey instrument. One entity collects the data by using the same survey instrument and method. There is a link to these decennial census and other census data products. Regular and timely data are available in 1-, 3-, 5-, and 10-year intervals.

The census also provides small-area data at the traffic analysis zone (TAZ) and block-group levels. Data at each of these levels have historically been used by smaller metropolitan planning organizations (MPOs) because the mix of characteristics was the only local data available. NCHRP 08-79 is examining improved handling of small samples and perturbed data by data suppression. There is a presentation on this project Wednesday afternoon. More geographic detail on work trips by TAZ would be beneficial.

There are many things we would not know without census data. We would not know that there was a drop in the percentage of African-American households without vehicles, from 31percent in 1990 to about 24 percent in 2000 and 20 percent in 2010. We would not know that the number of workers over 65 years of age rose by more than 21 percent while the population in that group increased 12 percent. Without the census, we would not know that Hispanics carpool at a rate double that for non-Hispanics, 23 percent versus 11 percent. Without the census, we would not know that new immigrants are more likely to carpool, walk and bike, and take transit to work.

Without the census, we would not know that commute times are becoming longer. The portion of workers who reach their jobs in less than 20 minutes dropped to 47 percent in 2000 and 42 percent in 2010, after being about 50 percent for decades. A large portion of the population still leaves home between 6:30 a.m. and 8:00 a.m., but that number has fallen from 43 percent to 38 percent during the past two decades. At the same time the share of people leaving before 6:00 a.m. has grown from approximately 9 percent to almost 13 percent.

Without the census, we would not know that from 1990 to 2000, 64 percent of the growth in metropolitan commuting was in flows from suburb to suburb. We also would not know that 27 percent of commuters work outside their county of residence and half of the workers added during the decade worked outside their county.

Census data are used in numerous ways. Census transportation data are used in performance management for livability and sustainability. Census data give us direct measures of workers who use nonmotorized and transit travel modes for commute trips. The census provides data on access to work including mode share and travel time. It provides data on mobility, including vehicle availability and transit accessibility. It also provides data on livable community profiles, including sociodemographics and employment.

Census data are also of use in examining the land use and transportation interface. Census data allow analysis of the interaction of population density, employment density, and transportation networks. Examples include examining the percentage of the population whose commute is less than 20 minutes or the percentage greater than 60 minutes, to explore the job–housing balance. Census data can be used to examine the percentage of the people who work outside their county or city of residence. Census data are also used in the analyses of congestion, infrastructure investment decisions, operations, social equity, and public involvement.

In examining census data, we have heard similar themes over the years. There are concerns over the timeliness of data, the availability of data at a granular level, and the need for easier access. Other ongoing concerns include reliability, disclosure issues, and costs.

A number of issues and topics need to be considered as we look to the future. The topics include increasing reliance on geographic information systems (GIS) for data access and analysis, the need for transportation data that are less expensive, and the need to maintain the sample size of the ACS to ensure the flow of information. There is also a need to improve response rates. Other topics are the need for new approaches to making microdata records from the ACS available for geographies smaller than a public use microdata area, which may require limiting the number of variables, but could change the way the Census Transportation Planning Products program is designed. New data collection mechanisms offer promise for origin–destination patterns, but the advantage of the ACS is the direct measurement of sociodemographic characteristics, and home and work locations combined with travel behavior. There is a need for a better understanding of the validity of longitudinal employer–household dynamics and how it can best be used. Finally, it would be helpful to have joint research projects between the Census Bureau and transportation agencies.

STATE AND METROPOLITAN PLANNING ORGANIZATION DATA NEEDS: MEETING FEDERAL REQUIREMENTS AND LOCAL APPLICATIONS

Martin Tuttle

Similar to in other parts of the country, the economy in California is unprecedently challenged. Caltrans and other state agencies are doing more with less. Census data, as well as data from other sources, along with technology can help. The theme of my presentation is better data for better decisions. Working together, we can maximize the use and benefits of data, including making better decisions.

In addition to an uncertain economy, California faces other challenges. The population of the state is projected to reach 54 million by 2040. The population is getting older, with the aging of the baby boom generation. Climate change and air quality are concerns, especially for coastal communities. The state continues to deal with chronic budget deficits. All of these challenges affect the transportation system.

Recent legislation also influences the transportation system. The Global Warming Solutions Act [or Assembly Bill 32 (AB 32)] requires reducing 20 percent of GHG emissions by 2020. Senate Bill 375 (SB 375) tasks the MPOs in the state with reducing GHG emissions through a sustainable communities strategy that coordinates land use and transportation in their regions. SB 391 tasks Caltrans with developing a transportation plan to reduce GHG emissions statewide.

Caltrans is responding to these and other requirements by developing the California Interregional Blueprint. MPOs in the state have successfully used blueprints. When I was director of the Sacramento MPO, we developed the first blueprint, linking land use and transportation. Blueprints provide the opportunity to make transportation investments in advance of land use changes, rather than transportation improvements following land use changes.

The California Interregional Blueprint is integrating statewide modal plans and programs with new technologies and tools. It is building on regional transportation plans and sustainable community strategies. The blueprint is developing more robust modeling tools to analyze different scenarios and strategies.

Linking transportation and land use planning has not always been easy in the past. It is now becoming the norm in California. Better transportation infrastructure investments can be made when they are linked to land uses. The California Interregional Blueprint will enhance future decision making.

State-level and regional-level activities are on parallel tracks and are being coordinated. Regions are developing sustainable communities strategies and land use plans. Caltrans is developing the Statewide Interregional Blueprint, which builds on the regional plans and focuses on the statewide transportation network. The interim plan will be completed by December 2012. The performance of the plan will be measured to assess GHG emissions reduction.

The statewide model framework includes the statewide travel demand model, which is a critical element in estimating GHG emissions. We have also developed a new tool, Caltrans Earth, to visualize GIS data layers. It includes the available parcel-level data in the state in a GIS format and uses the Google Earth platform. Data from Caltrans, other state agencies, MPOs and regions, and local governments are included. There is an overlay of aerial photography.

Caltrans Earth provides a large library of GIS information. It is currently being used by 7,000 Caltrans employees and will be available to other groups by mid-2012. We are realizing extensive benefits from the system, which is being used for a wide range of projects. For example, Caltrans employees examining curb cuts for compliance with the Americans with Disabilities Act are using Caltrans Earth, saving time and reducing traffic disruptions.

The statewide model framework includes the California Household Travel Survey. For the first time, the statewide household travel survey is being coordinated with 18 MPOs and several state agencies, including the Energy Commission. The sample includes approximately 57,000 surveys. We have been able to leverage funding from multiple sources and provide benefits to all groups. We hope this effort will lead to enhanced coordination and cooperation on other projects. The state travel demand model is being enhanced to add future years, and a statewide freight model is under development. We hope to develop a statewide integrated transportation land use and economic model in the future.

BROAD VIEW OF CENSUS SURVEYS FOR RELIABLE, HIGH-QUALITY DATA FOR SMALL GEOGRAPHIC AREAS

Vincent Barabba

I was asked to provide a broad view of the reliability and data quality of the census surveys for small geographic areas. I will focus on the broad view, as many of you in the audience are more knowledgeable about the accuracy and quality of specific data.

It is important to begin by defining "reliable." The best graphical representation I found was on the website Experiment-Resources.com, which is illustrated in Figure 1 (page 16). The key element of the definition is the relationship between

FIGURE 1 Defining reliability and validity.
(Source: Experiment-Resources.com)

reliability and validity. The first point to remember from this figure is to avoid data that reliably and consistently provide the same information time and time again when that information is not right. The Census Bureau has been facing this issue for a long time. We want information that comes from questions and collection methods that consider the changing times and provide the right and relevant answers.

I think we need to revise one of my favorite quotes by John Tukey from the 1960s to deal with this issue. The initial quote is "Far better an approximate answer to the right question, which is often vague; than the exact answer to the wrong question, which can always be made precise."

Although this quote has remained relevant for 50 years, I would offer the following revisions because of changing 21st-century conditions. "Far better an approximate and timely answer to the right question, which is often vague and sometimes modified to stay relevant, than the untimely exact answer to the wrong question, which can always be made to appear precise."

My generation was educated on the premise that with enough data we could make a good estimation of the probable outcome to a problem. I think the census long form responses were both reliable and a valid measurement of many of the important data elements in the United States for a long time, even beyond the date of the data that were collected. This was the case when the pace of change in the United States was evolutionary. Change is no longer evolutionary in this country or in the world.

At that time we also used the phrase "What is past is prologue," which was borrowed from Shakespeare and *The Tempest*. I think this phrase is also no longer accurate, and I would suggest that today, "What is past . . . is not as likely . . . to be prologue." This change raises a big question related to the census data.

One issue is that the long form has been taken away as a source of data. It was of great value, especially in the evolutionary period. The long form data did have limitations, however. The fact that the long form reports came from a large sample size provided data with a low sampling error. The other error, which is often overlooked, is non-sampling error. Non-sampling error focuses on whether the respondent provided an accurate response and whether the response was recorded and processed accurately.

These issues remind me of a paper written in the 1970s by Tore Delanius. He described the two types of errors in a fairly dramatic way. "Strain at a gnat and swallow a camel, or the problem of measuring sampling and nonsampling errors." He pointed out that "the sampling error plays the role of the gnat, sometimes malformed, while the non-sampling error plays the role of the camel, often of unknown size and always of unwieldy shape."

Anyone who has examined the issues associated with sampling error realizes the difficulties involved. The issue of non-sampling error has become more important recently as a result of rapid changes. Even though the long form had a smaller sampling error, it took up to 3 years to make the data publicly available because of screening errors and checking the data for accuracy. During the evolutionary period, the data were still valid 3 years later; however, I am not sure that is the case today.

The ACS has a larger sampling error, but the data can be processed much quicker. Interviews for the ACS are conducted continuously rather than every 10 years. The interview process is conducted by a permanent, extensively trained, and experienced field force, rather than the hardworking temporary employees who were provided training for the census. A relatively straightforward trade-off has been made. The sampling error, which we are capable of measuring, has been increased as a result of the smaller ACS sampling size. The level of confidence in the resulting number is known. On the basis of preliminary work done at the Census Bureau, it is felt that problems with the non-sampling error, that is, making sure the answer is recorded and tabulated properly, have been reduced.

More important, because the ACSs are conducted continuously, the timing of the response can be tied to changing conditions surrounding the respondent at the time the questions are answered. Another trade-off is being made in ensuring that the right question is being asked. The ultimate wish of any forecaster is to have a data set that covers a long period of time and includes responses to a consistent set of questions. That wish is appropriate, as long as the questions continue to be relevant to the travel issues being faced by the respondent specifically and by society in general.

As a member of the California Citizens Redistricting Commission, I have had the opportunity to attend public hearings throughout the state. These meetings and traveling to different parts of the state highlight how rapidly things are changing. I am not sure there is one question that can address all of the issues facing society today.

I think the trade-off of the long form for the ACS was reasonable. There is much hard work under way at the Census Bureau and within the different user communities providing insights on ways to improve the ACS, from a data collection and recording point of view and from the perspective of the ultimate user. As Wes Churchman, a distinguished professor at the University of California, Berkeley said, "the value of information is in its use, not its collection."

My suggestion to those of you who want to see the ACS improved is to be a constructive critic, with the focus on the word "constructive." The following portion of a poem written by Edna St. Vincent Millay in 1941 conveys what I mean by constructive criticism: "Upon this gifted age, in its dark hour, / Rains from the sky a meteoric shower / Of facts...they lie unquestioned, uncombined. Wisdom enough to leech us of our ill / Is daily spun; but there exists no loom / To weave it into fabric."

I found Martin Tuttle's description of Caltrans Earth interesting. You are in essence creating the equivalent of a loom. You have to determine which trends will be brought together. That is the approach and attitude that is necessary for our society as we begin to learn to live not with evolutionary change, but with chaotic, complex, and constantly changing issues. We cannot control what happens in our community, state, and country or with global connectivity. We need to learn to live with proximate information to really relevant questions.

Census Bureau Data Delivery and Research Activities

Alison Fields, *U.S. Census Bureau*
Jonette Kreideweis, *Minnesota Department of Transportation (retired), presiding*

This general session featured Alison Fields from the Census Bureau. She provided an update on projects and products from the Census Bureau's Journey-to-Work and Migration Statistics Branch.

CENSUS BUREAU UPDATE

Alison Fields

My comments focus on the activities of the Census Bureau's Journey-to-Work and Migration Statistics Branch. The branch is responsible for commuting, migration, and place-of-birth data.

I will highlight upcoming product release dates, commuting data and the American Community Survey (ACS), and improvements in the ACS sampling design. I will also describe the flows products from the ACS, the Equal Employment Opportunity (EEO) special tabulation, and the ACS data on disability. The ACS data on disability were unfortunately missing from the 3-year Census Transportation Planning Products (CTPP) as a result of a change in the question that caused a break in series. These data will also subsequently be unavailable for the 5-year CTPP.

The 1-year ACS data were released in September 2011. A commuting report based on the 2009 ACS data was also released in September. As the Deputy Director mentioned earlier today, the 3-year ACS data are being released on October 27, 2011. Various short information briefs will be released throughout the coming months on the 3-year data. Topics covered in these briefs include characteristics of school-age children with disabilities, the food stamp program, the foreign-born population, and recent immigrants with science and math degrees. The 5-year data will be released in December 2011. The 2010 ACS data use the 2010 census geography and population estimates.

The Deputy Director mentioned Migration Day, which is all about our branch. Migration Day is coming up (held on November 15, 2011). The most recent mobility data from the Current Population Survey (CPS) will be released, along with a joint report on migration based on ACS and CPS data. The most recent state-to-state migration flow data will also be released, along with a brief on people who continue to live in the state in which they were born. Further, a preview will be provided on the first products to be released on the 5-year ACS county-to-county migration flows. The previews will highlight how the flow products are being reenvisioned.

The geographic summary levels for the ACS include 953 areas in the residence-based tabulations for metropolitan statistical areas, 3,221 counties, approximately 70,000 tracts, and approximately 210,000 block groups. The 5-year estimates now include all counties and tracts. Block groups are provided for the first time.

Summary File 1 presents the most detailed characteristics for the total population in the Census Bureau's standard pretabulated data products. Detailed tables are available on age, sex, race categories and for the Hispanic or Latino origin, households, families, relationship to householder, housing occupancy and tenure, and group quarters. Counts are also available for the nine race categories and Hispanic or Latino origin groups.

Summary File 1 was released for each of the 50 states, the District of Columbia, and Puerto Rico. Most tables include data at the block- or census-tract level. Some tables are repeated for the nine race categories and for the Hispanic or Latino origin groups. Summary File 1 was released on a state-by-state basis from June 16 through August 25, 2011, on American FactFinder and the FTP site. All the Summary File 1 state data and the national data are to be released together on a single DVD. The entire national Summary File 1 will be released this week (in October 2011). It will include new geographic levels, including regions, divisions, and areas that cross state boundaries.

Summary File 1, the Urban/Rural Update File, will provide users with urban and rural population and housing unit counts, down to the block level, and characteristics for urbanized areas and urban clusters. This file is to be released on American FactFinder and the FTP site in October 2012. Summary File 1, Redefined Core Based Statistical Areas (CBSAs) Update File, will contain the same data tables as the state files for the redefined CBSAs. The Office of Management and Budget (OMB) is responsible for redefining the CBSAs on the basis of the 2010 census. The Redefined CBSAs File will be released on American FactFinder and the FTP site in August 2013. A commuting flow product will be released before August 2013.

The 2010 Census Briefs provide the Census Bureau's first analysis of the 2010 census population and housing data. The briefs highlight key topics, including the geographic distribution of population and housing data. The briefs are being released on an ongoing basis. Census Briefs on population distribution and change, race and Hispanic origin, age and sex composition, and the Hispanic population have been released. Two more briefs are to be released on September 29, 2011. These briefs address the white population and the black population.

Briefs on housing characteristics and congressional apportionment are to be released in November 2011. The Census Briefs are being released on the Internet and are available in print. An ACS product, similar to the Census Summary File 3 (SF 3) product, will be available in 2013. It will provide data from Native American and Alaskan areas.

A number of improvements are being made in the ACS sampling design. Data users and stakeholders have repeatedly requested larger sample sizes for more accurate data. As part of a budget initiative for FY 2011, the Census Bureau did receive funding for a sample expansion from 2.9 million addresses per year to 3.54 million addresses per year. The sample increase began for mail-out surveys in June 2011, for computer-assisted telephone interviewing in July 2011, and for computer-assisted personal interviewing in August 2011.

The Census Bureau has also taken another approach to increasing data accuracy, specifically for small areas. The sampling rates in smaller tracts and governmental units have been increased, and the sampling rates in larger tracks have been decreased slightly. We expect this reallocation to provide estimates that are more accurate. This reallocation of the sample began in January 2011. The combined sample expansion and sample reallocation provides for improvements in the coefficients of variation (CVs) for a sample statistic by tract size. The anticipated results include much less variability in the tract-level CV distribution, with smaller areas having significantly better estimates.

There is an anomaly at the national level for the 5-year data. A user note is currently available on the data documentation website. The measure of error (MOE) for several estimates at the national level is significantly larger than for the single and 3-year data. The extensive analysis conducted by the Census Bureau indicates that the data are correct. The increase in the MOE is being caused by a specific step used in the weighting factor. The note provides more information on the situation and contact information if there are questions.

Synthetic data are being used to improve the group quarters data in the ACS. The sample will still be stratified by areas with group quarters below 15,000 and those 15,000 and above. Our branch is currently reviewing the test synthetic data because the flow data tend to highlight data anomalies related to group quarters. The group quarters synthetic data will be implemented with the next round of ACS data.

Commuting data and other detailed demographic, social, economic, and housing data are now being collected with the ACS. The work commute data that used to be part of the census long form are now part of the ACS. The ACS questions related to commuting are the same as those on the 2000 census long form. The means of transportation have 12 categories, including worked at home. Respondents are asked to choose only one mode, so multimodal trips are not captured, such as driving to a park-and-ride lot and taking the bus or light rail transit. Other questions focus on the number of occupants per vehicle for carpooling, the time leaving home for work, and the one-way travel time in minutes. Other questions capture the place of work, with respondents being asked for the address of their workplace or the nearest intersection. These responses are important for flow and origin–destination data.

Commuting in the United States: 2009 is the Census Bureau's first ACS-based commuting report. It focuses on travel time, means of transportation, time of departure, and place of work for the nation and for metropolitan areas. *Public Transportation Usage Among U.S. Workers: 2008 and 2009* is also part of the ACS Briefs series.

The Deputy Director discussed the Census Bureau reorganization this morning. As part of the reorganization, our division name has been changed to the Social, Economic, and Housing Statistics Division. As I mentioned, our branch deals with two sets of flow data—commuting flow data and migration flow data. As a result, consideration of potential ACS products tends to focus on using consistent packaging and tables. The American FactFinder currently cannot support flow products. The challenge is disseminating useful and accessible flow data that protect the confidentiality of all respondents.

The journey-to-work data focus on three geography types. These types are residence geography, where people live; workplace geography, where people work; and workflow geography, where people work and where they commute from.

Confidentiality issues have become more central to data production since the release of the migration data package on the 2000 census. Previous efforts to present flow data included large matrices that provided counts linking origins and destinations. The number of characteristics is limited to protect confidentiality. The tables are cumbersome and difficult for American FactFinder to accommodate. Standard tables present data for only one end of the flow. More characteristics are available, but they are of limited utility as a result of less geographic specificity about the other end of the flow.

It took a considerable effort to maintain the workplace geography tables so that users can calculate pseudoflows. A similar approach was used with the migration data. Tables presenting characteristics by home state allow users to develop pseudomigration flows. The tables are limited because they can be developed only for certain characteristics. The geography on the workplace and the state of previous residence is restricted to what can be geocoded by the bureau. The migration questions were recently changed, allowing coding at the block level. Coding the journey-to-work data at the block level has never been possible. The extended place of work allocation edits have shown considerable improvements, and the process is expected to be used with the creation of the 2006–2010 CTPP data. The process is also expected to be implemented with the standard ACS products beginning in 2012.

A Census Bureau internal flows working group was formed to create new and modified flow products that will become part of the standard production. The focus is on data content and dissemination methods. The considerations for new product development include user needs, Disclosure Review Board (DRB) rules, and the content of existing products. Examples of existing flow products include county-level flows to support metropolitan definitions and CTPP and EEO special tabulations.

The proposed approach for presenting flows is to alternate characteristics across years. A new set of flow data based on the 5-year ACS data at the county-to-county level would be released each year. According to the DRB requirements, the initial release, which is expected in the spring of 2013, will include counts only for county-level flows to support metropolitan and micropolitan area definitions. Each subsequent annual release in a 5-year period will include flows crossed with a small number of characteristics that will vary by year, achieving a set of flow tables by multiple characteristics during the course of 5 years. The DRB indicated that no more than four characteristics can be presented. The DRB has also indicated that the same four characteristics cannot be presented each year. The same four characteristics can be presented only every nonoverlapping 5 years. We expect the DRB will place the same restrictions on the commuting package when it is presented. For example, the 2006–2010 migration county-to-county flow table will also include a single cross section of age, sex, race, and Hispanic origin. We will not be able to present those same four variables for another 5 years.

With respect to the commuting flow in general, our priority has been the AASHTO request for a special tabulation. We do not expect to release any other commuting flow products in the interim. The place-of-work flows are used to support delineation of metropolitan and micropolitan statistical areas by OMB. The Census Bureau has updated metropolitan definitions on the basis of new population totals and commuting patterns since the 1950s. Metropolitan and micropolitan areas are made up of a county or counties with a large population nucleus and adjacent counties that have a high degree of integration with that nucleus. Integration between counties is measured by county-to-county commuter flows.

The county-to-county residence-to-workplace and workplace-to-residence flows based on the 5-year ACS 2006–2010 will be released to support metropolitan and micropolitan area definitions. The flows will be released every 5 years with no characteristics.

Other forthcoming products include the special report on home-based workers. This document is the Census Bureau’s first report to focus on workers who work at home. Data from the ACS and the Survey of Income and Program Participation (SIPP) are being used in the report. Emphasis is placed on industry, occupation, and class of worker. A report examining home-based workers, self-employed, teleworking, and travel on different days will be released. The SIPP is also being reenvisioned. There was an opportunity to add questions, including questions on the use of multiple transportation modes.

The ACS-based daytime population estimates represent another forthcoming product. These table packages present ACS-based daytime population estimates for various geographies. Reports on migration and commuting flows will also be available soon. These table packages and accompanying reports describe basic trends and present summary-level statistics for migration and commuting flows.

The ACS Office had a contract with David Wong of George Mason University for the development of the ACS Mapping Tool. The tool considers MOEs when thematic maps are created from ACS data. More information on the mapping tool and the ArcGIS extension is available for download at http://gesg.gmu.edu/census/index.htm. We have also been asked to participate in this contract to help develop an ArcGIS level, which would use the migration and the commuting flows and integrate the MOEs used with those flows. The resulting product would be available to the public.

We have used congestion data from the Texas Transportation Institute's *Urban Mobility Report* and data from the ACS to examine commuting expenses across metropolitan areas given commute times and commute lengths. Staff from the bureau presented a paper on the analysis at the Joint Statistical Meeting, and it will be part of the Brookings Institution release on November 7, 2011, in which the Census Bureau will be presenting the most recent data on the supplementary poverty measures. It is our intent to continue with the work on the SIPP, which is the primary means to model the supplementary poverty measure on the Central Provident Fund. This analysis examines the use of transportation and the trade-offs between transportation and housing affordability.

The commuting flows also provide the base for the daytime population estimates. We have been given a development opportunity in the Census Bureau to reenvision the daytime population estimates. The Census 2000 estimates were released shortly after the decennial. We are now developing ACS daytime-based estimates. We are exploring new data visualization techniques as part of this process. This work will probably lead to similar data visualization methods with the flow products.

The EEO special tabulation is an external benchmark for conducting comparisons between the racial, ethnic, and sex composition of each employer's workforce with its available labor market. It is used by organizations to develop and update their affirmative action plans.

The Census Bureau has created these tabulations for agencies to help measure and enforce compliance with civil rights laws since 1970. As a special product, the tabulations have expanded in scope after the 1980, 1990, and 2000 censuses.

The tabulation includes detailed occupation by sex, race, Hispanic origin, age, educational attainment, and other variables at various geography levels including states, metropolitan areas, and counties.

Four federal agencies together have requested Census to produce this special tabulation. These agencies are the Equal Employment Opportunity Commission, the Department of Justice Employment Litigation Section of the Civil Rights Division, the Department of Labor Office of Federal Contract Compliance Programs, and the Office of Personnel Management. The tabulation is run in addition to the normal products and is designed to best fit the needs of the sponsoring agencies.

The EEO special tabulation, which is referred to as the EEO file, is a tabulation of the census data on the civilian labor force 16 years of age and older. The tabulation is a group of unique table sets by place of residence, place of work, and residence-to-worksite commuting flows. The tables include occupation by sex, race, Hispanic origin, and citizenship for detailed geography. Other characteristics included are industry, age, educational attainment, and median earnings.

The EEO special tabulation is an ACS 2006 to 2010 five-year file, including MOEs. It will be based on the Standard Occupational Classification categories. It includes residence, worksite, and commuter flow tables. The residence category includes a new table for unemployed with no work experience in the past 5 years. Data on nationality will also be available. The data will be disseminated through American FactFinder. Flows are expected to be included. The consortium had provided funding to the American FactFinder to integrate the flows into the American FactFinder to ensure that all the EEO data are disseminated from a single source. This integration has not yet been completed, however. It will be available at the national, state, and county levels, as well as county clusters, CBSAs, and places with populations of 50,000 or greater.

All EEO tables are scheduled to be available in late 2012. The administrative contracting officer has already approached us about the potential to integrate all of the migration and commuting flows into the American FactFinder if the American FactFinder is able to accomplish this integration and dissemination. We are waiting to see whether the integration of the EEO into the American FactFinder is successful. If it is a success, releasing flows on the American FactFinder will probably follow.

Disability information from the ACS is still available. Information is available from the just released 2008–2010 data. There have, however, been changes in the questions on hearing and vision difficulties, physical activities, cognitive functioning, basic activities of daily living, and instrumental activities of daily living. The questions on employment disabilities were dropped.

The concept labels were changed between the 2007 and the 2008 questions. The 2007 questions focused on disabilities related to sensory, physical, mental, self-care, going outside the home, and employment. The 2008 questions focused on disabilities related to hearing, vision, cognitive, ambulatory, self-care, and independent living.

Differences between 2007 and 2008 included decreases in cognitive, ambulatory, and self-care difficulties, and slight increases in independent living difficulty. These differences are magnified among the population of individuals 65 years of age and older. The lower rates of disabilities may be a consequence of changes in the questionnaire and may not reflect actual change. The Census Bureau advises not to make comparisons between the 2007 and 2008 data.

The National Institutes of Health has raised concerns over deleting the employment disabilities questions. As a result, a new work disability question may be added in the future. The hope is that the new questions can be added without influencing comparability of the other questions.

The following resources for census products are available:

- General assistance with AFF—301-763-INFO (4636) or 800-923-8282;
- FTP Technical Support—301-763-2626;
- Journey to Work and Migration Statistics Branch—301-763-2454;
- FAQs—https://ask.census.gov/app/answers/list/search/1/kw/znewaff; and
- Online tutorials—http://factfinder2.census.gov/help/en/american_factfinder_help.htm#.

Environmental and Social Equity Issues

Frank Wen, *Southern California Association of Governments*
Gregory Gould, *Natural Resources Defense Council*
Adnan Sheikh, *Georgia Institute of Technology*
Sara Khoeini, *Georgia Institute of Technology*
Sundaram Vedala, *Georgia Institute of Technology*
Vetri Venthan Elango, *Georgia Institute of Technology*
Randall Guensler, *Georgia Institute of Technology*
Penelope Weinberger, *AASHTO, presiding*

This breakout session focused on the use of census data to analyze environmental and social equity issues. Frank Wen of the Southern California Association of Governments described the development and use of a model to assess the effects of transit-oriented developments (TODs) in Southern California. Gregory Gould of the National Resources Defense Council discussed a study examining the disparity in income and race of populations living near busy roadways. Adnan Sheikh of the Georgia Institute of Technology summarized a study examining the demographic characteristics of different user groups on I-85 in Atlanta.

USING PERFORMANCE INDICATORS TO MONITOR GROWTH VISION PROGRESS IN TRANSIT-ORIENTED COMMUNITIES

Frank Wen

Frank Wen discussed the development and use of a model examining the effects of TODs and transit-oriented communities (TOCs) on travel patterns, travel behavior, and greenhouse gas (GHG) emissions at the Southern California Association of Governments (SCAG). He provided an overview of recent legislation in California addressing GHG emissions, the study objectives, the model, and the preliminary results. He covered the following topics in his presentation:

- Visioning processes are used by regional planners to develop regional land use–transportation scenarios. Visioning is a highly community-oriented planning technique used to create regional land use and transportation goals. It involves gathering participants and stakeholders to form a consensus vision. It can be used to identify preferred types of development and growth patterns.

- SCAG launched a Compass Blueprint Visioning program in 2000. The four key principles guiding future development and growth were mobility, getting where we want to go; livability, creating positive communities; prosperity, maintaining the long-term health of the region; and sustainability, promoting the efficient use of natural resources. The demographic trends considered in the process were the aging of the population, households without children, single-person households, and housing location and preference.
- California Senate Bill (SB) 375 became law in 2009. It requires achieving specified GHG emissions reduction targets in 2020 and 2035. These reductions are to come from automobiles and light-duty trucks and through land use and related policies. It also implements a small portion of Assembly Bill (AB) 32, the Global Warming Solutions Act. SB 375 further requires integrating the regional transportation plan (RTP) with other regional plans and processes, including the sustainable communities strategy (SCS). It also requires a regional housing needs assessment, and it offers California Environmental Quality Act (CEQA) streamlining provisions.
- AB 32 requires California to reduce GHG emissions to 1990 levels by 2020. The goal of SB 375 is to reduce GHG emissions from automobiles and light-duty trucks. It requires the California Air Resources Board to identify a reduction target for planning horizon years 2020 and 2035 for the region. It also requires that an SCS be developed as part of the RTP.
- Land uses affect energy, vehicle miles traveled (VMT), health, air quality, and GHG emissions. Land use also affects water use, urban runoff, and water quality. Integrated land use planning for sustainable developments considers all of these factors.
- Implementing land use strategies can result in numerous benefits. Examples of land use strategies include the jobs–housing balance, higher-density/mixed-use developments, pedestrian friendly developments, and TODs. The preservation of resource areas and industrial and brownfield conversion represent other strategies. Potential benefits include more transit and walk/bike trips, fewer and shorter automobile trips, and lower VMT and congestion. Other benefits are more affordable housing, improved air quality, less runoff and improved water quality, decreased energy and water consumption, and better public health.
- SB 375 promotes a transit priority project (TPP) as an approach to reduce GHG emissions in the RTP. Examples of TPP requirements include high residential density, mixed use, and locating developments within one-half mile of major transit stops and in high-quality transit corridors. A TPP is generally considered as a TOD project. Various TPPs will be provided by CEQA streamlining.
- The SCAG Growth Vision program encourages TOD types of community development. More growth is expected in residential and commercial areas near

major transit stations and other identified transit centers. It is important for planners to monitor and assess the progress of the Vision program. It is also important to develop tools to assess the effects of transportation GHG emissions from TOC development.

- There are challenges to more concentrated development patterns. Examples of the risks include the risk of gentrification; the displacement of transit-dependent or core riders, including minorities and low-income residents; and incompatible land uses in buffer areas near major roads, freeways, and transit stations. Other possible challenges are high development and infrastructure costs, community resistance, and reducing congestion, VMT, and pollution at the regional level, but increasing them at the local level.
- Significant investments have been made in passenger rail services in the region during the past 20 years. Ridership has grown in response to these improvements. Additional improvements to existing systems and new services are planned for 2035.
- The following questions were examined in this analysis: Will TOD work in Southern California? How should the performance of TOD project areas be monitored? How should the transportation and travel effects from TOC developments for local and regional planning be assessed?
- The first study objective was to evaluate whether TOC areas are moving toward more desirable, sustainable, and livable communities and the likely effects of those changes. The analysis applied block-group data from the 2000 census and the 2005–2009 American Community Survey (ACS) and calculated a set of performance indicators for TOCs and the other areas. Trends between the two time periods were evaluated to demonstrate the effects of the TOC areas.
- The second objective was to understand the social and travel characteristics of the households in the TOC areas. By using a disaggregated data set procured from the 2009 National Household Travel Survey (NHTS), the interlinks between demographic, economic, and travel characteristics of the households that remain in the TOC areas and in the SCAG region were analyzed.
- The third study objective was to develop a model to predict the effects on travel and transportation from TOC developments. A disaggregated data set from the 2009 NHTS was used to develop a statistical model to link densities with various transportation and travel outcomes, such as automobile ownership, vehicle trips, and VMT.
- Performance indicators were developed for growth, economic, sustainability, equity, and transportation factors. Data were collected for 125 TOCs. The TOCs were defined as one-half-mile buffer zones of 125 commuter rail and urban rail stations. The communities were identified by census block groups and NHTS households.

- The analysis indicated that the growth rates of population and households in TOC areas were at least 10 percent higher than those in the entire SCAG region. The households and population in the TOC areas represent approximately 3 percent to 4 percent of the total region.
- The median household income in the TOC areas was lower than the regional average. The growth rates for the workers and jobs in the TOC areas were faster than those in the entire region, however. The analysis indicated that the type of worker occupation or employed industry may affect the economic indexes.
- No dominant difference in age distribution between the SCAG region and the TOC areas were found between the two time points. The share of Hispanic population is about 13 percent higher in the TOC areas than in the SCAG region.
- The TOC areas demonstrated higher shares of zero-vehicle households than did the SCAG region, although the share is declining in the TOC areas. The average number of vehicles per household increased by 13 percent in the TOC areas and increased by 8 percent in the region.
- An ANOVA was applied to test the mean difference between 2000 and 2005–2009 data. The Tukey approach was used to highlight the major differences. Significant changes were found in vehicle use, density, and education-related variables. The TOCs were examined by rail type—urban rail and commuter rail. TOCs with commuter rail had a significant change in the number of households, and TOCs with urban rail demonstrated significant changes in vehicle use, employment density, and education-related variables.
- No direct measure from the census or ACS exists to analyze transportation-related indicators. The Caltrans Transportation System Information supports the 2009 NHTS California add-on data. With approximately 6,700 households and 15,000 individual samples, the 2009 NHTS data set provides valuable and sufficient observations for analyzing demographic and travel characteristics in the SCAG region and the TOC areas. The NHTS households within quarter-mile, one-half-mile, and 1-mile buffer zones from the 125 TOC stations were examined.
- This analysis of the TOC household characteristics demonstrated that households in the TOC areas reflect smaller household size, higher percentages of single-person households and households without children, and more households with workers than in the SCAG region as a whole.
- The travel characteristics of households in TOCs reflected lower travel levels and lower drive-alone levels than in the region as a whole. Households in TOCs reflected higher shared nonmotorized and transit modes and a lower shared-vehicle mode.
- The share of Hispanic and non-Hispanic households in TOCs is evenly split at 50–50. Compared with non-Hispanic households, Hispanic households have larger household size and lower household income. Compared with the SCAG

region, both the Hispanic and non-Hispanic population in the TOCs showed a similar pattern of fewer total trips and lower VMT.

- Compared with the SCAG region, the TOC households had smaller numbers of vehicles. Approximately 20 percent of the TOC households did not own a car, which is double the percent of the SCAG region. Vehicles are less available in TOC households. The lower availability level may reflect a lower need for personal vehicles as a result of transit availability.
- The total commuting distance is shorter for TOC workers. Commuting VMT is also shorter for TOC workers than for workers in the SCAG region. Approximately half of the commuting distance was made by automobiles for TOC workers, compared with 86 percent in the SCAG region. A self-selection bias may exist in these differences. Living in higher-density neighborhoods (TOCs) induces a shorter commuting distance, while commuting time is almost the same.
- By using 2009 NHTS data, SCAG developed statistical models and a sustainability tool to analyze the effect of land use on VMT and travel. The three-tiered model includes a vehicle ownership model, a vehicle trip-making model, and a VMT model. The model was adjusted by adding a TOC dummy variable. The model results showed that the TOC dummy coefficient is significant. The performance of TOC areas in regard to VMT and other transportation indicators was tested by applying SCAG 2008 data and 2035 forecast data to the model.
- The preliminary model results indicate that the TOC areas will experience significant reductions in household vehicle ownership and VMT per household, but increased transit use. At the same time, the percentage of walking trips may be slightly reduced.
- The key focus of the study was determining whether the TOC areas are moving toward more desirable, sustainable, and livable communities. The analysis using census and ACS data demonstrated significant but small changes in household growth and land use density. The NHTS and econometric analyses showed that the TOC areas may contain some significant benefits for the SCAG region, including shorter commutes, less reliance on personal vehicles, and more transit use. Future analysis should examine whether demographic change in TOCs, such as gentrification, affects travel patterns and travel behavior.

INCOME, RACE, AND ODDS OF LIVING ALONG BUSY ROADWAYS: USING CENSUS DATA TO CONSIDER EQUITY AND ENVIRONMENTAL JUSTICE IN REGIONAL TRANSPORTATION PLANNING

Gregory Gould

Gregory Gould discussed a research project examining the disparity in income and race of populations living near busy roadways for each county in the United States. The study also analyzed the relationship between observed disparities and proximity to a road and the road's traffic volume. He described other literature and studies examining this topic, the project method, the results, and future research. He covered the following topics in his presentation:

- Regional air quality has greatly improved during the past several decades, but local air quality hot spots remain a significant health risk for many communities. Locations near roadways are one of the most significant air pollution hot spots. Available research indicates that concentrations of dangerous air pollutants are elevated along busy roadways and that residents living near these roadways are at greater risk of negative health outcomes, including respiratory diseases and cancer. Some studies suggest that low-income and minority populations have a higher risk of negative health outcomes from exposure to air pollutants because they tend to live adjacent to heavily traveled roadways.
- A GIS program was used to create several distance buffers along each busy roadway in the United States. Buffers were drawn at 100-meter (m) intervals extending out to 500 m around four categories of busy roads—50,000 to 100,000 annual average daily traffic (AADT), 100,000 to 150,000 AADT, 150,000 to 200,000 AADT, and greater than 200,000 AADT for every county in the United States. FHWA Highway Performance Monitoring System data were used to identify the traffic volume on each road segment. For each county, the median household income and the population proportion of each racial group were estimated for each distance buffer for roads in each traffic volume category, and the results were compared with the county's median income and racial profile. The demographic profiles of each distance–traffic buffer were estimated by overlaying block-level race and block-group-level income data from the 2000 census with the buffers.
- Other studies have found concentrations of vehicle-related air pollutants within 300 m of busy roadways. This analysis indicates that more than 14 million people, or more than 5 percent of the U.S. population, live within 300 m of roads carrying greater than 50,000 vehicle trips per day. Low-income and minority residents are also overrepresented near roadways. The results show that income and racial disparities generally increase closer to busier roads.
- The results vary widely across, and even within, different regions of the country, however. For example, on average in the United States, the proportion of Latinos and Hispanics in the population living within 100 m, 300 m, and 500 m of roads with traffic volumes between 150,000 and 200,000 AADT is 25 percent, 14 percent, and 11 percent greater, respectively, than the proportion of Latinos and Hispanics in each county. Similar patterns exist for other minority groups, low-income households, and at the other levels of traffic volume. Disparities for individual counties can be nonexistent or much greater.

- The two main findings from the study are that many people live very close to very busy roads and that individuals living near busy roads are more likely to belong to lower-income households or a racial minority. Currently, very few monitors in the national ambient air quality monitoring network are located near busy roads. Because the national ambient air quality monitoring network is one of the main triggers of regulatory action to reduce air pollution from motor vehicles in specific regions, the population living near major roadways may not be adequately protected by the current monitoring network and the provisions of the Clean Air Act. The income and racial disparities in the near-roadway population observed in this study raise environmental justice concerns.
- One potential solution is to develop a more robust air quality monitoring network that requires monitors along busy roadways in each region. It may be impractical to locate enough monitors to measure each unique highway population because emissions exposure along highways is very complex and variable. A more refined monitoring network combined with high-resolution air dispersion modeling could offer a feasible solution. For new developments, monitoring and modeling could be used to determine safe distances from highway facilities before buildings are constructed. To assess the population exposure and equity of different transportation alternatives, air dispersion modeling and the more robust monitoring data could also be used when a regional transportation plan is developed.

LANE DEMOGRAPHICS: HIGH-OCCUPANCY VEHICLE AND GENERAL PURPOSE USE ALONG I-85 CORRIDOR IN METRO ATLANTA, GEORGIA

Adnan Sheikh, Sara Khoeini, Sundaram Vedala, Vetri Venthan Elango, and Randall Guensler

Adnan Sheikh discussed a study examining the demographic characteristics of uses on the high-occupancy vehicle (HOV) lanes and general purpose freeway lanes on I-85 in the Atlanta, Georgia, metropolitan area. He described the I-85 project converting the HOV lanes to high-occupancy toll (HOT) lanes, study objectives, data collection activities, preliminary results of the before license plate survey, and future activities. He covered the following points in his presentation:

- The I-85 HOT lanes represent a joint project of the Georgia Department of Transportation (DOT) and the State Road and Tollway Authority. The $110 million project is funded through the U.S. DOT Congestion Reduction Demonstration (CRD) Program. The project converts 15.5 miles of existing HOV lanes on I-85 to HOT lanes. The vehicle-occupancy requirement was increased from two persons (2+) per vehicle to three or more persons (3+) per vehicle as part of the HOT conversion. The project also includes dynamic pricing with automated tolling. The HOT lanes began operations on October 1, 2011.

- An evaluation of HOV to HOT conversion is being conducted by the Georgia Institute of Technology (Georgia Tech). A national evaluation of the CRD projects is also being conducted. The evaluation conducted by Georgia Tech is examining the demographic characteristics of users of HOV, HOT, and general purpose freeway lanes. Changes in vehicle occupancy levels are also being monitored. Census and ACS data are being used for the demographics analysis.
- Data are also being collected from license plate surveys of vehicles using the HOV, HOT, and general purpose freeway lanes. High-definition cameras are capturing license plates in two lanes each at five sites along I-85. The data collection covers 3 days a week in the morning and afternoon peak periods. The license plate data are processed to obtain the census block group identification associated with each vehicle. Data collection waves occur every season for 2 years. Seven students are also collecting vehicle occupancy counts.
- The project is examining whether there are demographic differences between users of the HOV, HOT, and general purpose freeway lanes in the I-85 corridor. Demographic data from the 5-year ACS are linked to the census blocks identified from the license plate survey. The ACS data include median incomes and average household size by block group. The ACS findings will be compared with future survey results. Only the before data are currently available. Future survey results will help examine possible HOT equity concerns and the prevalence of carpools formed with family members using the HOV and HOT lanes.
- The initial data set included approximately 465,000 records for the fall 2010 wave. Many of these records did not link to valid license plates, however, as a result of difficulties in transcribing the license plates from the recordings. After incomplete plates were removed, 176,700 records remained. Of these, there were 85,000 unique license plates across all lanes, including 59,000 license plates in the HOV and the adjacent general purpose freeway lane. The license plates were geoprocessed for use in GIS and matched to census block groups. Data from the 2005–2009 five-year ACS were used to examine median incomes, average household size, and average vehicle ownership.
- Results from the initial license plate surveys indicate that HOV lane users have slightly lower incomes and fewer vehicles than those using the general purpose freeway lanes. The mean household income for HOV lane users was $66,757, compared with $69,107 for general purpose freeway lane users. The mean household size for HOV lane users was 3.0 persons per household, and the mean vehicle ownership was 1.99 vehicles per household. The corresponding figures for general purpose freeway lane users were 3.02 persons per household and 2.02 vehicles per household.

- Some limitations with the analysis include the age and precision of the data and possible processing errors resulting from incorrect and incomplete plates. Further, the addresses are based on vehicle registration, which may not reflect the residence of the actual driver or current residences.
- Ongoing and future research activities include adding updated ACS data and continuing to process the preimplementation data. The postimplementation analysis will also be initiated. Surveys will be conducted to provide more precise data, including mode changes. Related projects include travel time studies, Peach Pass saturation studies, and price sensitivity analyses.

Demographic and Travel Forecasting

Behruz Paschai, *North Central Texas Council of Governments*
Kathleen Yu, *North Central Texas Council of Governments*
Arash Mirzaei, *North Central Texas Council of Governments*
Mara Kaminowitz, *Baltimore Metropolitan Council*
Wu Sun, *San Diego Association of Governments*
Ziyang Ouyang, *San Diego Association of Governments*
Eddie Janowicz, *San Diego Association of Governments*
Nick Cohn, *TomTom*
Thomas J. Kane, *Thomas J. Kane Consulting, presiding*

Speakers in this breakout session discussed the use of census data in demographic and travel forecasting. Behruz Paschai of the North Central Texas Council of Governments (NCTCOG) described the use of census data in the Dallas–Fort Worth, Texas, area. Mara Kaminowitz of the Baltimore Metropolitan Council (BMC) summarized the use of the new ACS small-area estimates in the Baltimore, Maryland, area. Wu Sun discussed modeling activities under way at the San Diego Association of Governments (SANDAG) in California. Nick Cohn of TomTom highlighted the use of traffic data from commercial vendors in planning and modeling.

USE OF ACS AND DECENNIAL CENSUS DATA PRODUCTS IN DEMOGRAPHIC FORECASTING PROCESS AT NCTCOG

Behruz Paschai, Kathleen Yu, and Arash Mirzaei

Behruz Paschai discussed the use of data from the 2010 census and the American Community Survey (ACS) in the demographic forecasting process at NCTCOG, which covers the Dallas–Fort Worth, Texas, metropolitan area. He covered the following topics in his presentation:

- The NCTCOG modeling area includes 12 counties and covers approximately 10,000 square miles. There are 230 member cities in NCTCOG. The population of the region has grown from approximately 5.1 million in 2000 to approximately 6.4 million in 2010, an increase of 24 percent.
- Data from the 2000 and 2010 census and the ACS were used in the development of a demographic forecasting model at NCTCOG in 2010. The gravity-based demographic forecasting model was recently recalibrated and validated according

to the employment and household data available for the lag and base years of 2000 and 2005, respectively. The model results were later validated for 2010 by using the 2010 census and the first 5-year ACS released in December 2010.

- The 2000 household data set was created on the basis of the decennial census data and was modified according to local input. However, the 2000 employment and 2005 household and employment data sets were constructed on the basis of an in-house development monitoring program. This product tracks residential and commercial developments in the region and the inputs provided by local governments. The local generation of the 2005 data was in 2007 before the 1-year ACS became available.
- The NCTCOG demographic forecasting model uses a software platform composed of the gravity land use model (G-LUM) and disaggregate residential allocation model (DRAM) and the employment allocation model (EMPAL). The G-LUM was developed by faculty at the University of Texas at Austin, and its development was funded by the Texas Department of Transportation (DOT). It is based on available DRAM/EMPAL documentation and provides a clarification of DRAM/EMPAL assumptions. G-LUM is open source and allows in-house development capabilities. It is written in MATLAB2 script and has been applied in the Austin and Waco, Texas, regions, as well as the NCTCOG region. The DRAM/EMPAL was developed by S.H. Putman Associates, Inc., and has been used by NCTCOG to develop demographic forecasts for a number of years. It is a black-box application with a proprietary source code. There are limitations on the number of forecast districts in the original application. The developer's involvement is required in the modeling and calibration process and at some level for reviewing results.
- A number of local issues are related to the use of census data. These issues include some local entities disagreeing with the census and the ACS data products, the development of independent population forecasts by local entities, and the general disagreement among demographers with the use of ACS data products, despite the lack of better data sources. There is also pressure from some local entities to review and adjust the small-area forecasts on the basis of their input and to reach a certain percentage of growth in the region. Further, some local entities disagree with the NCTCOG forecasts. Additional issues include the possible effects of NCTCOG's demographic forecasts on potential developers, the lack of sensitivity of NCTCOG's demographic forecasts to possible incentive-based growth strategies implemented by local entities, and unknown and dynamic incentive programs.
- The ACS was also used in disaggregating 242 demographic forecast districts into 5,303 smaller zones. The disaggregation process used a suitability index for each zone for assigning a relative weight. The 5-year ACS was used to evaluate the performance of the disaggregation process at the block-group level.

- The modeling process begins with the lag year and the base year households and employment, lag year land use, and lag year accessibility. The data feed into the household and employment calibration, which is applicable for 5-year-increment forecasts. The next step is the household and employment model validation. If the validation passes, the household and employment forecasts are generated in 5-year increments. If the validation fails, adjustments are made to the model and the calibration and validation steps are repeated.
- A data set from the Perryman Group was accepted as the source for regional household population and employment control totals in all calibration and forecast years. This data set is internally consistent and regularly updated. Further, the household population compared well against ACS and census data. The employment forecast totals were used without adjustments. The household populations by employment ratios are also reasonable. The household sizes were developed for converting household populations to households.
- One issue encountered was the difference between the average household size reported in the 2000 census for North Central Texas and the household size reported in the 2005–2007 three-year ACS data. The average household size according to data from the 2000 census was 2.70, compared with 2.81 for the 2005–2007 ACS. Differences in the household size growth trend in North Texas with trends in the rest of the country were also noted. Household size trends were examined by county. The household size assumptions included that by 2050, the household size in the rural counties will become similar to the average household size of suburban counties in 2005, which is 2.8.

ADVENTURES IN ACS: USING THE 2005–2009 ACS SMALL-AREA DATA IN REGIONAL TRANSPORTATION PLANNING

Mara Kaminowitz

Mara Kaminowitz discussed the new ACS small-area estimates and their use by BMC. She described the various projects, future projects, and some of the challenges associated with working with data that may have large margins of error. She covered the following topics in her presentation:

- BMC is the host agency for the Baltimore Regional Transportation Board, which is the metropolitan planning organization (MPO) for the Baltimore, Maryland, region. The MPO region includes the city of Baltimore and the five surrounding counties. The mission of BMC is to identify regional interests and develop programs that will improve the quality of life in the area, as well as further the region’s economic interests. Transportation programs are a major focus of BMC, along with travel forecasting, demographic analysis, GIS, environmental programs, and other activities.

- The new ACS data were used to develop the preliminary 2010 traffic analysis zone (TAZ) delineations. Although the delineations ultimately had to be based on 2010 census population data, there really was not enough time to conduct a thorough job between receiving the 2010 data and the time the TAZ boundaries were due. The block-group-level population data were used to obtain an idea of the current population of each TAZ to initiate changes.
- One of BMC's largest nontransportation roles is to serve as a source of regional data. These activities are conducted both by BMC and by the Pratt Library's Regional Information Center. The BMC is the go-to place for maps and for raw data derived from census products. Requests are not strictly related to transportation. Most regional information requests are for conventional data related to population change, income, commuting patterns, and related activities. A wealth of available data are posted on the BMC website.
- BMC also provides interactive community profiles. These profiles are based on the Google Maps interface, allowing users to zoom in to their community and obtain a variety of information. Implementation of an ArcGIS Server during the next few months is planned to allow for the creation of more sophisticated maps and expand the topics presented online. The geography used in the map is called regional planning districts, which were created by BMC on the basis of census tracts. This approach allows for the use of census and ACS data for small areas while maintaining names and approximate community boundaries that are familiar to local users.
- The ACS data were used in transportation planning to examine limited English proficiency (LEP) access to public transportation. This analysis was used internally, and maps were provided to the Maryland Transit Administration. Language ability was examined and cross-referenced with the languages spoken. This information is used to ensure that non-English signage in public transportation facilities is in the correct language and in the communities that need it. A public use microdata sample (PUMS) analysis that cross-references the ability to speak English with public transit use can also be provided. In the future, ACS data will also be used to examine whether future road and transit plans meet the needs of emerging immigrant communities.
- The ACS data are used in equity planning and environmental justice analysis as well. Minority and income data are included in project reports to allow the local jurisdictions and other interested parties to examine the effects of projects and programs on vulnerable communities.
- One of the main challenges with using ACS data is dealing with margins of error and perceived accuracy. The ACS has published a lot of information on the calculation methods and how to approach the data. It is hard to determine how much of a margin of error is too large, however. The BMC region varies widely in population and population characteristics. As a result, any small-area ACS tables

used will have high margins of error. Data tables cannot be discounted solely on the presence of high margins of error or that data would not be used at all.

- As expected, the higher the population in an area, the lower the margins of error. To examine this issue, data from the 2005–2009 ACS estimates, the 2000 census, and the 2010 census were compared for 34 block groups and tracts through the region that were identical in 2000 and 2010. On the basis of a recommendation from the ACS, population characteristics, rather than population counts, were examined. Three data sets for white alone and black alone were mapped on a scatter plot. The relationship of the ACS data to the 2000 and 2010 values was examined.
- Assumptions were made in examining the data. If there was a strong linear trend between 2000 and 2010—either strong growth or decline—and the ACS value and its error bar fell in the middle, the data were assumed to be similar. If the data were not linear, but the 2000 and 2010 values fell somewhere on the error bar, it was assumed that the ACS value must fall within the same area. When neither of these conditions was met, the data were considered questionable.
- Depending on the data examined, 30 percent to 35 percent of the tracts or block groups in the sample were questionable. Although ACS values might be accurate, the differences make them uncertain. For 6 percent to 12 percent of the sample, the divergence calls its accuracy into question. For example, one census tract grew by 200 percent and then lost 90 percent of those gains. There did not seem to be an improvement in accuracy going from block groups to tracts. The coefficient of variation, which is how big the margin of error is in relation to the value, also did not seem to predict which block groups and tracts were questionable. For those that were questionable, the relative relationship was largely preserved. If the value was not quite within the expected range, the data were still close enough to get a sense of whether the value was higher or lower than its neighbor. Although this examination was a simple exploratory process, it does suggest a number of questions. It would be of benefit if the Census Bureau and the ACS user community evaluated this issue further—the reliability of the data values and whether the stated margins of error are giving users an accurate picture of how trustworthy the data are.
- A number of approaches are used to address these concerns. One approach is to aggregate the data to tracts or regional planning districts. Disaggregating topics too much is discouraged. Another approach is to generalize when appropriate. For example, rather than provide the actual number of countries of origin of immigrants in an area of Baltimore that did not have a lot of immigrants, the top three countries for each neighborhood were listed. Portraying the data in relative terms from lesser to greater values instead of specific numbers is another approach. This technique is particularly useful when dealing with statistical outliers from derived data such as percentages or density.

- Educating others on the nature of the data and the limitations is important. For example, explaining the new 5-year data estimates in print and noting the margins of error are important. It is important to remember that professionals using the data are responsible for explaining the limitations of data and not using data that is felt to be inaccurate or wrong. It is also important to remember that the data provided, even with numerous caveats and disclaimers, will be used by others and attributed to you and your organization. Data take on a life of their own—so using only data you feel are reliable is important.
- Other challenges with the use of data from the Census Bureau and other national sources include explaining the 5-year estimate and placing it in the context of other 1-year to 10-year reports. Explaining why some data are outdated and other data are not is another challenge. In addition, no matter what data are included in the ACS or any other sources, requests are often made for data not available from any sources. Comparing change over small areas over time is also a problem. Major changes at the block level have occurred in the BMC region and comparing the 2000 ACS to the 2010 data at the tract level or smaller is virtually impossible.
- Even with all these challenges and limitations, the ACS and census provide good sources of data for transportation planning. Initiatives under way and planned at BMC include equity planning projects, including examining transit access for non-English speakers, improving public transportation access to low-income individuals and the elderly, and examining transit access for the disabled.
- Other future projects using small-area data include expanding the community profiles with an enhanced web component; examining commuting patterns and access to public transportation, particularly for transit-oriented developments; and validating the travel demand model. Another project is examining the possibility of using small-area ACS data for population forecasting.

USING CENSUS DATA IN DEVELOPMENT OF AN ACTIVITY-BASED MODEL: THE SAN DIEGO EXPERIENCE

Wu Sun, Ziyang Ouyang, and Eddie Janowicz

Wu Sun discussed the development of an activity-based model (ABM) at SANDAG. He described the San Diego ABM and the use of census data and data from other sources in the development of SANDAG ABM. He covered the following topics in his presentation:

- The SANDAG modeling group maintains a suite of models, including transportation models and land use models. SANDAG has an existing traditional four-step travel forecasting model and is developing an ABM. There are differences between ABMs and traditional four-step models. One difference is aggregate versus disaggregate approaches, with ABMs simulating each household and person's travel during a 24-hour period. A second difference is behavioral versus nonbehavioral, with ABMs putting a human face on a model. Each person's travel decision is based on an individual's sociodemographic characteristics. ABMs represent advanced method with much finer details related to the social demographic, the spatial dimension, and the temporal dimension. ABMs can be used to assess broad policy applications, including congestion pricing and toll policies, transportation demand management policies, equity analyses, and HOV policies. ABMs can also be used to examine transit fare policies, environmental policies, and border cross-cutting policies.

- The key SANDAG ABM dimensions include the spatial dimension of 22,000 master geographic reference areas, the temporal dimension of 30 minutes, and the sociodemographic dimensions of household attributes and person attributes. Other segmentations include 10 activity types, 25 modes, eight person types, five household income types, and seven occupation types.
- The model development steps include model estimation, model calibration, model validation, and model application. A model is a mathematical representation of travel behaviors. The model estimation step identifies the parameters of the mathematical formulas from the survey data. In the model calibration step, the estimated model is compared against the survey and other relevant data sources, including census data sources, to ensure that the parameters are correct. In the model validation step, the estimated model results are compared with road counts. In the model application step, the finalized model is applied for policy analysis. Census and ACS data sources were used in the San Diego ABM in the model estimation and the model calibrations steps.
- A variety of local, state, and national data sources were used in the development of the San Diego ABM. These sources include the household travel behavior survey, the transit onboard survey, network impedances, built environment data, and local land use data. Other sources include the parking inventory and survey, special market surveys, highway and transit counts, local sociodemographic data, and census and ACS data. The local sociodemographic data include estimates of household size, income, gender, race, and other variables. The census micro public use samples and the census or ACS summary files were also used.
- Census data are important to the ABM primarily because it adopts a disaggregate modeling approach that uses detailed sociodemographic variables. These variables include income, household size, household composition, age, gender, occupation, education level, and other factors. The census or the ACS is the source for the data. Census data are used in the population synthesizer, model estimation step, and model calibration step.
- The population synthesizer (PopSyn) contains a list of households and population in the modeled region. For example, the Model Year 2011 for the San

Diego region includes 1.2 million households and 3 million people. The PopSyn is needed because the ABM models each household and person's travel behavior.

- Two types of census data are needed in the PopSyn. First, PUMS data, which contain detailed sociodemographic information, are needed. The PUMS data are expanded to represent the region as a whole, providing the synthesized sociodemographic characteristics for the region. Second, the census and ACS summary files are used in the validation procedure to check the synthesized population. Control variables include household size, household income, presence of children, group quarters, household dwelling units, age, gender, and race.
- In the model estimation step, population data, such as population density or population distribution by age, gender, and race, are normally estimated locally. However, local estimates are based on census data. For example, the preschooler population is included in the school location choice model, population density is included in the mode choice models, employment by industry is included in the work location choice model, and employment density is included in the mode choice models.
- In the model calibration step, the model estimates are compared with survey and other relevant data sources. The San Diego ABM used Census Transportation Planning Products (CTPP) data for work trips. The automobile ownership model uses observed data from the 2000 CTPP. The work location choice model uses observed data from the 2000 CTPP worker flow.
- The data from the various sources use different geographies. Examples include census geographies, SANDAG transportation modeling geographies, and SANDAG land use modeling geographies. The various geographies cause some data consistency issues, primarily because the SANDAG and census geography do not always match.
- Other data sources could potentially help in building the San Diego ABM. These data sources include the National Household Travel Survey, the California Household Travel Survey, longitudinal employer–household dynamics, and continuous SANDAG surveys. SANDAG plans to incorporate these data into future modeling efforts.

GPS PROBE DATA FOR TRANSPORTATION PLANNING

Nick Cohn

Nick Cohn discussed the use of GPS probe data from commercial vendors, such as TomTom, in transportation planning and travel forecasting. He described the data collection methods, products, and services provided by TomTom, and possible transportation planning applications. He covered the following topics in his presentation:

- GPS probe data provide a new method for measuring speeds, travel times, and road performance. Probe devices in vehicles, which include cellular phones and more common GPS devices, provide the data. Vehicles with these devices travel everywhere on the road network. They are not limited to roadside infrastructure to communicate data. Traffic speeds can be measured everywhere probe vehicles are traveling, providing a number of major advantages when compared with traditional data collection methods.
- The approach does not require any installation or maintenance of roadside equipment, saving costs and avoiding any interruptions to traffic flow. GPS probe data provide information on the entire road network, not just sections where measurement infrastructure has been installed. It provides accurate data on any complex trajectory and is most dense where there is congestion. Further, it is environmentally friendly and can replace a large amount of fieldwork with desk work.
- Historical traffic information is collected by millions of navigation device users who share their usage statistics. These users voluntarily and anonymously report data for each of their journeys. As a result, TomTom now has a database containing more than 4 trillion measurements, collected since 2007 and growing continually.
- TomTom's historical traffic database has a number of unique characteristics. The database contains a very large sample size, filled by millions of different drivers. Because measurements are continually being made, the information is not vulnerable to bad weather conditions on one or two surveyed days. The database is filled primarily with information from passenger cars as opposed to delivery fleets or trucks. The historical GPS speed measurements are matched to TomTom's high-quality road map by individual road segment. The road segments are quite detailed, generally ranging in length from approximately 1 meter (1 yard) to 1.6 kilometers (1 mile). Access to this map-matched, historical traffic speed data is available via a web portal.
- Historical traffic products include speed profiles, custom travel times and custom area analysis, and custom probe counts. Speed profiles can provide better route-time planning with proven long-term knowledge. Examples of benefits from the use of speed profiles for estimated time of arrival are improved fastest route calculation and time-specific forecasting of travel times on road links. Examples using these data include a personal navigation device and smartphone navigation, Internet mobility portals, and logistics scheduling for standard and just-in-time delivery planning.
- Custom travel times provide detailed analyses of road or route traffic performance. Benefits include a detailed understanding of route travel times by time of day and an ability to study the effects of network changes in before and after studies. Examples of uses include travel time reliability studies, traffic engi-

neering analyses, background information for major infrastructure changes, and model validation.

- Custom probe counts and area analyses focus on historical travel times and speeds over a broad network of roads. Possible benefits include detailed travel times by time of day and the extraction of wide networks to shapefiles. Custom area analysis results may be used with performance measures, traffic engineering, and model validation.
- The traffic stats Internet portal provides online access to historical traffic information. Features of the portal include access to TomTom historical traffic products from any Internet-enabled computer anywhere and anytime, with reports available within 24 hours. Data can be downloaded for use in other applications. Custom travel times and custom area analysis are currently offered.
- Another planning applicant is examining origins and destinations on the basis of GPS navigation data. A large database is available for this use. Trips can be distilled on the basis of a clear set of criteria. Routes used for trips are known. The data are completely anonymous. There are challenges with this approach, however. These challenges include complete anonymity, no data on trip purpose, and no household characteristics. There are data on some known biases for consumer versus fleet use patterns.
- Research applications examined visitors to the Burlington Mall in Burlington, Massachussetts, by household income of origin census block. Steps in the analysis included attaching individual GPS measurements to block polygons at the beginning and ending trip times, identifying weekday and weekend travel, and identifying stop time and dwell time. The census block statistics were attached. Next steps in the project include identifying all the systematic biases and creating correction and expansion methods. Another step is validating the use of the census data. Additional steps are identifying the most valuable use cases and defining standardized output forms.

Examples of current projects under way in Europe include examining outdoor advertising in the Netherlands and analyzing the use of a toll road in France. TomTom data are being used in the United States in a modeling study in California and in modeling and planning efforts in Virginia.

- The current challenge is to make more aspects of the data usable. The potential for using the GPS origin and destination data for transportation planning is great, as doing so should offer a lower cost and higher sample size than large origin and destination field surveys. There are certainly some biases because of the source of the data (navigation users), and there are limitations because the data are anonymous. Research examining methods to use census data to create origin–destination tables that include some socioeconomic characteristics is under way.

Travel Modes

Transit, Bike, Walk

Caroline Ferris, *TranSystems*
Henning Eichler, *Southern California Regional Rail Authority*
Robert E. Bush, *HDR Engineering*
Ho-Ling Hwang, *Oak Ridge National Laboratory*
Ken Cervenka, *Federal Transit Administration, presiding*

This breakout session examined the use of census data in modeling transit, bike, and walk trips. Caroline Ferris of TranSystems discussed using Local Employment Dynamics (LED) data with FTA programs focusing on job access. Henning Eichler of the Southern California Regional Rail Authority described the use of Longitudinal Employer–Household Dynamics (LEHD) to estimate transit market sheds. Robert Bush of HDR Engineering summarized the use of a transit prosperity model in North Carolina. Ho-Ling Hwang of the Oak Ridge National Laboratory described using census data in a study of nonmotorized travel markets.

USE OF LED DATA IN ESTIMATING JOBS ACCESSED THROUGH FEDERAL JOB ACCESS AND REVERSE COMMUTE PROGRAM

Caroline Ferris

Caroline Ferris discussed the development of a process to use LED data to access the FTA Job Access and Reverse Commute (JARC) program. She described the LED data, the JARC program, JARC performance measures, and the method. She covered the following topics in her presentation:

- The LED uses modern statistical and computing techniques to combine federal and state administrative data on employers and employees with core Census Bureau censuses and surveys. OnTheMap is a web-based mapping and reporting application that shows where workers are employed and where they live. It also provides companion reports on age, earnings, industry distributions, as well as information on race, ethnicity, and educational attainment.
- FTA established the JARC grant program in 1999. The program has two major goals. The first is to help low-income individuals overcome the transportation barriers to find and keep jobs. The second is to address challenges of accessing suburban jobs. JARC grantees provide FTA with annual updates on service delivery and program performance.

- The JARC performance measures focus on one-way trips and jobs accessed. One-way trips are the standard transit statistic that is easy for most organizations to collect and report. One-way trips do not relate directly to the project goals of helping people get to work, however. Jobs accessed are difficult to measure, especially for transit organizations, and are not always directly related to service. It is an intuitive measure that relates directly to the JARC goals of helping people get to work, however.
- The challenge of the project was to find a way for JARC grantees to collect and report data about jobs accessed. The project goal was to develop a relationship between jobs made accessible and one or more standard transit statistics. The solution was to develop a jobs density factor to estimate jobs accessed.
- The project method included a number of steps. First, a sample of JARC services was selected. Second, route maps were obtained and a geographic information system (GIS) was used to digitize the routes and create buffer areas. The buffer areas were uploaded to OnTheMap and the desired reports in buffer areas were created. Jobs in buffer areas were calculated by North American Industry Classification System codes. The results were adjusted for low-wage jobs and for jobs likely to be reached.
- The process was highlighted by using an example. A random sample of routes from the full list of services was generated first. The route maps or written directions were located for each route in the sample. The routes were digitized in a GIS program, and a 1,320-foot buffer was created around each route. The buffer file was uploaded to OnTheMap. The OnTheMap results were displayed graphically. The jobs reported created were identified and exported to Excel. The file was adjusted for low-wage jobs, and jobs likely to be reached were determined.
- Results of the FY 2009 analysis identified approximately 52 million jobs. Approximately half of those jobs were within the buffer of fixed route transit services. Approximately 41 percent were within the buffer of demand-responsive services, 5 percent were within the buffer of shuttle services, and 4 percent were within the buffer of flex services. Approximately 48 percent of the identified low-wage jobs were within the buffer of fixed-route services, and 43 percent were within the buffer of demand- responsive services. A total of 75 percent of the identified likely to be reached jobs were within the buffer of fixed-route services, compared with 13 percent for demand-responsive services, 9 percent for flex services, and 3 percent for shuttle services.
- A few factors need to be considered in using the method. The jobs density factor was designed to work on a national aggregate level. The data cannot be unrolled for local estimates. Unique local conditions may not be reflected. There are also gaps in LED during the analysis time frame.

ANALYZING THE MARKET SHARE OF COMMUTER RAIL STATIONS BY USING LEHD DATA

Henning Eichler

Henning Eichler discussed the use of LEHD data to examine the market share for Metrolink travel sheds. He described the Metrolink system, the identification of Metrolink travel sheds, the method applying LEHD data to analyze competitive commute trips, and analyzing the results at the corridor level. He covered the following topics in his presentation:

- The project focused on answering three questions: What is the size of Metrolink's commute market? What share of the market does Metrolink capture? What is the latent demand, and what factors influence it?
- Metrolink is southern California's regional commuter rail system. It includes seven routes serving six counties, with 55 stations, 512 route miles, and 164 weekday trains. Metrolink averages 42,000 weekday trips.
- The Metrolink travel shed represents a competitive market. The station catchment area method includes analyzing travel data for current riders and defining station catchment areas according to trip origins–destinations. An origin–destination survey of current riders was conducted, and a GIS was used to analyze home and work locations for Metrolink riders. The sample size was large enough to provide enough observations for each station. The catchment area covers no less than 90 percent of trip origins or destinations for a particular station. The catchment areas are based on census tract or travel analysis zone geography.
- A total of 55 home catchment areas were identified for Metrolink stations. The average size of the catchment area is 73 square miles. The station access mode was 87 percent who drove alone. The average travel distance from home was 6 miles, and the median was 3 miles. A total of 55 work catchment areas were identified. The average size of the work catchment areas was 32 square miles. The station access mode was 54 percent transit and 23 percent walk/bike. The average travel distance to work was 5 miles, and the median was 1.5 miles.
- The LEHD provides the opportunity to estimate latent demand. OnTheMap and downloaded data were used in the analysis. The data include home and work census block codes and selected age, income, and industry classifications. The project linked LEHD data with station catchment areas. Approximately 6 million records of primary jobs were in
- Metrolink travel sheds. Each record includes work and residence census block code. Census blocks were aggregated by station catchment areas.
- Competitive commute trips for jobs within the Metrolink travel shed were examined. Approximately 25 percent of the trips have a residence and a work location within a travel shed and require no Metrolink transfer. Approximately 10

percent of the trips have a residence and a work location within a travel shed but require a Metrolink transfer. Uncompetitive commute trips were also identified. Approximately 4 percent have spatially uncompetitive origin–destination patterns, and 13 percent have a residence and work location within the same station catchment area and therefore do not need Metrolink services. Approximately 48 percent of workers commute from outside the Metrolink travel shed.

- Trip tables for all origin–destination pairs were developed for the corridor analysis by using ridership and LEHD commute trip data. The market share was calculated for each origin–destination pair.
- Metrolink's market share is 1.1 percent systemwide. It ranges from 0.6 percent to 2.7 percent by corridor. Factors influencing market share include the central business district share, cost, and travel time. Other factors are service availability, station access, and parking availability at stations.
- The level of service (LOS) analysis is a tool to evaluate station characteristics from the customer perspective. This approach is described in the TCRP *Transit Capacity and Quality of Service Manual*. The LOS ratings were coded numerically, weighted by LEHD data, and averaged for each station to help identify service quality issues.

USING CENSUS DATA TO IDENTIFY AREAS OF HIGH-TRANSIT PROPENSITY

Robert E. Bush

Robert Bush discussed a study developing and applying a transit propensity model. He discussed the development of the model, use of census data in the model, and application of the model in Wake County, North Carolina. He covered the following topics in his presentation:

- *TCRP Report 28: Transit Markets of the Future: The Challenge of Change* was the principal source for the method used in the analysis. The characteristics of transit riders were examined and analyzed at the county level. The factors of highest transit propensity were identified. Different operating environments were also considered.
- The demographic factors examined include zero vehicle housing units, employment disabilities, and mobility limitations. Other demographic factors examined were minorities, recent immigrants, low-income households, and females. All of these characteristics were found to be relevant when income is controlled for.
- The analysis focused on journey-to-work data, as little data are available on nonwork trips. The analysis was based on 1990 census data. Recent trends indicate some changes from these data. The analysis identified traditional transit

markets. There is little to no information on higher-income commuters. The analysis did not control for service levels and considered only the home end. Further, density factors were not considered.

- Other factors that were considered, but not included in the analysis, were younger and older workers, education levels, renters, and nonlicensed drivers. Data are not available at the block-group level for these factors. Education appears to be significant in larger markets with higher incomes equaling higher ridership. But these trends are likely related to service levels, particularly with rail systems. Renters and nonlicensed drivers have been found to be significant in some studies, but not controlled for income.
- The propensity model includes seven weighted demographic factors—zero-vehicle housing units (217), mobility limitations (174), work disability (66), minorities, (65), recent immigrants (39), low-income households (38), and females (31). The weights are based on relative propensity. An eighth factor for population density was added with a weight of 370. A composite score for each block group was calculated by multiplying indexes by weights. The results were grouped into five categories.
- The process was applied in Wake County, North Carolina, which is part of the Triangle Region. The region includes three counties—Wake, Durham, and Orange. It also includes four principal cities, Raleigh, Cary, Durham, and Chapel Hill, and three universities, North Carolina State, University of North Carolina–Chapel Hill, and Duke University. Transit agencies in the area include Capital Area Transit, Triangle Transit Authority, Cary Transit, and North Carolina State's Wolfline.
- The analysis examined the population density, the percentage of housing units with zero automobiles, and the percentage of the population with mobility limitations. Other characteristics examined included the percentage of the population with work disability, percentage of the nonwhite population, percentage of recent immigrants, percentage of households with annual incomes of less than $20,000, and percentage of females.
- The transit propensity was calculated and mapped by using GIS for Wake County and the city of Raleigh. The transit service coverage for Raleigh was mapped with GIS and compared with the transit propensity. The analysis indicated that most of the areas of very high transit propensity had transit service. Areas with high transit service propensity but no transit service were identified for further analysis.
- Two files from the 1990 census data were used. Three factors—density, minority population, and females—were used from Summary File 1 (SF 1). Five factors—mobility limitations, work disability, recent immigrants, low-income households, and zero-vehicle housing units—were used from SF 3. One issue is

the way the American Community Survey (ACS) will influence the use of these data in the future.

- Enhancements to the process are being considered. One enhancement is to adjust for other transit services, including rail, commuter bus, and demand-responsive service. Other enhancements include adding recent service level changes and examining exogenous factors, such as gasoline prices.

SKETCH PLANNING TOOL FOR NONMOTORIZED TRAVEL

Ho-Ling Hwang

Ho-Ling Hwang discussed a project conducted for FHWA's Office of Planning characterizing the market potential for nonmotorized travel. She described the study objectives, data sources, and preliminary analysis. She covered the following topics in her presentation:

- The study objectives are to use National Household Travel Survey (NHTS) data and integrate the data with ACS and other data sources to identify factors that make communities walkable and bikeable and to understand why walking and biking are integrated into some communities and not others. A follow-up optional element is to develop a GIS-based visualization planning tool with factors identified previously with a selected neighborhood for the prototype case study.
- The project uses census data directly and indirectly. Data from the 2009 NHTS national data set and the NHTS add-on data are used. Demographic data and journey-to-work data from the ACS are used. The census master address file and the Topologically Integrated Geographic Encoding and Referencing System (TIGER) are also used.
- The NHTS characteristics on walk and bike trips were examined and used. The results of this analysis indicate that households with children, households with higher incomes, households with workers, and persons with higher education levels are more likely to take walk/bike trips.
- The NHTS trip distribution by age group indicates that bike trips are made largely by the school age groups, specifically the 5- to 10-year-old age group and the 11- to 15-year-old age group. Children in the 11– to 15-year-old age group made more bike and walk trips than younger children. The youngest groups, 5 years of age to 20 years of age, took more than 46 percent of the total bike trips and about 25 percent of the total walk trips; this group took about 21 percent of the total trips when all modes are considered. According to the ACS walk and bike commuting trips data, walking accounted for more than 20 percent of the nonpersonal vehicle mode share category in commute trips nationally, whereas biking is only about 4 percent. Block-group data on walk-to-work and bike-to-work were mapped as part of the study.

- Census demographic data and TIGER boundary files were used in the study. For example, maps highlighting population density for all persons and persons ages 5 to 17 by block group were produced. Data derived with the census TIGER shapefiles allow for calculating the roadway density and intersection density in a given geographic boundary. Roadway density in miles or roadway per square mile of the block group can be displayed graphically. Intersection density, defined as the number of intersections per square mile of the block group, can also be displayed graphically.
- Walkability is being measured at the block-group level. Walk score measures provided by WalkScore.com are being used. The scores are based on distances to amenities, such as grocery stores, restaurants, shopping, coffee shops, banks, parks, schools, bookstores, and entertainment. The mode share of walking-to-work—the percent of commute trips made by walking—is based on the ACS 2005–2009 five-year data on workers 16 years of age and older.
- The dependent variables being considered in the modeling efforts include trip rate, trip length, and mode share. Trip rate variables include the number of walk trips per capita in a block group; the number of walk trips, including egress/access to transit, per capita in the block group; and the number of bike trips per capita in a census tract. Trip length variables are bike miles per capita in a census tract and walk miles per capita in a block group. The mode share variable is non-home-based walk trips.
- As described, census data play a major role in this research. The modeling effort is continuing. The workplace employment data are currently being acquired for the study areas. Investigating the potential entropy variables to measure the land use mix is also continuing. The final report for the Phase I study will be completed by the end of January 2012.

Data Synthesis and Evolving Tools

Dmitry Messen, *Houston–Galveston Area Council*
Rolf Moeckel, *Parsons Brinckerhoff*
Ram M. Pendyala, *Arizona State University*
Keith L. Killough, *Arizona Department of Transportation, presiding*

Speakers in this breakout session described the use of new data analysis tools and techniques. Dmitry Messen of the Houston–Galveston Area Council (H-GAC) in Texas discussed the application of a socioeconomic model using public use microdata sample (PUMS) data in the Houston area. Rolf Moeckel of Parsons Brinckerhoff reported on a new integrated simulation land use model. Ram Pendyala of Arizona State University discussed the use of microsimulation models, including the use of synthetic populations.

USE OF ACS PUMS DATA IN DEMOGRAPHIC AND ECONOMIC FORECASTING

Dmitry Messen

Dmitry Messen discussed the development of a socioeconomic forecasting model at H-GAC. He described the H-GAC model, household evolution model, economic model, and desirable enhancements to the American Community Survey (ACS) PUMS data. He covered the following topics in his presentation:

- H-GAC is using a two-tiered framework for the socioeconomic forecasting model. A macroapproach is used on the regional control totals and a microapproach is used on the local land use. The tiers are linked, but there is not a feedback loop. Fully disaggregate microsimulation is used. Key model attributes include path dependency, consistency, realism, transparency, modularity, and expandability.
- The macrolevel or regional control totals focus on people and jobs over time. An aspatial process is used, with intraregional location having no influence on the dynamics. The microlevel or parcel-level land use focuses on change with parcels over time. A spatial process is used. Intraregional location has everything to do with the real estate markets.
- The macromodel population and employment dynamics create demand estimates for new buildings, including residential housing and workplaces, by simulating a list of movers. These movers include new residents and individuals

relocating within the region. The micromodel parcel dynamics create a supply of new buildings by simulating a list of new buildings. The movers, which include households and jobs, are matched with buildings via a choice/assignment model. In the population-centric model, employment and jobs are endogenous. The region is the level of geography modeled. The household evolution model includes simulation of personal biological events, such as aging, surviving, and giving birth; personal social events, such as marriage, divorce, and children leaving their parents' household; and household social events, such as migration and relocation.

- Base-year person and household inputs include age, ethnicity, gender, marital status, and household role. Parameters include survival, fertility, household formation and dissolution, and migration and relocation. The household evolution model addresses the importance of households. Household formation and dissolution have explicit and immediate consequences for the housing market. In the model, one household equals one housing unit. Housing units are vacated because of death, marriage, and out-migration. The demand for housing units results from marriage, divorce, in-migration, and relocation. The synthesis applies only to the base year. The model maintains a full accounting of people, households, and implied housing units without linking them to specific sites.
- Data from the ACS PUMS and the state demographer are used for various events, such as births, deaths, marriage, divorce, children leaving the household, and relocations. Marriage and divorce in the ACS PUMS are segmented by age, race, and gender. The model uses the PUMS for all Texas residents.
- Household migration is included in the ACS PUMS. The public use microdata area (PUMA)—subcounty, county, or group of counties—identifies the current residence. The MIGSP (state or county) and MIGPUMA (county or group of counties) identify the previous residence for households not living in the same house 1 year ago. Current residence is defined in regard to the MIGPUMAs, and the region is defined in regard to the MIGPUMAs. Current year households are those currently residing in the region. Previous year households are those who resided in the region 1 year ago. Previous year households equal households that did not move plus households relocated within the region, minus domestic in-migration, plus domestic out-migration, minus foreign in-migration. The ACS PUMS-based migration rates and the domestic in-migration rates for the H-GAC region were used.
- The employment model begins with the population and then considers the labor force, employed and unemployed workers, jobs, and jobs by sector. Data for the employment model come from the ACS PUMS, the ACS CTPP, and the Bureau of Labor Statistics.
- In summary, the model provides an effective representation of biological and social dynamics for the population and households in the region. Jobs are balanced with people. Links and constraints are more important than absolute levels.

The model provides the ability to run scenarios, such as regional or national recessions, closed borders, and matching external employment forecasts.

- Enhancements to the ACS PUMS would be beneficial in the development and use of the model. These enhancements include a larger sample size; half-year indicators, as the census estimates are for midyear; and more data on non-marriage-based and nonfamily households, particularly on change over time.

INTEGRATED TRANSPORTATION LAND USE MODELING: USING PUBLICLY AVAILABLE DATA SOURCES FOR ADVANCED MODELING

Rolf Moeckel

Rolf Moeckel described the development of a new land model called the simple land use orchestrator (SILO). He described the need for land use modeling, the integrated simulation model concept, the SILO elements, and a prototype application of the SILO model in the Minneapolis–St. Paul, Minnesota, area. He covered the following points in his presentation:

- Interaction of land use and transportation can be thought of as a circle. Transportation influences accessibility, which influences land use, which results in activities, which require transportation. It is known that transportation and land use influence each other. The extent and direction of this influence is often guessed at because of the lack of modeling capabilities and actual data. Integrated modeling can have a number of benefits. It can assist decision makers and agency staff. Integrated modeling also enhances transparency and improves public participation.
- The concept of integrated simulation models begins with land use, which influences transportation demand. Dynamic traffic assignment is generated from the transportation demand. The dynamic traffic assignment generates environmental effects. There is a transportation feedback loop from the dynamic traffic assignment to land use and an environmental feedback loop from the environmental effects to land use.
- Motivations for the SILO concept include fully representing the land use transportation feedback cycle and developing a tool that can be implemented with tight schedules, smaller budgets, or both. The ability to analyze smart growth strategies, simulate GHG emissions, and run many scenarios in a short time frame represents other motivations.
- The design parameters for SILO include the use of a microsimulation model that is policy sensitive. It uses zones for spatial resolution, but can accommodate raster or parcel databases. It includes a user-friendly geographic user interface. The required data are publicly available. Short run times are desirable, including approximately 10 minutes for one simulation period in a large region of 5 million people.

- Examples of possible policy scenarios include zoning changes, job–housing balances, and alternative growth assumptions. Other potential scenarios are increased use of mixed developments and transit-oriented developments. Changes in transportation costs represent still another possible scenario.
- Data input includes essential data and desirable data. Essential PUMS data include households by size, income, and automobile ownership; persons by age and gender; and dwellings by type and price. Other essential data are nonresidential floor space by type and price and forecasts of population and employment growth for the study area. Desirable data include preferences for location choice and demographic transition probabilities. National averages can be used if these data are not available.
- Output indicators include population by age, household size, and automobile ownership. Other output indicators are employment by type, dwelling vacancy, nonresidential floor space vacancy, change in housing costs, and land use consumption.
- The prototype application was conducted for the Metropolitan Council, the metropolitan planning organization for the Minneapolis–St. Paul area. The prototype application required the development of a synthetic population. As a microsimulation, SILO requires microscopic data. Because these data are not publicly available, a synthetic population has been created. Microscopic PUMS is used to create microdata. Microdata cover households, persons, dwellings, employment, and nonresidential floor space.
- Household microdata include household size and the number of automobiles. These data are linked to every person in a household and to all dwellings. Microdata for individuals include age, gender, education, and occupation. Other individual microdata are driver's license, workplace zone, income, and relationship (single, married, child). Dwellings microdata include zone, type, area in square footage, quality (complete plumbing/complete kitchen), and monthly cost (rent or mortgage).
- The PUMS 2000 data 5 percent sample was used to increase spatial resolution. There are 20 PUMA regions in the Minneapolis–St. Paul area. Every PUMS record provides information on the household, the dwelling, and all persons in the household. Every PUMS record has a weight that is used to expand the PUMS data to the full population.
- The PUMS 5 percent sample file for Minnesota has 359,332 records, with 37,341 of these records in the Minneapolis–St. Paul study area. Expanding these data generates 1,017,153 households and 2,405,601 persons. Households by size, age, and PUMA were graphed and examined, as were dwellings by type and PUMA. The validation of synthetic population identified some issues. The PUMS defines age as 0 to 90 and greater. A need exists to generate age distribution for

91, 92, 93, and older. There is also a need to ignore records with Dwelling Type 10, which includes boats, recreation vehicles, and vans, which are irrelevant for land use modeling, and with Dwelling Type 999, which is unknown. The next step in the prototype application is calibrating SILO for the Minneapolis–St. Paul area. The ACS data for 2010 will be used to validate the land use model for 10 years of simulation.

POPULATION SYNTHESIS FOR TRAVEL DEMAND MODELING: DATA NEEDS AND APPLICATION CASE STUDIES

Ram M. Pendyala

Ram Pendyala discussed the use of microsimulation models for travel demand modeling. He described the benefits of microsimulation modeling, the use of synthetic population in the models, and data needs for the models. He covered the following topics in his presentation:

- Interest in microsimulation models for travel demand forecasting is increasing. Microsimulation models simulate travel at the level of the individual decision maker, while recognizing interdependencies among activities, trips, persons, time, and space. Microsimulation models of travel are increasingly based on the activity-based paradigm of travel behavior. This approach provides explicit recognition of the derived nature of travel demand. It also provides enhanced representation of time–space interactions and constraints.
- Activity-based microsimulation modeling approaches offer the ability to address emerging policy questions of interest. By simulating activities and travel at the level of the individual traveler, the models are able to address the effects of greenhouse gas emissions–reduction targets, flexible working arrangements, and information and communication technologies. These models also allow for the analysis of interactions between microscale land use changes and travel, pricing-based policies, and nonmotorized transportation mode enhancements.
- Disaggregate household and person sociodemographic data for the entire population of the model region are needed for activity-based microsimulation modeling. These data for the entire population are generally not available, leading to the need to synthesize a regional population from known statistical distributions of the population. Disaggregate data for a sample of the population are available from PUMS and travel surveys. Marginal distributions for the entire region are available from the census summary files and agency forecasts. Population synthesis involves generating a synthetic population by expanding the disaggregate sample data to mirror known aggregate distributions of household and person variables of interest.

- The standard iterative proportional fitting (IPF) based procedure is based on Beckman et al. from 1996. The procedure includes a number of steps. These steps are selecting household-level control variables, obtaining the marginal distributions on these variables from census summary files, and generating a seed matrix of the joint distribution from a microdata sample data set based on PUMS data and travel surveys. Another step is expanding the seed matrix by using an IPF-based procedure to match the given marginal control totals while maintaining the joint distribution implied by the seed matrix. The selection of probabilities is estimated for households in the microdata sample. Households are drawn by using the selection probabilities to match the expanded cell frequencies. The resulting synthetic population is checked for goodness-of-fit, and households are redrawn if necessary. The synthetic population is composed of all individuals in the synthesized households.
- A number of motivations exist for enhancements to this process. The key limitations of the standard IPF-based procedure include that the controls are only for household attributes and not person attributes, the synthetic populations fail to match the distributions of person characteristics of interest, and the method ignores differences in household composition among households in a cell. These limitations result in the need to reassign weights to sample households according to household composition. This issue has been recognized by researchers, and a number of solutions have been proposed.
- PopGen represents a new population synthesizer. It incorporates a new iterative proportional updating (IPU) algorithm for estimating household weights. The algorithm estimates sample household weights such that both household and person distributions are matched. A simple, practical, and computationally tractable algorithm with an intuitive interpretation is used. The basic idea behind the IPU algorithm in PopGen is to reallocate weights among sample households of a type to account for differences in household composition.
- Step 1 in the PopGen method is to estimate household- and person-type constraints. Inputs in this step are household and person sample data and household- and person-level marginal distributions. The IPF is run to estimate the household- and person-type constraints.
- Step 2 is to estimate household weights by using household and person sample data and the household- and person-type constraints from Step 1. The procedure is to run the IPU algorithm to estimate sample household weights that satisfy both household- and person-type constraints.
- Step 3 is to generate the synthetic population by using household and person sample data and the household weights from Step 2. The method applies rounding procedures to obtain the frequency of different household types in the synthetic population and estimates household selection probabilities by using the computed weights. Other steps are to draw sample households according to selection prob-

abilities for each household to match cell frequencies and to repeat the process until a synthetic population with the best fit is obtained.

- PopGen terminology includes household type, person type, and a measure-of-fit. Household type should not be confused with a household attribute. It refers to a combination of household-level variables of interest. It represents a cell in the joint distribution of a set of household-level variables. Person type is formed by a combination of multiple person-level variables of interest. A measure-of-fit measures the absolute relative deviation between the IPU-adjusted cell frequency and the IPF-estimated household- and person-type constraints.
- The synthetic population generation process can be divided into three steps: estimating whole frequencies, calculating selection probabilities, and drawing households.
- The IPF-estimated household-type constraints provide target frequencies. Rounding procedures are used to convert decimal values to whole frequencies. Rounding procedures implemented in PopGen include arithmetic rounding, bucket rounding, and stochastic rounding.
- Synthetic households are drawn probabilistically on the basis of IPU-estimated weights. Selection probabilities are estimated for each household type that needs to be synthesized. No additional adjustments to match person constraints are needed. The individuals from the synthetic households make up the synthetic population.
- Rounded frequencies and the selection probabilities from earlier steps are used to generate a synthetic population. For each household type, the corresponding selection probabilities are used to draw households. The persons in the drawn households make up the synthetic population for the target year. Because the drawing procedure is probabilistic, the fit of the synthetic population is checked. The drawing procedure is repeated until a synthetic population with the best fit is obtained.
- A goodness-of-fit measure is used to check the match against person-level distributions. The corresponding *p*-value represents the level of confidence at which the synthetic population matches the given constraints. A synthetic population is drawn repeatedly until a desired *p*-value is achieved or a maximum number of draws is reached. The maximum number of draws is user specified and is dependent on the geographic context.
- In the population evolution model, often the marginals are generated at the individual geography level for specific forecast years. The population is synthesized for each forecast year, and the activity-based travel demand model is applied to estimate the demand. Issues with this model include the estimation of travel demand for intermediate years and the reflection of underlying socioeconomic and demographic processes across forecast years.

- Instead of generating a synthetic population for every forecast year, the base year synthetic population can be evolved annually to obtain the population for any future year. Synthetic households and persons are subjected to a host of socioeconomic and demographic evolutionary processes, including immigration and emigration, person-level life-cycle events, and household-level changes over time. Population evolution allows intermediate-year simulations while reflecting dynamics across forecast years.
- Case studies from the Southern California Association of Governments in the Los Angeles area and the Baltimore Metropolitan Council (BMC) in Baltimore were developed. The data used for the control variables and the resulting synthesized population were presented. The synthesized data were compared with the census PUMS and the 3-year ACS PUMS. The synthesized population was close to the census PUMS and the 3-year ACS PUMS data. Data sources for the immigration, age, mortality, fertility, education, individuals moving out of the area, employment, household formation, and household dissolution were presented. The use of the data in the simulation model was also discussed.
- A population evolution prototype was developed and implemented for the BMC region. The prototype considers the basic dimensions of household- and person-level attributes of interest. A host of other household- and person-level socioeconomic and demographic processes are of interest. These include the formation of nonfamily households, vehicle fleet composition and evolution, bicycle ownership, mobility options, and the availability of information and communications technologies. There is a need to better understand evolutionary processes and to enhance their representation.
- Other challenges associated with the population evolution include the availability of data, modeling simultaneous choices, endogeneity of choices, and accounting for interpersonal dependencies in the population evolution choice dimensions. The sequencing and hierarchy of choice dimensions is another challenge.
- In summary, the state of the practice is moving toward disaggregate microsimulation modeling of travel demand. A need exists to generate a synthetic population for the base year and to evolve populations for subsequent years to apply microsimulation models of travel demand. The current census information is adequate for the base year synthetic population generation. The applicability of census data to estimate models of various socioeconomic and demographic evolutionary processes is limited, however.
- Researchers often seek alternative survey resources to model these processes. This need may potentially introduce survey biases. Data from the ACS offer a valuable opportunity to capture the dynamics of households annually. A possible approach is to create a "panel PUMS sample" that can be traced over long periods of time, with households rotating in and out of the panel sample. Another approach is to use estimate models of evolution from a single consistent data source and to introduce history dependency.

People Move, Jobs Change

Transportation Impacts of Population Dynamics

Harold M. Brazil, *Metropolitan Transportation Commission*
Shimon Israel, *Metropolitan Transportation Commission*
Charles L. Purvis, *Metropolitan Transportation Commission (retired)*
Evelyn Blumenberg, *University of California, Los Angeles*
Madeline Wander, *University of California, Los Angeles*
Edward J. Spar, *COPAFS*
Ken Hodges, *Nielsen Claritas*
Jennifer Toth, *Arizona Department of Transportation, presiding*

This breakout session focused on changing racial and ethnic characteristics in urban areas in California. Shimon Israel of the Metropolitan Transportation Commission discussed the ethnic composition of limited English proficient (LEP) households in the San Francisco Bay area. Evelyn Blumenberg of the University of California, Los Angeles, reported on a study examining mode choice by Hispanics living in Hispanic neighborhoods. Ken Hodges of Nielsen Claritas described using the journey-to-work tables in the 2005–2009 five-year drive-alone (ACS) to examine means of transportation to work and travel time to work.

TRANSPORTATION CHARACTERISTICS OF THE LEP POPULATION IN THE SAN FRANCISCO BAY AREA

Harold M. Brazil, Shimon Israel, and Charles L. Purvis

Shimon Israel discussed the transportation characteristics of the population and linguistically isolated households (LIHs) in the United States, California, the Bay Area, and San Francisco. He reviewed federal guidelines related to LEP speakers, presented information on the journey-to-work mode for LEP households, and described the approaches being used by transit agencies in the Bay Area to provide information to LEP populations. He covered the following topics in his presentation:

- Title VI of the Civil Rights Act of 1964 prohibits discrimination on the basis of national origin. Executive Order 13166, signed by President Lyndon Johnson, provided further protection for limited-English-speaking populations.
- LEP is defined as pertaining to individuals who speak English less than "very well." There are questions in the ACS addressing LEP speakers. The initial

question asks, “Does this person speak a language other than English at home?” If the response is yes, the individual is asked to identify the language. The individual is also asked, “How well does this person speak English?” The possible responses are “very well,” “well,” “not well,” “not at all.” Responses of “well,” “not well,” and “not at all” are considered to indicate LEP.

- The 2005–2009 ACS provided the data for the study. The ACS standard tabulations, public use microdata sample (PUMS), and integrated public use microdata series (IPUMS) USA from the Minnesota Population Center were all used in the analysis.
- California, including the Bay Area and San Francisco, has higher levels of LEP speakers than the United States as a whole. Approximately 9 percent of the U.S. population is classified as LEP, compared with 20 percent for California, 18 percent from the Bay Area, and 23 percent for San Francisco.
- Spanish is the majority language spoken by LEP speakers at homes in the United States, at 60 percent. The remaining LEP languages are Korean, 7 percent; Chinese, 6 percent; Vietnamese, 3 percent; Tagalog, 2 percent; and other, 21 percent. The breakdown for California is slightly different. Spanish still represents the majority language used by LEP speakers, at 69 percent. The remaining LEP languages in California are Chinese, 8 percent; Vietnamese, 4 percent; Tagalog, 4 percent; Korean 3 percent; and other, 12 percent.
- The Bay Area and San Francisco reflect even more LEP diversity. In the Bay Area, Spanish represents 46 percent of LEP speakers, followed by 31 percent Chinese, 8 percent Vietnamese, 6 percent Tagalog, 2 percent Russian, 2 percent Korean, and 5 percent other. In San Francisco, Chinese represents 50 percent of LEP speakers, followed by 29 percent Spanish, 6 percent Tagalog, 5 percent Russian, 4 percent Vietnamese, 2 percent Korean, and 4 percent other. The location of those different LEP speaking groups were displayed by using the geographic information system (GIS) for the San Francisco Bay area.
- Identifying LEP speakers represents a person-level characteristic. Identifying an LIH represents a household-level characteristic and transportation characteristic. An LIH is defined as a household in which no one over 13 speaks English “very well.” Information on LIH transportation characteristics includes the average household size, mean household income, mean vehicles per household, and share of households with zero vehicles.
- The average household size for LIHs is larger than non-LIHs in the United States, California, the Bay Area, and San Francisco. For the United States, the average household size for LIHs is 2.9 compared with 2.5 for non-LIHs. In San Francisco, the difference is 2.3 for LIHs and 2.1 for non-LIHs.
- The mean household income for LIHs is lower than for non-LIHs in the United States, California, the Bay Area, and San Francisco. The mean household income for LIHs is $40,000 for the United States, $44,000 for California, $56,000

for the Bay Area, and $43,000 for San Francisco. In comparison, the mean family income for non-LIHs is $72,000 for the United States, $88,000 for California, $107,000 for the Bay Area, and $109,000 for San Francisco.

- The mean vehicles per household is also lower for LIHs. The mean vehicles per household for LIHs is 1.3 for the United States, 1.4 for California, 1.4 for the Bay Area, and 0.7 for San Francisco. In comparison, the mean vehicles per household for non-LIHs is 1.8 for the United States, 1.9 for California, 1.9 for the Bay Area, and 1.1 for San Francisco.
- The share of households with zero vehicles is also higher for LIHs. The share of LIHs with zero vehicles is 23 percent for the United States, 18 percent for California, 20 percent for the Bay Area, and 53 percent for San Francisco. The share of non-LIHs with zero vehicles is 8 per cent for the United States, 6 percent for California, 8 percent for the Bay Area, and 27 percent for San Francisco.
- The journey-to-work mode for LEP workers was also examined. Information on LEP workers in the Bay Area by language spoken was examined for carpools and transit. LEP workers have a lower drive-alone mode share than non-LEP workers. For the United States, 58 percent of LEP workers drove alone, compared with 78 percent for non-LEP workers. The percentages were generally similar for California (61 percent for LEP workers and 76 percent for non-LEP workers) and the Bay Area (59 percent for LEP workers and 69 percent for non-LEP workers). In San Francisco, LEP workers and non-LEP workers had much lower levels of drive-alone mode use, with 35 percent and 40 percent, respectively.
- LEP workers reported higher carpool mode shares than non-LEP workers. Carpooling was used by 21 percent of LEP workers in the United States, 20 percent in California, 18 percent in the Bay Area, and 11 percent in San Francisco. In comparison, 10 percent of non-LEP workers in the United States and California carpooled, and 9 percent in the Bay Area and 7 percent in San Francisco carpooled.
- LEP workers reported higher transit mode shares than non-LEP workers. Transit mode share among LEPs was 11 percent in the United States, 9 percent in California, and 12 percent in the Bay Area. The corresponding transit mode share among non-LEP workers was 4 percent in the United States, 4 percent in California, and 10 percent in the Bay Area. San Francisco had higher transit use among both groups with 39 percent of LEP workers and 31 percent of non-LEP workers using transit.
- Information on the Bay Area LEP worker mode share by language spoken at home shows higher levels of carpooling over taking the bus by Spanish-, Vietnamese-, and Korean-speaking households. Spanish-speaking households had the highest level of carpooling at 22 percent. Chinese-speaking households had slightly higher transit use, 17 percent, than carpool use, 15 percent. Tagalog-

speaking households were fairly evenly split, with 17 percent reporting the carpool mode share and 16 percent reporting transit use. The mode split for all workers in the Bay Area was 11 percent carpool and 10 percent transit. All LEP worker households were above these figures for both carpool and transit use except Vietnamese- and Korean-speaking households, which had 5 percent and 4 percent transit mode shares, respectively.

- The LEP strategies used by the major Bay Area transit agencies were examined. The major transit agencies all have telephone services, printed material, and websites in multiple languages. Most use media ads and electronic ticket machines in multiple languages. Two transit agencies have in-person assistance.

LATINO NEW URBANISM: EXPLORING THE LINK BETWEEN HISPANIC MODE CHOICE AND RESIDENTIAL LOCATION IN HISPANIC NEIGHBORHOODS

Evelyn Blumenberg and Madeline Wander

Evelyn Blumenberg discussed a recent study examining mode choice by Hispanic people living in Hispanic neighborhoods. She described the growth in the Hispanic population, previous studies exploring the travel behavior of Hispanics, and preliminary results from the current study. She covered the following points in her presentation:

- The study examines whether residents of Latino neighborhoods are less likely to drive alone than are nonresidents. This focus is predicated on the notion that the Latino culture is compatible with "new urbanist" communities—compact, dense, and walkable neighborhoods that contain a mix of activities such as employment, housing, shopping, and recreation. These types of urban environments ought to be conducive to the use of alternative travel modes, including carpooling, transit, and walking.
- The concept of "Latino new urbanism" is gaining attention in the popular media and professional journals. It has also been the focus of recent conferences and meetings. There are a number of reasons for studying the travel behavior of Latinos. First, the Hispanic population is large. Second, it is growing. Third, it is concentrated in particular areas of the country. Fourth, it is growing rapidly in new areas of the country. And fifth, the Hispanic population is ethnically diverse and, therefore, complicated.
- The 2006–2008 ACS provides data on race and ethnicity in the United States. White households account for approximately 60 percent of the total households in the country, compared with 15 percent Hispanic, 12 percent black, 4 percent Asian, and 3 percent other. The Hispanic population is forecast to increase to approximately 102.6 million by 2050, which is a significant increase from the

9.6 million recorded in the 1970 census. The Hispanic population represented 5 percent of the total population in 1970 and 13 percent in 2000. It is projected to represent 24 percent of the total population by 2050.

- Examining the percent Hispanic population by county in 2006 and the percent change in Hispanic population by county from 2000 to 2006 illustrates growth throughout the country, not just in the traditional areas of Hispanic immigration. Mexican is the major Hispanic origin according to data from the ACS, with 64 percent. Other Hispanic origins include Puerto Rican, 9 percent; South American, 5 percent; Central American, 5 percent; Dominican, 3 percent; Cuban, 3 percent; and other Hispanic, 8 percent.
- Results from previous studies indicate that Hispanics are less likely to travel by driving alone compared with non-Hispanic whites and are more likely to carpool. These trends were particularly strong for immigrants. Previous studies indicate that Hispanics are less likely to have driver's licenses than non-Hispanic whites, particularly Hispanic women. With time in the United States, Hispanic immigrants tend to assimilate to driving, similar to other segments of the population. For example, commute modes among Hispanics from 1980 to 2000 reflect increasing solo driving and declining carpool, transit, and nonmotorized transportation. Data from the 2006 to 2008 ACS on commute mode by race/ethnicity indicate that Hispanics have a lower percentage of drive-alone commuters and a higher percentage of carpoolers than white, black, or Asian population groups.
- A number of factors may influence a relationship between Latino neighborhoods and travel modes. First, the spatial proximity to other families of the same ethnic, religious, or cultural group might allow adults to easily find and coordinate with carpooling partners. Ethnic employment niches might mean that residents of Latino neighborhoods are more likely to travel to common employment destinations. Latinos living in ethnic enclaves—a subset of ethnic neighborhoods that includes ethnic-specific businesses and services—may travel relatively short distances and be more inclined to use alternative modes of travel.
- This study used 2006 to 2008 ACS data on a sample of metropolitan area workers and a multinomial logistic model to predict the likelihood of commuting by carpooling, public transit, and nonmotorized modes over driving alone. The method controlled for individual, household, and neighborhood characteristics. Neighborhood characteristics were examined by PUMA. Latino neighborhoods were identified by the Hispanic share of the PUMA population relative to the Hispanic share of the population in the metropolitan area. Residential density and the percentage of the housing stock built before 1940 in the central city were examined. Hispanic residents living in Latino neighborhoods represent a match between ethnicity of resident and neighborhood.
- Multiple models were examined for all Hispanics, Hispanic immigrants, and recent Hispanic immigrants. Approximately 40 percent of Hispanics are foreign

born compared with 12.5 percent of the total population. The travel characteristics of Hispanics in six metropolitan areas—Los Angeles–Long Beach, California; New York–New Jersey; Chicago, Illinois; Houston, Texas; Riverside–San Bernardino, California; and Miami, Florida—were examined.

- Hispanic commuters are more likely to travel by carpool and public transit compared with non-Hispanic white commuters. This finding holds for five of the six metropolitan areas. The one exception is Miami, where Hispanic workers are less likely to commute by public transit than non-Hispanic whites, even controlling for household income.
- In Los Angeles, Chicago, and Miami, there is a positive relationship between living in a Hispanic neighborhood and carpool use. In Houston and Riverside–San Bernardino, Hispanic workers who live in Hispanic neighborhoods are more likely to carpool.
- In Los Angeles, New York, and Riverside–San Bernardino, there is a positive relationship between living in a Hispanic neighborhood and transit use. In both Houston and Riverside–San Bernardino, Hispanic workers who live in Hispanic neighborhoods are more likely to use public transit.
- In all six metropolitan areas, Hispanic workers are less likely to commute by nonmotorized modes than are non-Hispanic whites. With one exception, there is a negative relationship between residence in a Hispanic neighborhood and nonmotorized travel. In Miami, workers in Hispanic neighborhoods are more likely to commute by nonmotorized modes.
- The findings from the national analysis indicate that there appears to be a strong positive relationship between Latino neighborhoods and commuting by carpooling, one that is enhanced by a match between residents and the ethnic neighborhood and is suggestive of the role of ethnic social ties in travel. The transit findings appear relevant only for recent immigrants; they may move to ethnic neighborhoods to take advantage of local opportunities and services, including transit services.
- The findings further indicate that in areas with very high percentages of Hispanics, such as Los Angeles or Miami, the match between residents and the ethnic neighborhood may make little difference. In metropolitan statistical areas (MSAs) with high levels of transit service, public transit may outcompete automobiles. Carpools may play less of a role. In MSAs with low levels of transit service, Latinos may move to Latino neighborhoods to take advantage of local opportunities and services.
- The study findings indicate implications for policy, data, and research. From a policy standpoint, matching local opportunities to local residents and enhancing public transit services in ethnic neighborhoods to help facilitate adjustment to the United States are two options. From a data standpoint, smaller geographic units

are needed, such as census tracts and block groups, rather than PUMAs. Data are needed on nonwork travel, as it is possible that the neighborhood effects would be more substantial if we examined nonwork travel. Additional research refining the models used in this study and examining whether similar trends hold for Asian and African-American neighborhoods would be beneficial. Exploring the variations within metropolitan areas would also be beneficial.

PREVIEW OF SMALL-AREA TRANSPORTATION DATA FROM THE AMERICAN COMMUNITY SURVEY

Edward J. Spar and Ken Hodges

Ken Hodges discussed an initial examination of the 2005–2009 ACS 5-year journey-to-work tables for means of transportation to work and travel time to work. The analysis explored the distribution of the ACS sample across block groups, as well as the MOEs reported for the journey-to-work tables. He covered the following topics in his presentation:

- The 2005–2009 ACS 5-year data include 11.1 billion estimates or cells. These ACS data are the first ACS data for small counties, cities, towns, census tracts, and block groups. Data for TAZs are not included.
- The ACS-based CTPP is subject to severe restrictions because of disclosure concerns. These restrictions include the wholesale suppression of the TAZ level. To provide an alternative, AASHTO contracted with the Census Bureau for the production of synthetic estimates. These synthetic estimates are created by perturbing ACS microdata records by using a method developed by Westat. These synthetic estimates will permit the release of TAZ data without suppression. The synthetic CTPP data will not be available until 2013.
- A preliminary analysis of the ACS journey-to-work data for block groups was conducted. The analysis focused only on the basic tables for means of transportation to work and travel time to work. The ACS sample is smaller than the long form sample, even after 5 years of data collection. The 2000 long form included 115,904,641 total households and 18,345,474 sample households. The sample household accounted for 15.8 percent of the total households. In comparison, the 2005–2009 ACS included 127,699,712 total households and 14,450,288 sample households. The sample households accounted for 11.3 percent of the total households. In addition, there were 9,658,438, or 7.6 percent, unweighted households.
- More than 90 percent of the block groups in the ACS sample have fewer than 100 interviews. Many measures of error (MOEs) are large. They provide a guide to uncertainty rather than a measure of actual error. More than MOEs are needed to judge ACS, but it is important to consider the MOEs.

- The MOEs and the cell values were examined for the means of transportation to work data. Overall, the MOEs are usually larger than the cell values, but the cell values are often zero. For the ACS, there are large MOEs for values of zero. It is important to remember that for means of transportation, drive-alone dominates in many areas, so many cells should be zero; there are few subway riders in Idaho.
- As an example, the means of transportation to work were examined for the block group including the conference hotel and the block group including the Beckman Center. The number of workers in each transportation category and the reported MOE were examined as were data from the 2000 census long form. Travel time to work was also examined. A dominant pattern with the travel time data is lacking; 15 to 19 minutes is possible anywhere. The MOEs are often greater than the cell values for block groups, even in cases in which the cell values are nonzero. For any given cell, the ACS has more block groups with values of zero. This result appears to be a byproduct of the smaller ACS sample. It is apparent especially with travel time to work data.
- The ACS data on means of transportation to work were examined for subways or elevated transit modes and ferryboat, which represent uncommon modes. Results indicated that the ACS data for the top 15 block groups for subway or elevated transit modes and ferry are in expected areas.
- A block group in Larimer, Colorado, had 618 bicycle trips in the 5-year ACS, with a 967 MOE. The census long form had zero bicycle trips for the block group. One possible scenario explaining the high number of bicycle trips is that the ACS picked up one bicycle trip in an interview and weighted it up to 618 bicycle trips. The overestimation in this block group compensates for underestimation in other block groups. Many block groups may have one or two bicycle trips that the ACS does not capture. The excess in this block group improves the estimates for block-group aggregations. One question is how this situation will be addressed in the synthetic data.
- A broader assessment of the ACS block-group data was conducted by using an index of dissimilarity (IOD). The IOD is the difference between two percent distributions. The bigger the IOD, the less similar are the distributions. The IOD was calculated and compared across all block groups for the 5-year ACS and the 2000 census long form.
- For all block groups the IOD for the means of transportation to work was 15.1 percent and the travel time to work was 30.7 percent. For block groups with stable populations, which were defined as less than 5 percent change between 2000 and 2010, the IOD for groups with means of transportation to work was 14.2 and the IOD for travel time to work was 29.7. For block groups with 100 or more ACS interviews, the IOD for means of transportation was 8.2 and the IOD

for travel time to work was 18.3. For block groups in counties with populations of 500,000 and greater, the IOD for means of transportation was 17.2 and the IOD for travel time to work was 32.2.

- The comparisons of block group IOD for the ACS and census were consistently lower for the means than the travel time. Results for the stable block groups were about the same as total block groups. The results were lower for block groups with 100 or more interviews and higher for block groups in large counties, which tend to be the areas of most interest to transportation planners.
- One factor contributing to these results is that the ACS sample is stronger in small towns than in block groups because governments have priority over statistical geographies. Examining the percentage of households interviewed in different geographies illustrates this point. For all block groups, 9.3 percent of the households were interviewed. For block groups in counties with populations of 500,000 or greater, 7.7 percent of the households were interviewed. In block groups in counties of 20,000 or less, 16.1 percent of the households were interviewed. In comparison, 37.8 of the households in Taylor, North Dakota, and 48.9 percent of the households in Washington Island, Wisconsin, were interviewed. There are more places than block groups with 1 to 49 households and more block groups than places with 1,000 or more households.
- The ACS can be used to help track rapid growth in areas. Although not official estimates, the ACS does provide population and household totals. For example, the block group including the site of the former Denver Stapleton Airport had no households in the 2000 census. The 2010 census listed 4,084 households, and the 2005–2009 five-year ACS listed 2,897 households.
- Overall, the various analyses suggest mixed reactions for the ACS 5-year data for small areas. There are larger errors than in the census long form and many zero cells. Rare characteristics are sometimes weighted to unrealistically high levels to compensate for zero cells. There is a trade-off of accuracy for block groups versus aggregations. The samples are stronger in small towns than in block groups. The ACS may be weak in large cities and metropolitan areas. The ACS does add value in rapid growth areas.
- The analysis indicates it is too early to pass final judgment on the ACS. The value of the ACS depends on the application. At this early stage, the ACS appears good enough to merit user support, but not so good that it can absorb cuts. There is room for improvement, but it would be a loss if the ACS were eliminated.

How Far Can You Go in This Vehicle?

Exploring the Utility and Limits of Census Data for Transportation Planning and Research

---

Sara Khoeini, *Georgia Institute of Technology*
Sundaram Vedala, *Georgia Institute of Technology*
Vetri Venthan Elango, *Georgia Institute of Technology*
Randall Guensler, *Georgia Institute of Technology*
Xin Ye, *University of Maryland, College Park*
Sabyasachee Mishra, *University of Maryland, College Park*
Fred Ducca, *University of Maryland, College Park*
Subrat Mahapatra, *Maryland State Highway Administration*
Xin Wang, *Old Dominion University*
Asad J. Khattak, *Old Dominion University*
Juyin Chen, *Virginia Department of Transportation*
Jinghua Xu, *Parsons Brinckerhoff*
Rosella Picado, *PB Consult, Inc.*
Dawn McKinstry, *PB Consult, Inc.*
Robert Santos, *The Urban Institute, presiding*

Speakers in this breakout session described different applications of census data in transportation planning and research. Sara Khoeini of the Georgia Institute of Technology discussed comparing the socioeconomic data from a 2006 survey of vehicles using I-85 in Atlanta, Georgia, with data from the 2000 census and the 2005–2009 five-year American Community Survey (ACS). Xin Ye of the University of Maryland summarized the use of census data in statewide modeling in Maryland. Xin Wang of Old Dominion University reported on a study examining the use of aggregate data based on census boundaries for travel demand forecasting. Jinghua Xu discussed validating person trip tables for transit demand forecasting.

USING 2007 LICENSE PLATE STUDY TO EVALUATE CENSUS SENSITIVITY BETWEEN 2000 AND 2010

Sara Khoeini, Sundaram Vedala, Vetri Venthan Elango, and Randall Guensler

Sara Khoeini discussed a study comparing socioeconomic data obtained from a 2006 license plate survey of vehicles traveling on I-85 in Atlanta, Georgia, with data from the 2000 census and the 2005–2009 five-year ACS. She described the study process and highlighted a few results of the analysis. She covered the following topics in her presentation:

- Data from license plate surveys are a valuable source of information for approximating sociodemographic characteristics and travel behavior of transportation facility users. The available sources of sociodemographic data linked to registered license plates determine the benefits as well as limitations of license plate studies, however. A question is whether the data are adequate for transportation planning.
- The census long form has traditionally been the most comprehensive database for sociodemographic characteristics. Data from the ACS are being more widely used in transportation analysis as a result of the discontinuation of the long form in the 2010 census. The census long form covered almost 17 percent of the total U.S. population in less than a year, while the ACS annual surveys cover only 1 percent of the U.S. population each year. Because the 1 percent sample is not a proper statistical representation of the population across small geographic areas, such as block groups, the 2005–2009 ACS 5-year summary file was used. Although the accuracy of the 5-year summary file is higher than that of the 3-year and 1-year files, the fact that the data are collected during a period of time instead of at a point of time makes time series analysis more challenging.
- In the summer of 2006, license plate data were collected in five locations on I-95 in Atlanta. The license plate data collection was performed to assess the sociodemographic characteristics of the commuter shed for the I-85 corridor. The existing high-occupancy vehicle (HOV) lanes on I-85 were expanded to high-occupancy toll (HOT) lanes on October 1, 2011. After the license plates were matched to the census block groups by using the geographic information system, the sociodemographic characteristics from the 2000 census and the 2005–2009 ACS 5-year summary file were assigned to the associated households. Household size, household income, vehicle availability, means of transportation, travel time to work, and time leaving home were selected for comparison between the census and the 5-year ACS.
- Household income was adjusted for inflation, as $1 in 2000 was worth $1.25 in 2009. On average, the median household income in the 2000 census was $9,000 higher compared with the ACS. Only 18 percent of the records had less than $5,000 difference between the census and the ACS. The recession may influence the ACS data.
- Household size was also examined. A total of 88 percent of the households have had less than a 0.2 difference in household size between the census and the ACS. On average, household size from the census is 0.13 higher compared with the ACS. The Atlanta metropolitan area population increased about 18 percent from the 2000 census to the 2010 census.

- The weighted vehicle available average for the census was 1.90 vehicles per household. The weighted vehicle availability average for the ACS was 1.88 vehicles per household.
- The ACS provides a greater level of detail in the travel time to work categories. The weighted average travel time to work from the 2000 census was 32.27 minutes. The weighted average travel time to work from the ACS was 31.83 minutes. The percentages of trips under 30 minutes are slightly higher in the ACS, and the percentages of trips 45 minutes and longer are also slightly higher.
- There are differences between the 2000 census and the 5-year ACS in the reported time leaving home for work. The percent of people leaving home in the 6:00 a.m.-to-6:30 a.m., 6:30 a.m.-to-7:00 a.m., and 7:00 a.m.-to-7:00 a.m. time periods is higher in the census than in the ACS. The ACS reflects higher percentages of people leaving home in the 12:00 a.m.-to-5:00 a.m., 5:00 a.m.-to-5:30 a.m., 5:30 a.m.-to-6:00 a.m., 8:00 a.m.-to-8:30 a.m., 8:30 a.m.-to-9:00 a.m., and 9:00 a.m.-to-10:30 a.m. time periods.
- In conclusion, the comparison indicates that there is a major difference between the census and 5-year ACS for household income and minor differences for the other attributes. These differences may reflect the different time periods, different sampling methods, and different sample sizes. The statistical tests applied rejected the equality of the distributions. The analysis indicates that the 5-year ACS data are good for developing overall trends nationally and at the state level, but that there is a need for local detailed surveys for transportation planning and project development purposes.

SOCIOECONOMIC DATA RECONCILIATION PROCEDURE IN THE MARYLAND STATEWIDE TRANSPORTATION MODEL: CHALLENGES, CURRENT SOLUTIONS, AND FUTURE STEPS

Xin Ye, Sabyasachee Mishra, Fred Ducca, and Subrat Mahapatra

Xin Ye described the use of census data in the Maryland Statewide Transportation Model (MSTM). He provided an overview of the MSTM, use of census data in the model, and preliminary results. He covered the following points in his presentation:

- The MSTM goal is to support improved planning decisions for an effective multimodal transportation system in Maryland by providing reliable and consistent travel forecasts and analysis capabilities. The MSTM objectives include estimating travel demand in non–metropolitan planning organization (MPO) regions of the state and in corridors spanning multiple MPOs. Modeling freight and intercity transit modes, including commuter rail and intercity bus, represents another objective. Analyzing regional transportation scenarios resulting from land use changes, network changes, and policy changes is another objective. Assisting MPO modeling efforts with external inputs, providing a tool for connecting available MPO models, and assisting with developing the long-range transportation plan, highway needs identification, and a study of highway performance measures represent still other objectives. The intent of the MSTM is to provide reasonable estimates on an aggregate level for planning, performance measurement, and decision making.

- The University of Maryland and Parsons Brinckerhoff submitted a scope to the Maryland State Highway Administration to develop a statewide transportation model in late 2006. A proof of concept model was developed in 2008 under Phase 1 of the project. The model was calibrated with National Household Travel Survey (NHTS) data in Phase 2. Further enhancements were made with the long-distance travel, destination choice, and mode choice models in Phase 3. The development of the draft final MSTM Version 1.0 is complete. A review by the technical committee and stakeholders is under way.
- The MSTM uses a three-level approach. The levels include regional, statewide, and national data. Both regional and national data flow into the statewide model. The model uses data on personal travel and trucks. The national, state, and MPO land use forecasts link to the socioeconomic data reconciliation, which feeds into the person travel and the truck components. Data from the NHTS feed the person travel model, and the Freight Analysis Framework 2 feeds the truck model.
- The overall study area encompasses 64 counties, including 24 counties in Maryland, 19 counties in Virginia, nine counties in Pennsylvania, eight counties in West Virginia, three counties in Delaware, and the District of Columbia. The statewide model zone (SMZ) map includes 1,588 SMZs. A total of 1,151 of these SMZs are in Maryland. There are 99 SMZs in the BMC region and 401 SMZs in the Metropolitan Washington Council of Governments (MWCOG) region.
- The primary socioeconomic data are from the Baltimore Metropolitan Council, MWCOG, the Pennsylvania Department of Transportation (DOT), the Virginia Department of Transportation, and the Delaware Department of Transportation. Data from the ACS, public use microdata sample (PUMS), and Longitudinal Employer–Household Dynamics (LEHD) are also used.
- The production model uses the joint distribution of household size and household income and the joint distribution of household size and the number of workers. Data from the ACS 2005–2009 estimates on household size, number of workers, and household income are used. PUMS data providing joint distribution by PUMAs are also used. An iterative proportional fitting is used to estimate joint distribution.

- The trip attraction model uses LEHD data at the census block level. The 20 different types are aggregated into the four categories of office, retail, industry, and other. Household and employment density for 2007 and 2030 are included. The modeled vehicle miles traveled (VMT) and the Highway Performance Monitoring System VMT were compared by county.
- In conclusion, census-related data components included in the modeling effort were the ACS, PUMS data for joint distribution of workers and household size by income category, and LEHD data for employment. These data were very useful for the MSTM.

CREATING LAND USE AND BUILT ENVIRONMENT VARIABLES IN BUFFERS SURROUNDING HOUSEHOLDS: ACCURACY OF GEOCOORDINATE IMPUTATION

Xin Wang, Asad J. Khattak, and Juyin Chen

Xin Wang discussed a research study examining the challenge of using aggregate data based on census boundaries for travel demand forecasting and transportation planning. He described the study process and preliminary results. His presentation covered the following topics:

- Understanding travel behavior is critical for the development of travel demand models. Census data sociodemographic and economic context variables are increasingly being used in modeling travel demand. Given the new developments in spatial analysis, integrating built environment variables in the estimation of disaggregate travel demand models is gaining momentum. To better understand the associations between travelers' behavior and their residential location and surrounding land uses and infrastructure, this study calculated built environment characteristics surrounding residential locations and used the variables in travel behavior models.
- To create new variables in buffers surrounding a residence, exact geocoordinates of survey respondents' residences are needed. Conventional travel behavior surveys such as the NHTS and census surveys do not publicly reveal the exact residential location of respondents because of confidentiality concerns. If the respondent's residential zone location is known, then zonal average sociodemographic measurements can be used as correlates in models to represent the average land use variables surrounding a specific residential location. The use of zonal averages can create measurement errors and reduce the local variation that may exist in reality, however.
- Geocoordinate imputation can overcome the lack of an exact geocoordinate location of a household; it can assign households to an exact geocoordinate location (latitude–longitude). The study explored the accuracy of geocoordinate

assignments. Data from household travel surveys conducted in the Triangle Area of North Carolina during 2003 and in Charlotte in 2006 were used to study this issue. Exact geocoordinates of residences were available to the research team and were compared with synthetically assigned residential locations. All households were also geocoded at the traffic analysis zone (TAZ) and to the census block levels.

- Given the number of residences in each TAZ or block, the same numbers of residences were assigned to a location in the TAZ or block semirandomly by constraining them to existing roadways excluding freeways, bridges, and ramps. The imputation results were compared statistically by examining roadway lengths within buffers created for synthetically assigned residences versus actual residential locations. If the random assignments can create equivalent synthetic residences that are not significantly different statistically from real residences in regard to roadway lengths in local neighborhoods, then the randomly assigned latitude–longitude can be used to replace actual residential locations. By doing so, the confidentiality issue can be overcome.
- Results indicate that synthetic assignment of residences using TAZ information is not equivalent to having actual household locations when the roadway length in 0.25-mile buffers around residences is analyzed. More specifically, there is a statistically significant difference at the 5 percent level between actual residences and synthetically assigned residences, in regard to the roadway length in 0.25-mile buffers around residences. The difference is not statistically significant for larger buffer sizes, however, including 0.5-mile and 0.75-mile buffers. Larger differences between synthetic assignments and actual residential locations are concentrated in suburban areas, where TAZs tend to be larger.
- Synthetic assignments are being conducted at smaller geographic units, including census blocks, and the roadway length is being recalculated in 0.25-mile buffers. The initial results for block-level synthetic assignments show that using smaller geographic units can reduce the differences in surrounding roadway lengths between actual and synthetic residential locations. Whether synthetically assigned residences can be used as a substitute for actual residences depends on their location. The Charlotte data showed that there was no statistically significant difference between synthetically assigned residences and actual residential locations in block average roadway length in the 0.25-mile buffers. The difference was statistically significant for the Triangle Area, however. The inconsistency between these two study areas may be partly the result of their different urbanization levels, given that the differences between roadway lengths in buffers are concentrated in suburban areas.
- The study assessed the extent of errors when the synthetic geocoordinate imputation method is used. Overall, results suggest that the extent of errors in geoimputation depends on the geographical unit used to assign residences, as well

as the buffer sizes around the residences. There is a trade-off between the analysis needs and computation burden because finer-level geoimputation increases the computation burden dramatically. In addition, geoimputation works better in urban areas than in suburban areas. To generate built environment variables, such as roadway length around residential locations, TAZ-based geoimputation is accurate enough to create synthetic residences in an urban area. For suburban areas, finer levels of geographical units may be needed to obtain more accurate results. To increase the accuracy of geoimputation, further efforts are needed to add more constraints on synthetic assignment, such as residential density. Future research is needed to evaluate whether using built environment variables obtained from geoimputation rather than using zonal averages can improve model estimation.

VALIDATION OF PERSON TRIP TABLES FOR TRANSIT FORECASTING

Jinghua Xu, Rosella Picado, and Dawn McKinstry

Jinghua Xu discussed the FTA New Starts model update and the verification of home-based work (HBW) person trip tables. He used the Orange County–Los Angeles metropolitan areas as an example. He covered the following topics in his presentation:

- In most metropolitan areas in the United States, approximately one-half of the weekday transit ridership consists of people traveling to and from work. The trip distribution modeling step needs to match this observed pattern to validate a behaviorally sound mode choice model. Home interview survey data are often used to validate the origin–destination distribution of trips. These surveys may not have a large enough sample to provide a comprehensive picture of worker flows in a region, however. The Census Transportation Planning Products (CTPP) worker flow data provide advantages in regard to availability and geographic coverage. The data provide sound travel information for checking the validity of HBW trip tables. Transit onboard surveys of riders can also be used to provide information on observed transit trip distributions and other trip characteristics.
- The purpose of the update of the FTA New Starts model is to update and calibrate the mode choice model following FTA guidelines. FTA requires that models used for New Starts forecasting provide a coherent picture of travel behavior, reproduce current travel patterns, and predict rational responses to transportation system changes.
- The Orange County Transit Activity Model was updated to reflect FTA requirements. The modeling area focuses on Orange County, but also includes Los Angeles, part of Riverside and San Bernardino, and Ventura Counties in California. In Orange County, transit service is mostly local bus, but has important

long-distance commuter rail markets. Los Angeles County has much more intensive transit service, which tends to dominate the transit calibration.

- The accuracy of the estimated person trip tables was verified in the mode choice model. The CTPP data were compared with the estimated trip distribution patterns geographically and by different demographic characteristics. The validation of HBW person trip tables was based on the CTPP worker flow data. The HBW person trip tables were validated across two travel market dimensions—household income level and district level. The tables were validated by income, workers per household, household size, distance, and other variables. One issue encountered with this process was the high level of suppression of worker flows by household income at the TAZ level. To overcome this issue, two sets of the CTPP worker flows were used. These two sets were the CTPP worker flows reported at the county level, by household income, and the CTPP journey-to-work worker flows reported at the TAZ level, by household income, which were further expanded to match the control margins.
- Validation of HBW person trip tables was conducted at the county level and the district level. A district is defined as an aggregation of TAZs that make up a major travel market. It provides the geographical dimension to compare and contrast the trip distribution and mode choice forecasts across major travel markets.
- Another indicator for verifying the trip distribution is checking the reasonableness of the transit mode share, which is calculated as the ratio of the observed transit trips over modeled "available" person trips at the district level. Distance-based transit mode share also provides an important indicator for examining the transit travel patterns. The reasonableness of the transit mode share can be checked to validate the distance-based trip length distribution. Distance-based mode share is calculated as the ratio of the observed transit trips over person trips by distance.
- The HBW person trip table was updated to address the issues of too many trips produced in/attracted to one area while not enough trips produced in/attracted to another area and poor validation of the distance-based trip length distribution. Census data were used in the household income submodel adjustment, the works in household submodel adjustment, and the household size submodel adjustment of the HBW trip production.
- The seed classification of households by income is in the household income submodel, adjusted so that the total proportions of households by income group per county better match the CTPP 2000 observed percentages. The workers in the household classification submodel can be updated to better match the CTPP proportions. The combined effect of these enhancements is that the production percentages by county are close to the CTPP data.

USING CENSUS DATA FOR TRANSPORTATION APPLICATIONS

- The HBW trip attraction component was adjusted on the basis of other data sources and the combined effect. In regard to HBW trip distribution, to reproduce the CTPP worker flow distance distribution, the HBW gravity model needs to be calibrated to match distance-based targets. The model uses a modified mode choice logsum as the travel impedance. Calibration to logsum-based targets is unlikely to yield the desired distance trip lengths. The approach used was to map the logsums to distance ranges and iteratively adjust the friction factors corresponding to specific distance ranges until the desired trip length distribution was obtained.
- In conclusion, the use of census data in the update of the FTA New Starts model provides sound travel information for checking the validity of HBW trip tables. The census data provide inherent advantages in regard to availability and geographic coverage.

Research Activities

Tom Krenzke, *Westat*
Brian McKenzie, *U.S. Census Bureau*
Bruce D. Spear, *Cambridge Systematics, Inc.*
Ron S. Jarmin, *U.S. Census Bureau*
Elaine Murakami, *Federal Highway Administration, presiding*

Speakers in this general session described recent research projects using census data. Tom Krenzke of Westat discussed applying statistical disclosure control to American Community Survey (ACS) microdata before Census Transportation Planning Products (CTPP) estimates are generated. Brian McKenzie of the Census Bureau described using place-of-work allocation for workers with missing employment addresses in the ACS. Bruce Spear of Cambridge Systematics, Inc., summarized the results of a recent NCHRP project, Improving Employment Data for Transportation Planning. The Census Bureau's data products for transportation planning were reviewed by Ron S. Jarmin.

APPLYING STATISTICAL DISCLOSURE CONTROL TO AMERICAN COMMUNITY SURVEY MICRODATA BEFORE GENERATING CTPP ESTIMATES

Tom Krenzke

My presentation focuses on applying statistical disclosure control to American Community Survey (ACS) microdata to generate Census Transportation Planning Products (CTPP) estimates. It is based on NCHRP project 08-79, which examined methods to perturb microdata for the generation of CTPP tables from 5-year ACS data.

I will summarize key elements from the executive summary from the NCHRP 08-79 project. I will provide an overview of the results of the preliminary investigations. I will also describe the general approach to the perturbation process and highlight some of the evaluation results. I will close by suggesting a few future research needs and potential improvements to the process.

The CTPP tables are divided into three parts. Parts 1 and 2 are residence- and workplace-based tables. Part 3, which includes the journey-to-work flow tables, is the central component of the CTPP. An outcome of this research is that more than 90 million perturbed tables consisting mainly of tables of counts, means, and medians will be generated.

A number of challenges were encountered with the project. First, there are more than 140,000 traffic analysis zones (TAZs) with the ACS sample. Transportation planners can piece the TAZs together in different ways, depending on the planning study. There are approximately 7 million TAZ-to-TAZ flows from the journey-to-work tables and only approximately 10 million workers in the ACS 5-year sample. This situation results in sparse tables in most TAZ-to-TAZ flows.

A second challenge relates to the Census Bureau disclosure rule, which is the Rule of 3. This rule requires at least three records in the marginal of the table. With this rule, cell suppression was used in 2000.

Third, the ACS 5-year sample is approximately 50 percent of the 2000 census long form sample. There were expectations that this difference would lead to more cell suppression. Our investigation indicated that 90 percent of the TAZs would be affected by the disclosure rules and that 30 to 50 percent of all microdata records would be affected. Because of these challenges, an alternative approach was sought.

On the basis of this initial analysis, three project research objectives were identified. The first objective was to develop an alternative approach to cell suppression that is operationally practical for the Census Bureau ACS Operations staff to implement. The second objective was to satisfy the transportation analytical needs, as represented by the TRB panel and planners. The third was to satisfy the disclosure rules established by the Census Bureau Disclosure Review Board (DRB).

The preliminary investigations focused on identifying the disclosure concerns. I have already mentioned the concerns with small geography and TAZ-to-TAZ flow tables. Another contributor to risk concerns is detailed categories of variables, such as industries. The major concern, however, is an intruder approach called "table linking." The means of transportation (MOT) is a common thread. For example, assume that we have three tables, MOT*A, MOT*B, and MOT*C for a TAZ. For a small category of MOT, someone could link tables to form a "microdata record" consisting of TAZ MOT A, B, and C. This record can be matched to the ACS public use microdata sample PUMS by using public use microdata area (PUMA) MOT A, B, and C to obtain more information and to put TAZ on the PUMA, which would be a violation of the census disclosure. Because of this concern, and the sparseness of the tables, we consider this to be both a tabular problem and a microdata problem.

A critical assessment of plausible approaches was conducted. Tabular approaches, which generate the tables first and then modify the result, were considered. Because of the large number of tables, the different types of tables and variables, and the sparseness of tables, three microdata approaches were analyzed. The three approaches evaluated were a semiparametric model-assisted approach, a parametric model-based approach, and a constrained hot deck (HD) approach. The constrained HD approach limits the amount of change that can occur to the target variable that is being perturbed. It groups records into cells, which are formed according to categories of the target records, as well as the sampling weights, covariants, and localities. The constrained HD approach can be conducted at the combined TAZ levels. A “without replacement” draw occurs from within the cell, which becomes a perturbed value. This approach provided good results.

The general approach divided the tables into Set A and Set B tables to retain as much of the original ACS data as possible. The Set A tables are not subject to DRB thresholds and will be generated from original ACS weights and original ACS variables. The usual rounding rules apply. The Set B tables are initially subjected to DRB thresholds. Special adjusted weights and perturbed variables are applied. The data are determined by the DRB to have acceptable protection from disclosure, so the threshold rules are lifted. Because Set A comes from the raw data and Set B comes from the perturbed data, the margins for the same variable that exists in Set A and Set B will not match exactly.

As mentioned, the Set B tables will be generated from the perturbed file. The Set B tables are provided in Appendix C of the NCHRP 08-79 Final Report. Since all Set B tables are generated from the same perturbed file, the property of additivity is retained. For example, all tables are generated for all localities from the TAZ level all the way up to the state level. In addition, the results for all TAZs in a state should equal the state-level table.

In addition, because all of the Set B tables are generated from the same perturbed file, the property of consistency is retained. This approach also allows consistency among collapsed versions of a table series. Consistency also exists for all flows for a table; for example, those in Table 33204 in Part 3 involving residence TAZ B, if added together, will produce the same results as Table 13204 for residence TAZ B from Part 1. Some tables are generated for large geographies only. Large geographies include county, PUMA, and state levels.

The general approach for perturbing the ACS microdata for the Set B tables involves a number of steps. The process begins with an initial risk analysis. The key objective of this step is to identify high-risk data values by running the tables and applying the DRB rules. Data values that contribute to violating the rules are flagged, and risk strata are formed to classify values from high-, medium-, and low-risk groups.

The next step is the data perturbation step. High-risk data values are perturbed in this step. This step was conducted with the three evaluation approaches mentioned previously. After the data perturbation approach is conducted, the weights are recalculated by using a raking approach, with some adjustments to the sub-PUMA level. A vast array of checks and summaries are produced from risk and utility measures.

There were four test sites, three approaches, two full replacement amounts, replacement of all data values, and partial replacement using risk strata and targeting the highest risk values. The partial replacement approach was accepted by the DRB. Five runs for each scenario were conducted to provide an indication of the amount of variation in the results. A total of 120 runs were conducted. There was very little variation across the runs.

After processing the 120 runs from data replacement through the raking process, the perturbed data and ACS data were compared by cell means, weighted cell counts, standard errors, paired correlations, and multivariate associations. The travel model outputs were also compared. The objective of this task was to conduct a reasonableness check to determine whether the performance of the perturbed ACS CTPP tabulations was no worse than the performance of the raw tabulations when compared against typical model outputs. The perturbed ACS data were compared directly with travel model outputs from the four test sites.

In the evaluation of cell means, the perturbed mean travel time on the *Y* axis was compared with the raw mean travel time on the *X* axis. The analysis was conducted for the constrained HD, semiparametric, and parametric approaches. The results indicate that the constrained HD approach is superior. In the evaluation of weighted cell counts, the perturbed weighted cell counts on the *Y* axis were compared with the raw weighted cell counts on the *X* axis. The constrained HD approach again provided the superior results.

To summarize the evaluation conclusions, the constrained HD approach performed best for each evaluation measure. However, it applies to ordinal variables only. The semiparametric approach was selected for the unordered categorical variables. For the travel model output comparison, the performance of the perturbed ACS tabulations was equal to that of the raw ACS when compared with the model output. A validation phase was conducted, which confirmed the results from the development phase. The nationwide operational test was successfully processed. The processing CPU time took approximately a day.

Westat is conducting future disclosure-related research for the Census Bureau. This research includes examining high-risk data values in the microdata, examining methods to perturb spatial outliers and nonspatial outliers, and evaluating scenarios to balance the use of weights, model predictions, and between-locality variance when data are perturbed. Westat is also evaluating variance estimation approaches that account for perturbation error and assessing the effect of perturbation on important research questions.

The NCHRP project identified possible methods for improving data dissemination, which must consider the Census Bureau disclosure rules. Applying the NCHRP 08-79 approach would need to be modified to retain more interaction terms among the 20 CTPP variables. More perturbation will most likely be necessary to allow a microdata set to be made available. Other options include proposing a higher level of geography, such as the use of transportation analysis districts. The use of research data centers could also be considered to assist with microdata dissemination.

Another possibility is to create an online analytic system to use partially perturbed ACS microdata. This approach would allow dynamic table queries from underlying microdata. It could provide remote public access. The Census Bureau Microdata Access System and Westat's WesDAX provide examples of this approach. Typically, there are threshold rules for universe and table specifications, table denials, subsampling and/or perturbation, and rounding. Where the data would reside and where the system would reside are key issues that need to be addressed.

The NCHRP project illustrated that the perturbation of microdata approach is a feasible alternative to self-suppression for the upcoming generation of tables from the 2006–2010 ACS. The good news is that the data will be available for all CTPP tables.

Finally, I would like to acknowledge and thank Guy Rousseau, who chaired the NCHRP project panel and other members of the panel, Nanda Srinivasan, the NCHRP Project Officer, and Laura Zayatz, who was the point of contact for the Census Bureau.

OVERVIEW OF PLACE OF WORK EXTENDED ALLOCATION FOR THE ACS

Brian McKenzie

My comments focus on the place of work extended allocation with the ACS. As part of the series of questions on commuting, the ACS asks respondents where they work and asks specifically for a workplace address. A portion of the respondent records does not provide sufficient information to geocode the respondent's workplace location to a block. In some cases, this information is missing entirely, and in other cases the response is ambiguous.

The place of work extended allocation is a set of processes to allocate the place of work block for sample ACS records that could not be coded to the block level during the Census Bureau's standard place of work allocation processes. It is used to increase the sample size and the completeness of the records. The increased number of records improves the accuracy at all geographies, but is especially crucial for smaller geographies such as block groups and tracts and for the CTPP TAZs. The place of work extended allocation was conducted on the Census 2000 records post hoc for the 2000 CTPP. Currently, it is being developed for use in the forthcoming 5-year CTPP, which will be released in 2013.

The place of work block could not be geocoded for about 24 percent of the records after the standard allocation procedures in the 2000 census. The extended allocation process coded another 13 percent of records to the place of work block level, accounting for approximately 2.5 million records. The extended allocation process is limited to records that fall within designated coding areas, that is, areas where response coverage is fairly high. These are essentially counties with the highest population.

For the ACS, after the standard ACS allocation procedures, the place of work block could not be geocoded for about 25 percent of the records. The standard allocation in ACS is that records without sufficient place of work information are imputed by matching to a donor record. A donor record is an individual with similar characteristics related to industry/occupation; travel time and travel mode; residence state, county, and tract; and place of work state, county, and place. Once matched, the record is imputed to the donor's place of work block.

Records that remain unmatched are entered into the place of work extended allocation. The extended allocation is slightly different from the standard allocation. It capitalizes on known characteristics of the respondent's industry, occupation, travel mode and travel time, and place of residence geography, but these matches do not necessarily have to be found with a single donor.

The extended allocation takes a slightly different approach. Rather than focusing on finding a specific person or respondent donor as a plausible match in the ACS data set, we look for clues that identify a block as a plausible match. We focus on industry/occupation, travel time, and travel mode. Therefore, the imputed place of work block shows at least one person who works in it who has the same industry and occupation, and is a plausible destination given the individual's place of residence, travel time, and travel mode, but the matches for these different characteristics do not have to come from a single donor. They can be found among multiple donors, as long as all characteristics are assigned a plausible match.

The place of work extended allocation includes three parts or phases. Part A matches jobs by industry and occupation. Part B matches trips by travel time and travel mode. Part C is the block assignment based on the distribution of workers for the industry/occupation.

Part A matches jobs by industry and occupation. The process finds blocks with the same detailed industry and occupation combination. There are 90 industry and 24 occupation categorizations. If a match is not found, more aggregated industry and occupation categories are examined until a match is found. There are seven levels of industry–occupation aggregations. This process creates a limited set of potential blocks.

As an example, a social worker in a hospital most likely works in a block where other social workers work in a hospital. The process would match the social worker to the appropriate industry–organization category.

Part B matches trips by private vehicle and transit. The set of potential place of work blocks is further limited by matching by travel mode and travel time. Private vehicle commuters are matched by using free-flow travel time estimates from the Bureau of Transportation Statistics. The process uses tract-to-tract travel time estimates to find all tracts that can be reached from the respondent's residence tract in the same amount of time as the respondent reported it took to get to work. The same process is used for public transportation commuters, but they are matched by using travel times with impedance factors applied to increase travel time. Preference is given to blocks in transit-accessible tracts.

The New York Metropolitan Transit Commission (NYMTC) transit travel time flow files were removed from the analysis. Although the New York City area has a much higher transit usage rate than other metropolitan areas, we feel that these differences will be sufficiently captured with ACS data. For consistency across places and standardization across years, the NYMTC files were deleted from the process, and ACS data will be used for that region. Bicycling and walking trips are also addressed. The process is simplified by removing tract adjacency for bicyclists and removing the modal supplement file for bicyclists and walkers.

Part C is the final block selection. If the final set of potential destination blocks has multiple blocks, then the process randomly selects one block by using probabilities based on the cumulative industry–occupation distribution across all potential blocks. For example, assume that there are only two blocks left in a set of potential blocks and 100 workers in the respondent's industry–organization combination. Block A has 90 workers and Block B has 10. Block A is selected with a probability of 0.9, and Block B is selected with a probability of 0.1. In other words, there is a 90 percent chance of selecting Block A, and a 10 percent chance of selecting Block B.

Ultimately, all records entered into the extended allocation system will be assigned a place of work block group. We expect that the most relaxed categorizations of industry/occupation and travel time will have to be used only for a small number of records.

The goals of the efficiency revisions are to make the edit completely in-house, use the most recent data, provide consistency across years, and provide for the efficient integration with the standard editing process. Other elements of the efficiency revisions include the NYMTC files with standard ACS data. Biking and walking will also be modified. The focus is on the most recent 5-year (2006–2010) files as inputs to process.

The implementation plan includes test runs of data to be completed in the fall of 2011. The results will be reviewed and analyzed. The back allocation of FY 2006–2010 will be completed by the ACS Office. The deadline for completion is the spring of 2012.

Future steps include updating the Bureau of Transportation Statistics tract-to-tract free flows. Consideration is being given to replacing the free flows with in-house tract-to-tract distance files or estimates based on average travel time responses. The plan is to integrate the place of work extended allocation into the routine processing of ACS records in 2012. Another future step is to include additional ACS files beyond the standard 5-year sample to increase coverage while retaining comparability.

EVALUATION OF DATA FROM THE QUARTERLY CENSUS OF EMPLOYMENT AND WAGES AND LONGITUDINAL EMPLOYMENT–HOUSEHOLD DYNAMICS–OnTheMap FOR TRANSPORTATION APPLICATIONS

Bruce D. Spear

My comments focus on the findings from NCHRP 8-36, Task 98, Improving Employment Data for Transportation Planning, which was sponsored by the AASHTO Standing Committee on Planning. The final report is posted at http://onlinepubs.trb.org/onlinepubs/nchrp/docs/NCHRP08-36(98)_FR.pdf. The research project was conducted by Cambridge Systematics, Inc.

Longitudinal Employment–Household Dynamics (LEHD) is a program to generate new information on workers and employers from existing data sources, as opposed to new surveys. The LEHD, which represents an innovative approach, was developed by the Census Bureau in coordination with state partners. The LEHD data are based solely on administrative records. All employment that is subject to state unemployment insurance laws is included.

There are two basic LEHD data products. The first is the quarterly workforce indicators (QWI). The QWI contains quarterly data on employment dynamics, including total employment, job creation, wages, and worker turnover. It is oriented toward groups interested in the dynamics of the labor force. The second data product is the LEHD origin–destination employment statistics (LODES). The LODES provides annual data on the locations and characteristics of workers by residence and workplace, and home-to-work flows. The LODES, which is probably of greater interest to the transportation community, was formerly called the LEHD OnTheMap.

The LODES supports the LEHD OnTheMap application, which is a web-based mapping and reporting application available at http://lehdmap.did.census.gov/. The LEHD OnTheMap application supports multiple analyses involving workplace and worker residence locations. These queries are supported at multiple levels of geography, including census blocks, tracts, zip codes, places, urbanized areas, counties, and states. Worker characteristics include industry type, age, income, gender, race, ethnicity, and education. Capabilities of the LEHD OnTheMap application include inflow–outflow analysis, home destination analysis, paired area analysis, and other data queries. The data are mapped and presented visually.

The LODES data files include residence area characteristics (RAC), workplace area characteristics (WAC), and origin–destination (OD) flows. The RAC files include the number and characteristics of workers summarized by residence geography and reporting year. The WAC files include the number and characteristics of workers summarized by workplace geography and reporting year. The OD files include the number and summary characteristics of workers who reside in one location and work in another.

The primary LODES data source for employer characteristics is the Quarterly Census of Employment and Wages (QCEW). The QCEW data are reported quarterly by employers to state employment security agencies. This source was formerly known as ES-202 data. For a number of years, the QCEW has been funded and managed by the Bureau of Labor Statistics (BLS). In the past decade, BLS has taken a much stronger interest in quality control, and funding has been provided to the state agencies to improve the quality of the data and how they are collected. The enhanced employer files contain data on total monthly employment and total wages by quarter for every employer. Employers with multiple worksites are encouraged, and in some states required, to complete multiple workplace reports to provide employment data by workplace.

A number of data sources are used to obtain the worker characteristics used in LODES. The personal characteristics file is derived from the Social Security Application File, which is also known as the Numident file. This file contains information on gender, date of birth, race, and citizenship. The composite person record (CPR) represents a second source of data. The CPR is derived from multiple federal sources, including the Internal Revenue Service (IRS), Medicare, and the Department of Housing and Urban Development. The data are based on a worker's Social Security number and include the worker's place of residence. In addition, employers are required to complete wage records as part of employment security measures. These wage records list individual employees by Social Security number. This record provides the critical link between workers and employers.

The LEHD data processing steps include actions by state agencies behind the firewall and BLS. The census statistical administrative record systems collect data from other agencies, including the IRS, Medicare, and the Selective Service. These data are also processed behind the firewall. The data are linked to an individual's Social Security number. At some point in the process, an individual's Social Security number is replaced by a protected identification key (PIK) to ensure individual privacy. The data are merged for each worker. Workers are matched to employers by using first-quarter wage records. Residence and workplace locations are geocoded for each worker. Combined records are subjected to disclosure proofing procedures to protect the identities of individual workers and employers. The two resulting data products are the QWI and the LODES.

Disclosure proofing is used because these are administrative records and because small geographies are involved. The LEHD records are subjected to disclosure proofing to protect worker/employer identities. A series of disclosure proofing techniques are used with the LEHD records. The census introduces a small amount of noise into some of the employer characteristics at the establishment level. The worker characteristics at the residence level are synthesized according to industry code, age, and earnings. The data are synthetic at the lowest level of geography. There is also some data suppression for small geographic units. Actual distributions are retained at more aggregate levels of geography and industry groups.

There are some data limitations with LODES. Some employment categories are excluded because of the use of unemployment insurance records. Self-employed and sole proprietors are excluded. Certain categories of workers who are covered by other unemployment systems, such as federal, military, and railroad workers, are excluded, as is employment exempt from unemployment insurance laws. The percent of workers in these exempt categories varies by state. For example, Florida has the highest percentage of self-employed workers and the District of Columbia has the highest percentage of federal workers. Although all states are now participating, data are missing for the states that recently joined. These include the District of Columbia, Massachusetts, New Hampshire, Puerto Rico, and the Virgin Islands.

The other data issue relates to multiple-worksite employers. Some multiworksite employers refuse to file multiple worksite reports. Without multiple worksite reports, all employees are reported at the primary worksite location, which obviously introduces errors in the workplace area characteristics and origin–destination databases. BLS examined noncompliance rates and found that about 5 percent of all employment nationwide is affected. The rates vary significantly from state to state. The noncompliance rates are slightly lower in states with mandatory multiple worksite reporting. The highest noncompliance rates are among local government agencies, such as school districts and other municipal agencies. State agencies have generally good compliance rates, as do federal agencies and private businesses. A table in the report presents the noncompliance rates by state.

A related issue is assigning workers to worksites. Only one state, Minnesota, requires employers to identify worksites on employee wage records, linking workers to specific worksites in the state. All other states simply require that employers provide a list of the total number of employees, but do not require that employees be assigned to specific worksites. Workers of multiworksite employers are assigned to worksites on the basis of a distribution model calibrated with Minnesota data. That situation is a possible concern.

As part of the research project, we compared LODES against the CTPP. It has been suggested by some groups that LODES and LEHD could be a replacement for the CTPP because of some of the issues with the CTPP. The study findings indicate that LODES is not a substitute for the CTPP. Most important, there are no trip characteristics in LODES; only the location of worker residence and workplaces and the potential for flow between them are provided. There are no data on travel mode, travel time, or departure time. At best, LODES is a supplement to the CTPP. Both LODES and the CTPP provided public sources of employment data for transportation planning. They have the same general characteristics of residence and workplace locations and OD flows for work trips.

We compared LODES and the 3-year CTPP by sample size, geographic coverage, employer/industry categories, and job categories. In regard to sample size, LODES has full enumeration for covered employment categories compared with an approximate 7 percent of households in the 3-year ACS sample. For geographic coverage, LODES excludes nonparticipating LED states, and the 3-year CTPP excludes counties with a population of less than 20,000. LODES excludes employers not subject to state unemployment insurance laws, whereas the CTPP includes all employers and industry sectors in the sample universe. LODES includes all jobs by workers in covered employment categories, whereas the CTPP excludes second jobs by workers with multiple jobs.

We found that both LODES and the CTPP typically underreport total employment by 15 percent to 17 percent nationally and by state. In the case of the LODES, the underreporting is attributed to the excluded employer categories and non-LED states. In the case of the 3-year CTPP, the underreporting is attributed to the secondary work trips and suppressed data. The 3-year CTPP produces significantly higher OD trip rates than does the LODES on a county-by-county basis but distributes them over many fewer county pairs.

We examined LODES and the 2000 CTPP at the tract level in the Minneapolis, Minnesota, and the Kansas City, Missouri, metropolitan areas. We found characteristics similar to those just described. The 2000 CTPP produced higher OD trip rates than LODES but distributed them over many fewer tract-to-tract pairs. The differences in flow rates between common tracts were much smaller. Differences in employment for individual tracts could generally be attributed to missing employment categories in LODES and new developments occurring after 2000.

We compared the work trip length distributions for LODES, the 2000 CTPP, and the 3-year CTPP. For the most part, LODES produces more trips in the longer trip categories and fewer trips in the shorter trip categories. Factors that may contribute to these differences are the exclusion of self-employed individuals, many of whom work at home, in the LODES data, and the ability of LODES to pick up low-frequency O-D pairs, which are more likely to occur between zones separated by longer distances.

The general findings from the study indicate that LODES is a good source of data on the distribution of home-to-work flows. LODES provides more comprehensive and current data than the CTPP, and it is updated annually. However, the study points out that LODES data should be used carefully and supplemented with local knowledge, especially to examine missing employment categories and multiple worksite employers.

Research needs identified in the study included better documentation of LEHD processes, especially related to the synthesis of worker characteristics for small geographies and the assignment of workers to worksites for multisite employers. Other identified research needs are further analysis of LODES OD distributions, alternative sources for missing employer data, and strategies for integrating LODES and the CTPP.

RESEARCH ON ACS AND ADMINISTRATIVE RECORDS

Ron S. Jarmin

The Census Bureau provides multiple data products relevant for transportation planning. The journey-to-work program is based on the self-reported place of work responses from the 2000 census long form and the ACS. The housing units are the frame of reference. The CTPP consists of estimated home-to-work flow data that are also focused on housing units.

The LEHD links administrative records and survey data. The administrative records include the QCEW and unemployment insurance wage records. The LEHD contains administrative person-level data on place of residence and demographic data from surveys and administrative data. Jobs are the frame of reference for the LEHD.

The LEHD is released via the OnTheMap and QWI products. OnTheMap is the product most relevant to transportation research and planning and consists of primary job data with a work and home location. Precise geography is possible on the public use product because of the use of advance disclosure avoidance techniques from the establishment data and the household- or jobs-level data.

The difference in the frame of reference—housing units for the CTPP and jobs for the LEHD—is fundamental to the way the data can be used and what can be learned from the data. Comparisons of CTPP and OnTheMap data products have been conducted to help users better understand both data sets. The two data sets are intended to be complementary rather than competitors. The LEHD was not originally intended to be a transportation application. It was developed for economic analyses, but it has been expanded for additional applications, including transportation. Examples of questions we get include why do LEHD data contain more long-distance commutes than CTPP data and how should these data be interpreted for use in transportation planning?

There is a new Census Bureau initiative to evaluate ACS and LEHD transportation data. The research project to link and compare ACS and LEHD microdata is funded out of the Improving Operational Efficiency program for FY 2011 that Deputy Director Commissioner Mesenbourg mentioned in his remarks. The goals of the project include integrating confidential microdata at a job level on an exploratory basis, including matching workers from the ACS to the workers in the LEHD, better understanding current data products by way of microdata comparisons, communicating the findings and guidelines to data users, and identifying opportunities to improve data quality. The project team includes personnel from the journey-to-work team and the LEHD team.

The plan for data integration and evaluation includes a number of major steps in constructing an integrated jobs frame. A first step is placing a PIK on ACS person records, which is needed to link the ACS data and the LEHD data. Because the ACS does not collect an individual's Social Security number, other administrative records have to be used. A PIK can be identified for 93 percent of the individuals in the ACS. We match ACS employment responses to LEHD jobs on the basis of two steps. First, the ACS person records are matched to LEHD workers via the PIK. Second, the ACS employment response is linked to an LEHD job associated with the worker according to the ACS interview date, employer name, and place of work. A final step is to evaluate whether the resulting integrated jobs frame is representative of each data set.

Potential microdata comparisons include examining the correspondence of ACS place of work responses with LEHD establishment geography and imputations and examining the distribution of ACS and LEHD home-to-work commute distances. Another analysis is measuring the sensitivity of the correspondence to differences in input data and data processing methods. This sensitivity analysis will focus on the ACS employment response versus the primary job in the LEHD, disclosure avoidance techniques, and place of work imputation. In addition, we will examine how these microdata comparisons relate to comparisons made with public-use data.

There are access options for analyzing more detailed data. The census products reflect the difficult balance between maintaining respondent confidentiality while providing data that are useful for statistical analyses. The Microdata Analysis System (MAS) should be available soon, although I do not think it will contain LEHD, ACS, or CTPP data. The MAS allows users to receive results from statistical analyses without having direct access to the data. It limits analyses to those that satisfy confidentiality rules. It allows multiple types of analysis, including cross tabulations and regression. The research data centers (RDCs) also provide the potential for full access to confidential microdata based on required approval. The RDCs provide the transportation data user community with an avenue to assist with the project evaluating LEHD and ACS data. It is an approved project, so it may be possible to add researchers with additional ideas on examining the data.

Power Users' Forum

Making Large Data Sets Useful and Usable for Transportation Agencies

---

Nanda Srinivasan, *Transportation Research Board*
Ken Hodges, *Nielsen Claritas*
Vincent Bernardin, Jr., *Bernardin, Lochmueller and Associates, Inc.*
Kevin Tierney, *Bird's Hill Research*
Ed Christopher, *Federal Highway Administration, presiding*

Private- and public-sector representatives involved in the identification, application, and dissemination of census and other sociodemographic data for the support of transportation planning shared their insights in this session. They responded to questions on available and planned census and census-derived data products, alternative data sources, and the state of the practice in transferring, extrapolating, and synthesizing sociodemographic and economic data.

How would you rate the American Community Survey (ACS)? Is it more of a success or is it more of a failure?

Ken Hodges: I think it is premature to pass final judgment on the ACS. If I had to choose between unqualified success and unqualified failure, I would say the ACS has been a success. I think we need to take a long-term view of the ACS. I was not in favor of the initial proposals for the ACS a decade ago. I think the Census Bureau listened to the comments from the data user community and the current product is better than the early proposals. We still face challenges in maintaining the availability of data at the block-group level. At this point, I would suggest the ACS does more things right than wrong, but improvements are still needed.

It is also important to remember that we are comparing the ACS with the 2000 long form. I am not sure that had there been a long form as part of the 2010 census it would have had as large a sample, as strong a response, and data quality comparable with the 2000 long form.

Vincent Bernardin, Jr.: I echo many of Ken's comments. The smaller sample size of the ACS, compared with the long form, is a challenge, especially in the small geographies. There are some gains in having more current information, however. The upcoming increase in the sample size and the improvements in the sampling strategy should help address some of the issues with the small geographies. These changes are signs that the ACS is listening to the data user community and reflect a move in a positive direction.

Kevin Tierney: The Census Bureau delivered what it said it would with the ACS. The ACS has much higher levels of standard errors with better quality and more reliable individual data records. The adequacy of the ACS is a different question, which is yet to be determined.

Nanda Srinivasan: The answer depends on whom you talk to. I think the Census Bureau personnel would say the ACS is a success. The ACS represents a new approach and a first for the bureau. It is a good addition to the products available from the Census Bureau. The margin of errors and the sample size are concerns for the data user community, however. There are also issues with timeliness. Data are available 3 years after the 5th year of data collection for the 5-year ACS. The Census Bureau indicated that the ACS was designed to have a standard error that was 33 percent of the census long form estimate. I think we are over that level with some variables, such as mode to work. We do not need the census or the ACS to confirm that most people drive alone to work. We know that. We need better data on the lesser used/lesser dominant modes from the ACS and the census, such as transit, carpooling, biking, and walking, especially for smaller metropolitan planning organizations.

From your perspective how useful have the ACS data been for the different ways you and the transportation community use them?

Srinivasan: A number of National Cooperative Highway Research Program (NCHRP) projects use ACS data directly or indirectly. The most recent project that used the ACS data in a limited way was NCHRP Project 8-61, which focused on travel demand parameters and techniques. NCHRP Project 08-78, which focuses on estimating bicycling and walking, is expected to use ACS data. NCHRP Project 08-83, addressing project-level forecasting, is also expected to use ACS data. In essence, any research project that needs demographic data will use ACS data.

Hodges: Most of our work focuses on the production of demographic estimates; the ACS has been used in some of the simpler analyses. The ACS has not been used yet for more ambitious analyses, such as those addressing income or housing value, partially because we are still working on a transition from the methods used with the long form. An interesting aspect of the ACS is that we are making projections from a base that is continually moving.

Bernardin: I have used ACS data, including the basic demographic data and synthetic population data from the ACS PUMS data, in a few travel model developments. At a basic level, the ACS provides the data necessary for the fundamental elements of travel demand forecasting models. I have experienced some of the issues with the differences in incomes reported in the ACS compared with the incomes reported in the census. This issue was recently encountered in the update of an automobile ownership model. I would suggest that using the ACS, especially in the initial applications, requires us to be thoughtful.

Tierney: I think the transition to the ACS has reinforced that we need to treat data as data, not as facts. We tended to treat data from the 2000 census long form, which was based on a sample, as fact. We are still learning about the ACS sampling method and the ACS data. A challenge going forward with the ACS is the ability to explain to others the limitations and nuances of the data.

Are there challenges in using the multiyear averages (period estimates) for the "older" models or methods? How have you overcome them?

Bernardin: We are examining other data sources to supplement the ACS data. We are also evaluating the ACS data on the basis of other sources. For example, automobile ownership data could be obtained from state departments of motor vehicles. I was encouraged by the comments from the representatives from the Census Bureau that consideration is being given to combining the ACS and the administrative records. I hope the ACS continues to improve so we do not need to use supplemental data from other sources.

What are you doing to supplement or complement the utility of the ACS data? Local data sources? Other national data sets?

Hodges: Use of the data at the individual block-group level reveals some strange results, largely because the sample is so thin in some block groups. To address this issue we are considering using other ACS data. There are approximately 208,000 block groups nationwide. For each block group and for each table in the ACS that we are using, we are carrying two distributions. One is the distribution for the table as published by the ACS. The second is a composite of the data for that block group as published by the ACS and the published data for every block group that shares a border with that block group. This second distribution provides a contextual measure. If it is a block group with a large number of interviews, such as more than 100 interviews, we will assign 100 percent weight to the data published. As the number of interviews declines, we will assign greater weight to the composite or contextual distribution. For example, if a block group has only 30 interviews, we might assign only 30 percent weight to the ACS data as published, and 70 percent weight to the contextual distribution. This approach is based on the premise that the patterns exhibited by the block groups in the neighborhood are more representative than a distribution based on a very small number of interviews.

Srinivasan: Most travel models use four or five modes in the mode choice component. The ACS data can be aggregated into the appropriate categories for use in a specific model. The margins of error may be smaller for aggregated data. Other national data sets may also be appropriate for use in transportation. These data sets include the Bureau of Labor Statistics Surveys, the Housing and Urban Development American Housing Survey, and the National Household Travel Survey (NHTS).

What are the most useful CTPP data products?

Srinivasan: The flow data set are by far the most used data set according to the number of support calls I received in my 9 years as a support contractor for FHWA. Flow by income is also used often. Data on travel mode, travel time, housing, and income, and departure time, to a lesser extent, are used frequently.

Bernardin: I would agree that the flow data are the most used, followed by travel time. Flow data by income and by mode to work are also used. I would say that the least used are some of the demographic cross tabulations that the Census Transportation Planning Products (CTPP) provides. These cross tabulations could probably be done by using the public use microdata sample.

Tierney: I agree that the flow data set, travel time, and mode are the most used. I use the demographic cross tabulations and would disagree that they are the least useful. I appreciate the three-way tabulations provided by the CTPP. They are of use in travel demand modeling and developing trip generation models. The flow data set is what makes the CTPP a unique and useful data product. The challenge with the CTPP has been with the disclosure limitations, which have restricted the ability to provide the flow data the transportation community really wants. The work Westat is doing on developing a method to obtain a complete flow table without worrying about which cells are redacted for disclosure reasons will be very beneficial.

If you had to identify one like and one dislike with the CTPP data products, what would they be?

Hodges: We are not really users of the CTPP because we cannot use traffic analysis zones (TAZs). We would like to be able to use it. After the 1990 CTPP was released, we contracted with the Census Bureau to produce a series of special tabulations for summary tape file tables by place of work. We made the same request after the 2000 CTPP, but it did not work out. I understand the transportation data community uses TAZs, but we cannot use them.

Are there any improvements that you would suggest for the CTPP data products?

Bernardin: The traffic analysis districts (TADs) are good. We are looking forward to using them as they provide good coverage and a good level of aggregation. There has been a gap between the TAZ level and the county level. The TADs provide a good level between the TAZs and the counties.

Have you used the Longitudinal Employer–Household Dynamics (LEHD) and if so, how have you used it?

Srinivasan: The LEHD has been used with a few NCHRP projects, such as NCHRP 08-36/Tasks 98 and 81. Checking it with local data sources may be beneficial.

Hodges: I have not kept current with some of the recent developments with the LEHD. It does appear to have some potential, but I have not had the opportunity to use it.

Bernardin: I was fortunate to work on a project in a region where the regional agency had two proprietary sources of employment data. I was able to compare the LEHD data with the data from these two sources. I also had three regions with one proprietary employment data source each and LEHD data. I am very hopeful for the LEHD and the potential it offers. My experience in trying to use it, however, has been frustrating. I have no delusions that the proprietary data sets are perfect; there are issues with them. Having the disaggregate microdata from the proprietary sources provides a better understanding of these issues and possible limitations. We do not have access to the disaggregate microdata with the LEHD.

In the one region where I had the two proprietary data sources, they tracked each other quite well if I examined the geographic distribution of employment. The two were pretty highly correlated even if I divided the data by industry, with correlation coefficients in the range of .85 or .9. When I conducted the same comparison with the LEHD, the correlation coefficients dropped to about .6, which is substantially worse. I think a significant part of this difference is the proprietorship. Several speakers have noted the concerns over recording federal employment, but I think a bigger issue concerns the nonwage and salary private employment. It has been difficult to justify the use of LEHD, even though it is free, over proprietary data sets given some of these concerns.

I also had data for multiple years, so I was able to examine changes in employment over time. The comparisons I mentioned were at the TAZ level. The proprietary data sets showed a clear pattern of employment growth and decline over time in the region, which could be verified in many cases with major changes such as the closing of large plants. The LEHD data did not reflect these same trends. I realize that ongoing improvements have been made in the LEHD and these types of issues may have been addressed. When I ran the correlations in the change in employment, the LEHD data had a correlation coefficient of .1, which would suggest there are some serious issues with the stability of the LEHD data over time. Improved disclosure proofing techniques may be able to address this issue. I think there is great potential for the LEHD, but my attempts to use it thus far have been frustrating.

Tierney: The one thing I would add to the discussion of the LEHD is that we need to remember that state labor departments provide the data. The basic data quality is dependent on the state labor departments, so the data or quality may vary from state to state.

Srinivasan: It is also important to remember the size of the LEHD, which includes full records on 75 percent to 80 percent of all workers in the country. If there is a good model to provide a current workplace, there is the potential to obtain flows by combining the workplace and the home residence.

In thinking about data gaps and shortcomings—what specific tools or data sets have you used to address them? Are there any parallel experiences that we can benefit from? Have you found any shortcomings that are not fixable with some type of bridge or other data?

Bernardin: The income issues I mentioned previously continue to be a concern. I am not sure anyone has developed a good solution. It is a topic worth additional discussion.

Hodges: Some of the tables in the ACS provide less category detail than was available from the long form data. This change results in a major data gap. For example, we have numerous clients who use the table on household income by age of householder. I think the long form provided seven age categories for the householder, compared with four categories in the ACS. It is difficult to do anything meaningful with just four age categories.

Srinivasan: The lack of information concerning nonwork trips has been mentioned a number of times during the conference. Although the National Household Travel Survey provides data on nonwork trips, interest was expressed in using the Global Positioning System (GPS) and other technologies to obtain additional data on nonwork travel. Research on methods to collect and analyze travel data from GPS and other technologies would be beneficial. NCHRP Project 08-89 on GPS and travel behavior was recently initiated. The private sector has efforts under way in this area. ESRI has a website that provides GPS travel data from a number of sources.

Bernardin: Most of the basic elements are available from the 5-year releases, which is more than the long form. We have talked about accuracy concerns resulting from the small sample size. I hope the larger sample size and the improved sampling strategy will improve the block-group-level estimates. It would be beneficial to have block-level data, even with large margins of errors. It is important to remember that the transportation data community is very concerned about location. The geographic specificity of the data at small levels of geography is critical to the quality of the job we can do for transportation planning purposes.

CLOSING SESSION

Improving Transportation Through Better Demographic Information

Reports from Breakout Groups

Ed Christopher, *Federal Highway Administration*
Kristen Rohanna, *San Diego Association of Governments*
Guy Rousseau, *Atlanta Regional Commission*
Nathan Erlbaum, *New York State Department of Transportation*
Keith Killough, *Arizona Department of Transportation*
Kathleen Lindquist, *Washington State Department of Transportation*
Thomas J. Kane, *Thomas J. Kane Consulting*
Alan E. Pisarski, *Consultant*
Elaine Murakami, *Federal Highway Administration*
Alison K. Fields, *U.S. Census Bureau*
Steven E. Polzin, *University of South Florida*
Jonette Kreideweis, *Minnesota Department of Transportation (retired), presiding*

Speakers in this general session summarized the key points from the breakout groups held earlier in the day. The following are the topics for the five breakout groups:

A. Data Content Specifications;
B. Integration with Other Sources and Private-Sector Data;
C. Data Dissemination, Data Access Tools, and Models;
D. Funding and Institutional Arrangements Relating to the Census and Alternative Sources; and
E. Research and Professional Development.

Each breakout group had a leader and a coleader to facilitate the discussion, take notes, and report back to the full group during this session. Participants in the breakout groups discussed issues and opportunities, research needs, and follow-up activities.

BREAKOUT GROUP A
DATA CONTENT SPECIFICATIONS (GEOGRAPHIC PRECISION, CURRENCY, ACCURACY, CONTENT)

Ed Christopher, Leader
Kristen Rohanna, Coleader

The topic for Group A was data content specification—geographic precision, currency, accuracy, and content. The group was made up mainly of individuals from metropolitan planning organizations (MPOs). As a result, the conversation and discussion focused on the data needs of MPOs.

Many participants indicated that the census product they used the most is the 5-year American Community Survey (ACS) tract and block-group data. They would like to use block data more. Many noted they use the most recent available data and recently released products. The 3-year ACS is used to examine county-level and large-city flows. A need was cited for 5 years of data for small areas.

Participants identified a number of different uses of census and ACS data. Data used included population counts and household counts from the 2010 census, the 3-year ACS flows, and the ACS household size. The 2010 census household and population data provide the base for forecasts and estimates. Issues associated with the ACS household size data were discussed. The calculated ACS household size did not match the published household size data. The frequency of different data products was discussed. Many participants noted the need for different tabulations every year, rather than full data sets and tabulations every year. A full data set every 3 years or every 5 years was identified as a viable alternative that would meet most needs.

The group spent a lot of time discussing ACS household size data. The ACS provides data on the number of persons per household. If you divide the total number of people by the total number of households, however, you arrive at a different number. Although this difference is to be expected given the weighting of the data, the calculated persons per household data tend to pass the reasonableness test and appear to be more appropriate than the ACS persons per household data. Examining this issue through a research project was suggested.

Participants identified household income as one of the most frequently used data elements. An issue was identified with household income in the 5-year ACS data at the block-group level. The block-group-level data are provided uncollapsed. Aggregating the data and making collapsed categories can be problematic. Issues are encountered in aggregating the cells. It was noted that the Census Bureau has provided a footnote that more than four categories should not be collapsed together because the formula for calculating the margin of error will not compute correctly. Identifying examples of aggregating zonal ACS data for multiple block groups is needed.

Participants discussed concerns about using census data and data from other sources and providing data to other agencies and the public when there are questions of accuracy. These situations can put MPO staff in the position of having to defend data that they have questions about.

The need to incorporate the 2000 population geography of the traffic analysis districts (TADs) into the standard ACS product line was discussed. It is just another level of geography that is built from census tracts. Participants could not identify any barriers to this request, which should be a no-cost item. Encouraging the transportation data community to promote this request could benefit all data user groups.

The group spent a lot of time discussing accuracy levels and margins of error. The Missouri State Data Center is color coding the margins of error included in the various data tables. The margin of error categories highlighted are less than 15 percent of the estimate, 15 percent to 35 percent of the estimate, and greater than 35 percent of the estimate. Participants discussed what margins of error could be accepted and what is a good margin of error. It was suggested that the accepted margin of error may vary depending on the variable and the level of geography. It was also noted that different terms are used, including margin of error, coefficient of variation, and standard error. Additional research on this topic would be beneficial.

We discussed possible data sets and products that are not currently being provided by the Census Bureau. An asymmetrical work flow for multiple geographies was suggested as one desired product. The 5-year ACS asymmetrical flows with constant geography are being provided. Using different size zones for residential data and workplace data is one option that would be worth exploring. The need for four or five dimensional tables in the Census Transportation Planning Products (CTPP) was also suggested. Another suggestion was providing a fully perturbed data set for any type of cross tabulation at any geographic level. This data set could be used for modeling and many other applications.

A number of existing data gaps and data needs were identified by participants. Environmental justice data at small areas, including disability, environmental quality data, and public health data, represent some examples of these needs. Having better active transportation data, including bicycle and pedestrian data, was also identified as a need. It was noted that federal grant applications often require data that are not available. Better communication and coordination between federal agencies related to data requirements and availability were suggested.

Participants also discussed employment data, noting that Longitudinal Employer–Household Dynamics (LEHD) data are not enough. Additional data on the universe of employees, the distribution of earnings by the North American Industry Classification System, the distribution of wages and earnings at the workplace, a variable identifying the relationship between individual earnings and households income, and Equal Employment Opportunity–like tabulations at smaller geographic levels to provide a better indication of transportation options by race and occupation were identified as beneficial.

Improving the American FactFinder tool was noted as important. With continuous sampling, the idea of considering the potential cost savings by not providing the same data products every year was suggested. The cost savings could be used to fund larger sample sizes.

BREAKOUT GROUP B
INTEGRATION WITH OTHER SOURCES AND PRIVATE-SECTOR DATA

Guy Rousseau, Leader
Nathan Erlbaum, Coleader

Participants in the breakout group came from universities, a small MPO, a large MPO, a large department of transportation, a consulting firm, and a telematics data provider. The breakout group identified and discussed four major issues. These issues addressed alternative types of data sources, private-sector data and potential risks associated with the lack of standardization and availability across all geographies, possible obstacles associated with the use of private-sector data and methods to overcome the obstacles, and potential uses of LEHD data, including combining LEHD data with ACS data.

The discussion during the first session focused on using private-sector data, including data from business establishments and credit collection agencies. Using marketing files with household characteristics available from private sources was discussed, along with the potential risks associated with using these data for addressing transportation issues. Many participants indicated that there were good working relationships with state labor departments. Employment and socioeconomic data from these departments are used in transportation planning and modeling. Using CTPP data in rural transit planning was noted. Other participants noted that CTPP data were used to validate private-sector data and that the cross tabulations were of interest. The ability to share private-sector data with others to validate results was noted as important.

During the second session participants spent more time discussing key transportation problems and how the data from different sources can best be used to identify and analyze possible solutions. An observation was made that there are many stable measures provided by the National Household Travel Survey (NHTS). Further, the CTPP, ACS, and LEHD can be used to provide aggregate geographic-level controls for expanding sample data resources. It was noted that the data being used today are partly a function of the legacy tools and models used in transportation planning, project selection, and operations.

Real-time telematics data, including data provided by GPS and other devices by the private and public sectors, bypass the traditional trip generation, trip distribution, and trip assignment process because they are already assigned to the network. Questions with these data include the sample size, the potential of self-selection biases, and other potential issues. Approaches for using data not intentionally collected for transportation uses to answer transportation-related questions were discussed. Potential issues and risks associated with using these data were noted. One participant noted the challenges of providing needed data to decision makers resulting from business cycle influences changing data collection schedules.

Research on using data from the private sector and emerging technologies in transportation planning and operations was identified as needed. Telematics and information technologies are changing rapidly. There is much more interest in visualization today and showing people the results of different alternatives graphically. The historical static approach of presenting information does not recognize available technologies.

It was suggested that we need to better understand how GPS and other telematics data can be used as a substitute for traditional transportation data, in combination with the census and other data sources, and as a complete replacement for existing data sources. Telematics data can illustrate the real-time daily traffic flows and adaptive behavior of motorists. Research to better understand the benefits and the limitations of these data would be beneficial.

Two graduate students from Georgia Tech University were in the breakout group. They are working on a very interesting project using employment data and credit reporting agency data from a private firm. The credit reporting agency file includes extensive data on household and individual characteristics. The project is examining how these data can be linked to other data sources, such as vehicle registration data from the Georgia Department of Motor Vehicles, and how to extract behavioral life-cycle cluster information. The desired result is a pseudo-total travel survey–type record at a much lower cost than a traditional household travel survey. This approach could provide timely data and address many of the issues with traditional household surveys. How it all fits together is still a question that deserves a closer look.

It appears that private-sector data providers and telematics data providers are willing and interested in working with public-sector agencies, universities, and other groups to explore opportunities to use their data in transportation planning, project development, operations, and evaluation. Follow-up was suggested to explore how these data sources can be used and how they might be combined with data from the census, NHTS, CTPP, and other sources to make significant advancements.

BREAKOUT GROUP C
DATA DISSEMINATION, DATA ACCESS TOOLS, AND MODELS

Keith Killough, Leader
Kathleen Lindquist, Coleader

Breakout Group C focused on models, data access tools, and data dissemination. We discussed issues and opportunities related to these topics, possible research projects to address identified concerns, and stakeholders for the research.

The first issue discussed was potential improvements to data disclosure protection methods. A possible research project related to this issue is to identify the concerns, examine their validity, and develop approaches for addressing valid concerns. The need for more research on the utility of perturbation versus suppression was discussed. The stakeholders for these issues and research include MPOs, state departments of transportation, and AASHTO.

The second issue discussed was updating employer referencing systems. We discussed the need for improved coordination between stakeholders to avoid duplication of effort in compiling employer listings. A research synthesis examining how various regions are compiling and updating employer referencing lists would be beneficial. Possible coordination with the Quarterly Census of Employment & Wages process could be explored as part of the research synthesis. The major stakeholders for this research are MPOs.

The third issue discussed was estimating network travel distances. Participants noted that network travel distances or proxies are currently available from a number of sources, but that travel times by mode are equally important to many users. It was suggested that in the long term, private data sources, such as TomTom and INRIX, may provide the best option for obtaining needed data. Research on current methods and possible future directions was identified as beneficial. State departments of transportation and MPOs were identified as the key stakeholders for this research and data.

Coding workplaces when insufficient address data are provided represents the fourth issue we discussed. The suggestion to resolve this issue was cross-checking multiple sources. The identified stakeholders for this information were state departments of transportation and MPOs.

Participants discussed the adequacy of current methods to disseminate the data. It was suggested that making the data available on the Internet was critical. It was noted that the American FactFinder is not very user friendly. It was further suggested that obtaining a 3.0 version of American FactFinder would benefit all groups. The CTPP Access Tool was noted as good, especially the Beyond 20/20 data dissemination and visualization software. Many participants felt the CTPP listserv works well and provides a community-based peer-to-peer system. It was further suggested that a wiki could be created and used to share information and best practices. Other topics discussed included the CTPP Status Report, the need for face-to-face training, an electronic guidebook, a Bureau of Transportation Statistics (BTS) bookstore, LEHD training, and the ACS Retrieval Tool.

Participants noted that this conference and the breakout groups provide an excellent example of the benefits of face-to-face interaction. People shared their experiences during the discussion of various issues. Often when one person described a problem in his or her area, someone else was able to share information on how he or she had addressed the same or similar issues. The exchange of ideas and information was excellent.

Participants discussed possible process enhancements that could be made at the Census Bureau or elsewhere to obtain higher-quality data. It was suggested that implementing TAD geography would be beneficial. Better responsiveness to inquiries was also suggested, including enabling one-on-one contacts with Census Bureau staff. MPOs were identified as the key stakeholder for this support, but state departments of transportation and other census user groups could also benefit.

Another issue discussed in the breakout group concerned whether or not the disclosure process was producing usable data. As noted by other speakers, concerns about accuracy are sometimes raised. Users may never know until there is a question about the data from outside groups. Clearly and visibly noting any data concerns or data limitations was suggested as an important improvement. The user community was identified as the stakeholders for these improvements.

Participants discussed the potential for further improvements to the place of work allocation. The need for additional validation was suggested. The state of Minnesota, which requires firms to accurately report multisite branches, was noted as a good model for other states. Stakeholders for this topic were state departments of transportation, MPOs, and other groups.

Similar to other breakout groups, we discussed the roles technology can play to promote dissemination, analysis, and presentation of census data for transportation planning, operation, and research. One suggestion was to fund the university transportation centers to act as data clearinghouses, providing one-stop locations for available data. It was noted that stewardship of the data is needed to ensure providing long-term, ongoing availability. The need for a permanent cloud storage or a data repository was suggested. The need for a Census Bureau equivalent of the BTS long-term archival library and management system was also discussed, along with considering a web-based data language and organizing system, such as the Hypertext Preprocessor. The U.S. Department of Transportation (DOT) was suggested as the main stakeholder for this concept.

The final topic we discussed related to making microlevel census data more easily and inexpensively available to the transportation research community. Online training and National Transit Institute/National Highway Institute–type training was suggested. The stakeholders identified for this training included TRB, the Association of Metropolitan Planning Organizations, the National Association of Regional Councils, the American Public Transportation Association, the American Planning Association, and AASHTO. A final long-term theme discussed by the participants was using technology to support data collection and data sharing, including interaction between different user groups.

BREAKOUT GROUP D

FUNDING AND INSTITUTIONAL ARRANGEMENTS RELATING TO THE CENSUS AND ALTERNATIVE SOURCES

Thomas J. Kane, Leader
Alan E. Pisarski, Coleader

Breakout Group D discussed funding, institutional arrangements, and cooperation. Thinking of the agencies involved in the collection and dissemination of census data for transportation purposes as a triangle, with the Census Bureau, the U.S. DOT, and AASHTO at the points of the triangle, was suggested. There are multiple issues from a communication and a coordination point of view. These issues are internal to the agencies and organizations and external. The external communication and coordination issues can be considered both bilateral and trilateral.

Participants discussed internal issues in all three groups. Many participants noted that fostering improved coordination in the Census Bureau could be beneficial to overcome internal stovepipes. A key example is the need to enhance communication and cooperation between the journey-to-work program and the LEHD program. Earlier speakers from the Census Bureau indicated that improvements in coordination between these two programs are being made. We look forward to seeing the results of this enhanced communication and coordination.

Another internal issue is better integrating the data to respond to the congressional mandates. Many participants were concerned about what are, in effect, unfunded mandates. The group discussed performance measurement, performance monitoring, and the different aspects of the CTPP that are not now being considered programmatic responsibilities.

The desirability of funding the transportation data programs on a more substantive and ongoing basis was discussed. Most of the transportation data programs are in the research portion of the U.S. DOT reauthorization. Moving the transportation programs into ongoing programs, rather than the research portion of the reauthorization, might provide a more substantial and stable funding environment.

Some participants observed that the U.S. DOT could better represent the transportation data community in dealing with other federal agencies. Many of these agencies have an interest in transportation data and use transportation data, but have not necessarily funded or supported any data collection elements. It was suggested that the U.S. DOT might take a more proactive role in gaining cooperation and support, including financial support where appropriate. A key point is that the transportation data community is not alone in the need for good data. Many other federal agencies also rely on transportation and census data.

More collaboration within the AASHTO programs was also noted as important. The AASHTO programs are expanding in response to federal mandates and the needs of the various states. Realizing and coordinating the different data demands of these programs are important.

We discussed the opportunities for a peer exchange or a synthesis to better understand and document how state departments of transportation and MPOs are using, applying, and coordinating purchases of private-sector data. This topic was also discussed in other breakout groups. Organizing the peer exchange or synthesis is an important institutional issue. It was suggested that AASHTO is the logical group to promote a better understanding of the full complement of data patterns and data users. These partners and users include state departments of transportation, MPOs, cities, counties, regional agencies, transit agencies, and other organizations. All of these groups would need to be involved in a possible peer exchange.

There are other issues associated with the bilateral relationships—the relationships between the U.S. DOT and AASHTO, between the U.S. DOT and the Census Bureau, and between the Census Bureau and AASHTO. Participants discussed the potential to strengthen the working relationship between the U.S. DOT and the Census Bureau. Many noted that although there is an excellent working relationship between these two agencies, it could be broadened, strengthened, and expanded to higher levels. The same holds true for enhancing the working relationship between AASHTO and the Census Bureau. The need for regular communication at the top levels of all three organizations was discussed.

Funding issues, including obtaining needed resources for the transportation data programs and possible flexible funding, were discussed. Moving toward the standardizing of census products represents one approach to addressing some of the funding issues. The weakness and limitations in the ability to transfer funds between programs in an agency, as well as between agencies, was identified as a potential issue. A variety of possible funding sources and combinations were identified and discussed, including promoting funding flexibility and pooled fund projects.

A key theme through all of the items discussed was the importance of collaboration and the importance of relationships. That message—collaboration and relationships—is critical for moving forward.

BREAKOUT GROUP E
RESEARCH AND PROFESSIONAL DEVELOPMENT

Elaine Murakami, Leader
Alison K. Fields, Coleader

Breakout Group E was tasked with examining research and professional development. The breakout group included representatives from state departments of transportation, MPOs, consultants, universities, the Census Bureau, and FHWA. Even though there was only one person who used census data for transit planning, there was a lot of discussion about mode choice, including transit, walk, and bike modes.

The charge to this breakout group was to consider avenues of research that could be implemented in the next 3 to 5 years to inform the transportation community. Participants discussed research priorities and professional development needs. Many of the topics discussed were similar to those noted by previous speakers.

Participants noted that census data are often taken for granted. People assume that census data will always be available, that they will always be reliable, and that they will include small areas. The increasing use of census data for social justice, environmental justice, and limited English proficiency analyses was discussed, as well as the increasing use for population synthesis, microsimulation, and the activity-based model at large MPOs. The CTPP three-way cross tabulations for households and the ACS public use microdata sample (PUMS) are of growing importance in these efforts.

For most small MPOs, the census data are the only source of population, socioeconomic, and travel data. The home-to-work travel flows from the ACS and the CTPP are critical components for the traditional four-step travel demand models. Many of these areas have not conducted household travel surveys for more than 25 years. As a result, there is an increasing reliance on census data, the NHTS, and transferable parameters from other areas for travel demand models.

It was noted that the flow data from the CTPP are in high demand and are valued by the transportation community and other groups. There may be opportunities for the transportation community to work with other groups using the flow data to build partnerships to support the ACS and to leverage other data sources. There may also be opportunities to coordinate with the growing interest in freight data in many areas. Another opportunity discussed by participants is the use of private-sector data and other alternative sources. Private-sector data from INRIX, TomTom, GM On-Star, smartphone tracking, credit card tracking, and social media, such as Facebook and Twitter, may provide viable options to be explored. Possible issues related to self-selection bias and coverage bias could be examined with these data sources.

Participants identified and discussed a number of research needs. One research topic focused on synthetic microdata with a system for disclosure proofed user-defined tables. The 2006–2010 CTPP will include more than 200 tables, some of which will require the use of partially synthetic records. Although some tables are used by many agencies and planners, other tables are used rarely. Some participants noted it would be beneficial to develop a new system that would protect microdata behind a firewall and establish a tabulation interface with disclosure checking. A GIS component for data visualization could also be beneficial.

One key issue to be researched is whether or not an ACS-based microdata summary system can include all the ACS variables or whether a microdata summary system with a more limited set of variables for transportation would be more cost-effective and easier to implement. Some existing research projects were discussed at the conference, including research at the Census Bureau with the current population survey that Laura Zayatz is doing and work at Westat conducted for the National Center for Health Statistics. A test using the ACS data is not yet being conducted, but the Census Bureau test system and disclosure limitations rules would be applicable to ACS and other census data sets.

Many participants would also like to see contextual variables added to the microdata records, including neighborhood characteristics related to transit accessibility and land use. Integrating the microdata access program with GIS was also viewed as important. Because individual confidentiality must be protected, the public use microdata area (PUMA) geography linked to the PUMS is large. For many transportation planning applications, the PUMA geography limits the ability to understand land use, such as job density and shopping accessibility, and transportation accessibility effects on travel choice, including travel mode choice (transit, walk, and bike).

Participants discussed the potential to attach small-areas summary information that would be valuable for transportation planning applications to individual microdata records without compromising confidentiality. It was noted that research is needed to identify and prioritize contextual variables and then to test how to modify the variables (e.g., ranks or scores, not discrete values) to maintain confidentiality while enhancing the utility of the microdata for transportation data users.

To implement the microdata access program, many participants felt that it would be beneficial to reach out to other users of the ACS flow data at the Census Bureau. The feasibility of an ACS microdata set that includes all of the variables, as well as maintaining a microdata set with a limited number of variables for smaller geographies, could be explored.

The second research topic focuses on combining administrative records with the ACS. Conference speakers discussed some of the work the Census Bureau has initiated on combining the ACS with administrative records, including the LEHD and statistical administrative record systems, and potentially relying on the use of the Census Bureau Business Register. Ron Jarmin noted that the research data center (RDC) program could be used to involve transportation planners working as research partners. It was noted that because the research project is already under way, the approval process for additional research should not take as long as previous attempts at conducting transportation-related research at RDCs. It was suggested that the Census Bureau team working on this topic also consider combining other sources, including private data, the NHTS, and other public data. Many participants felt that this project is very important and the transportation data community would like to be involved and assist if possible.

Another research topic is to examine all travel, not just commute trips. Additional data, including private data sources, would be needed in this effort. Participants were very interested in the role new technologies can play in the collection, analysis, and display of data on all types of travel. The upcoming International Conference on Travel Surveys in Chile may provide additional insight into new technologies for conducting travel surveys. There was interest in the use of smartphones with GPS and in-vehicle navigation systems for capturing data on daily travel and examining travel over time. Traditional surveys have obtained travel information for individuals for 1 or 2 days. Multiday data are needed for modeling and other analyses.

Participants also discussed the need for freight data. Transportation planners rely on the census and the ACS for personal travel data and commute trips. Trucking firms and telematics companies represent a source of travel data. Purchasing data from these providers and combining them with census data and data from other sources represent an option that could be explored further.

The need for high-quality, detailed geography employment data was discussed as another research area. The results from the CTPP and private sources, such as Dun & Bradstreet, InfoUSA, and LEHD origin–destination employment statistics (LODES), can vary considerably at the detailed geographic scale. Participants identified a number of questions related to these data. Why is the correlation between LODES and private-employment lists at the block level so low? Why are the trip length distributions between CTPP and NHTS so different from LODES? What is the imputation and allocation method used for LODES? What is the best way to identify, address, and correct specific types of known problems, such as headquarter concentrations where multiple worksites are not included, such as school district headquarters? Research addressing these questions could be beneficial.

Developing better relationships with the private sector to enable the fusion of public and private data sources was also identified as a need. Private-sector data that have the potential to provide detailed origin–destination matrices may lack a link to personal and household characteristics that is typically obtained from the census–ACS data. This link between the sociodemographic and specific travel behavior is critical to travel demand forecasts and microsimulation. Therefore, it was suggested that methods to use these private- and public-sector data sets are needed. This fusing would require innovative approaches to partnerships and cost sharing, not just technical innovations. Research on this topic could be beneficial.

Another research topic focuses on evaluating the effectiveness and utility of the synthesized CTPP by using the method outlined in NCHRP Project 08-70, which was discussed in one of the general sessions. The CTPP data product for 2006–2010 ACS data will provide synthesized data that will enable the Census Bureau to provide perturbed data at detailed geographic delineations that conform to Census Bureau disclosure limitation requirements. The perturbation procedures were developed and will be applied by Westat. After the CTPP is released and data users have had the ability to use the data, it would be instructive to evaluate how much the data perturbation has affected common transportation planning applications.

Participants identified a number of training and professional development activities. A key theme of many of these efforts was knowledge transfer. Staff turnover continues to be a concern at many transportation agencies, especially smaller MPOs. With the aging of the baby boom generation, attracting new professionals to the transportation planning and modeling field is a priority at many agencies. It was noted that there are more professionals with GIS expertise in transportation planning, but attracting professionals with expertise in computer science was also suggested as important.

Much of this discussion focused on improving communication among professionals in different agencies and recognizing generational differences. The discussion of improving communication focused on getting personnel from state departments of transportation and MPOs to work together to ensure that the right people have the right information about data, training and conferences, and research findings. This improved communication would increase consistency in planning and the use of data for transportation planning. The discussion of recognizing generational differences focused on the fact that newer employers may be more GIS savvy but lack the experience to judge the appropriateness or reasonableness of the data. Spatial data are increasingly easier to access but are not always correct. The ability to use the tools does not imply the skills needed in evaluating the data.

One training need identified was developing a way to describe data synthesis in a way that is understandable and that addresses privacy protected data. Protecting individual confidentiality is important. Identifying a method that is palatable and understandable would be beneficial.

Developing a Transportation Data 101 manual and delivering it through a wiki was suggested. The wikis used for the On-line Travel Survey Manual (www.travelsurveymanual.org) could serve as an example for the Transportation Data

101 manual. The manual would provide an overview of the basic data used in transportation planning, including the census, the ACS, CTPP, the NHTS, LEHD, household travel surveys, traffic counts, and other sources. Finally, participants discussed the need to better check market and brand synthetic data, explain the issues of privacy and disclosure, and describe their use. Adding a CTPP e-learning module would be a good way to address this need in the short term. A National Highway Institute training course on this topic might also be appropriate.

Training on privacy protected/disclosure proofed (synthetic) data was noted as needed. The 2006–2010 CTPP will have disclosure proofing applied so that data at the census tract and TAZ level can be released. It was suggested that the implementation of the NCHRP 08-79 project needs to be explained simply, with a discussion of data accessibility versus data suppression, privacy protection, and usability and reliability. Expanding the CTPP e-learning to include a specific module on this topic would be beneficial. Some participants noted that it would be useful to find a way to describe data synthesis with language that is positive, does not raise questions about believability, and focuses on privacy protection.

The need for training, including exercises, for judging data reasonability and utility was also suggested. Topics that could be included are how to evaluate whether data should be used, how to evaluate whether data sources do not agree with each other, and what to use if the sources do not agree. Another suggested training included highlighting best practices in data use by developing one or two case studies to document the data activities at an MPO or a state department of transportation for planning applications. Updating a document from the U.S. DOT's Travel Model Improvement Program completed in the late 1990s that showcased the Portland Metro data program was suggested.

Developing training in best practices in human capacity management was also suggested. This training could include methods such as using teams and cross training to ensure that institutional knowledge is not lost. Another suggested idea was to use outside experts to provide training.

Finally, a suggested outreach activity was to develop an "elevator speech" or short list of discussion points that enumerate the highlights of critical data elements. This product could be in the form of a brochure to ensure that executives and upper management know the value of reliable sociodemographic travel data for completing planning activities.

CONCLUDING OBSERVATIONS

Steven E. Polzin

The summaries highlight the dynamic discussions on a wide range of topics. There are a number of good suggestions for research needs, training, and follow-up activities.

This conference is a little like putting together a jigsaw puzzle. We began by putting the puzzle pieces on the table. During the course of the conference, we got the pieces all turned up, we found the edge pieces, and we sorted the pieces by color. We have not quite completed the puzzle, however. We just heard all the good discussions and suggestions from the breakout sessions, but we have not quite completed an action agenda. The challenge ahead of all of us is to discern and learn the key messages of our discussions. We need to follow up in a time frame and in a way that leads to improvements.

It is interesting to consider that this conference on the census is being held in 2011. Previous conferences have been held in the middle of the decade between censuses. The conference is earlier because there is a realization that the country and world are changing rapidly. If the old order were still in existence, the 2010 census would have included the long form and the results would not be available yet. Vince Barabba did an excellent job of highlighting the change from the previous evolutionary pace and the effect of this change on the census.

There have been a number of common themes over the past half century, including the aging of the baby boom generation and the increasing urbanization and suburbanization of the country. Other themes focus on women joining the workforce, increases in automobile ownership, rapid per capita growth in travel, and high migration and immigration levels. Some relatively consistent travel behaviors accompany those changes. There was a consistent shift away from transit, walking, and biking and toward single-occupant vehicle use. There was rapid growth in urban travel and increases in trip lengths.

I think we are at a point at which many of these trends have changed. We will see continual changes in the future. I think there will be a new normal after the current economic downturn is over. The country and the world are changing. We need to use this change in discussing the importance of having census data and other data for good decision making. We need current and accurate data to reflect the changing conditions in the United States and the world.

We often lose track of how dynamic change is. Rapid employment growth is currently occurring in North Dakota. Florida recently experienced an annual population decline. We will have different transportation needs going forward. All of the evidence seems to indicate that we have had a plateauing of personal travel on a per capita basis, which will influence future planning.

The need for timely and accurate data is clear. This need exists in a very challenging time in which there are numerous demands on scarce resources. We also face issues with confidentiality, response rates, and new technologies. However, the need to make well-informed decisions is as strong as or stronger than ever. Good data can be a foundation for good decisions, and good decisions are a bipartisan need in a time of limited resources. A strong case can be made for the importance of accurate and timely data for transportation planning and decision making.

As a data community, we need to be careful that we do not get caught in a mind-set in which we need specific types of data or data elements because we have always had them. We need to step back on occasion and consider what data we would need if we were beginning with a clean slate, as well as what tools and models we would need. Further, does every community, regardless of its size, rate of growth, typography, geography, and culture, need the same set of tools and data?

We also need to consider how we can increase the productivity of data collection and processing and help control the cost of data. The data community has done an excellent job with listservs to help educate practitioners, websites to help visualize and share data, and innovative strategies for collecting real-time information. We have not always been good at setting priorities or making trade-offs in response to limited resources and changing conditions, however. For example, we heard a suggestion to consider sacrificing the frequency of certain data variables for an increase in sample size. Is this a trade-off we are willing to make?

In regard to census-specific challenges, we have heard concerns with Fact-Finder 2. Concerns were also voiced in response to the spring 2013 date for the release of the 2006–2010 CTPP flow data. We empathize with the challenges facing the Census Bureau, but there are critical needs for these and other data.

We need to be careful that our data needs are not just a desire to provide input to travel forecasting models or data for the academic research community. We need to be able to link data needs to decision-making issues. We cannot become obsessed with blank cells, measures of error, and small geography data cell values that appear counterintuitive. We need to be able to trace the significance of data quality and precision to the recommendations and the decisions that are being made. We need to be careful about what we ask for, and we need to be able to trace the data we ask for to their application and how they influence decisions.

There was a lot of discussion about institutional relationships. Institutional issues are not unique to census data. Alan Pisarski and I served on the national committee examining data needs. Institutional issues, communication concerns, relationships, and coordination are common among many data programs. The data community is a relatively small community, and much of what we do is built on individual relationships. The power of a single person, and that person's personality, passion, and perseverance, can make the difference between progress and failure. Formal institutional relationships that surround the collection and distribution of data are often fragile. We do not have the luxury of strong institutional structures and a deep bench of human capacity that can insulate us from the performance of an individual. As individuals, we are all important to positive working relationships and moving forward.

The potential for using private-sector data and other alternative data sources was discussed extensively. I think we are making progress moving in that direction, and I think we will see major steps in the next decade. The use of these alternative sources of data will depend somewhat on how transportation programs are governed in the future and the authority of various levels of government. If decisions are made at the federal level, the need for standardized data and criteria will be greater. If funding and decision making are devolved to lower levels of government, there is less need for standardized data and approaches, resulting in different data needs and criteria in different areas.

I have been an advocate of the LEHD because the web tool provides numerous benefits. As the LEHD becomes more established, it needs to be more transparent and user friendly. Explaining the limitations and where the LEHD works and does not work is needed. The Census Bureau and the transportation data community can partner to make the LEHD a much more powerful and useful analysis tool.

Participants

Judy Althoff, *Regional Transportation Commission of Washoe County*
Peter Bang, *Regional Transportation Commission of Washoe County*
Vincent Barabba, *Market Insight Corporation*
Vince Bernardin, *Bernardin, Lochmueller & Associates*
Evelyn Blumenberg, *University of California, Los Angeles*
Harold Brazil, *Metropolitan Transit Commission*
Jonathan Brooks, *Texas Transportation Institute*
Robert Bush, *HDR Engineering, Inc.*
Ken Cervenka, *Federal Transit Administration*
Juyin Chen, *Virginia Department of Transportation*
Shih-Miao Chih, *Oak Ridge National Laboratory*
Simon Choi, *Southern California Association of Governments*
Ed Christopher, *Federal Highway Administration*
Mary Kay Christopher, *MKC Associates*
Chaushie Chu, *Los Angeles Metro*
Kathy Chung, *Fresno Council of Governments*
Nick Cohn, *TomTom*
Kathleen Dove, *California Department of Transportation*
Henning Eichler, *Southern California Regional Rail Authority/Metrolink*
Patrick Eliot, *Mountainland Association of Governments*
Patti Ellis, *Texas Transportation Insitutue*
Nathan Erlbaum, *New York State Department of Transportation*
Caroline Ferris, *TranSystems*
Scott Festin, *Wasatch Front Regional Council*
Alison Fields, *U.S. Census Bureau*
Kevin Garcia, *Northwestern Indiana Regional Planning Commission*
Tina Glover, *Sacramento Area Council of Governments*
Susan Gorski, *Michigan Department of Transportation*
Gregory Gould, *Natural Resources Defense Council*
Matt Hardy, *American Association of State Highway and Transportation Officials*
Laine Heltebridle, *Pennsylvania Department of Transportation*
Mell Henderson, *Mid-America Regional Council*
Ken Hodges, *Nielsen*
Hsi-Hwa Hu, *Southern California Association of Governments*
Pat Hu, *Research and Innovative Technology Administration*
Jaclyn Hubersberger, *Transportation Research Board*

PARTICIPANTS

Leta Huntsinger, *Durham–Chapel Hill Area MPO*
Ho-Ling Hwang, *Oak Ridge National Laboratory*
Shimon Israel, *Metropolitan Transportation Commission*
Edward Janowicz, *San Diego Association of Governments*
Ron Jarmin, *U.S. Census Bureau*
Mara Kaminowitz, *Baltimore Metropolitan Council*
Tom Kane, *Thomas J. Kane Consulting*
Asad Khattak, *Old Dominion University*
Sara Khoeini, *Georgia Institue of Technology*
Keith Killough, *Arizona Department of Transportation*
Jonette Kreideweis, *Consultant*
Thomas Krenzke, *Westat*
Josephine Kressner, *Georgia Institute of Technology*
Brian Lasagna, *Butte County Association of Governments*
Catherine Lawson, *University at Albany*
Kathy Lindquist, *Washington State Department of Transportation*
Yuh Wen Ling, *U.S. Department of Transportation*
Liang Long, *Cambridge Systematics, Inc.*
Evan Lowry, *Charlotte–Mecklenburg Planning Department*
Thi Bich Thuy Luong, *Institute of Transportation Studies*
Jianming Ma, *Texas Department of Transportation*
Gregory Macfarlane, *Georgia Institute of Technology*
Doug MacIvor, *California Department of Transportation*
Lora Mattox, *City of Billings/Yellowstone County*
Brad McAllester, *Los Angeles Metro*
Brian McKenzie, *U.S. Census Bureau*
Phil Mescher, *Iowa Department of Transportation*
Thomas Mesenbourg, *U.S. Census Bureau*
Dmitry Messen, *Houston–Galveston Area Council*
Arash Mirzaei, *North Central Texas Council of Governments*
Rolf Moeckel, *Parsons Brinckerhoff*
Elaine Murakami, *Federal Highway Administration*
Felix Nwoko, *City of Durham, Transportation Division*
Janet Oakley, *American Association of State Highway and Transportation Officials*
Ziying Ouyang, *San Diego Association of Governments*
Thomas Palmerlee, *Transportation Research Board*
Nokil Park, *Des Moines Area Metropolitan Planning Organization*
Behruz Paschai, *North Central Texas Council of Governments*
Ram Pendyala, *Arizona State University*
Lani Pennington, *Oregon Departtment of Transportation*

Alan Pisarski, *Consultant*
Steven Polzin, *Center for Urban Transportation Research, University of South Florida*
Cory Pope, *Utah Department of Transportation*
Melanie Rapino, *U.S. Census Bureau*
Clara Reschovsky, *Metropolitan Washington Council of Governments*
Kristen Rohanna, *San Diego Council of Governments*
Guy Rousseau, *Atlanta Regional Commission*
SungHo Ryu, *Southern California Association of Governments*
Saleem Salameh, *Kyova Interstate Planning Commission*
Robert Santos, *Urban Institute*
Heidi Schallberg, *Mid-America Regional Council*
Robert Schiffer, *Cambridge Systematics, Inc.*
Elizabeth Schrey, *West Florida Regional Planning Council*
John Sharp, *Association of Central Oklahoma Governments*
Adnan Sheikh, *Georgia Institute of Technology*
Huiwei Shen, *Florida Department of Transportation*
Frank Southworth, *Oak Ridge National Laboratory*
Edward Spar, *Council of Professional Associations on Federal Statistics*
Bruce Spear, *Cambridge Systematics, Inc.*
Nanda Srinivasan, *Transportation Research Board*
Wu Sun, *San Diego Association of Governments*
Kevin Tierney, *Bird's Hill Research*
Mary Lynn Tischer, *Federal Highway Administration*
Jennifer Toth, *Arizona Department of Transportation*
Charles Trainor, *Community Planning Association of SW Idaho*
Katherine Turnbull, *Texas Transportation Institute*
Martin Tuttle, *California Department of Transportation*
Krishnan Viswanathan, *Wilbur Smith Associates*
Madeline Wander, *University of California, Los Angeles*
Xin Wang, *Old Dominion University*
Kendra Watkins, *Mid-Regional Council of Governments*
Penelope Weinberger, *American Association of State Highway and Transportation Officials*
Frank Wen, *Southern California Association of Governments*
Lynn Wombold, *ESRI*
Beth Xie, *Regional Transportation Commission of Southern Nevada*
Jinghua Xu, *Parsons Brinckerhoff*
Ming Hsun Yang, *Institute of Transportation Studies*
Xin Ye, *University of Maryland, College Park*
Supin Yoder, *Federal Highway Administration*

TRB

TRANSPORTATION REASEARCH BOARD
500 Fifth Street, Nw
Washington, DC 20001

www.TRB.org