Volume 8 Number 3, 2005
ISSN 1094-8848

SPECIAL ISSUE ON THE NATIONAL HOUSEHOLD TRAVEL SURVEY

JOURNAL OF TRANSPORTATION AND STATISTICS

U.S. Department of Transportation
Research and Innovative Technology Administration
Bureau of Transportation Statistics

JOURNAL OF TRANSPORTATION AND STATISTICS

A PEER-REVIEWED JOURNAL

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 8 Number 3, 2005
ISSN 1094-8848

U.S. Department of Transportation
Research and Innovative Technology Administration
Bureau of Transportation Statistics

The *Journal of Transportation and Statistics* releases three numbered issues a year and is published by the

Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
400 7th Street SW, Room 4117
Washington, DC 20590
USA
journal@bts.gov

Subscription information

To receive a complimentary subscription:

mail Product Orders
Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
400 7th Street SW, Room 4117
Washington, DC 20590
USA

phone 202.366.DATA

internet www.bts.dot.gov

Information Service

email answers@bts.gov

phone 800.853.1351

Cover and text design Susan JZ Hoffmeyer

The Secretary of Transportation has determined that the publication of this periodical is necessary in the transaction of the public business required by law of this Department.

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 8 Number 3
2005

Contents

Introduction to the Special Issue on the 2001 National Household Travel Survey

BACKGROUND

Policymakers rely on transportation statistics, including data on personal travel behavior, to formulate strategic transportation policies and to improve the safety and efficiency of the U.S. transportation system. Data on personal travel trends are needed to examine the reliability, efficiency, capacity, and flexibility of the nation's transportation system to meet current and future demands; to assess the feasibility and efficiency of alternative congestion-alleviating technologies (e.g., high-speed rail, magnetically levitated trains, and intelligent vehicle and highway systems); to evaluate the merits of alternative transportation investment programs; and to assess the energy use and air quality impacts of various policies.

To address these data needs, the U.S. Department of Transportation (USDOT) initiated an effort to collect detailed data on personal travel. In 1969, the first Nationwide Personal Transportation Survey (NPTS) was conducted. The survey was repeated in 1977, 1983, 1990, and 1995. In 2001, an expanded survey included both daily and long-distance travel. In essence, the 2001 survey combined the NPTS and the 1995 American Travel Survey and was renamed the National Household Travel Survey (NHTS).

Three USDOT agencies sponsored the 2001 survey: the Federal Highway Administration, the Bureau of Transportation Statistics, and the National Highway Traffic Safety Administration. The primary objective of the survey was to collect trip-based data on the nature and characteristics of personal travel so that the relationships between the characteristics of personal travel and the demographics of the traveler could be established. Commercial travel was not part of the survey.

For the 2001 survey, the national sample consisted of 26,000 households. Four states and five planning areas purchased over 43,000 additional samples. These "add-ons" increased the number of sample households in these state/planning areas so that trip rates and travel statistics could be estimated more reliably at that geographic level.

Interviews were attempted with all members of sample households. Each was asked to provide detailed information on their daily and long-distance travel. They were also asked to provide information about their household, its members, and vehicles. Data about every one-way trip taken by each household member during a designated 24-hour period (the household's designated *travel day*) and data describing every roundtrip of 50 miles or more from home taken by each household member during a four-week period (the household's designated *travel period*) were collected.

SUMMARY OF TOPICS IN THIS SPECIAL ISSUE

This special issue begins with Erlbaum's comprehensive analysis of the quality of the 2001 NHTS data and comparisons between 2001 NHTS data and data from other sources (e.g., traffic count programs and administrative records). Although Erlbaum's analysis was based primarily on the New York add-on sample, it helps illuminate how 2001 NHTS data can be used and/or integrated with other data sources.

Understanding the interactions between land use and travel is critical to designing balanced transportation systems. With the wealth of information in the 2001 NHTS, Scuderi and Clifton used a Bayesian approach to improve, in relation to previous research, complex quantitative relationships between land use and transportation. Specifically, the authors apply Bayesian belief networks (BBNs) to analyze complex spatial systems (e.g., the urban environment). They demonstrate that this approach does not rely on ad hoc statistical models or assumptions. As such, there is no need to characterize variables as independent or dependent. Although limited results are presented, this paper identifies future opportunities where BBNs could provide insights to effectively address complex transportation issues.

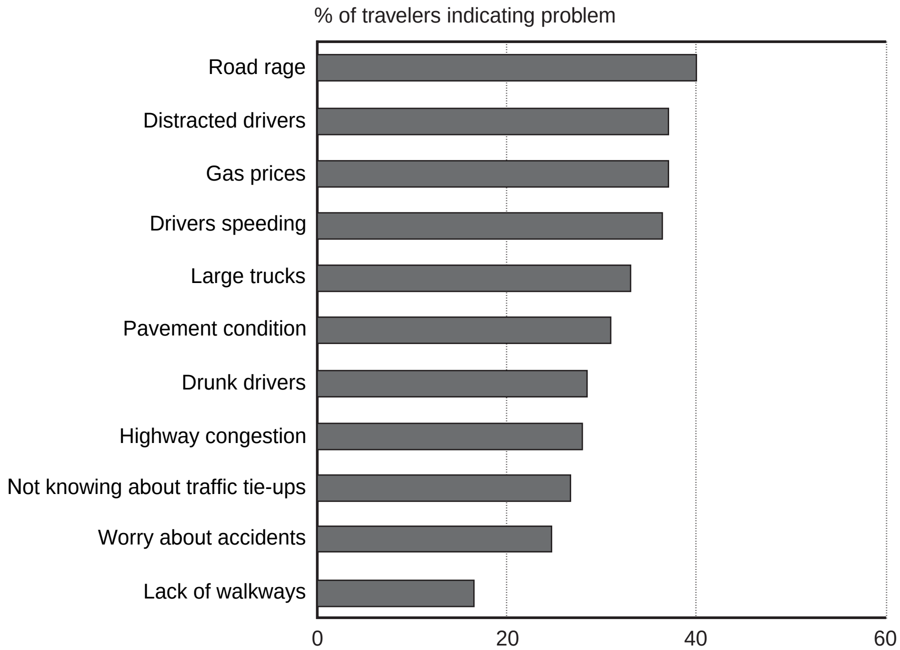
Polzin and Chu analyze 2001 NHTS data and other data sources to take a closer look at the national trends in public transportation mode share. Their research shows that as the growth in overall national travel has slowed, transit use appears to have fluctuated. This was the case in terms of both absolute trips and transit's share of overall travel. The research also identifies the shortcomings and differences of the various data sources for determining transit use and mode share trends. The authors recommend synthesizing multiple measures of mode share to identify data needs and to provide a knowledge base to inform national policy deliberations.

Special attention was given in the 2001 survey to prompt the responders about their walk and bike trips, contributing to a 60% increase in reported walking trips from 1995 to 2001. Using 2001 NHTS data for the Baltimore metropolitan region, Targa and Clifton present an empirical analysis of the effects of several land-use, urban form, and neighborhood-level design attributes, as well as traveler attitudes/perceptions of the urban system, on the frequency of walking and the share of walking trips relative to total trips. Their results suggest that people who walk more frequently live in neighborhoods with higher densities, more diverse land-use mixes, better street connectivity, and better access to bus transit lines.

While the NHTS is an invaluable resource to estimate household travel statistics at the *national* level, it is not advisable to use NHTS data to estimate travel statistics for geographic areas smaller than a Census Division. To meet the data needs for areas smaller than a Census Division, Mei et al. used a Bayesian updating approach to estimate travel statistics at the statewide level and for counties and rural areas.

At a time when budgets do not allow the collection of adequate sample sizes, Stopher and his colleagues demonstrate that Monte Carlo simulation with Bayesian updating could be a reasonable alternative to a full-scale household

FIGURE 1 Concerns About the Quality of Transportation Services: 2001



Source: Oak Ridge National Laboratory.

travel survey. Furthermore, the preliminary results obtained by the authors suggest that 2001 NHTS data could be suitable for Monte Carlo simulation of household tours.

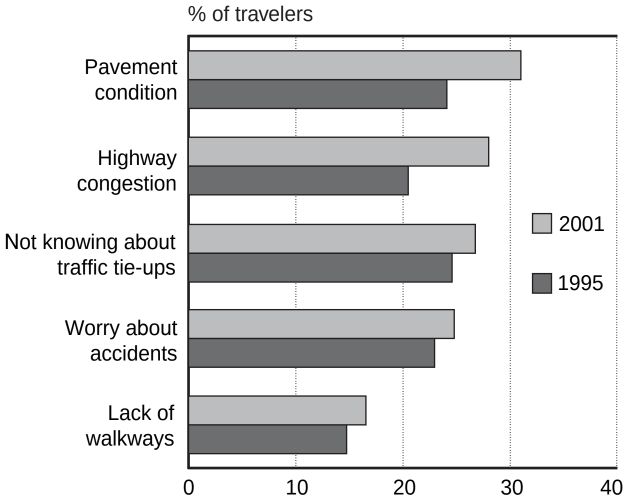
The last paper in this Special Issue, by Sharp and Murakami, suggests some methodological considerations for future household surveys. In light of technology trends (e.g., cell phones and web utilities) and sociodemographic changes, the authors offer considerations to ensure data quality, increase response rates, and sustain survey efficiency. These considerations can assist the transportation community in designing and implementing future surveys in a more effective and efficient manner.

OTHER CORE DATA

Space limitations in this Special Issue precluded comprehensive coverage of the 2001 NHTS. A few examples of core data we could not include are the amount of time Americans spend in their vehicles on a typical day, how Americans view the quality of our nation's transportation services, how these views have changed over time, and how and what different subpopulations require of the nation's transportation services. For example, Americans considered road rage as the most serious transportation problem in 2001, followed by distracted drivers and high gasoline prices (figure 1).

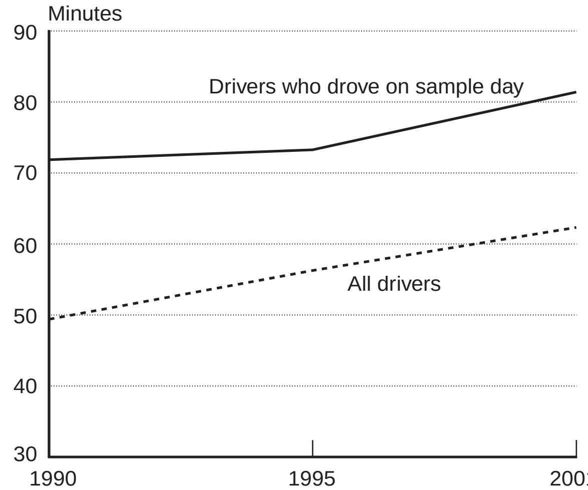
Americans felt that the quality of our nation's transportation services had not improved since the 1995 survey. For those quality aspects included in both the 1995 and 2001 surveys, pavement conditions and highway congestion were the top two concerns (figure 2). The growing discontent over highway congestion was partially substantiated by the increasing number of minutes spent driving, which rose to more than an hour a day for a typical driver (figure 3).

FIGURE 2 Percentages of Travelers Indicating a Significant Transportation Problem



Source: Oak Ridge National Laboratory.

FIGURE 3 Minutes Spent Driving per Day



Source: Oak Ridge National Laboratory.

Readers can explore the potential of the 2001 NHTS by visiting its website (http://nhts.ornl.gov) and by sharing research findings with others. Conference papers are available at http://www.trb.org/Conferences/NHTS/Program.pdf.

ACKNOWLEDGMENTS

The articles in this Special Issue were among the many presentations made at the *National Household Travel Survey Conference: Understanding Our Nation's Travel*, held at the National Academy of Sciences in Washington, DC, November 1–2, 2004. This conference served as a forum for the various data users of the national survey to discuss analysis and findings from the recently released 2001 NHTS. We wish to thank the Transportation Research Board for organizing, and the Bureau of Transportation Statistics and other agencies for providing funding and support to, the conference.

This issue would not have been possible without the dedication, expertise, and objectivity of the numerous referees. Their insight and careful reviews helped advance the quality of every article. We also wish to thank the journal's Editor-in-Chief, Peg Young, and the publishing staff: Marsha Fenn, Alpha Glass, and Dorinda Edmondson. Their advice, commitment, and professionalism made it much easier to undertake the editing and compilation of this issue.

Pat Hu
Guest Editor
Center for Transportation Analysis
Oak Ridge National Laboratory

Lee Giesbrecht and Joy Sharp
Guest Co-Editors
Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation

Assessment of 2001 New York State NHTS Add-On Data Using Empirical and Auditable Data Sources

NATHAN ERLBAUM

Policy and Strategy Division
Office of Policy and Performance
New York State Department of Transportation
50 Wolf Road, 6th Floor
Albany, New York 12232

ABSTRACT

This study assesses how well the 2001 National Household Travel Survey (NHTS) estimates compare with other sources of comparable data from the census or administrative records. Comparisons of a number of NHTS measures are made with benchmark data sources and brief findings presented. Traffic count-based vehicle-miles of travel (VMT) estimates of residential travel show that monthly patterns for survey VMT are inconsistent with observed statewide ground-count estimates; residential-based VMT is not comparable to total VMT, but by using additional data to specify nonresidential and commercial VMT, it is possible to reach the traffic count estimate for total VMT. For transit ridership, NHTS person trips by subway correspond well to Metropolitan Transportation Authority reports of subway trips. There is general agreement for estimates of workers, but some geographies within New York State show statistical differences between the two surveys. For status of drivers with DMV licenses, agreement exists, but some strata and gender groupings show statistical differences between the two measures. For the number of registered household vehicles, for most strata, the census estimates are within the NHTS 95%

KEYWORDS: National Household Travel Survey (NHTS), confidence interval, census, quality assessment.

lower bound and the estimate. The study presents findings and recommendations based on these comparisons, as well as observations as to how important a role standard error and confidence interval play in the analysis of survey results.

INTRODUCTION

Conducting and analyzing a household travel survey is a fairly common mechanism for obtaining information about travel behavior and characteristics of the household, including its members, their trip making activity, and vehicle usage. The 2001 National Household Travel Survey (NHTS) is the latest in a series of roughly quinquennial residential-based household travel surveys undertaken by the Federal Highway Administration (FHWA).

This study assesses the quality and usefulness of the data for the New York State Department of Transportation's (NYSDOT's) purposes. Comparisons to other data sources used by the NYSDOT are made and differences noted. The paper focuses on a discussion and assessment of selected survey measures from the 2001 NHTS NYS add-on and how well they compare with data drawn from the 2000 Decennial Census, the 1997 Vehicle Inventory and Usage Survey (VIUS), transit operator annual reports, summary data from continuous traffic counting sites, and other sources available to NYSDOT.

The paper examines highway travel from two perspectives: 1) temporal trends in the NHTS are compared with NYSDOT ground-count-based estimates for the same period; and 2) combining the estimates of NHTS residential travel and VIUS commercial travel with NYSDOT ground-count-based estimates of statewide travel.

The NHTS is a list-assisted random digit dialing survey designed to yield an equal probability sample of households with telephones. The survey is effectively a metropolitan/nonmetropolitan area survey. In a national probabilistic sample of over 25,000 U.S. households, New York State would be represented by its share of the national population or about 1,600 samples. The 1,600 national samples would normally be drawn primarily from New York City and its suburban counties and the Buffalo metropolitan area, because that is where the largest share of the state's population resides.

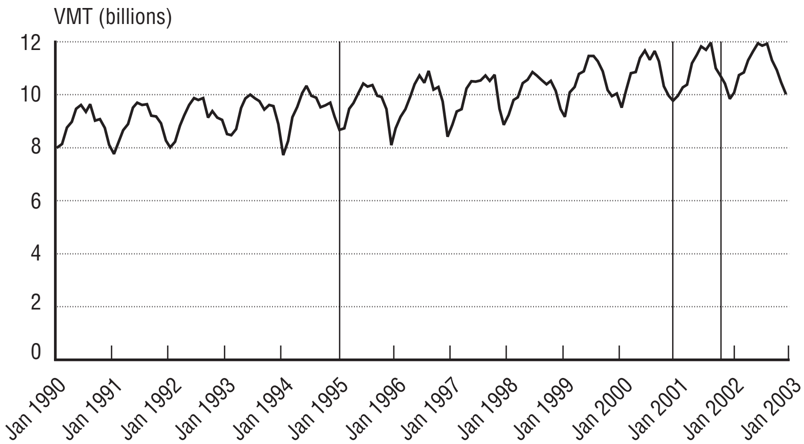
In 1995 and 2001, NYSDOT participated with FHWA as an add-on, purchasing over 11,000 additional household samples. The additional households enable the state to examine travel behavior in the state with greater statistical reliability. The add-on sample also enables NYSDOT to look at the primary urban counties within each metropolitan area, treat the 12 counties in the lower Hudson Valley as if they were each separate areas, and assess the counties not included as urban as part of a non-urban aggregate, thereby providing greater understanding of similarities and differences in travel behavior in different areas of the state.

In 1995, the Research Triangle Institute conducted the survey; in 2001, Westat conducted the survey. While the survey format and national sample size is essentially similar, differences exist between the two surveys with respect to nonresponse adjustment, access and egress modes for public transportation, the method for annualizing the odometer readings, and perhaps, most importantly, how the events of September 11, 2001 (9/11) impacted the survey and/or reflect some type of permanent travel behavior alteration in NYS.

The primary objectives of this report are to:

- Validate that the survey temporal distribution of residential household personal travel over the year was consistent with the distribution obtained from ground-count data-collection efforts and that it also exhibited similar patterns in the post 9/11 period.
- Understand how residential household personal vehicle-miles of travel (VMT) might have remained constant between 1995 and 2001 in light of the 13% increase in statewide VMT over the same period.
- Validate that the NHTS as a total travel survey adequately reflected public transit ridership and the significant growth in unlinked transit trips reported by the Metropolitan Transportation Authority (MTA) in the New York metropolitan region between 1995 and 2001.
- Assess how well the NHTS estimates for workers, drivers, and vehicles available in households compared with the 2000 Census and motor vehicle records.

FIGURE 1 Monthly Vehicle-Miles of Travel (VMT) Estimate



Source: New York State Department of Transportation, Office of Policy and Performance.

TEMPORAL TRENDS IN TRAVEL SURVEY VS. GROUND-COUNT ESTIMATES OF VMT

In this section, the monthly distribution of vehicle travel is compared using two sources—the continuous monitoring ground-count data supplied to NYS and the 1995 Nationwide Personal Transportation Survey (NPTS) and 2001 NHTS (USDOT 1995 and 2001). Some differences in these two sources are worth noting.

Temporal ground counts illustrate trends in highway travel for residents and nonresidents: the use of private and commercial vehicles as well as public transit vehicles. The counts reflect intrastate travel as well as cross-state and interstate travel and include both work-related and discretionary travel.

The vehicle trip data from the NPTS/NHTS reflect residential household personal travel. A yearlong summary of all vehicle trips for all purposes was sampled, with each household reporting for a single "travel day." In reality, these estimates reflect primarily local vehicle travel—travel that probably does not change much from day to day as people go about their lives and daily business and errands. The survey does not include commercial travel, travel by visitors to NYS residents, or information about nonresidents who work or travel into or through NYS on a daily basis.

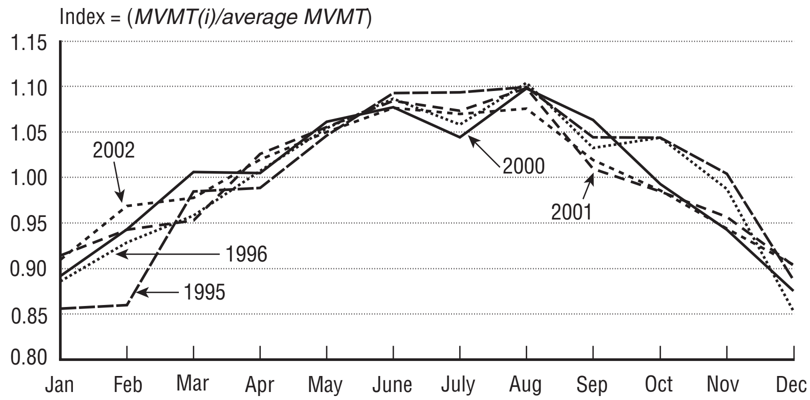
A plot of monthly estimates for statewide VMT based on data from the continuous monitoring sites within NYS is shown in figure 1. A vertical line is provided for reference to indicate January 1995 and 2001, and September 2001.

The slope of travel shows an increasing trend from 1995 through 2002. The arc-like pattern within the years may reflect the impact of weather on the travel season. However, the overall trend nonetheless shows upward change from a 1995 summer peak of 10.5 billion VMT to a summer peak of almost 12 billion VMT in 2001, reflecting an annualized increase in vehicle travel of 2.25% per year over the 6 year period (NYSDOT OPP 2003d).

Examination of the ground-count data on vehicle travel for the survey years and those between the 1995 NPTS and the 2001 NHTS (figure 2) provides the following temporal presentation. In all five years, there is a significant decline in vehicle travel in September relative to August as vacations end and school starts each fall. In 1995 and 1996, there were increases in October, which may be more a reflection of unseasonably warm weather causing travel to remain higher. Figure 3 shows a comparison of the VMT estimates for the United States (for 2001/2002) and NYS (1995/1996 and 2001/2002) and illustrates a similar pattern (USDOT 2003).

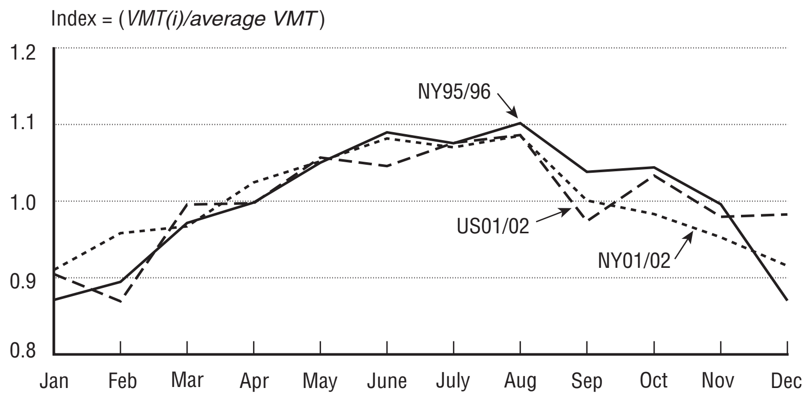
Perhaps more interesting is what happens when the NPTS/NHTS survey is used as a surrogate for vehicle travel. In the NPTS/NHTS series, the number of completed samples associated with each calendar month was weighted to reflect each month as 1/12 of the total number of samples. If the survey period spanned the same month, say there are two

FIGURE 2 Temporal Distribution of Monthly Vehicle-Miles of Travel (MVMT)



Source: New York State Department of Transportation, Office of Policy and Performance.

FIGURE 3 Average 2001/2002 and 1995/1996 Statewide Vehicle-Miles of Travel (VMT) Temporal Distribution



Source: New York State Department of Transportation, Office of Policy and Performance; and U.S. Department of Transportation, Federal Highway Administration.

months of April samples, then the sum of those samples was combined and made 1/12 of the total. As noted in the ground-count data, vehicle travel is not evenly distributed across the months. The weighting method should correctly estimate the pattern of monthly travel across the year.

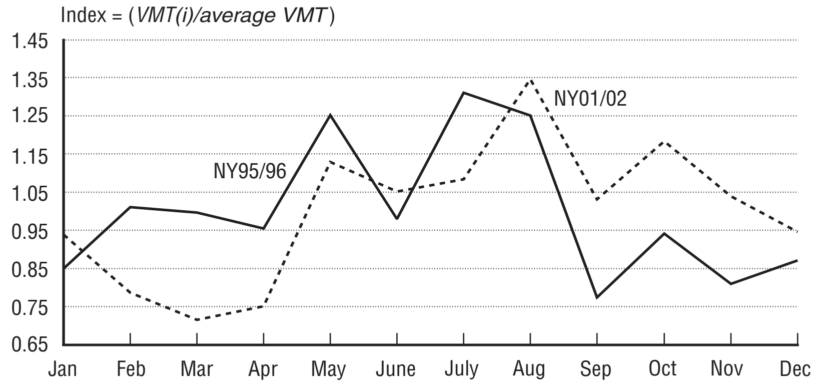
Figure 4 is based on the statewide summary for NYS of vehicle trips from the NPTS/NHTS add-on data by month. On the whole, the curves tend to reflect a seasonal pattern, but are not quite in agreement with observed statewide travel from ground counts (NYSDOT OPP 2003e). It is possible that the trough in vehicle travel from February to April in the 2001 NHTS is consistent with the decline noted in seasonal travel due to weather. Equally worthy of note is the more significant decline in September between the two surveys when compared with the actual ground-count data, which may reflect the events of 9/11. More importantly is the depth of the decline in 1995/1996, which is more likely related to response rate and survey temporal and other adjustments rather than catastrophic events.

Weather Effects

Assuming that the propensity to travel is directly related to how temperate the weather is, figure 5 shows heating degree days.[1] In the figure, December, January, and February 1995/1996 were colder than the winter of 2001/2002, reflecting a pattern consistent with the ground-count-based travel data (NYSERDA 2004).

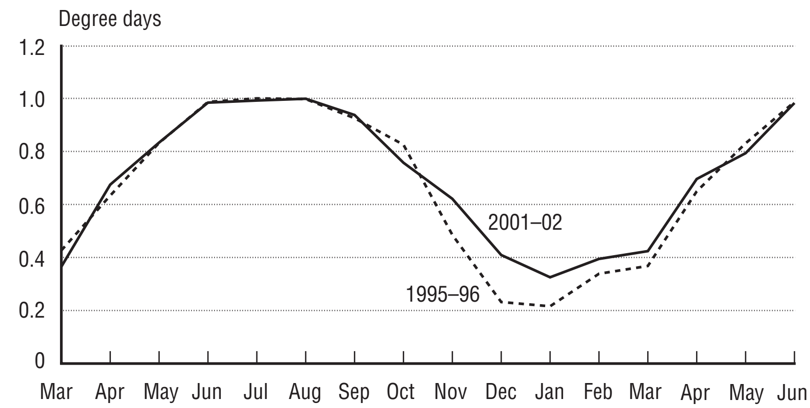
[1] A degree day = 65 minus the days average temperature ((high + low) / 2).

FIGURE 4 Survey of Statewide Vehicle-Miles of Travel (VMT) Temporal Distribution



Source: New York State Department of Transportation, Office of Policy and Performance.

FIGURE 5 Heating Degree Days



Source: New York State Energy Research and Development Authority.

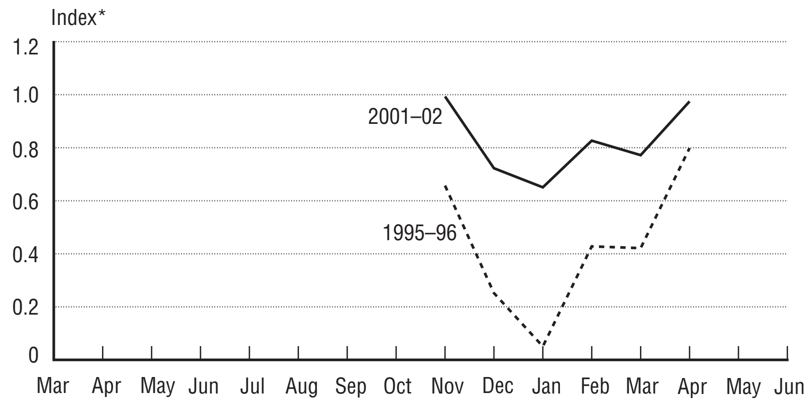
Temperature severity alone does not preclude travel. The amount of precipitation and snowfall during the winter may provide more of an explanation for a variation in travel patterns across years.

In figure 6, the population weighted statewide snowfall by NYSDOT residency was adjusted to reflect its potential impact on the propensity to travel. The graph would suggest that the snowfall impact on travel during the winter months of 1995/1996 was more severe than in 2001/2002. However, both this observation and the ground-count data appear to contradict the survey's estimate of monthly seasonal travel (NYSDOT OOM 2003). Perhaps what the survey is showing is that people still go about their activities even when it snows. It could also be that the amount, duration, and timing of the snow event is more likely to impact travel than just the total snowfall amount. At the present time, access to specific data to test this hypothesis is unavailable within NYSDOT.

Effects of September 11, 2001

Clearly, 9/11 had an impact on the 2001 NHTS; however, the impact is unclear. Anecdotally, it has been suggested that the public hunkered down and stayed home after 9/11 and the anthrax scares; this is not reinforced by ground-count observations. In an examination of the impacts of 9/11, NYSDOT found that the severity of the vehicular traffic impacts were short in duration, with daily travel returning to almost normal by late October and highway travel in general returning to annual seasonal patterns by the end of the year.

FIGURE 6 Population Weighted Snowfall Adjusted for Impact on Propensity to Travel



$$\text{* Index} = \left(40 - \sum_{counties} (\text{precipitation in inches} \bullet (\text{county pop} / (\text{state pop})))\right) / 40$$

where 40 represents the rounded statewide average for December, January, and February 1995–1996, the months where precipitation is likely to be snow.

Source: New York State Department of Transportation (NYSDOT), Office of Policy and Performance, based on data from NYSDOT, Office of Operations Management.

If travel in 2000 is considered a normal year, then the effect of 9/11 on travel was a 6% reduction statewide in that month. October travel showed a 2% reduction. However, by November and December 2001, highway travel resumed at a greater value than in 2000 (NYSDOT OPP 2002). This is perhaps, in part, due to the avoidance of air travel and the shift to highway for longer distance trips over the Thanksgiving and Christmas holidays and more temperate weather.

Another observation of this study was that the farther the travel was from “ground zero,” the less the impact. The most significant impacts were observed where specific traffic restrictions were in place or where facilities were closed. Air travel showed significant long-term declines, with observational and anecdotal evidence showing that long-distance vehicle traffic was increasing in the short-range air travel corridor.

In first quarter of 2002, the decline in survey base travel for NYS residents can perhaps be explained by the economic impacts of much less business travel and the loss of jobs due to the deepening recession. Another possibility is the national-level survey adjustment for monthly variation may have in and of itself masked the actual travel for NYS.

ASSESSING TOTAL HIGHWAY TRAVEL: SURVEY VS. GROUND COUNT

It may not be possible to expect a resident household personal travel survey such as the NPTS/NHTS to form the basis for assessing temporal variation in vehicular travel. Limitations in sample size, weighting, and nonresponse adjustments may work effectively to address issues associated with the sampling unit; however, it is not clear how to correctly adjust for design effect variables. It is equally unclear how to account for all of the other vehicular travel that is not residential household personal transportation traveling to, from, and across NYS that is measured by ground-count means.

The following discussion attempts to assess how the survey estimate of total residential-based travel may exist within the context of a ground-count-based estimate of total highway travel in NYS. The approach relies on related surveys, studies, pseudo and empirical data, and a number of enabling assumptions based on observation, anecdotal data, and local conditions.

When examining survey data, it is important to take into consideration the impact of sampling error. Sample size has a considerable impact on sampling error. While a large number of samples were taken in NYS as part of the NHTS add-on, it is possible that sampling error is sufficiently large to preclude detection of small changes between 1995 and 2001. In order to examine this, consideration must be given to the confidence interval associated with an estimated value of a measure (e.g., VMT) and view that measure as being within a certain range of values above or below the true value. In the case of a survey such as the NPTS/NHTS, the estimated value is represented by the weighted sample data.

Assuming that the 2001 NHTS produces an unbiased estimate of residential household VMT (because it is from a random sample), we can be confident with 95% certainty that the survey estimate does not differ from the true residential-based VMT by more than twice the standard error in either direction. In 2001, the estimated statewide residential-based VMT value was 95.2 billion. There is a 95% certainty that the true value of residential-based VMT lies in the interval from 91.6 billion to 98.8 billion VMT, a relative error of 1.9%. The estimated VMT for 1995 was 95.6 billion. There is a 95% certainty that the true value of statewide residential-based VMT lies in the interval from 91.2 billion to 100.0 billion VMT.

Looking at the sampling error on the estimate of the VMT from both the 2001 and 1995 surveys, there is a 95% certainty that the true value of residential-based VMT lies between 89.2 billion and 100.9 billion VMT (USDOT 1995 and 2001). Therefore, no statistically significant difference exists in the two estimates.

However, given the inability to discern any difference, we could interpret this as follows:

- residential household-based VMT may have grown by as much as 5 billion, or 5% relative to 1995;
- the 1995 estimate could be an overstatement of 6 billion;
- true growth could be as large as 11 billion.

For the 2001 NHTS in NYS, a much more detailed analysis for the different urban strata shows that the relative errors are much larger, ranging from +/–2.6 to +/–19.7%, compared with +/–2.0% for the state as a whole.

Assessing Travel

Given the level of uncertainty with the survey's estimate of residential household VMT, can it be resolved with the ground-count estimate of travel? Consider the following example using the 2001 NHTS and the 1997 VIUS. With careful examination of the VIUS for estimates of personal and nonpersonal transportation VMT and with adjustments based on the assumptions noted, it is possible to construct an estimate using the upper bounds of the confidence limit almost equal to the 130 billion statewide ground-count VMT estimate in 2001 (NYSDOT OPP 2003a).

Table 1 shows a substitution of the 2001 NHTS trucks that are available for personal use within a household for those in the VIUS. Trucks that may be considered available for commercial use at an establishment in the VIUS are then added to the NHTS. This allows the NHTS to specify residential-based VMT and the VIUS to specify resident-based commercial VMT. Using a series of assumptions and other data sources, adjustments are made to compensate for the following: survey error; difference in time period; the flow of both personal and commercial vehicles in, out, and across the state; NYS public transit and school buses; and the recognition that NYS is a net importer of goods and services for vehicles not registered in the state.

The Difficulty with Small Trucks

Pickups, vans, and other truck vehicles that can be used for transporting both people and goods are difficult to quantify from any source. Small home-based businesses can use a vehicle for both personal and commercial travel in the same day, sometimes in the same trip. The NHTS asks about commercial vehicle use and obtains the occupation code of the vehicle user, but if more than 10 commercial trips are made on the travel day (e.g., taxis or police cars) the survey asks the respondent to report only the trips for personal use of the vehicle.

In New York State as in other states, pickups, small vans, and sport utility vehicles (SUVs) may be registered as either cars or trucks depending on usage (e.g., pickup trucks with a permanently attached cap are registered as standard passenger series vehicles). In the 1997 VIUS, passenger car files

TABLE 1 Total Travel by Residents, Businesses, and Vehicles

VIUS97 categories	VIUS97 VMT	NHTS01 categories	NHTS01 VMT	VIUS97 VMT adjusted to 2001		95% upper CL VMT estimate	Notes
	Private and commercial truck VMT (000,000)		Residential household personal VMT (000,000)	Estimated personal VMT (000,000)	Estimated commercial VMT (000,000)	VMT (000,000)	
		Total	95,207.1			131,002.5	Sum + f,g,i
		Auto	62,795.9			67,415.2	a
		Motorcycle	315.1			471.3	a
Total	36,396.6				37,203.9		
Heavy trucks	3,998.7				16,267.5	16,267.5	d,e,h
Light trucks	32,397.9	**Light truck**	32,081.9	25,449.4	20,936.4		
Pickup	10,663.9	**Pickup**	9,119.8	8,424.6	6,851.3	15,276.0	b,c,e,h
Panel/van	3,526.8	**Not pickup**	22,962.0	17,024.8	14,085.1	31,109.9	b,c,e,h
Minivan	7,311.3	**Van**	9,538.2				
		Other truck	2,108.5				
SUV	9,902.5	**SUV**	11,315.3				
Station wagon	993.3						
		RV	14.2				

Key: CFS = Commodity Flow Survey; CL = confidence limit; DOT = Department of Transportation; NHTS = National Household Travel Survey; NYS = New York State; RV = recreational vehicle; VIUS = Vehicle Inventory and Use Survey; VMT = vehicle-miles of travel.

Notes:

a. Assume 95% upper CL estimate of NHTS01 personal VMT.

b. NHTS01/VIUS97 ratios for personal travel for pickup, nonpickup used to grow 2001 commercial component of (VIUS stratum 1,2).

c. Assume VMT contribution of non-NYS pickup + panel/van to be at least equal to NYS value.

d. Assume that VIUS97 stratum 3–4 vehicles are illogical as personal transportation choice and are considered commercial use.

e. Assume VIUS97 heavy-truck VMT estimate for surrounding states is equal to NYS estimate, since CFS shows 70% of all NYS origin movements are less than 50 miles; add to this the Reebie TranSearch 2001 primary truck shipment assigned network VMT in NYS with equal empty backhaul estimate (no estimate for secondary and/or transshipments are then made).

f. NYSDOT operating assistance-based bus transit VMT (000,000), which assumes 10% addition for deadhead miles 257.9

g. NYS school bus VMT (000,000) 204.8

h. Census provisions for state-based relative error computation are not provided. U.S. error is significantly less than that for the NYS sample. Twice the value of the U.S. relative error for truck-miles by operating class were used for NYS tables of annual miles by *opclas* and *samtyp* to compute upper 95% CL estimate.

i. Total includes 14.2 million VMT for RV not included in any of the other table rows.

Sources:

New York State, Department of Education, estimates of school bus VMT, available at http://www.schoolbusfleet.com and http://www.schoolbusfleet.com/Stats/pdf/SBFFB03p29-30.pdf, Total Route Mileage.

New York State Department of Transportation, Office of Policy and Performance data.

______. Passenger Transportation Division, Annual report on public transportation assistance programs and specialized tabulation on bus VMT.

New York State Thruway data.

Reebie & Associates, TranSearch data, Custom tabulations, April 2003.

U.S. Department of Commerce, U.S. Census Bureau, Vehicle Inventory and Use Survey data.

______. U.S. census data.

U.S. Department of Transportation, Federal Highway Administration, National Household Travel Survey data.

U.S. Department of Transportation, Research and Innovative Technology Administration, Bureau of Transportation Statistics, Commodity Flow Survey data.

were searched and any such vehicles were included in the VIUS sampling frame along with truck registrations. Therefore, the 1997 VIUS contains both personal use and commercial use trucks. This is also the case in the 1995 NPTS and 2001 NHTS. The survey does not ask for the type of registration for household-based vehicles.

In this comparison the non-NYS pickup/panel/van estimate is assumed to be at least equal to the NYS value. In analyzing the 1995 NPTS NYS add-on, the NYS metropolitan area strata were compared against the nation as a whole. Based on this analysis, we found that with the exception of New York County (Manhattan) and the remainder of New York County; the rest of the state had travel characteristics similar to that of the nation. Additionally, given the substitution effect of pickups, panels, vans, and SUVs for autos and their usage similarities, this assumption is necessary to address in migration of nonresident vehicles in border areas.

Travel from Outside NYS

For nonresident travel, other assumptions are necessary given the expanse of the multistate New York City (NYC) labor market area, where residents and businesses in northern New Jersey and western Connecticut regularly engage in travel and business activities in NYC and its suburban counties. Equally important is the considerable daily passenger and truck traffic in western New York between Canada and NYS, which is clearly evident by Canadian-New York border crossing counts, neither of which are measured by the VIUS nor the NHTS.

Discussions with staff at the NYS Thruway (TWY) indicate that consultant studies done in the mid-1990s showed a nonresident presence on the TWY in excess of 30% (Maynus 2004). The reader should note that the TWY is mostly a rural road skirting many of the major urban areas it traverses and carries less than 10% of the state's VMT. More recent studies in 2004 in the highly urbanized NYC metro area related to the TWY Tappen Zee Bridge/I-287 corridor indicate a significantly higher proportion of nonresident usage. For three primary facilities that cross the Hudson River (where tolls are collected in the east bound direction only)—the Tappen Zee Bridge, the George Washington Bridge, and the Lincoln Tunnel—the daily nonresident share of the vehicular flow was 67.5%.

To address nonpersonal vehicle usage, a number of assumptions were also made. Doubling the assignment of Reebie TransSearch fully laden trucks conservatively accounts for empty backhauls and less-than-truckload movements by large long-distance trucks to, from, and across NYS (Reebie 2001). The VIUS estimate of nonresident truck movements (given the number of vehicles crossing the Hudson River) is, at a minimum, at least equivalent to that of NYS and, absent other data, is distributed equally across all vehicle categories.

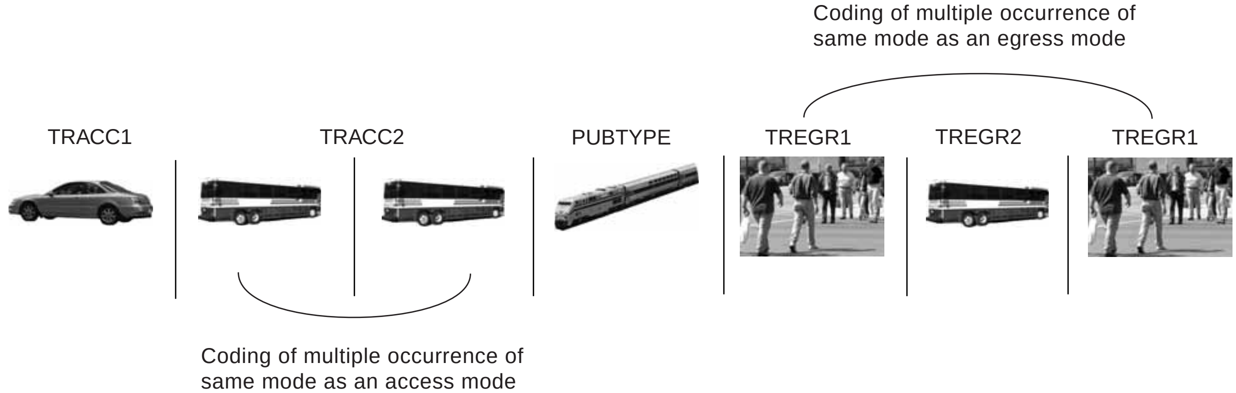
Adjustments for the VIUS relative error using national values will understate the error associated with the smaller NYS sample, hence twice the national relative error is assumed for NYS.[2] It should also be noted that because NYS is a net importer of goods and services as demonstrated by the Commodity Flow Survey (CFS) data,[3] the VIUS will not adequately represent the vehicles entering NYS. Equally important are adjustments for things that cannot be measured, are addressed based on anecdotal data, or that will likely overstate the longer distance movements of trucks.

Taking all of the above issues and assumptions into consideration and the reality that the ground-count-based estimate of 130 billion statewide VMT may in and of itself have perhaps a +/–5% error or be +/–6.5 billion VMT off, we can reach the 130 billion statewide VMT estimate. By iteratively back solving and/or adjusting assumptions, the estimates also lie within the range of 123 billion to 136 billion VMT.

It may be possible, therefore, to make estimates with a variety of survey resources that come close to ground-count-based estimates of total statewide VMT as reported in the Highway Performance Monitoring System (HPMS).[4] This approach is tenuous at best, as there are equally problematic issues associated with ground-count expansion, vehicle classification, other issues that affect the HPMS, and estimation of travel not adequately covered by existing surveys that are crucial in resolving the

[2] See http://www.census.gov/econ/www/viusmain.html.
[3] See http://www.bts.gov/programs/commodity_flow_survey/.
[4] See http://www.fhwa.dot.gov/policy/ohpi/hpms/.

FIGURE 7 Understanding Main Mode and Access and Egress Mode Change Relationships with Respect to Access and Egress 1–5 Coding



Source: New York State Department of Transportation, Office of Policy and Performance.

regional travel impacts for a bridge state. It is at least possible to accept that residential household personal travel may indeed be at or near the upper confidence level shown in the NHTS and that the nonresident movement may be accounting for the growth observed through ground counts.

ASSESSING PUBLIC TRANSPORTATION RIDERSHIP

The NHTS collects data on trips by all modes of travel, and New York transit trips are well represented in the survey. This section focuses on a comparison of transit ridership between the 1995 NPTS and 2001 NHTS NYS add-ons in relation to reported transit ridership. In order to do so, some discussion of the difference in survey collection for transit trips is necessary. There are some definitional differences in how transit operator ridership may be reported.

- If every time a rider changed modes a fare was required, these individual trips would look like revenue trips collected.
- If free transfers between modes are allowed, then the number of modal trips may differ from revenue-based trips.

Public transportation operators providing regularly scheduled transit services typically report passenger revenue separately and passenger ridership in the form of unlinked trips, thereby accounting for each time a person boards a vehicle. Consider the example where a transit fare card is used and the traveler desires to go from point A to point B, as illustrated in figure 7. The traveler or rider begins at the origin with an auto trip to a bus station, takes a bus trip, and makes a free transfer to another bus. The rider then takes a commuter rail trip, walks to another bus, and arrives at his or her destination. This could be reported on the fare card as four separate unlinked transit trips but in reality represents the collection of three separate fares. In household travel surveys, it becomes very important to pay close attention to what is reported as a transit trip, especially if one desires to compare survey results with auditable transit operator statistics.

The NPTS/NHTS uses the following question sequence to determine the origin and destination and mode used:

- Where are you?
- Where did you go next?
- What mode did you use?

When the mode reported is public transportation, a single main mode is then determined based on the longest distance in the link (although some respondents may have reported the longest time segment) or what the respondent identifies as the main public transportation mode for the trip. Mode changes are recorded as access and egress modes to the main mode (2001 NHTS), or as segmented transit trips (1995 NPTS).

The approach in the 1995 NPTS was to determine if public transportation was used at anytime on a trip, identify the main mode of public transpor-

TABLE 2 Comparison of Survey and MTA Estimates of Transit Trips

	NPTS/NHTS		MTA
	Person trips (000) main mode[1]	Person trips (000) main mode segmented or access/egress trips	Unlinked trips (000)
1995	2,762,973	2,992,455	1,740,655
2001	2,943,206	3,390,937	2,317,786
% change	6.5%	13.3%	33.2%

[1] Includes bus, train, and school bus.

Source: New York State Department of Transportation, NPTS/NHTS tabulations; Metropolitan Transportation Authority (MTA) ridership statistics.

tation, and then classify up to four segments for the use of other modes of transportation. For example, in a trip where an individual drives to the train station, takes a bus, transfers to a bus, takes commuter rail, walks, takes a bus, and then walks to a destination, commuter rail is the longest segment and is coded as the main mode for the transit trip. This public transit trip has more than one segment and would show up in the segmented transit trip file.

Since all transit trips have walk access and egress, these would typically not be coded unless the walk trip was of significant length or between transit modes. A commuter rail trip would show up in the travel day file, and in the segmented file one would find for segments 1–4: auto, bus, bus, and commuter rail. The second bus trip might be lost or the bus-to-bus transfer may get coded as one bus trip because the mode did not change. Although each transit trip would be assumed to have at least three segments (walk access, vehicle trip, and walk egress) a large majority of trips in the segmented file obtained two or fewer segments.

In the 2001 NHTS, a different approach to recording public transit trips was employed. There was no segmented transit trip file for recording the multiple modes associated with transit use. Instead, when public transit (PUBTYPE) was used, the respondent was asked to identify the main transit mode. However, up to five access (TRACC1–5) and five egress (TREGR1–5) modes for this main mode were also available. If successive modes were the same then they were coded as one, because the issue is mode capture and a bus-to-bus change is still bus. Figure 7 illustrates the 2001 method.

Based on the above discussion, it becomes clear that the two surveys are not directly comparable in their estimate of public transit ridership. On the whole, most public transit trips had less than three segments in 1995, indicating that even walk access and egress were poorly reported. Good detail on multiple transit mode trips is not available from the 1995 NPTS.

Table 2 presents a comparison of the relative growth in the number of personal trips on passenger transit when reported as the main mode or the main mode with segmented or access/egress transit trips in the 1995 NPTS and 2001 NHTS for NYS. It also shows the average 1995/1996 (to correspond with the survey period) and 2001/2002 Metropolitan Transportation Authority (MTA) unlinked passenger trips (NYSDOT PTD 2003b). A significant difference can be seen in what was reported in the surveys and what was measured by MTA. It should be noted that, between the 1995 NPTS and the 2001 NHTS, MTA introduced a fare card and a variety of different fare policies to encourage free transfers or offpeak discounts. These policies may have contributed to the large increase in unlinked trips (or modal boardings) reported by MTA.

MTA carries the bulk of all transit passenger trips in NYS. In calendar year 2001, the NYSDOT Passenger Transportation Division reported that MTA provided service for approximately 92% of the statewide transit passengers (NYSDOT PTD 2003a). Adding private operators downstate to MTA's share shows that service was provided to approximately 98.6% of all transit passengers in NYS.

A comparison with MTA operational data is very important (NYSDOT PTD 2003c). With the introduction of the MTA fare card, many bus trips provide free transfers, in part accounting for the increase in ridership due to greater system flexibility. However, from a fare card perspective there is no change in the way subway riders are reported. Sub-

TABLE 3 Comparison of Survey Estimates for Subway Ridership with Actual System Statistics

	1995	2001	Percentage change
System statistics			
MTA subway riders (000)	1,107.9	1,412.5	
Estimated PATH riders from PATH survey (000) (c,d)	57.1	60.0	
Survey estimates			
Main mode subway (000)		1,283.5	
Access/egress 1: subway–main mode subway (000)		83.9	
Access/egress 2–5: subway–main mode subway (000) (a)		6.6	
Access/egress: main mode not subway (000)		35.3	
Total survey: subway		1,409.4	
No subway segments: main mode subway (000)	352.2		
Subway segments: main mode subway (000)	646.3		
Consecutive subway segments: main mode subway (000) (b)	136.2		
Total survey: subway	1,134.7		24.2%

Assumptions:

(a) Consecutive access/egress by subway suggest PATH in 2001.

(b) Consecutive segments include subway transfer and PATH.

(c) In 2001, PATH + subway could be 60,000–80,000 trips per weekday, annualized to approximately 60,000 per day.

(d) Census 1990–2000 county work flow indicates 9.8% growth in New Jersey workers in New York; assume 1995–2000 = 5% then 1995 PATH riders = 57,100.

Sources: New York State Department of Transportation, NPTS/NHTS tabulations.
Metropolitan Transportation Authority (MTA) ridership statistics.
Port Authority Trans-Hudson Corporation (PATH), Monthly passenger traffic by station, various years, obtained by special request of New York State Department of Transportation, Passenger Transportation Division.

way ridership reporting in 1995 and 2001 allows free transfers between trains without being recorded as a boarding, very much the same way that bus transfers are currently counted. Therefore, a more focused analysis of subway ridership statewide was undertaken.

Table 3 presents 1995 NPTS and 2001 NHTS subway ridership for NYS. Subway ridership is unique to the five New York City boroughs. Port Authority Trans-Hudson (PATH) service between Manhattan and New Jersey is essentially subway-like, but NYC residents are astute enough to recognize PATH as a separate and different mode and would likely indicate it as such.

The way the surveys are coded, it is possible to identify whether the trip used public transportation on any portion. It is assumed in 2001 that the occurrence of subway access/egress to *subway as a main mode* may indicate that PATH was used. In 1995 two consecutive segments of subway would indicate the same or a subway transfer.

Data from a 2001 PATH transit survey show origins and destinations based on stops (Eng-Wong Taub 2003). While the PATH survey does not indicate transfers to the subway, a modest assumption of 60,000 trips, based on examination of origin and destination stations, assumed a transfer to the subway from PATH. Since the actual number of PATH to subway riders is not precisely known, nor is it possible to estimate the number of nonresidents who may arrive in NYC by other means for business and tourism, these are not unreasonable assumptions for this analysis. The 1995 value for this number was taken as a reduction in the 2001 value by half of the change in decennial census county workflow from New Jersey to New York, which was 9.8% between 1990 and 2000.

Given these assumptions for comparability in subway ridership between the two surveys, we may conclude that the actual public transit ridership represented by the survey and that from operator records are relatively close in terms of percentage change (26.4% vs. 24.2%). Undertaking this analysis using the NHTS confidence intervals for these same data would most likely indicate that the survey estimates easily accommodate the operator statistics for subway ridership. It is then possible that the survey may provide a representative estimate of transit ridership when the problem of unlinked trips is controlled.

ENUMERATING THE WORKFORCE

A U.S. Census Bureau report (Clark et al. 2003) makes the following observations in the executive summary with respect to employment:

- Lower counts of employed people (and the civilian labor force) in censuses than in the Current Population Survey (CPS) extend back to 1950, but in 2000 the differences between the census and the CPS were larger than in the past. The 2000 employment data may be influenced by anomalous data for individuals in group quarters. (For a discussion of employment data for group quarter populations, see USDOC 2000, pp. 960–961.)
- The 2000 census estimate of the number of employed people was about 5% lower than the CPS estimate. But the 2000 census estimate of the number of unemployed people was over 50% higher than the CPS estimate.
- The 2000 census estimate of the labor force participation rate was 2.1% lower than the CPS estimate. The Census unemployment rate was 2.1% higher than the CPS.

It is possible that during the collection of the 2000 census the temporary field interviewers concentrated more on getting "complete count" data and, therefore, were less likely to get all long-form questions completed, resulting in a lot of missing data that was later filled in by imputation. Examination of SF3 Table P132: "Imputation of Work Status for Persons Age 16 and Over for New York State," shows that the numbers generally hover around 12% (NYSDOT OPP 2003b). Table 4 indicates that the percent imputation can be higher for the aggregated county data associated with each of the NHTS add-on strata in NYS. In fact, within the five boroughs of New York City, which represent a population of over 8 million persons, the level of worker imputation is 10% or higher.

An internal analysis conducted by FHWA of specific census tracts in Washington, DC, found tracts where the percent imputed varied from 30% to 88%, especially in poorer neighborhoods and among specific racial groups (Murakami 2003).

Taking these observations into consideration, along with the fact that the five boroughs of New York City comprise a very racially and economically diverse area of the state, the accuracy of the census estimate for the number of workers is very important. This is especially so when surveys rely on the decennial census for controls (NYSDOT OPP 2003b).

In a sample survey like the NHTS, the number of workers is an effect variable resulting from questions asked of members of the household during the interview. The sample estimate of the number of workers from the NHTS must be examined within the context of the confidence interval. Similarly, census long form measures, such as workers, are also obtained through sampling, and it is equally important to estimate the confidence interval for the 2000 Census SF3 (USDOC 2000).

Table 5 compares the 2000 census estimate and the confidence interval for the universe of workers ages 16 and over with the NHTS estimate and confidence interval for the variable "worker." In 17 of the 23 strata shown, the census and NHTS 95% confidence intervals are mutually exclusive, suggesting truly different numbers.

The nature of the survey instruments, question wording, and the timeframe of the census (2000) and the NHTS (2001/2002) clearly offer the potential for differences. The events of September 11, 2001, and the severity of the economic collapse that led to job loss in the NYC metro area and state as a whole would suggest lower NHTS values. However, such is not the case; in every county, the NHTS estimate of workers is higher.

In the 2000 decennial census question 21 asks: Last week, did this person do any work for either pay or profit? Question 21b asks: Last week, was this person temporarily absent from a job or business? The response to questions 21 and 21b, along with related response logic, forms the basis for identifying whether or not someone is a worker.

TABLE 4 Census Estimate of Workers and Worker Imputation

NHTS-NYSDOT strata	2000 census table P30		Imputed workers table P132
	Workers	Share	Percent
Albany, Rensselaer, Saratoga, Schenectady	384,047	4.7%	9.4%
Warren, Washington	56,203	0.7%	10.3%
Herkimer, Oneida	129,422	1.6%	12.6%
Onondaga	211,646	2.6%	8.6%
Tompkins	47,394	0.6%	6.7%
Monroe	345,019	4.2%	8.0%
Erie, Niagara	520,350	6.3%	10.2%
Chemung	38,451	0.5%	13.6%
Dutchess	128,437	1.6%	7.7%
Broome, Tioga	113,884	1.4%	12.2%
Orange	152,489	1.9%	16.7%
Bronx	415,075	5.1%	15.8%
Kings	901,027	11.0%	13.9%
New York	753,114	9.2%	14.4%
Queens	931,709	11.3%	9.8%
Richmond	191,145	2.3%	10.0%
Nassau	619,586	7.5%	10.4%
Suffolk	670,406	8.2%	12.1%
Putnam	48,167	0.6%	9.6%
Rockland	132,302	1.6%	11.9%
Westchester	425,052	5.2%	10.8%
Rest of state	996,991	12.1%	8.1%
State total	8,211,916	100.0%	12.2%

Source: New York State Department of Transportation (NYSDOT), census tabulations.

In the 2001 NHTS, the questioning was slightly different. There were several questions and responses that led to the determination of the worker status of the respondent:

1. During the household interview this question was asked: Does this household member have a job?
2. Later on in the personal interview, the primary activity was determined: Was the person working? Temporarily absent from a job or business? Looking for work? A homemaker? Going to school? Retired? or Doing something else?
3. If the activity was something other than working or being temporarily absent from a job, the NHTS used the same wording as the census; that is, the person was asked "Last week, did you do any work for pay or profit?"

The results of this line of questioning are shown in table 6.

The questioning in the NHTS attempts to avoid the worker underreporting problem that was felt to exist in the 1995 NPTS. However, exactly what "having a job" means to the respondent is self-determined. If the census question by design underestimates the number of workers relative to the CPS, then the basic definition of "what is a worker" is the real question. By allowing the respondent to determine the definition of "having a job" and by asking the question in multiple places, the NHTS may identify a set of part-time, occasional, or otherwise uncounted workers that the census may not.

In table 6 from the NHTS for New York State as a whole, those who work or were temporarily absent from a job or business (close to the census definition) represent an estimated population of 8,352,459, which is very similar to the census estimate of 8,211,916. However, as stated before, these estimates must be looked at in the context of error

TABLE 5 Census and NHTS Estimates of Workers

NHTS-NYSDOT strata	2000 census SF3 table P30 Lower 95%	Workers	Upper 95%	2001 NHTS variable = worker Lower 95%	Workers	Upper 95%	Census-NHTS confidence interval overlap
Albany, Rensselaer, Saratoga, Schenectady	381,851	384,047	386,243	402,683	416,380	430,077	No
Warren, Washington	55,458	56,203	56,948	56,810	61,247	65,684	Yes
Herkimer, Oneida	128,254	129,422	130,590	134,830	144,507	154,183	No
Onondaga	210,242	211,646	213,050	203,777	218,723	233,668	Yes
Tompkins	46,721	47,394	48,067	46,343	48,973	51,603	Yes
Monroe	342,923	345,019	347,115	343,361	364,167	384,974	Yes
Erie, Niagara	517,646	520,350	523,054	554,490	589,200	623,910	No
Chemung	37,796	38,451	39,106	42,123	45,266	48,409	No
Dutchess	127,128	128,437	129,746	133,397	142,028	150,659	No
Broome, Tioga	112,801	113,884	114,967	110,639	118,314	125,988	Yes
Orange	151,088	152,489	153,890	162,456	172,916	183,376	No
Bronx	412,293	415,075	417,857	594,454	641,846	689,238	No
Kings	897,127	901,027	904,927	1,061,196	1,166,320	1,271,444	No
New York	749,362	753,114	756,866	788,268	852,703	917,138	No
Queens	928,761	931,709	934,657	1,071,511	1,163,783	1,256,055	No
Richmond	189,487	191,145	192,803	204,679	219,920	235,162	No
Nassau	617,127	619,586	622,045	678,739	722,195	765,650	No
Suffolk	667,918	670,406	672,894	675,181	727,581	779,982	No
Putnam	47,536	48,167	48,798	48,882	52,554	562,257	Yes
Rockland	131,018	132,302	133,586	135,888	1,463,409	156,792	No
Westchester	422,684	425,052	427,420	471,282	500,862	530,441	No
Rest of state	993,749	996,991	1,000,233	1,084,806	1,129,430	1,174,054	No
State total	8,201,020	8,211,916	8,222,812	9,479,227	9,645,253	9,811,279	No

Source: New York State Department of Transportation (NYSDOT), census and NHTS tabulations.

and confidence limits (table 7). This table presents nine strata where the 95% confidence intervals are mutually exclusive (i.e., the estimates are statistically different). In the remaining 14 strata, there is no significant difference between the two numbers (i.e., statistically they are the same estimate).

Clearly the concept of worker in the NHTS shows that many people do not have full- or part-time jobs and do some other type of activity during the week, yet they reported that they worked and were compensated. This type of questioning may explain some of the problems between the decennial census and the CPS, as well as illustrate the effects of differences in question wording, survey instrument design and administration, the difference between job and worker for the respondent, the impact of effect variables that are not controlled for, and the need to clearly evaluate survey estimates within the context of confidence limits.

DRIVERS AND DRIVER LICENSES

In this section, the relationship between "driver licenses in force" from the New York State Department of Motor Vehicle (DMV) and the effect variable, "driver," in the NHTS will be examined.[5] Given that the NHTS is weighted by age, race, and sex, the logical assumption is that the account of drivers would correspond well with that of DMV. The DMV licenses in-force summary by gender for 2001 reflects all persons who have a valid driver license at the end of the year (which is approximately midpoint through the survey).

[5] This section is based on NYSDMV (2001).

TABLE 6 NHTS Worker Question Categories

NHTS NYSDOT-strata	Working	Temporarily absent from job or business	Other activity but work for pay or profit	Determined to be workers
Albany, Rensselaer, Saratoga, Schenectady	335,017	25,465	53,431	416,380
Warren, Washington	48,646	4,257	7,916	61,247
Herkimer, Oneida	117,384	10,925	15,984	144,507
Onondaga	177,641	16,067	24,208	218,723
Tompkins	36,486	2,598	9,547	48,973
Monroe	296,739	19,114	45,875	364,167
Erie, Niagara	453,740	48,697	84,020	589,200
Chemung	35,510	3,595	6,040	45,266
Dutchess	107,623	10,506	22,905	142,028
Broome, Tioga	99,276	5,815	11,885	118,314
Orange	143,655	8,701	18,751	172,916
Bronx	492,448	44,177	101,227	641,846
Kings	927,935	74,510	153,808	1,166,320
New York	689,552	61,296	99,064	852,703
Queens	949,720	75,361	137,169	1,163,783
Richmond	182,183	14,550	22,964	219,920
Nassau	583,508	33,653	102,151	722,195
Suffolk	558,922	78,449	86,444	727,581
Putnam	42,006	3,791	6,720	52,554
Rockland	117,271	7,636	20,156	146,340
Westchester	411,270	21,351	65,028	500,862
Rest of state	889,608	83,774	148,968	1,129,430
State total	7,696,142	654,289	1,244,261	9,645,253

Note: Each column has been separately derived from the sample; confidence limits are not shown.

Source: New York State Department of Transportation (NYSDOT), census tabulations.

The NHTS does not specifically ask if each driver holds a current and valid license; the respondent is simply asked whether the person is a driver. Therefore, the NHTS may count people whose licenses may have been suspended, people who have licenses but are no longer driving, or people who are licensed out of state but may be residing temporarily in NYS.

In table 8, it is possible to see that sample size is critical to how well the number of drivers estimated by the NHTS corresponds with the number of in-force driver licenses in each stratum. At the statewide level, only the total number of NHTS female drivers is statistically different from DMV in-force licenses. On a strata basis, however, the correspondence is very different. Of the 23 strata shown, 13 are statistically different for male drivers in the NHTS vs. DMV, 8 are different for female drivers (7 of the 8 are the same strata as for males), and 12 are statistically different for all drivers.

Interestingly, there is no apparent pattern for the correspondence between the two sources or the driver categories (male, female, total). Even in the Albany, Rensselaer, Saratoga, Schenectady stratum, which has a considerable oversample relative to other strata (at the request of the metropolitan planning organization (MPO)), there is mixed correspondence.

The correspondence seems to be better in small to moderate strata rather than very urban strata in the NYC area. Clearly, the very low number of persons with DMV driver licenses in New York County

TABLE 7 Census and "Question-Equivalent" NHTS Workers

NHTS NYSDOT-strata	2000 census SF3 table P30 Lower 95%	2000 census SF3 table P30 Workers	2000 census SF3 table P30 Upper 95%	2001 NHTS variable = worker (working or temporarily absent from job or business) Lower 95%	2001 NHTS variable = worker (working or temporarily absent from job or business) Workers	2001 NHTS variable = worker (working or temporarily absent from job or business) Upper 95%	Census-NHTS confidence interval overlap
Albany, Rensselaer, Saratoga, Schenectady	381,851	384,047	386,243	346,089	360,482	374,876	No
Warren, Washington	55,458	56,203	56,948	48,546	52,903	57,260	Yes
Herkimer, Oneida	128,254	129,422	130,590	118,759	128,310	137,860	Yes
Onondaga	210,242	211,646	213,050	179,377	193,708	208,040	No
Tompkins	46,721	47,394	48,067	36,027	39,084	42,140	No
Monroe	342,923	345,019	347,115	296,440	315,853	335,267	No
Erie, Niagara	517,646	520,350	523,054	470,225	502,437	534,649	Yes
Chemung	37,796	38,451	39,106	36,233	39,105	41,977	Yes
Dutchess	127,128	128,437	129,746	110,109	118,129	126,150	No
Broome, Tioga	112,801	113,884	114,967	97,651	105,091	112,532	No
Orange	151,088	152,489	153,890	142,944	152,356	161,769	Yes
Bronx	412,293	415,075	417,857	489,917	536,625	583,332	No
Kings	897,127	901,027	904,927	899,206	1,002,444	1,105,683	Yes
New York	749,362	753,114	756,866	685,892	750,848	815,805	Yes
Queens	928,761	931,709	934,657	938,071	1,025,081	1,112,090	No
Richmond	189,487	191,145	192,803	181,476	196,733	211,990	Yes
Nassau	617,127	619,586	622,045	572,563	617,162	661,760	Yes
Suffolk	667,918	670,406	672,894	582,791	637,371	691,952	Yes
Putnam	47,536	48,167	48,798	41,817	45,798	49,778	Yes
Rockland	131,018	132,302	133,586	114,660	124,907	135,155	Yes
Westchester	422,684	425,052	427,420	401,115	432,621	464,126	Yes
Rest of state	993,749	996,991	1,000,233	927,456	973,382	1,019,309	Yes
State total	8,201,020	8,211,916	8,222,812	8,185,695	8,350,431	8,515,168	Yes

Source: New York State Department of Transportation (NYSDOT), census and NHTS tabulations.

(Manhattan) is understandable given the availability of mass transit services and the high cost of housing, operating, and garaging an automobile. Many of the residents are age-eligible to drive yet do not have a license, and this cuts across all age cohorts for this county. What is not so understandable is why the survey reported such a high number of drivers in this county. It is possible that the population is more transient and residents are licensed in other states, or that they really do not have legal licenses. Also worthy of note is the trend for the survey to reflect drivers more in line with population. DMV licenses may reflect the inherent residential density and spatial context that may contribute to more auto trips taking place.

The survey, however, estimates drivers at the state level quite well; the overall estimate for total drivers is almost an exact match with DMV licenses in force. The obvious conclusion then is that this survey effect variable when broken out by gender and geography is highly sensitive to sample size and sampling error.

It is also possible that this sensitivity extends to trip production as well, which would likely impact survey estimates of respondent VMT. Clearly, geographic-based age, sex, and race weighting may not adequately reflect the spatial living arrangements that density introduces into the dynamic associated with owning and driving a car.

TABLE 8 New York State DMV-Licensed Drivers vs. NHTS Drivers, by Gender

	Male drivers				DMV01 within NHTS CI
	DMV01	Lower 95%	Estimate	Upper 95%	
Albany, Rensselaer, Saratoga, Schenectady	280,475	257,927	267,438	276,949	False
Warren, Washington	45,869	42,587	45,260	47,933	True
Herkimer, Oneida	103,642	94,266	101,277	108,288	True
Onondaga	155,294	137,917	147,507	157,098	True
Tompkins	29,691	30,036	31,757	33,478	False
Monroe	248,741	221,927	235,905	249,883	True
Erie, Niagara	394,150	350,753	374,233	397,712	True
Chemung	30,982	28,949	30,879	32,808	True
Dutchess	100,740	82,411	88,971	95,531	False
Broome, Tioga	90,333	80,480	85,104	89,729	False
Orange	117,171	97,921	106,952	115,982	False
Bronx	241,987	261,536	304,612	347,688	False
Kings	492,165	602,794	666,556	730,318	False
New York	372,441	374,448	415,102	455,755	False
Queens	608,943	509,372	577,247	645,121	True
Richmond	145,391	133,428	143,771	154,114	True
Nassau	497,712	392,005	423,374	454,742	False
Suffolk	530,574	429,867	463,954	498,042	False
Putnam	38,132	29,326	32,125	34,924	False
Rockland	105,643	81,540	88,720	95,899	False
Westchester	325,830	308,378	328,924	349,471	True
Rest of state	805,915	740,027	770,021	800,016	False
State total	5,784,348	5,588,241	5,729,689	5,871,137	True

(continues on next page)

Previous NYSDOT Analysis

As part of NYSDOT's analysis of the 1995 NPTS, a series of analytical reports were prepared for each of the survey strata by the Center for Transportation Analysis at Oak Ridge National Laboratory (Hu and Young 1999). One report, "1995 New York NPTS: A Comparison Study," focused on comparing and contrasting the individual survey strata that corresponded to the primary urban counties in each metropolitan area. The intent of this report was to assess comparability in travel measures to make it easier for MPOs to benefit and draw from travel model updates and improvements done in areas with similar characteristics. One of the findings of this comparison study was that comparability in travel behavior measures was best for areas of similar tract-level population density.

VEHICLES AVAILABLE

Neither the census nor the NHTS asked if the vehicles available within the household were owned and/or registered to someone in the household. Nor did they ask if the vehicles were leased by someone else or how the vehicles were used (primarily for personal or for commercial use). Both surveys simply asked about the vehicles available for use, which makes comparison with registered vehicles difficult.

The total number of vehicles available in households is not directly available from the census, which gives the number of households categorized by the number of vehicles (zero, one, two, etc., up to six or more). By multiplying the number of households by the corresponding number of vehicles available, it is possible to estimate the number of vehicles available (NYSDOT OPP 2003c).

TABLE 8 New York State DMV-Licensed Drivers vs. NHTS Drivers, by Gender *(Continued)*

	Female drivers				DMV01 within NHTS CI
	DMV01	Lower 95%	Estimate	Upper 95%	
Albany, Rensselaer, Saratoga, Schenectady	283,048	267,124	276,728	286,332	True
Warren, Washington	45,786	42,237	45,050	47,864	True
Herkimer, Oneida	103,051	91,839	98,579	105,320	True
Onondaga	160,321	147,320	158,925	170,530	True
Tompkins	29,933	30,765	32,565	34,365	False
Monroe	257,826	247,356	261,293	275,230	True
Erie, Niagara	404,106	349,588	378,627	407,667	True
Chemung	31,756	30,415	32,283	34,151	True
Dutchess	98,431	84,391	89,580	94,770	False
Broome, Tioga	89,776	83,662	88,512	93,363	True
Orange	111,984	107,706	116,102	124,497	True
Bronx	166,736	191,071	224,495	257,918	False
Kings	334,302	362,881	430,247	497,613	False
New York	284,602	296,570	346,345	396,120	False
Queens	435,817	485,906	546,285	606,664	False
Richmond	131,671	112,441	124,420	136,400	True
Nassau	487,201	433,659	472,412	511,166	True
Suffolk	516,977	463,135	500,276	537,417	True
Putnam	36,338	30,403	32,641	34,880	False
Rockland	99,887	86,562	94,116	101,670	True
Westchester	314,073	280,901	306,127	331,353	True
Rest of state	789,956	756,547	781,933	807,319	True
State total	5,230,457	5,306,638	5,437,541	5,568,444	False

(continues on next page)

The NHTS enumerates all the vehicles in every household and provides a vehicle file with the make, model, and primary driver of each vehicle noted. In the 2001 NHTS, the Energy Information Agency (EIA) added the fuel type, fuel efficiency, and gas cost at the residential location to the vehicle file, expanding its usefulness.

The number of vehicles reported by the census and the NHTS should compare reasonably well against the number registered within the county. Unfortunately, the NYSDMV registers vehicles as "standard series" (mostly passenger cars), commercial (trucks, vans, pickups), and other categories that define specific vehicles and/or their use (trailers, taxi, rental, farm, etc.). Examination of the VIUS indicates that the bulk of the commercial vehicles that fell into the pickup, van, SUV, and other truck categories were being used for personal transportation, and, therefore, the NYSDMV standard series vehicles are comparable and can be used for this comparison (NYSDMV 2001).

Some difficulty still exists in figuring out how many cars may be in commercial use and how many commercial vehicles may be in personal use. In table 9, the NHTS data have been recoded to correspond to the census distribution (zero, one, two, three, four, and five or more vehicles). The NHTS survey estimate and its upper and lower confidence limits are shown. Since the census values are mostly within the NHTS confidence interval, the census confidence interval was not computed. In addition to these data, the NYSDMV registration data for standard series, commercial vehicles, and their sum are included for comparison purposes.

Unlike the previous discussion for drivers and driver licenses, vehicles available may indeed be

TABLE 8 New York State DMV-Licensed Drivers vs. NHTS Drivers, by Gender *(Continued)*

	All drivers				DMV01 within NHTS CI
	DMV01	Lower 95%	Estimate	Upper 95%	
Albany, Rensselaer, Saratoga, Schenectady	563,523	530,101	544,166	558,230	False
Warren, Washington	91,655	86,096	90,310	94,525	True
Herkimer, Oneida	206,693	188,389	199,856	211,323	True
Onondaga	315,615	289,545	306,432	323,319	True
Tompkins	59,624	61,579	64,322	67,065	False
Monroe	506,567	476,648	497,198	517,748	True
Erie, Niagara	798,256	709,260	752,860	796,461	False
Chemung	62,738	60,111	63,162	66,213	True
Dutchess	199,171	169,603	178,551	187,500	False
Broome, Tioga	180,109	166,092	173,617	181,141	True
Orange	229,155	209,529	223,053	236,577	True
Bronx	408,723	474,754	529,107	583,459	False
Kings	826,467	998,263	1,096,803	1,195,342	False
New York	657,043	685,856	761,447	837,037	False
Queens	1,044,760	1,027,069	1,123,532	1,219,995	True
Richmond	277,062	251,352	268,191	285,030	True
Nassau	984,913	844,067	895,786	947,505	False
Suffolk	1,047,551	915,582	964,230	1,012,878	False
Putnam	74,470	60,688	64,767	68,846	False
Rockland	205,530	171,434	182,836	194,238	False
Westchester	639,903	603,584	635,051	666,519	True
Rest of state	1,595,871	1,508,288	1,551,954	1,595,620	False
State total	11,014,805	10,958,833	11,167,231	11,375,629	True

Sources: New York State Department of Motor Vehicles (NYSDMV), data on driver licenses in force, 2001. NYS Department of Transportation, special tabulations prepared from the NHTS 2001 NYS add-on, 2001.

much more closely related to the basic sampling unit—the household in both the census and the NHTS. As part of the 2001 NHTS survey contract with NYSDOT, the survey vendor was asked to evaluate if disaggregated registration data by registration type provided better weighting than households for the vehicle file in the 1995 NPTS. As a result of their analysis, it was determined that households and registrations on the whole were both equal to the task for weighting the vehicle file. This would suggest then that the census, NHTS, and DMV data would be proportionally similar.

In table 9, the reader should note that within the census and NHTS estimates it is possible that there are non-NYS registered vehicles being counted. Examination of the table shows that for the most part the census estimate is essentially found between the 95% lower bound and the estimate for the NHTS.

There are two exceptions, the statewide total and the Albany, Rensselaer, Saratoga, Schenectady strata. In the case of the Albany strata, it is possible that the census upper confidence interval overlaps the lower bound of the NHTS. However, no explanation can be offered for the differences in the Albany strata and the statewide total. Equally interesting is that, for the most part, the standard series registrations alone correspond well except for four strata, and when these values are taken with the "Commercial" vehicles, the "Sum" falls within the NHTS confidence interval except for Suffolk and Rockland counties. It is possible that in these counties there were a greater proportion of commercial business vehicles. Most important to note is that the entire statewide NHTS confidence interval is significantly above the census 2000 estimate, perhaps due in part to the difference in the survey instrument and sample size.

TABLE 9 New York State DMV Registrations vs. Census and NHTS Vehicles Available in Households

	Number of vehicles available in the household				Registrations in force 2001		
		NHTS			NYSDMV		
	Census 2000	Lower 95%	Estimate	Upper 95%	Standard series	Commercial	Sum
Albany, Rensselaer, Saratoga, Schenectady	**501,917**	503,125	573,359	643,593	**489,277**	88,170	577,447
Warren, Washington	80,549	73,800	93,978	114,155	**73,205**	22,810	96,015
Herkimer, Oneida	179,980	167,147	210,345	253,544	**163,394**	38,975	202,369
Onondaga	274,445	226,959	279,419	331,879	267,756	51,186	318,942
Tompkins	55,606	46,943	61,604	76,264	48,245	10,378	58,623
Monroe	452,180	422,455	516,562	610,669	453,283	62,395	515,678
Erie, Niagara	693,540	652,666	808,079	963,492	669,293	108,131	777,424
Chemung	54,594	52,729	65,770	78,810	**52,577**	11,695	64,272
Dutchess	179,097	152,015	189,905	227,794	199,191	20,178	219,369
Broome, Tioga	158,915	142,554	179,913	217,271	152,283	32,102	184,385
Orange	199,400	181,699	229,386	277,073	221,397	27,563	248,960
Bronx	236,070	188,454	300,888	413,322	249,785	9,340	259,125
Kings	491,844	401,420	550,936	700,451	426,786	19,163	445,949
New York	191,879	169,621	275,494	381,366	229,715	13,655	243,370
Queens	700,593	514,598	699,430	884,262	700,531	33,017	733,548
Richmond	215,656	186,916	243,022	299,128	235,660	5,867	241,527
Nassau	805,034	683,508	857,709	1,031,909	919,400	31,536	950,936
Suffolk	906,998	709,666	894,318	1,078,971	1,027,031	82,077	**1,109,108**
Putnam	66,899	57,793	76,710	95,628	78,049	5,429	83,478
Rockland	165,577	131,697	165,457	199,218	196,050	8,693	**204,743**
Westchester	511,568	431,046	567,798	704,550	613,310	25,968	639,278
Rest of state	1,423,082	1,446,318	1,645,355	1,844,393	**1,275,992**	382,758	1,658,750
State total	**8,545,423**	8,960,670	9,485,437	10,010,204	**8,742,210**	1,091,086	9,833,296

Note: Values in bold are outside the NHTS lower and upper 95% confidence intervals for the sample strata shown.
Source: New York State (NYS) Department of Transportation, NHTS tabulations; and NYS Department of Motor Vehicles (DMV).

FINDINGS

Does the survey estimate of vehicle travel over time adequately match observed monthly VMT? No.

- In a comparison of survey and ground-count-based estimates of monthly VMT, the effects of seasonal variation alone did not explain the differences. Temperature trends and seasonal snowfall did not provide any additional explanatory data. Perhaps the duration and timing of weather events had more to do with impacting day travel than just the amount of snow that fell.
- Equally important is the fact that the NHTS is a residential household survey; it is not possible to assess the effects of nonresident and commercial travel to or through the state. The specific impacts of September 11, 2001, and the deepening recession on personal, business, and commercial travel are intricately woven into the fabric of daily travel reflected in ground counts.
- Also, the monthly ground-count data were not disaggregated by vehicle classification, state of origin, purpose (personal, business, or commercial travel) for either resident or nonresident vehicles, which would be necessary for a rigorous comparison of the survey results.

Can the apparent lack of change in survey estimates of residential household personal VMT for 1995 and 2001 be explained given the increase in the ground-count-based estimate of VMT? Yes.

- Survey estimates of effect variables require careful examination of the standard error and the 95% confidence limit. Sample size has a considerable impact on sampling error.
- Total VMT was comprised of residential and nonresidential personal and commercial travel. The NPTS/NHTS addressed residential personal household travel. The VIUS addressed residential vehicular (truck) travel. Both surveys occurred at different time intervals and had significantly different sample sizes and sampling universes. Both surveys lacked consistency in definitions for mode relative to how vehicles were registered and used. Neither survey addressed interstate personal or vehicular travel.
- Ground-count estimates of VMT included residential, nonresidential, personal, commercial, and interstate movements.
- With appropriate assumptions, it may be possible to illustrate that growth in count-based VMT is perhaps being driven by what can loosely be described as commercial vehicle travel. Further research into the potential impact of growing commercial vehicle travel, especially linked to shopping and home-based businesses, is warranted.
- Lastly and perhaps most important, survey sampling in the NHTS, VIUS, and CFS is typically administered for the resident population, domiciled registered vehicles, or the shipper state of origin. In multistate labor market areas where a regional context is required, the inability to adequately assess in some manner the net migration within, into, and across the labor market by state severely hinders the ability to understand the complete travel picture, especially as measured by what is on the road.

Will a survey adequately reflect public transportation ridership? Yes, in certain cases.

- Outside of New York City and the surrounding region, defining transit may be a simple undertaking; that is, in some cases only one mode is available (e.g., a bus). NYC offers a wide variety of transit, both publicly and privately operated, and riders may take one or more modes and/or transfer within the same mode. Also, nonresidents (from New Jersey or Connecticut) enter the city to work every day and are not included in any residential survey. Survey estimates of transit person trips tend to underestimate unlinked trips.
- Additionally there are significant definitional problems in analyzing public transit trips derived from a survey with respect to those trips reported by a transit operator. Transit operators collect revenue and monitor data on unlinked trips, which do not have a one-to-one relationship, especially when a sliding fare and transfers are readily available.
- Within the context of a narrowly defined segment of public transportation (MTA subway), it is possible that the survey estimate may correspond well with the operator report of unlinked trips.

Is there comparability between the census and the NHTS on "Who is a worker?" Yes, with consistent definitions and at the state level.

- The concept of *worker* in the NHTS indicates that many people do not have full- or part-time jobs, yet they report that they engage in some other type of activity for which they are compensated. The difference between the census and the NHTS estimates for NYS is about 1 million jobs.
- The decennial census and the NHTS are both surveys and, as such, are subject to concerns of sample size, standard error, and the need to evaluate survey estimates within the 95% confidence interval. Worker status is a household effect variable in both surveys. Equally important is question wording and the survey instrument and its administration.
- Careful separation of the NHTS response to best match that of the decennial census concept of worker shows that, for the state as a whole, there is agreement. However, for nine of the strata in NYS, the two estimates of worker are statistically different.
- It is clear that for the respondent, the census and the NHTS have very different concepts of worker compared with that of transportation analysts. By asking "Do you have a job?" which is self-defined in the NHTS, and then probing first for traditional work status and then for any activities

with pay or profit (the census question), it is possible that the NHTS reveals nontraditional or illegal employment (e.g., under-the-table employment or the underground workforce).

Does the survey adequately reflect drivers and driver licenses? Yes and no.

- On a statewide basis, the NHTS survey estimates for drivers matches NYSDMV total "driver licenses in force" quite well. When categorized by strata and gender, the results are mixed.
- The NHTS concept of driver may not be equivalent to licensed driver as reported by NYSDMV. The NHTS asks who a driver is without qualifying whether that person has a legal license.
- We can conclude that the survey estimate of drivers is an effect variable that is highly sensitive to sample size and sampling error. The sensitivity may also result from the spatial impact for travel opportunities due to settlement density, the availability of mass transit options, transient population, residents licensed in other states, or respondents without legal licenses.

Are the census and the NHTS estimates of the number of vehicles available within households in NYS accurate? Maybe.

- For most strata, the census estimate is essentially found between the NHTS 95% lower bound and the survey estimate.
- One of the problems in making this comparison is that the census does not adequately delineate vehicles that may be available within the household for use by type, registration, and usage (personal and nonpersonal or commercial). The census simply collects the number of households with zero to five or more vehicles that are available.
- One improvement to the NHTS would be a mechanism to match vehicle type with registration category. As part of the 2001 NHTS data collection and analysis, NYSDOT requested an assessment of whether DMV vehicle registration data by county was a better measure for weighting the vehicle file than the household weight used in 1995. The conclusion was that they were nearly equivalent. However, when the census and the NHTS are compared with NYSDMV registration categories the inconsistencies were problematic.

Lastly, both the census and the NHTS may be counting non-NYS registered vehicles.

REFERENCES

Clark, S.L., J. Iceland, T. Palumbo, K. Posey, and M. Weismantle. 2003. Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey. Suitland, MD: U.S. Department of Commerce, U.S. Census Bureau.

Eng-Wong Taub & Associates. 2003. Origin and Destination Data by Station, prepared for the Port Authority Trans-Hudson Corporation (PATH), 2001 PATH Passenger Travel Study. Obtained by special request of NYSDOT, Passenger Transportation Division.

P.S. Hu and J.R. Young. 1999. 1995 New York NPTS: A Comparison Study, prepared for NYSDOT, Planning and Strategy Group. November 23. Available at http://www.dot.state.ny.us/ttss/1995npts/nynpts95_comparison_study.pdf.

L. Maynus, New York State Thruway Authority. 2004. Personal communications.

E. Murakami, Federal Highway Administration, U.S. Department of Transportation. 2003. Personal communication. October 29.

New York State Department of Motor Vehicles (NYSDMV). 2001. Tabulations of driver licenses in force.

New York State Department of Transportation (NYSDOT), Passenger Transportation Division (PTD). 2003a. Metropolitan Transportation Authority fare structure data.

______. 2003b. Metropolitan Transportation Authority historical revenue and ridership data.

______. 2003c. Port Authority Trans-Hudson Corporation (PATH) operating statistics, from the 2001 National Transit Database.

New York State Department of Transportation (NYSDOT), Office of Policy and Performance (OPP). 2002. *Trends and Travel after September 11, 2001*. New York, NY.

______. 2003a. Customized tabulations using U.S. Department of Commerce, U.S. Census Bureau. 1998. *1997 Vehicle Inventory and use Survey*, digital micro data CD.

______. 2003b. Data extraction using U.S. Department of Commerce, U.S. Census Bureau, *2000 Census SF3 DVD* data: NYSDOT, Tables: P1—Total Population, P4—Percent of Population in the Sample, P8—Sex by Age, P30—Means of Transportation to Work for Workers 16 Years and Over, P43—Sex by Employment Status for the Population 16 Years and Over, P132—Imputation of Work Status in 1999 for the Population 16 Years and Over.

______. 2003c. Data extraction from U.S. Department of Commerce, U.S. Census Bureau. *2000 Census SF3 DVD*. Tables—H6: Occupancy Status; H44: Tenure by Vehicles Available.

______. 2003d. Estimates of historical monthly vehicle-miles of travel.

______. 2003e. Temporal distributions of vehicle-miles of travel, special tabulations from the 1995 NPTS and 2001 NHTS.

New York State Department of Transportation (NYSDOT), Office of Operations Management (OOM). 2003. Statewide snowfall reported by NYSDOT residencies. Internal reports and tabulations.

New York State Energy Research and Development Authority (NYSERDA). 2004. *Patterns and Trends, New York State Energy Profiles: 1989–2003*. Available at http://www.nyserda.org, as of October 2005.

Reebie & Associates. 2001. TranSearch Data. Custom tabulations. April 2003.

U.S. Department of Commerce (USDOC), U.S. Census Bureau. 2000. Chapter 8: Accuracy of the Data, Using Tables to Compute Standard Errors and Confidence Intervals. *Summary File 3, 2000 Census of Population and Housing,* Technical Documentation. Available at http://www.census.gov/prod/cen2000/doc/sf3.pdf.

U.S. Department of Transportation (USDOT), Federal Highway Administration. 1995. Nationwide Personal Transportation Survey, New York State add-on. Digital data files and custom tabulations.

______. 2001. National Household Transportation Survey, New York State add-on. Digital data files and custom tabulations.

______. 2003. *Highway Statistics.* Washington, DC. Estimates of U.S. monthly vehicle-miles of travel for 2001/2002, special tabulations.

Bayesian Approaches to Learning from Data: Using NHTS Data for the Analysis of Land Use and Travel Behavior

MARCO G. SCUDERI[1]
KELLY J. CLIFTON[2, *]

[1] Urban Studies and Planning
0109 Caroline Hall
University of Maryland
College Park, MD 20742

[2] Urban Studies and Planning
National Center for Smart Growth Research and Education
Preinkert Field House, Suite 1112
University of Maryland
College Park, MD 20742

ABSTRACT

This paper introduces the application of Bayesian belief networks (BBNs) to the investigation of the relationship between land use and travel behavior and emphasizes the use of 2001 National Household Travel Survey (NHTS) data. Bayesian statistics are used to reason under uncertainty and provide the basis for a methodological approach that does not require stringent a priori assumptions about the statistical model employed to analyze the data. For this reason, this method is appropriate for exploring new relationships between land use and travel behavior that may not be apparent using more traditional approaches. This study focuses on the utilization of the NHTS add-on data for the Baltimore metropolitan region. The paper provides an introduction to modeling relationships between variables based on the structures of BBNs, provides insight into the specific methodological constructs needed to analyze NHTS data, and develops the potential to contribute alternative insights into the land use-travel behavior relationship.

INTRODUCTION

This paper develops and tests a method to analytically derive a representation of land-use and travel

Email addresses:
* Corresponding author—kclifton@umd.edu
M. Scuderi—scudhome@yahoo.com

KEYWORDS: Bayesian belief networks, meta-heuristic search algorithms, travel behavior, land use, travel survey data.

behavior relationships using data from the National Household Travel Survey (NHTS). This research attempts to bridge the different existing theories, which tend to take a "top down" or inductive approach, by employing a complementary "bottom up" or deductive approach based on what the data may represent rather than how they may be analyzed. This approach, using Bayesian belief networks (BBNs), contributes a new and original method for the analysis of complex spatial-behavioral systems such as human interaction in the urban environment, and it presents an opportunity to expand further our theoretical knowledge in the area of land use and travel behavior.

Despite more than 20 years of intensive studies in this area, no unified theory exists to explain the interactions between land use and travel behavior. Conflicting results and frameworks remain a central theme in the debate about the possibility of recursive effects between the domains. Several issues arise due to these theoretical and empirical shortcomings. The impacts of land-use and transportation policy interventions on travel demand cannot be accurately gauged. Expensive transportation projects rarely result in accurate forecasts of future numbers of users, in part because forecasting methodologies do not adequately consider the effects of land-use changes resulting from transportation investments and how these changes alter travel demand (i.e., induced demand). This area of inquiry could benefit from a more specific and quantitative characterization of the relationships between land use and travel behavior than is available today.

Using 2001 NHTS add-on data for the Baltimore region, this paper proposes and tests a new approach to the analysis of the interactions between land use and transportation choices that does not require the design of statistical models prior to the analysis of the data. This approach is based on a process of knowledge discovery that uses BBNs to identify potential causal dependencies among variables, rather than imposing or assuming those relationships a priori. In this paper, not only does the output of BBNs become the foundation for an analytically oriented approach to model the land-use and travel behavior interaction, it also provides quantitative measures, in the form of conditional probability distributions, of how the various factors affect and interact with each other. The combination of Bayesian probability theory, graph theory, and geographic information systems gives a series of additional analytical perspectives to this problem. For example, in transit mode share, conditional probability distributions can be used to identify the highest probabilities of usage of a particular transit mode or the probabilities of obtaining a specific transit usage. The resulting maps also help distinguish among localities with similar urban characteristics but where the differing qualities of urban environments result in different behavioral responses.

The 2001 NHTS data, supplemented with local land-use data, offer a number of areas in which to test this Bayesian approach. For one, travel diary data provide a complete accounting of daily travel for all trip modes and purposes. Two, data from the add-on survey for the Baltimore region can be combined with a variety of land-use, urban form, transportation system, and community attributes. Three, the Baltimore region exhibits much variation in the urban environment, allowing for a robust study design. The variables proposed in this research design, including accessibility indices, land uses, and socioeconomic diversity, have all been identified in past studies as key factors in this relationship. The transport choices derived from the 2001 NHTS Baltimore add-on data are considered the behavioral response to social and economic conditions, transportation availability, and land-use characteristics.

The paper is organized as follows. First, the case for employing Bayesian approaches to land use-travel behavior research is made. Then, BBNs are explained in some detail with emphasis on the benefits of applying this data-driven approach to the topic of interest.

THE CASE FOR BAYESIAN APPROACHES

Traditional deductive research approaches suffer from a few drawbacks that can limit their ability to identify relationships. Statistical studies are often designed prior to data analysis (and sometimes data collection) and can be driven by theoretical assumptions. A distinction is made a priori about the nature of the relationships under investigation, including the direction and degree of the relationships between and among variables. It is a challenge to consider all of the complex phenomena and processes that influence travel behavior concurrently. Even more challenging is identifying the relationships among and between these factors. These underlying assumptions and model specifications, if incorrect or incomplete, can limit the findings and potentially mask important relationships.

Inductive reasoning breaks away from this deductive reasoning process and allows the analyst to directly query actual data for possible relationships among them so that the analyst can become more confident about the correct theoretical framework to use, one that could possibly be less fragmented and more universal than what is currently in use. BBNs provide a means to rise to this task because of their ability to assess an infinite number of relationships at the same time and their ability to present them in graphical form. In such an inductive environment, questions can be asked without the confinements dictated by specific statistical constructs or analytical methods.

Within the above set of relationships and behavioral decisions related to transportation outcomes, the linkages between daily activity participation and travel in the short term is of keen interest and is explored in more detail in this paper. The literature in this area presents a great number of differing conclusions using a variety of analytical approaches (e.g., Badoe and Miller 2000; Crane 2000; Ewing and Cervero 2001). The attempts to model these relationships are many; however, a robust behavioral framework is lacking (Waddell 2001). From a preliminary review of past studies, it appears that a great need still exists to study the relationship between land use and travel behavior because of its indetermination and the limitations of previous results, which can be identified as:

- The tendency to determine a priori the statistical output by selecting specifically diverse neighborhoods with contrasting characteristics, in order to prove that different land uses are associated with specific travel choices and vice versa.
- The difficulty of differentiating among qualitative properties of urban forms, which in this study can be resolved by observing and quantifying the human response to the built environment.
- Reliance on ad hoc statistical models based on the personal knowledge of specific researchers, inconclusive results, or excessive emphasis given to anecdotal and contradictory empirical evidence.

BBNs, however, are capable of addressing many of the shortcomings commonly found in existing approaches to the study of land use and travel behavior interactions. BBNs, also referred to as decision networks or probabilistic causal networks, have been quietly gaining momentum within the research community, mainly as a result of the great advantages obtained in the field of computer science, artificial intelligence, and automated learning (see Jensen 1996 and 2001 for an introduction). BBNs provide an easily understandable and easy-to-use environment for the analysis of complex spatial processes and the investigation of relationships between numerous variables. Still, the application of such a method would be meaningless without a comprehensive dataset that characterizes individual socioeconomic characteristics and captures individual preferences about transportation choices over a period of time. One potential data source is the NHTS and its information on American households, the individuals comprising them, and their transportation choices.

BBNs are a graphical representation of probabilistic causal information based on two components: a directed acyclic graph and a probability distribution (Glymour and Cooper 1999; Torres and Huber 2003). Nodes in the directed acyclic graph (DAG) represent stochastic variables and arcs represent directed stochastic dependencies among these variables. Thus, the graph provides a simple summary of the dependency structure relating the variables. This is an effective way to describe the overall dependency structure of a large number of variables, thus removing the limitation of examining the pair-wise associations of variables.

BBNs can also be used to reveal causal relationships among variables, which is an advantage when trying to gain an understanding of a problem domain, as in exploratory data analysis, and to predict the consequence of intervention. For example, Bayesian approaches are being used to predict credit card fraud and in the causal analysis of health issues. A classic example (Heckerman et al. 1995) looks at a marketing analyst trying to assess whether or not it is worthwhile to promote a specific advertisement in order to increase the sales of a product. The answer to this question depends on whether the advertisement is a cause for increased sales or not, and if so to what degree.

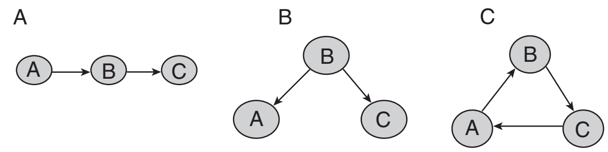
BBNs are an ideal representation for combining prior knowledge and data, because they combine both causal and probabilistic semantics. In many cases, real-world analysis benefits from prior knowledge and, in some cases, when data are incomplete or expensive, information from experts in the field is the only available source. Thus, it follows that a system that can integrate such prior knowledge into an analytical framework is a great advance.

OVERVIEW OF AUTOMATED LEARNING: THE BAYESIAN ALTERNATIVE

BBNs are computational objects able to represent compactly joint probability distributions by means of DAGs, which denote dependencies and independencies among variables as well as the conditional probability distributions of each variable, given its parents in the graph (Aliferis et al. 2003; Neapolitan 1990). The fundamental axiom of BBNs is the *Markov Condition* that allows for a concise factorization of the joint distribution and captures the main characteristic of causation in macroscopic systems, namely that causation is local (Glymour and Cooper 1999). In the graphs, nodes represent the variables, and the dependencies between variables are depicted as directional links from a parent node to a child node, which also correspond to conditional probabilities (Torres and Huber 2003).

Under uncertainty, the probability of B given A, $p(B|A)$, represents the strength of the link in the graphs. A simple example of a BBN graph is shown in figure 1, where both nodes A and B are parents of node C, the child. However, if node C is itself a parent of B as in the feedback loop of figure 1(C), then we do not know how node B and C behave; they may cooperate or counteract each other in various ways. For these reasons, BBNs do not yet model feedback processes even though the differential calculus required to implement this functionality is well understood.

FIGURE 1 Serial Connections (A), Diverging Connections (B), Feedback Loop (Cyclic Graph)



A BBN has the following properties:

- a set of variables and a set of directed edges between variables,
- each variable has a finite set of mutually exclusive states,
- the variables, together with the directed edges, form a DAG (a directed graph is acyclic if there is no directed path $A_1 \rightarrow \ldots \rightarrow A_n$ *s.t.* $A_1 = A_n$),
- for each variable A with parents $B_1,\ldots,B_n$, there is a potential table $p(A|B_1,\ldots,B_n)$ attached.

Any conditional dependence represented by an edge (or link) is quantified by the set of conditional distributions of the child variable given a configuration of the parent variables. In a statistical experiment where nodes represent stochastic variables $X = (X_1, X_2,\ldots,X_v)$, the conditional probability distribution is factorized as in:

$$p(x_{1k}, x_{2k},\ldots, x_{vk}) = \prod_{i=1}^{v} p(x_{ik}|\pi_{ij}),$$

where $(x_{1k}, x_{2k},\ldots, x_{vk})$ is a combination of values of the variables in X. For each i, the variable Π_i denotes the parents of X_i, while x_{ik} and π_{ij} denote the events $X_i = x_{ik}$, and $\Pi_i = \pi_{ij}$; the latter is the combination of values of the parent variable Π_i in the event $X = (X_1, X_2,\ldots, X_v)$.

While, traditionally, BBNs have been designed manually, one BBN represents all but one hypothetical dependency structure relating the variables. Many structures can be derived from the same set of data; thus, the analyst faces two problems: how to design the networks efficiently, and how to assess which one is better at encoding the relationship among the variables. In both cases, the latest advances in computer science and artificial intelligence now allow for the automatic learning of such structures by means of meta-heuristic search algorithms in which the subjectivity of individual beliefs is replaced by the tenets of probabilistic reasoning. Several commercial programs such as Bayesware Discoverer (http://www.bayesware.com) or Hugin (http://www.hugin.com) are now available to researchers. The authors of this paper use WinMine from Microsoft (http://www.winmine.com).

It is important to note that most search algorithms used to derive BBNs treat the data as a collection of cases where unique records are identified by a particular combination of values in the variables. Progressively, each case is read and compared with other cases in the dataset in order to derive the likelihood that a given event takes place in relation to the likelihood of other events in identical or similar cases.

STRUCTURE LEARNING

A generic model of the relationship among variables is little more than a starting point; what follows is a search for the best model that represents the relationship among the variables. This model is obtained by learning the structure of a BBN. In general terms, there are two approaches to learning these structures: constraint-based and search-and-score. They differ greatly, because the constraint-based approaches usually start with a fully connected graph and progressively remove the relational links connecting the variables if certain conditional independencies are measured in the data. This has the disadvantage that repeated independence tests lose statistical power and, therefore, this approach is used less often.

In the more commonly used search-and-score approach, the main step is a search through the space of all possible DAGs, which is intended to return one, or in some cases, a set of possible sample networks, which represent an approximation of the ideal dependency structure in the data. Unfortunately, the number of possible DAGs is a function of the number of nodes $G(n)$, and it is super-exponential with respect to n. There is no known closed form formula for $G(n)$, but the first few values for $n = 1,2,...,10$ are listed in table 1 (from Bayesware Discoverer). Because the number of possible networks is super-exponential in the number of nodes, it is not feasible to exhaustively examine the entire search space, so a local search algorithm (e.g., greedy hill climbing) or a global search algorithm (e.g., Markov Chain Monte Carlo—MCMC) is generally employed. The most basic procedure used for this task is the K2 algorithm, which tries to find the best structure by recursively selecting the best set of parents for each node independently. This implies that the total ordering of the variables is known, a situation that may not always be true. If the variable ordering is unknown, a search over the most likely orderings is usually more efficient than searching over DAGs (Friedman and Koller 2000).

TABLE 1 Number of Directed Acyclic Graphs (DAGs) as a Function of the Number of Nodes (*G*)

***G*(*n*)**	**DAGS**
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	1.1 x 10^9
8	7.8 x 10^11
9	1.2 x 10^15
10	4.2 x 10^18

In addition to the search procedure, the specifications for the scoring function are as follows: let the set $M = \{M_1, M_2, ..., M_3\}$ be a grouping of BBNs for the discrete random variables. With $p(M_h)$ denoting the prior probability of M_h for each $h = 1, \dots, g$, the typical solution to the model selection problem is to choose the network with the maximum posterior probability:

$$p(M_h|D) = \frac{p(D|M_h)p(M_h)}{p(D)}$$

The quantity $p(D|M_h)$ is the marginal likelihood that provides the Bayesian score with which to compare different models.

Selecting the network with the maximum posterior probability as derived by the search and score algorithm is quite a brute force approach to structure learning because of the need to generate and score all possible DAGs. This approach, however, provides a baseline for comparing the performance of other algorithms used to generate BBNs. More effective than the K2 algorithm, the hill-climbing algorithm searches all points in space and their nearest neighbors, defined as all "graphs that can be generated from the current graph by adding, deleting or reversing a single arc" (Chickering et al. 1997). It then moves to the neighbor that has the highest score, and if no neighbors have a higher score than the current point, the algorithm stops. The best practice is then to restart the procedure at a different point in space *n* number of times until the scores converge.

Another technique to automatically generate BBN structures is the MCMC algorithm that effectively searches the space of all possible DAGs, a property that characterizes it as being polynomial (not exponential) in the dimensionality of the search space. This makes the MCMC approach difficult for practical applications requiring the use of more than 10 variables.

Finally, the search and score approach used by the authors of this paper is the one developed by Chickering et al. (1997) at Microsoft Research. Similar to the hill-climbing approach, this algorithm adds, deletes, and reverses the possible arcs among the variables, but it does so in the context of decision graphs used to represent the relationship among each pair of variables. This algorithm also integrates aspects of the Expected Maximization algorithm, which requires the calculation of the *expected sufficient statistics* for the data. The expected sufficient statistics are then used to ensure the convergence of the results obtained using a dataset with missing values with the results generated from a complete dataset.

With this technique, the analysis begins with the observation that the local distribution for variable X_i in a dependency network is the conditional distribution $p(x_i|X\backslash x_i)$, which can be estimated by any number of probabilistic classification techniques (or regression techniques, if we were to consider continuous variables) such as generalized linear models, neural networks, probabilistic support-vector machine, or embedded regression/classification models (Heckerman et al. 2000). The method we chose in this case is a probabilistic decision tree where for each variable X_i in domain X, the classification algorithm independently estimates its local distribution from the data. Once all estimates for the local distributions are obtained, the structure of the Bayesian network can be constructed from the (in)dependencies encoded in these estimates (Heckerman et al. 2000). Each variable is modeled as a multinomial distribution and the learned decision tree corresponds to the Bayesian network.

The algorithm searches each row of data for unique combinations of categorical data. Each unique combination is called a "case" and it forms the basis of the following analytical steps, where the algorithm greedily grows decision trees using the Bayesian scoring criterion. This is a greedy algorithm that combines global search over the structure's relational links with local search over all of the nodes in the decision graphs. It begins with one node (variable) and evaluates its relationship to the other nodes (variables) by means of decision trees; then it scores the corresponding Bayesian structure based on its posterior probability of such a network considering the given cases. The procedure is as follows:

1. Score a generic network structure. For each node *x* (variable) in the graph:
2. Add every nondescendant that is not a parent of *x* to the parent set
3. For every possible operator *O* in the graph:
   i. Apply *O* to *BS*
   ii. Score the resulting structure
   iii. Un-apply *O*
4. Remove any parent that was added to *x* in step 3
5. If the best score from step ii is better than the current score
   a. Let *O* be the operator that resulted in the best score
   b. If *O* is a split operator (either complete or binary) on a node *x* that is not in its set of parents then add a new node to the parent set
   c. Apply *O* to *BS*
   d. Go to 1
6. Otherwise, return *BS*.

Three operators (*O*) are allowed:

- Complete split adds a child node to a set of parents,
- Binary split adds two children to a set of parents,
- Merge split combines two or more children in a single new node inheriting all of their parent nodes.

To learn a decision-tree structure for X_i, the search algorithm is initialized with a single root node having no children. Then, each leaf node is replaced with a binary split on some variable X_j in $X \setminus X_i$ until no such replacement increases the score of the tree. The binary split on X_j is a decision-tree node with two children: one of the children corresponds to a particular value of X_j, and the other child corresponds to *all other* values of X_j (Chickering et al. 1997).

BBN APPLICATION TO LAND USE AND TRANSPORTATION RESEARCH USING THE 2001 NHTS

As mentioned earlier, significant questions remain about the land-use and transportation relationships and their interdependencies. A variety of approaches and data sources have been applied to this problem with varying results and often with conflicting findings. BBNs, coupled with automatic learning, could provide new insight and perhaps offer a better approach to the analysis of this subject. The focus of the research reported here is to assess the effectiveness of such a method. Efforts have so far centered on testing survey data and analytical requirements of BBNs. Here, we pay particular attention to the use of 2001 NHTS data for the Baltimore metropolitan area.

This paper expands Torres and Huber's application of BBNs to research travel behavior questions (Torres and Huber 2003) by adding land-use variables and by employing a more advanced search algorithm. Torres and Huber investigated the use of BBNs to estimate travel mode choice as a function of socioeconomic variables only. Their approach made use of the K2 algorithm that is now obsolete, largely because it required the analysts to design a hypothetical BBN that was used as the starting point for the search algorithm. The method presented in this paper drops such requirements and is truly heuristic.

STUDY AREA AND DATA

The study area was the Baltimore metropolitan region, which covers the counties of Carroll, Howard, Anne Arundel, Baltimore County, Harford, and Baltimore City. Detailed data for sampled households and individuals were obtained from the 2001 NHTS Baltimore add-on survey. To these were added derived profiles of typical land-use patterns, socioeconomic characteristics, and road density for each tract and zip code in the study areas. The variables used in this analysis are shown in table 2.

All variables were reclassified into categorical form. In many cases, the number of classes within each variable was reduced to simplify the analysis. For example, the variable *age* for the respondents was reduced to four classes, with an important separation for teenagers at 16 years of age to reflect the possibility of acquiring a driving license. The response variable *race* was reclassified into four categories. More importantly, the race of the respondent was also assigned to the remaining members of the family, an assumption that might not always hold true. The personal income variable was created by first classifying the household income into 11 classes and then dividing the midway dollar amount associated with each class by the number of people living in a particular household. Transportation mode choices were reduced to just three classes: private vehicle, walking, and public transit. Private vehicle trips include the use of private cars, trucks, motorcycles, vanpooling, etc. Walking trips include bicycling, wheelchair mobility, jogging, and any other nonmotorized trip. The choice for transit included all public transportation systems except for ferry and water taxi, which given their limited presence in the data were not analyzed in this study.

The land-use variables were derived from the Maryland Property View Data; in particular, we used the 1997 Land-Use/Land Cover geographic information system information layer updated to the year 2000. Each land-use polygon was assigned to a zip or a tract and its boundaries reshaped to fit into such administrative units. Based on the total area of each administrative boundary, land-use variables were then calculated as a percentage of the total area and then reclassified into 10 discrete amounts of land-use covers for each type of residential, commercial, or other land use. The road network was subject to similar processing where each road segment was assigned to a tract or zip and its spatial length recalculated accordingly. A discrete

TABLE 2 Variable List

Variable	Description
Tract or zip	U.S. census tracts and five digit zip code tabulation areas (an aggregation of census blocks) were used to define the boundaries for local land-use characteristics
Driver status	Driver status of respondent: licensed, not licensed, or not appropriate
Worker status	Worker status of respondent: working, nonworking, or not appropriate
Age	Age of respondent: 0–5 years, 6–16 years, 17–65 years, 66 years and more
Vehicle count	Number of vehicles in household: 1–9 or more
Household size	Household size: 1–9 or more
Race	Race of the head of household: White, African American, Hispanic, other (including Asian)
Driver count	Number of drivers in household: 1–6 or more
Personal income	Household income divided by number of persons in household; 11 classes from $0–$100,000 or more
Transportation mode choice	Transportation mode choice among motorized options, walking and biking, or transit (including buses, metro, and rail)
% low residential	Low-density residential: detached single-family/duplex dwelling units, yards, and associated areas; areas of more than 90% single-family/duplex dwelling units, with lot sizes of less than 5 acres but at least half an acre (0.2 dwelling units/acre to 2 dwelling units/acre)
% medium residential	Medium-density residential: detached single-family/duplex, attached single-unit row housing, yards, and associated areas; areas of more than 90% single-family/duplex units and attached single-unit row housing, with lot sizes of less than half an acre but at least one-eighth acre (2 dwelling units/acre to 8 dwelling units/acre)
% high residential	High-density residential: attached single-unit row housing, garden apartments, high-rise apartments/condominiums, mobile homes, and trailer parks; areas of more than 90% high-density residential units, with more than 8 dwelling units per acre
% commercial	Commercial, retail, and wholesale services; areas used primarily for the sale of products and services, including associated yards and parking areas
% industrial	Manufacturing and industrial parks, including associated warehouses, storage yards, research laboratories, and parking areas
% vacant	Vacant land, such as bodies of water
% other	Other land uses
Road density index	Liner road length over square areas, 10 classes, with a lower value indicating denser road networks

ratio of the total road length within each administrative unit over the total areas for such units created an index of road density.

Land-use variables are available as continuous percentage values, but the decision was made to classify them into discrete categories, as was done for the other NHTS data. The resulting dataset can be organized in at least four different ways for analysis with BBNs. Each data framework has its own advantages and disadvantages as summarized below. In this paper, individual trip records were used as the unit of analysis for the transportation data.

1. **Individual trip records** allow for the maximum number of cases that the search score can use to generate the most compelling networks. For this study, 22,000 trip records were used to generate a model linking land-use variables, socioeconomic factors, and other variables to transportation mode choice. The drawback of trip-level analysis is that the total numbers of trips by mode cannot be analyzed.
2. **Spatial units such as tracts or zip codes** could also be used as the basic unit of analysis. For the study area here, there are just over 600 census tracts and just over 150 zip code areas covering the 6 counties. With this data structure, the characteristics of each spatial area could be summarized and transportation mode choice could be analyzed in terms of overall number of trips made by each mode.

For the technically inclined, this data structure is the transpose of the case above and although it results in a considerably lower number of records, it could be considered as a more geographically based approach. However, the number of trips in any given census tract or zip code may be limited due to the sampling structure of the NHTS and it may be insufficient to yield robust results.

3. **Individuals or households**, too, could form the basis for analysis. For the Baltimore add-on, there were approximately 7,800 individual records and 5,000 household records to analyze for the entire area of interest. An individual's full array of trips on the travel day could be the focus of analysis that would highlight an individual's autonomy in decisionmaking and the role of individual circumstances, resources, and constraints. Basing the analysis on households has the advantage of examining the full array of trips (or trips by specific modes) made at the household level, which may be the preferred decisionmaking unit and reflect shared resources and household responsibilities.
4. Finally, **trip tours** could be constructed and analyzed to understand the interdependencies that occur between a sequence of trips and their relation to personal, household, and land-use characteristics. Considerable effort would be required to construct trip tours, but this remains a very promising and relatively new area of investigation.

PRELIMINARY RESULTS

For this paper, the unit of analysis was Case 1 presented above—the individual trip. However, the land-use attributes for the trip origin were aggregated and tested at two geographic scales: census tracts and zip codes. As such, there were multiple resulting BBNs depending on the spatial unit of aggregation. The analysis was also carried out with and without all control variables, such as age, household size, and vehicle count, to investigate the influence of variables with considerably fewer discrete classes. Finally, the models were run with and without specifying variable ordering, such as the characterization of the variables as input, output, and super-groups. Table 3 summarizes the six model specifications.

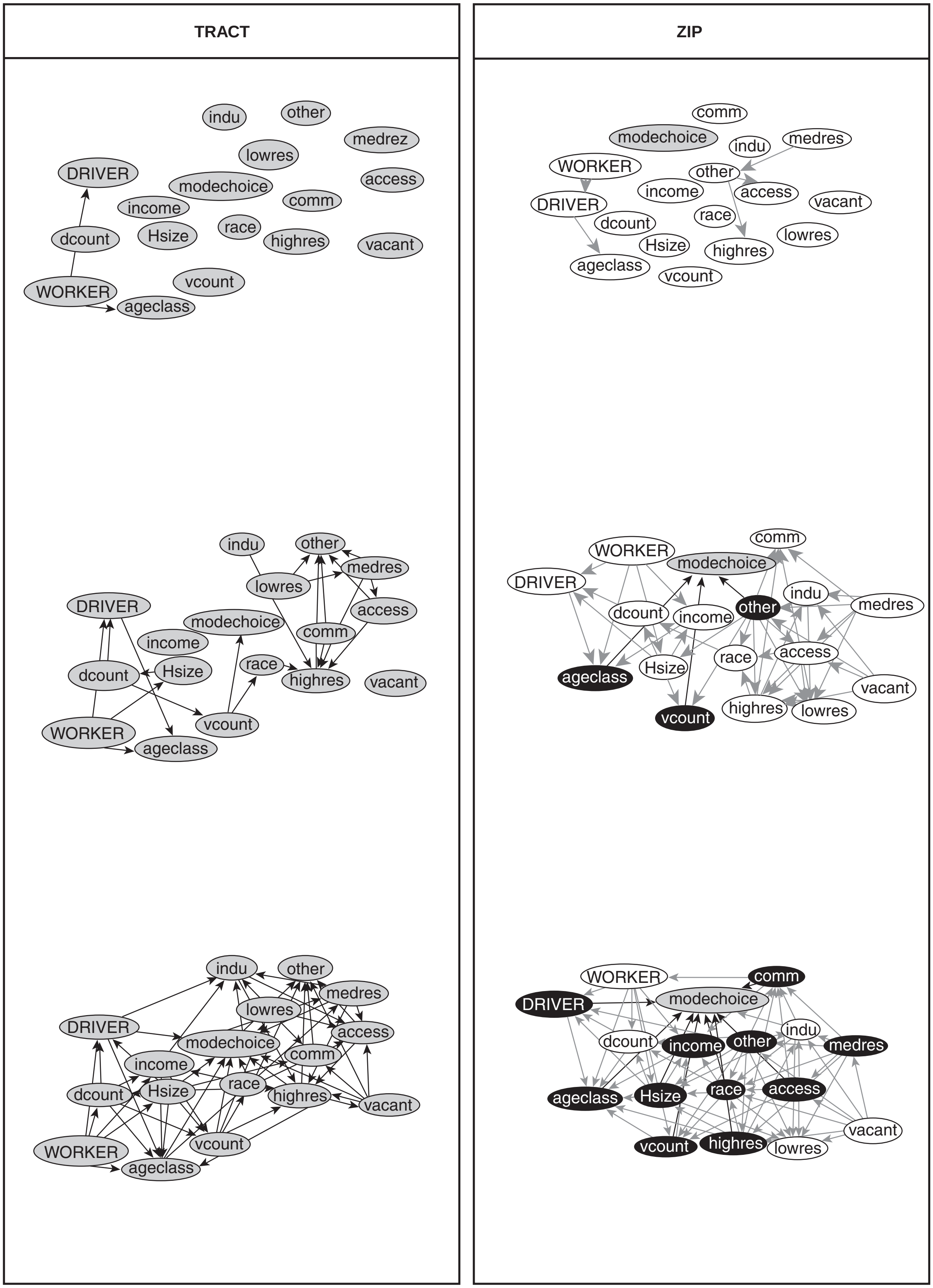
Models 1 and 4 created the most comprehensive results with a graphical representation of the relationship between land use, transportation choice, and all the control variables. In all cases, the graphs represent a relationship with a nondirected link having an arrow at both extremes. In the case of strong directional relationships that can be interpreted as causal relationships, the links show one single arrow pointing toward the child variable and originating from the parent node by which the child is influenced. All the relationships are quantitative in the sense that their strength is computed by the algorithm.

In figure 2, we present this strength in three sequential snapshots of the two models (where the land-use measures are calculated at both the tract and zip code level) that show first the strongest links, then the links with an average strength, and finally all links. In reality, the user can select the link strength as a continuum and obtain the appropriate display at any stage, a case that cannot be replicated on paper because of obvious space limitations. For the zip code model in figure 2, the analyst has selected the mode choice node as the one node of interest. The nodes in black are the parent set of the mode choice child, and they are presented in order of influential strength. In fact, the algorithm also distinguishes among variables predicted by and predictive of the variable of interest (mode choice) with appropriately colored nodes (not shown).

TABLE 3 Summary of Model Runs

By tract	All variables	No model input of any sort	Model 1
By tract	All variables	Model constrained	Model 2
By tract	Transportation land-use variables only	No model input of any sort	Model 3
By zip	All variables	No model input of any sort	Model 4
By zip	All variables	Model constrained	Model 5
By zip	Transportation land-use variables only	No model input of any sort	Model 6

FIGURE 2 Model Runs at the Tract and Zip Code Level



It is interesting to explore these outputs of the models in more detail. As the calculation of the land-use variables is moved from tract to zip level, it can be seen that land-use variables have a weaker influence on mode choice and in fact the percentage of vacant land is even excluded from our resulting BBN. These results can be explained by the fact that at coarser spatial aggregation, each spatial unit becomes more and more homogeneous compared with other polygonal areas and there is less variation in land use across tracts.

Although the goal of this paper is to present results as a proof of concept rather than an in-depth discussion of the land-use/travel behavior relationship, it is worth noting how the mode choice is influenced by the other variables. The strongest links associated with the choice of transportation are the availability of a private vehicle (condition sine qua non for driving), the driver status (having a license or not), age (another condition required to have a driving license), and how empty the landscape looks around the point of origin. This result is even more interesting if we consider that the land-use variable *other* includes agricultural land, which is critical in suburban or ex-urban conditions.

Figure 2 shows a limited sequence of how these links are progressively presented as part of the relationship structure; as the strength of the relationships weakens, we detect ethnicity, driver status, and the land-use variable of medium residential as also influencing mode choice. Household size, income, and number of commercial spaces were the least influential variables. These may be interpreted as important results, because they demonstrate that, despite their low income, poorer families also use private vehicles to a great extent. Weak relationships with income underscore the fact that low-income households rely on all modes of transportation, not just transit. It is only from the analysis of the conditional probability distribution (CPD) (table 4) that a broader interpretation for income is possible. Household size can be seen as a proxy for generating trips, but it is not a good predictor of mode choice and neither is the amount of commercial activity around the point of trip origin. This is a surprising result for those advocating mixed commercial uses around denser neighborhoods, but we will see later how this variable should in fact be grouped with other land-use variables.

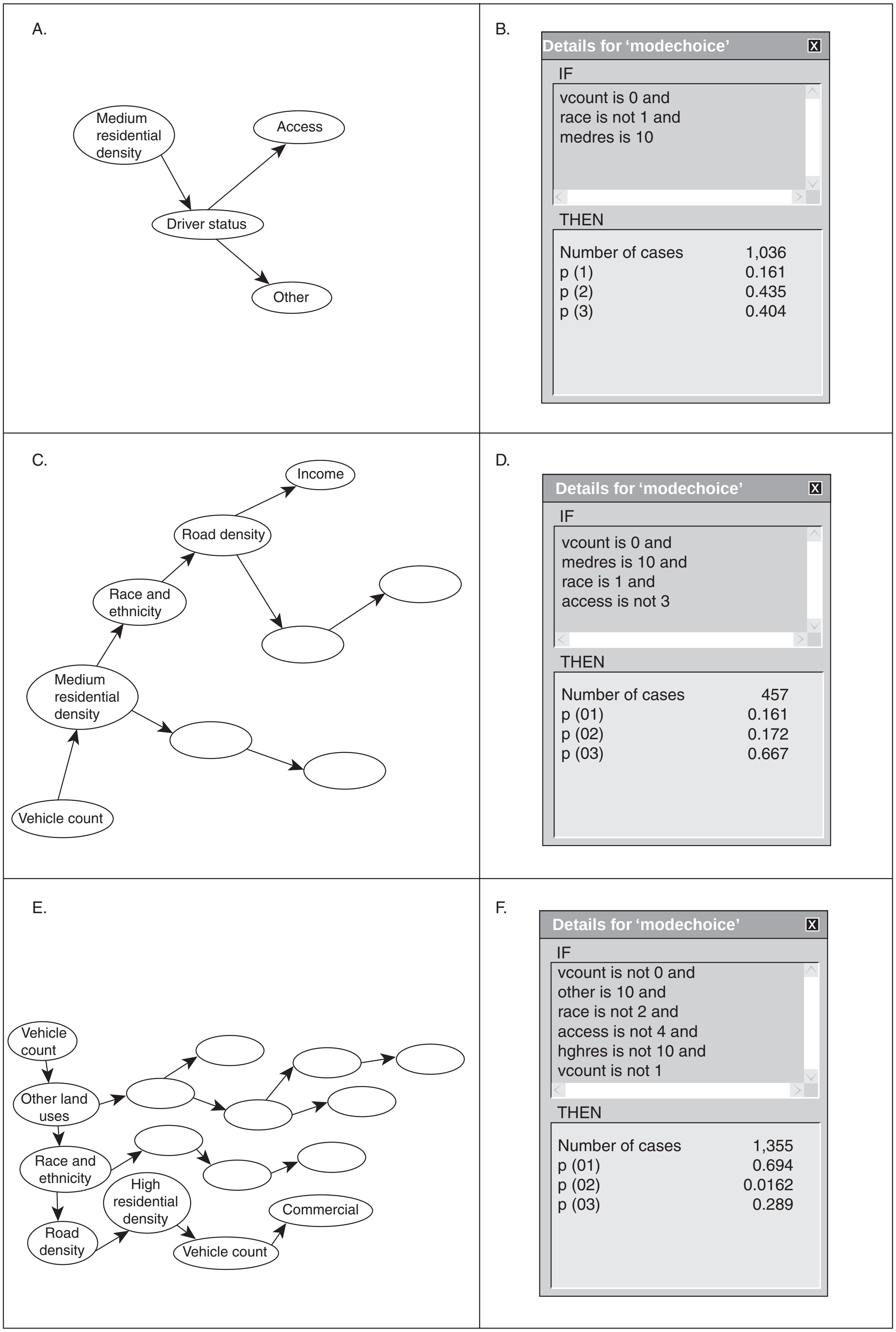
If the BBN outputs were limited solely to graphs, such as those presented in figure 2, the analysis would be little more than an intellectual exercise. However, each BBN algorithm provides the analysts with decision trees, based on a multinomial choice

TABLE 4 A Sample Conditional Probability Table (Version 1)

Land uses and income				Mode choices	
Income	% medium residential	% high residential	Probability motorized trip	Probability nonmotorized trip	Probability transit trip
0	20	60	86	6	8
0	30	10	88	5	7
0	10	60	60	11	29
0	10	70	60	12	28
2	40	20	86	6	8
2	20	70	60	11	29
4	20	50	37	17	46
5	10	80	60	11	29
6	10	50	60	11	29
6	90	10	86	6	8
8	10	90	37	17	46
8	90	10	88	6	6

Note: The probabilities of using a private vehicle are high regardless of the income level, which indicates that even persons with low incomes tend to use private vehicles for their mobility needs.

FIGURE 3 Decision Trees and Details of Probabilities of Mode Choice as a Function of Road Density, Income, and Percentage of Commercial Uses



distribution about how the various nodes behave in relation to each other in a quantitative way.

Figure 3A presents the decision tree associated with Model 1 when analyzing mode choice. Each node presents a binary split of one variable based on the conditions of the parent set of variables. For example, the trees can examine detailed questions such as: what are the probabilities that someone living in a medium-level residential area will choose driving versus transit as a function of vehicle availability and race? From the decision tree, one can see that for a vehicle count other than zero and for any race group other than white, the probability of a motorized trip is low; the probability of a nonmotorized trip is high; and the probability of a transit trip is medium. As we move toward the tree's end-leaves, these conditional probabilities are retained but the tree adds the case of no licensed driver available in the household, in which case the probability of transit trips almost doubles (from 30% to 60%).

Graphically, we have followed the path from the node *medres* (medium-density residential land-use) to *other* (other land uses) in figure 3A and, for each node, the probability information is presented as in figure 3B. The remaining details in figures 3C, 3D, 3E, and 3F zoom out to include all the paths of evidence from the strongest variable affecting mode choice (vehicle count) to the one variable of interest, in this case income and percentage of commercial land uses. In all cases, the probability of choosing one mode over another changes as the influence of a new variable is added to the set of parents.

The application of meta-search algorithms for the creation of BBNs results in graphs and decision trees. If using categorical data, it is also possible to calculate tables containing the CPD of each node and its parent set. The results look similar to tables 4, 5, and 6, where the probability of each mode choice is calculated as a function of the status of all the classes within the variables of driver count, vehicle count, percentage medium-density residential, and percentage high-density residential. The algorithm used in this research calculates the probability for the state of a class in all variables so the resulting tables are quite large. For example, in the case of the classes for variable driver count, which reports the number of drivers per household, these are assigned a probability of occurrence based on the occurrence of all other classes in all other variables. This is quite useful, but the algorithm has no knowledge that the land-use variables should all add up to 100% of the land-use cover for a given area. It follows that the CPD tables for land-use variables include situations where the occurrence of an 80% high-density residential area is compared with the occurrence of a 60% commercial land use, a case that clearly does not happen in reality.

One interesting outcome derived from the analysis of the six models' CPDs is that high probabilities of transit share and nonmotorized trips occurred either when the land-use variables have shown a large concentration of residential land use (as in a downtown area) or when there were small percentages of a mix of different land uses. This result would tend to quantitatively support the argument of those who favor mixed use as a means to improve transit ridership and abate private vehicle use and pollution.

TABLE 5 A Sample Conditional Probability Table (Version 2)

Land uses and vehicles					Mode choices	
No. drivers	Vehicle count	% medium residential	% high residential	Probability motorized trip	Probability nonmotorized trip	Probability transit trip
1	0	10	50	21	35	43
2	2	10	50	70	15	15
2	0	20	30	20	35	45
1	1	10	10	95	0	5
1	1	20	80	20	35	45

Notes: The algorithm properly finds that for those cases where drivers have a license but not a vehicle at their disposal, the probabilities of motorized trips are low. When a private vehicle is available, the probabilities of motorized trips are high but only if associated with small percentages of high residential land use.

TABLE 6 A Sample Conditional Probability Table (Version 3)

Land uses					Mode choices	
% medium residential	% high residential	% commercial	% other land uses	Probability motorized trip	Probability nonmotorized trip	Probability transit trip
10	10	20	60	90	5	5
10	10	40	10	52	16	33
10	40	30	20	69	18	13
10	80	10	0	62	8	30
20	10	10	10	78	11	11
30	10	10	10	78	11	11
10	20	50	10	53	14	33
10	30	20	10	53	14	33
10	40	20	10	63	7	30
20	60	10	10	78	10	12
10	30	30	10	52	15	33

Notes: Notice how a small mix of uses has high probabilities of transit and nonmotorized trips. The same happens with extremely high percentages of residential densities. Could this be the quantitative proof about land-use mixes?

DISCUSSION AND CONCLUSIONS

This paper presents the successful application of BBNs to the land-use/travel behavior relationship using data from the 2001 NHTS add-on, supplemented with local land-use and socioeconomic data. This "bottom up" or inductive approach can potentially contribute to the knowledge base by identifying relationships that might otherwise be masked by the limitations of traditional deductive approaches and by aiding in the development of theoretical models. The limited results presented here, however, were not meant to form the basis for theory building per se but rather demonstrate the utility of the NHTS data and the BBN method for future applications to theoretical and empirical investigations of transportation questions. In doing so, a number of advantages and limitations of this method were identified, as well as opportunities for future work.

The creation of BBNs provides the analyst with a model of the relationships among variables under study that is derived by means of meta-heuristic search methods. No statistical model needs to be specified a priori, and there is no need to characterize variables as independent or dependent. It provides quantitative assessments of the occurrences of specific outcomes based on the status of all other variables, and it allows for the study of complex problems based on how the data capture them.

The confidence that these graphs represent identify a real underlying relationship between land use and transportation remains to be tested. The "lift over marginal" log score provides information on how well the model fits the data. Also, as in all appropriate modeling attempts, it is possible to test the model on a subset of data to verify that its relationship construct and conditional probabilities still hold true.

Finally, there is no standard approach on how to compare the graphical results of a BBN with the quantities obtained by using traditional inferential statistics. One procedure proposed here is to derive elasticities by calculating the means for all variables and their associated regression coefficients using Bayesian inferential statistics. Once the elasticity for each variable has been established, simple comparisons could be made between the results obtained by means of heurist inductive reasoning and those derived by means of traditional deductive model building. Another approach is to translate the resulting BBNs into a discrete structural equation model of the standard error of the mean (SEM); this technique is a linear cross-sectional statistical approach that uses path analysis as its input. Analysts usually create the causal path among variables ad hoc, but the output of inductive reasoning, such as the BBN presented in this study, could be used as an independently derived variable path for SEM. Once again, elasticities could be derived to compare the results with other deductive studies.

This analysis of land-use and transportation interactions by means of BBNs has highlighted a number of important factors. In our analysis of the NHTS travel diary data, each trip was considered unique and was characterized by the land-use conditions of the tract or zip from which it originated. This assumption implies that each trip was treated independently of all other trips, even if some were originally taken as a part of a trip chain. In practice, the algorithm used in this analysis treated trips as discrete separate events, which is not always the case, as when multiple trips are made by the same person. This type of analysis is not necessarily based on the best assumptions but, as mentioned above, future analysis can be undertaken with the trip chains being explicitly considered as such. Furthermore, the aim in this research was to focus on the land-use conditions underlying the decision to use a particular trip mode, even as we recognize that interdependencies exist between sequential trips and their modal choices.

One issue with NHTS data was also related to the spatial limitation of the sample taken at the national and local level. Clearly such an analysis would have not been possible without the add-on data and the availability of records for about 25,000 trips in the Baltimore region. However, some issues were identified that relate to the spatial distribution of the respondents. Some tracts show as having no, or a low number of, trips originating from them, and repeating the analysis at a smaller geographic scale, such as the block level, will exacerbate this problem. This is not an issue when using journey-to-work data from the U.S. Census at the tract level, due to the more extensive household sampling of one in six households per tract (for the SF3 data).

A final issue relates to the scale of aggregation used in the analysis. The analysis at two different geographic scales, tract and zip code levels, is important to detect the sensitivity of dependencies between variables as a result of aggregations. These results are interesting, and, in the future, to investigate the effects from the Modifiable Aerial Unit Effect, the analysis will be carried out at four different geographic scales: traffic analysis zones, zip codes, census tracts, and block groups. Work is also underway to recreate the analysis with better measures of transit accessibility and to use the various administrative units, not trips, as the base records for the input database. The authors believe this to be a more geographical approach to the analysis of data, which would complement any analysis of survey data based on trips or personal information.

Finally, data mining applications using Bayesian approaches are in fact just one application area. Bayesian inferential approaches may also be used as modeling tools to create parameter estimates and develop forecasts. A worthwhile study would be to find the future transportation mode split, in light of infrastructure development—for example, to assess the impact of new transit lines or new bus routes. A network of relationships can be derived heuristically and directly from data. The second step is to collect sample data from the area in which the investments are to take place, to source the actual data with which to instantiate the Bayesian model. Inference is then the simple exercise of finding the posterior probability of each mode as a function of both the local data and the probabilities for the transportation choice parameters. This approach provides a simple and immediate local forecast of transportation mode split and the likelihoods for each mode; however, more accurate estimates can be obtained by slightly varying the instantiation values so that a number of equivalent posterior probability distributions can be sampled and a more robust simulation produced. This more complex method provides not only the probability of each mode split but also the probability distribution for such modes as other variables change.

It must be noted that unless data-collection efforts and surveys such as the NHTS continue to be carried out, the availability of large datasets required for the use of meta-heuristic algorithms will be limited, and thus the full potential of such a method in the field of planning might not be fully realized. This would be unfortunate, because there is a promising future for the application of Bayesian statistics and BBNs. Microsoft is already implementing these methodologies for data-mining functions in their flagship product, SQL. Academics in computer science are trying to implement algorithms that will specifically model feedback processes, and dynamic BBNs can be used to model changing relationships over time. For planners and transportation practitioners, the hope is that this method will provide us with the ability to gain more in-depth knowledge for the solution of complex issues.

REFERENCES

Aliferis, C., I. Tsamardinos, A. Statnikov, and L.E. Brown. 2003. Causal Explorer: A Probabilistic Network Learning Toolkit for Biomedical Discovery, in Proceedings of the 2003 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS), Las Vegas, Nevada, USA, June 23–26, 2003.

Badoe, D.A. and E.J. Miller 2000. Transportation Land-Use Interaction: Empirical Findings in North America, and Their Implications for Modelling. *Transportation Research Part D: Transport and Environment* 5(4):235–263.

Chickering, D.M., D. Heckerman, and C. Meek. 1997. A Bayesian Approach to Learning Bayesian Networks with Local Structure. *Microsoft Technical Report,* MSR-TR-97-07. Available at http://research.microsoft.com/research/pubs/.

Crane, R. 2000. The Influence of Urban Form on Travel: An Interpretive Review. *Journal of Planning Literature* 15(1):3–24, August.

Ewing, R. and R. Cervero. 2001. Travel and the Built Environment: A Synthesis. *Transportation Research Record* 1780: 87–122.

Friedman, N. and D. Koller. 2000. Being Bayesian About Network Structure. *Uncertainty in Artificial Intelligence.* Edited by C. Boutilier and M. Godszmidt. San Francisco, CA: Morgan Kaufmann Publishers, pp. 201–210.

Glymour, C. and G.F. Cooper (eds). 1999. *Computation, Causation, and Discovery.* Menlo Park, CA: AAAI Press/The MIT Press.

Heckerman, D., D.M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. 2000. Dependency Networks for Inference, Collaborative Filtering, and Data Visualization. *Journal of Machine Learning Research* 1:49–75

Heckerman, D., D. Geiger, and D.M. Chickering. 1995. Learning Baysian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning* 20: 197–243.

Jensen, F. 1996. *An Introduction to Bayesian Networks.* New York, NY: Springer.

_____. 2001. *Bayesian Networks and Decision Graphs.* New York, NY: Springer.

Neapolitan, R.E. 1990. *Probabilistic Reasoning in Expert Systems.* New York, NY: John Wiley and Sons.

Torres, F.J. and M. Huber. 2003. Learning a Causal Model from Household Survey Data Using a Bayesian Belief Network, presented at the Annual Meetings of the Transportation Research Board, Washington, DC.

Waddell, P. 2001. Towards a Behavioral Integration of Land Use and Transportation Modeling, presented at the 9th International Association for Travel Behavior Research Conference, Queensland, Australia.

A Closer Look at Public Transportation Mode Share Trends

STEVEN E. POLZIN*
XUEHAO CHU

Center for Urban Transportation Research
University of South Florida
4202 East Fowler Avenue, CUT100
Tampa, FL 33620-5375

ABSTRACT

Recent releases of census transportation information, American Housing Survey results, the National Household Travel Survey (NHTS), American Public Transportation Association ridership statistics, and Federal Highway Administration vehicle-miles of travel data provide opportunities for researchers and policy analysts to glean information on travel behavior trends in the United States. Several data sources, specifically the NHTS, shed light on changes in transit use and mode share trends at the national level. This paper looks at transit mode share trends with both field count and survey data results.

The research indicates that unlinked transit trips declined in the early 1990s followed by ridership growth through 2001, at which point ridership began declining again before rising in 2004. It is clear that transit has grown in terms of total trips, and its overall mode share has stabilized. As overall national travel growth has slowed, transit use appears to be fluctuating between positive and negative growth in terms of both absolute trips and transit's share of overall travel. The research also identifies the shortcomings and differences of the various data sources for determining transit use and mode share trends.

Email addresses:
* Corresponding author—polzin@cutr.usf.edu
X. Chu—xchu@cutr.usf.edu.

KEYWORDS: NHTS, transit ridership, mode share, commuting.

INTRODUCTION

Mode share and transit ridership trends are relevant to a number of policy deliberations. Policies and investments are often designed to increase transit ridership or mode share, and subsequent measures of ridership provide feedback on market response. Arguments are often made linking mode share trends and public funding for transit. This link between funding and trends in transit ridership has been presented in several different forms by both advocates and critics of various initiatives to fund public transportation. During the recent reauthorization of the federal surface transportation program, transit supporters, for example, used increases in transit ridership during the later half of the 1990s as a reason to support increased federal funding in this area (STPP 2002). Opponents, on the other hand, used the continued decline in the mode share of transit from the decennial census as a reason for reducing federal funding for transit (Cox and Utt 2002). Perceptions of transit ridership levels and trends can influence funding levels, research priorities, and investment decisions at all levels of government (Urban Mobility Corp. 2002).

Developing a clear understanding of what is actually occurring regarding transit use trends is highly dependent on what is measured and reported. Critical issues include:

- Is the focus on absolute ridership or transit mode share?
- Is the unit of measurement unlinked trips, linked trips, or passenger-miles?
- Is the data source observation (count data) or respondent-stated (survey data)?
  —For survey data, is the definition of use “actual” or “usual” mode?
  —For survey data, how large is the sample and what biases might exist?
  —For count data, how accurate and comprehensive is the measure?
- Are the sampling errors and nonsampling errors such that confidence can be placed in the estimates?

Many researchers have studied the trend in transit ridership and mode share, and others have studied factors and policies that influence trends in transit ridership and mode share. Much of the previous literature and policy debates used a single measurement of ridership. One exception is Pisarski (2003), who compares and contrasts the recent trend in transit mode share using information from the National Transit Database versus information from the decennial census. The research presented here goes beyond Pisarski (2003) by providing an overall look at the recent trend in transit’s modal share. This paper contributes to the literature and the policy debates by comparing a variety of data sources and measurements of transit’s mode share. In addition to presenting a comprehensive perspective of transit use trends, the paper comments on both the quality of the data and the implications of the trends in the context of ongoing policy deliberations regarding transit funding.

Releases of new data, including census transportation information, the American Housing Survey results, the National Household Travel Survey (NHTS), and regular updates to American Public Transportation Association (APTA) ridership statistics and Federal Highway Administration (FHWA) vehicle-miles of travel (VMT), provide opportunities for researchers, policy analysts, and others to glean information regarding travel behavior trends in the United States. This paper reviews various sources of data, including both survey data results and field count data, from which one can develop estimates of public transportation mode share trends. The analysis provides a richer understanding of mode share trends as well as insight into issues associated with the relationships between the various data sources. Most studies of mode share and transit ridership trends are motivated by a desire to understand causal factors underlying ridership (Joint Center for Political Studies 1985; TRB 2001; Millar 1999; Mason 1998).

In the era of the Intermodal Surface Transportation Efficiency Act of 1991, a plethora of studies targeted strategies for enhancing ridership (Taylor and McCullough 1998; Kain and Liu 1999; Stanley 1998; Project for Public Spaces 1999; Urbitran Associates 1999; Taylor and Haas 2002; Norman 2003; Schmidt 2001). All these initiatives benefit from a rich understanding of ridership and mode share trends.

ALTERNATIVE MEASURES OF TRANSIT MODE SHARE

Count-Based Measures

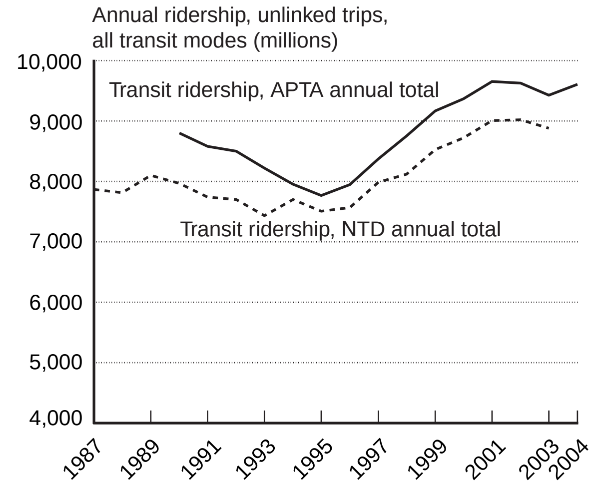
Figure 1 illustrates the reported transit ridership expressed as annual national total ridership on public transit. These trends are drawn from two data sources: APTA,[1] which receives quarterly vehicle boarding counts from members that are factored into a national total; and the National Transit Database (NTD),[2] which gathers annual sampled counts of ridership reported to the Federal Transit Administration by agencies receiving federal funds.

The numbers reported show meaningful positive increases in transit ridership of approximately 22% between 1995 and 2001. The 2002 and 2003 APTA data show a reversal of the trend as the economy slowed and related fare and service changes resulted in declines in ridership in those years. Data for 2004 indicate a recovery in ridership to 2001 and 2002 levels. For 2002, NTD data indicate a very slight increase in ridership and for 2003 show a decline similar to that shown by APTA. Both of these sources report measures of persons boarding transit vehicles (called *unlinked trips*). If a person has to board two or more vehicles to complete a trip to a destination, this is defined as a *linked trip*. Both data sources are subject to errors associated with farebox data-collection methods and neither contains the full universe of transit operators. However, they represent the best available aggregate count data and reasonable sources for understanding industry trends.

Figure 2 shows the most recent data on overall travel trends as measured in percentage change in VMT. The trend shows a declining growth rate over

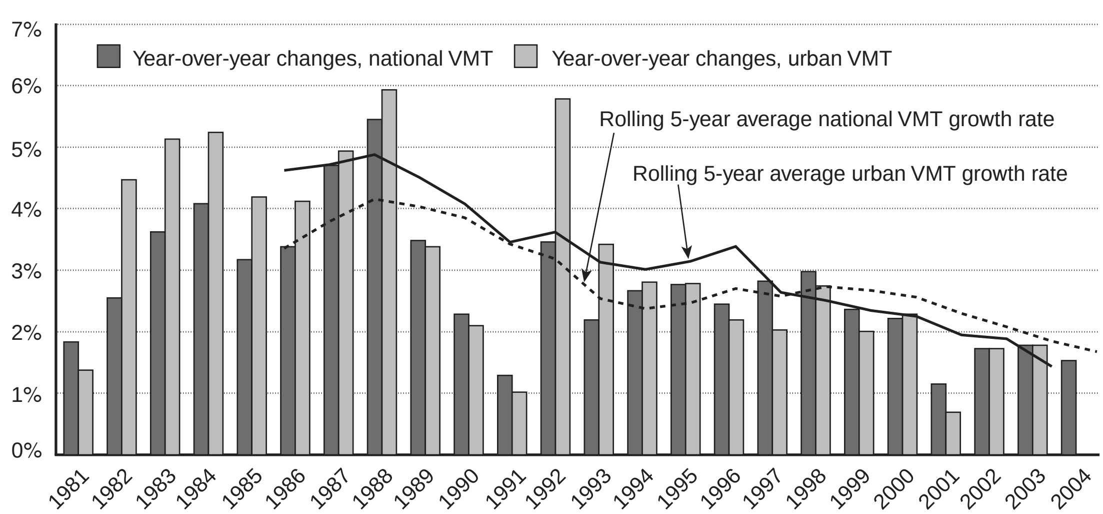
[1] See http://www.apta.com/research/stats/ridershp/#A3.
[2] See http://www.ntdprogram.com.

FIGURE 1 Transit Ridership Trends



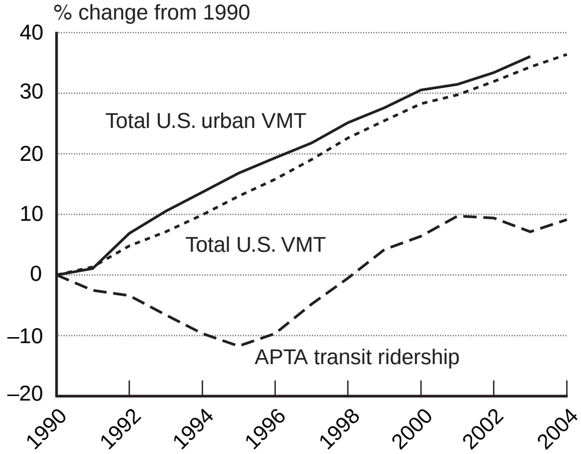
Sources: See the following websites: http://www.apta.com/research/stats/ridershp/#A3; http://www.ntdprogram.com/NTD/NTST/2003/PDFFiles/2003%20National%20Transit%20Summaries%20and%20Trends%20(NTST).pdf.

FIGURE 2 National VMT Trends



Source: Data assembled by the Federal Highway Administration.

FIGURE 3 Rates of Change in Transit Ridership and Vehicle-Miles of Travel (VMT)



Source: University of South Florida, Center for Urban Transportation Research, analysis of FHWA and APTA data.

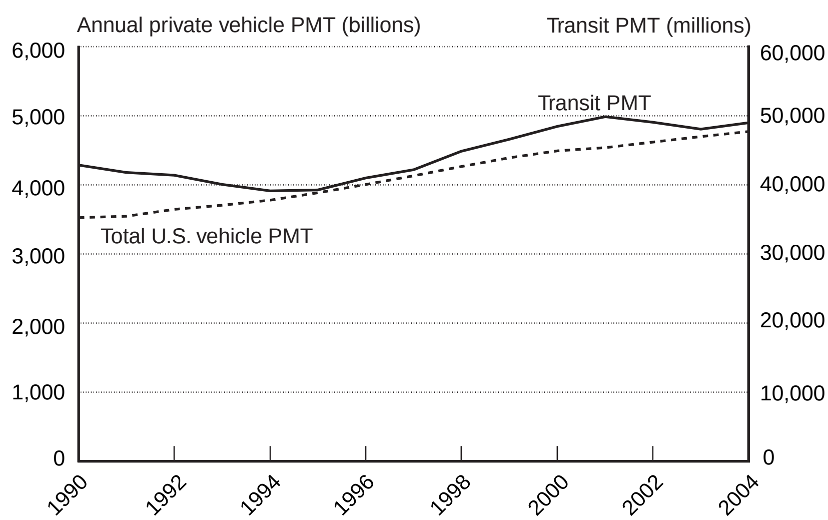
the past several years. The data, based on FHWA reporting of VMT through 2004, include total urban and national VMT. Urban VMT rates of change go from being higher than national totals, indicating a growing share of total VMT in urban areas, to a situation where total VMT outpaced urban VMT, indicating more rapid growth in non-urban areas. In both cases, the pace of VMT growth has clearly slowed.

Figure 3 displays the relative rates of change for VMT and transit ridership. More rapid rates of change for transit ridership indicate times when transit is likely to gain market share (assuming constant length transit trips, because the figure compares transit trips against vehicle-miles for auto). Based on this indicator, transit was losing market share between 1990 and 1995, gaining share from 1996 through approximately 2000, and subsequently losing share in more recent years.

Figure 4 indicates changes in person-miles of travel (PMT) for auto and transit, using an estimated measure of PMT. Transit PMT is estimated by multiplying trips measured by APTA by an average transit trip length developed yearly from NTD data. Auto VMT are converted into PMT by factoring VMT by vehicle occupancy. Vehicle occupancy uses NHTS data and interpolates between survey years. This enables the development of a measure of mode share that compares person-miles for privately operated vehicles versus public transit. It accounts for the differences in average trip length by mode and thus more accurately reflects travel by each mode. Because unlinked transit trips are significantly shorter than auto trips, the mode share calculation, based on PMT, is markedly lower than the level for trip-based count or survey measures.

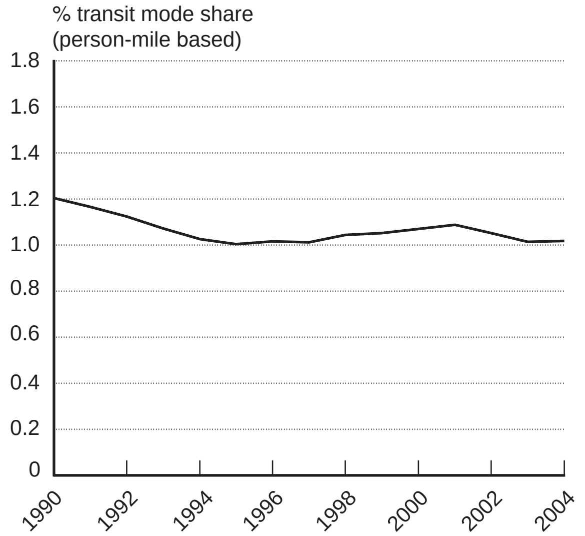
Figure 5 indicates a slight increase in PMT-based transit mode share from 1995 through 2001, with the trend reversing and showing a decline in share for transit after 2001. This measure shows the estimated 2003 mode share being at one of the lowest

FIGURE 4 Person-Miles of Travel (PMT) Trends for Vehicles and Public Transit



Note: Scales vary by a factor of 100.
Source: University of South Florida, Center for Urban Transportation Research, analysis of FHWA and APTA data.

FIGURE 5 PMT-Based Transit Mode Share



Source: University of South Florida, Center for Urban Transportation Research, calculations based on figure 4.

historical levels with approximately 1% of total national PMT being carried by public transit.

An alternative strategy for reporting PMT-based mode share is to use urban, noncommercial vehicle PMT, since this is the more comparable market for most transit services. Transit is not intended as an alternative for commercial/freight traffic or for intercity travel. Based on the share of VMT that is urban, approximately 60%, and factoring out commercial traffic from the measure of VMT, the values for PMT-based mode share increase by about 75% to approximately 1.75%. These adjusted comparisons give a clearer impression of the role that transit plays in urban personal mobility; however, the PMT-based indicator is relevant in the context of discussion of overall transportation investment policy.

Survey-Based Measures of Transit Mode Share

The previous section's derived mode share estimation is only one way to explore transit mode share trends. Other national survey data also provide insight into transit mode share.

Census Journey-to-Work

Journey-to-work mode share can be calculated from census data long-form information. These data are available for prior censuses and contain a large sample with a high response rate. The census data are based on a question that asks: "How did you usually get to work last week?" Guidance is provided to the respondent relating to multimodal trips where the dominant mode is to be noted as the primary mode, and how to handle multiple work trips, working away from the normal workplace location, etc. A detailed list of transit modes is defined including taxi, ferry, commuter rail, etc. Work at home is a category of response and is typically included in the denominator of the mode share calculations.

For the census, the spring delivery results in the respondent answering with respect to the narrowly defined timeframe and hence does not capture seasonal variation. The greatest sensitivity regarding the application of census data relates to whether or not the "usual trip" language impacts the interpretation of the results in contrast to other measures. A perception exists that transit may be an occasional mode for noncaptive travelers and hence usual mode measures might underrepresent actual everyday average use. This is discussed in more detail below.

American Community Survey (ACS)

As the planned replacement for the decennial long form, this annual smaller sample survey is similarly structured and has been in the pretest application stages before planned ongoing implementation starting in 2005. The commuting questions in this survey follow the census long-form language by querying about the most frequent mode in the reference week. Respondents are continually surveyed (unlike the census). Work-at-home respondents are included in the denominator. The available ACS results from the sample counties are counties that, with respect to transit mode share, are more transit-intensive. (The U.S. Department of Transportation (DOT) has evaluated the census long form mode share results for these same counties in comparison to national average mode shares to determine why the ACS has shown a somewhat higher mode share.)

American Housing Survey

This survey[3] is conducted by the Census Bureau for the Department of Housing and Urban Development. It collects data on the nation's housing: apartments, single-family homes, mobile homes, and

[3] See http://www.census.gov/hhes/www/ahs.html.

vacant housing units; and household characteristics, income, housing and neighborhood quality, housing costs, equipment and fuels, size of housing unit, journey to work, and recent moves. National data are collected in odd numbered years, and data for each of 47 selected metropolitan areas are collected about every 6 years. The national sample covers an average of 55,000 housing units. Each metropolitan area sample covers 4,100 or more housing units. The mode question is identical to that asked in the census long form, in the ACS, and in the 2001 NHTS person file.

Omnibus Household Survey

The Omnibus Household Survey,[4] conducted by the Bureau of Transportation Statistics, is a major data-collection exercise to assess customer satisfaction in fulfillment of the DOT Performance Plan. The survey asks supplementary questions every other month to address five DOT strategic goals: safety, mobility, economic growth, the human and natural environment, and national security. It asks general questions about satisfaction with the transportation system and public interactions with DOT agencies. Data for the survey come from telephone interviews of approximately 1,000 randomly selected households and are weighted to allow inferences about the non-institutionalized population aged 18 years or older currently living in the United States. The mode question, asked every other month, is stated as:

On a typical day in September, to get to work did you:

01) Walk
02) Drive or ride in a personal vehicle, not in a company car
03) Drive or ride in a carpool or vanpool
04) Use public transit
05) Drive or ride in a company car
06) Bicycle to work
07) Use a combination of modes
97) Other

Figure 6 shows the trends for these different national travel surveys. Each survey uses somewhat different sampling methods, definitions of terms, and reference time periods. Information about sample sizes, errors, and the specific questions is readily available from the respective survey agencies. These data suggest that over a longer period of time (e.g., comparing 1990 and 2000 data), the transit mode share has declined for census and household survey data sources. Survey information from the more recent years paints a somewhat less clear picture. Of particular interest is the NHTS. This source indicates a mode share of 1.59% of person-trips on transit. Differences in survey questions, mode classifications, and samples require modifications to the data to make meaningful comparisons to the prior years' data. Adjustments for sample and definition differences result in a mode share of 1.76%, closer to the 1.81% the 1995 survey. Thus, this data source suggests a very slight decline in overall mode share for transit in the past six years. This is discussed in more detail below.

COMPARING NHTS/NPTS TRENDS

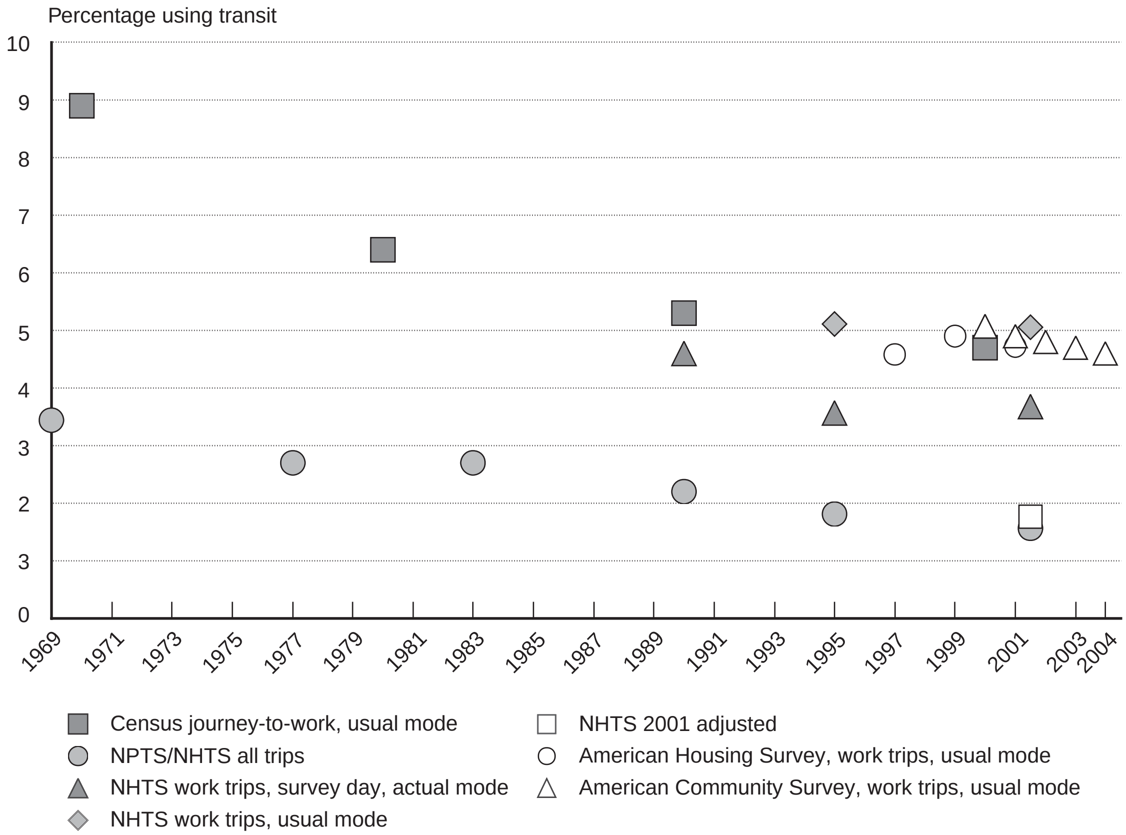
The survey methodology for carrying out the NHTS is refined with each application in order to provide the best possible data while still trying to preserve comparability over time. In comparing the 2001 NHTS with the 1995 Nationwide Personal Transportation survey (NPTS) transit mode share calculations, there were several subtle differences that needed to be accounted for to enhance the comparability of the estimates.

Use of an add-on sample. The 2001 national sample NHTS numbers produced a transit mode share of 1.561%. This, however, is not directly comparable to the 1995 number for a number of reasons. The 1995 NPTS included the add-on samples (add-on samples are larger samples above and beyond the original national sample purchased by specific geographies to address local needs), which, while factored to remain representative of national totals, nonetheless, produced a slightly different transit mode share. When the 2004 release of the NHTS database with the add-ons included became available, the transit mode share was again calculated and produced a slightly higher 1.591% share.

Adjustment for higher walking trip reporting. The 2001 NHTS was designed to try to do a better job of gathering information about walking trips. This included an additional probing question to spe-

[4] See http://www.bts.gov/omnibus_surveys/household_survey/.

FIGURE 6 Transit Mode Share Trends—Survey Data



cifically solicit information on walk travel. The result was a significant increase in reported walking trips presumed to be well beyond actual changes in the walk mode and a result of the change in survey design. This increase in total trips about which information was gathered had the effect of slightly depressing the transit mode share because the total trip denominator was now a larger number. To quantify the impact of this, it was assumed that the walk trip rate remained the same as in 1995 for purposes of estimating total trips. With this adjusted measure of total trips, the transit mode share would have been approximately 0.04% higher.

Definition of transit. The 2001 NHTS had a slightly different set of submodes that were classified as public transit. In 1995, intercity bus and courtesy bus were included in the calculation of transit mode share. The 2001 survey disaggregated the data to allow a closer estimation of what is typically referred to as public transportation. If the 2001 transit mode definition is adjusted to be most comparable to the 1995 data, transit mode share in 2001 shows an increase of approximately 0.065%.

Children under age five. The 1995 and prior surveys excluded trips by children under five years of age. This population segment travels only modest amounts and disproportionately less by transit. If the 2001 data are adjusted by removing children under five to be most comparable to 1995, the mode share for public transportation increased approximately 0.029%.

Collectively, these adjustments produce the mode share calculation to be used in comparison with the 1995 NPTS findings (summarized in table 1). It is important to understand that these adjustments are made only to increase the comparability between the 1995 and the 2001 survey numbers. In absolute terms, the 2001 NHTS directly calculated mode share number appears to be the more accurate reflection of actual transit share. The particular interest in exploring this issue in greater detail is to both allow a more comparable data trend analysis and specifically to explore the relative change in mode share between 1995 and 2001, as indicated in NHTS/NPTS data, versus the changes perceived and

TABLE 1 Summary of NHTS 2001 Mode Share Adjustments

	Percent
2001 NHTS transit mode share for all trips	1.561
Adjustment for add-on sample changes	0.030
Adjustment for walking share change	0.040
Adjustment for transit definition	0.065
Adjustment for inclusion of children under 5	0.029
Adjustment for using non-add-on 2001 to compare with add-on 1995 data	0.030
Adjusted 2001 NHTS public transit mode share (total)	1.755

calculated by looking at field data on ridership changes and calculated mode shares. This will be discussed in more detail later.

Table 2 presents a variety of different survey-based measures of transit mode share. These are for various points in time, various survey methods, and various trip purposes. Caution should be used when comparing these data items; however, the collective message can provide guidance to analysts regarding mode share trends.

Table 3 shows the transit share from the Omnibus Household Survey. These numbers should be used with caution for two reasons. The sample is small for measuring transit share, and the transit category may exclude transit used as part of a trip on a combination of modes. Thus, while these data may be interesting, we did not include them in figure 6 nor do we comment on them in detail.

NHTS "USUAL" VS. "ACTUAL" WORK TRIP MODE

One of the challenges in comparing mode shares measured across data sources is understanding the comparability of questions that inquire as to usual mode from those that seek information about a specific trip (actual mode). NHTS is unique in that both questions are asked of respondents, thus providing an opportunity to reflect on the differences. As indicated in table 2, NHTS data on actual work trip mode share is noticeably different, with actual mode share on transit being more than 1% lower than the usual mode measures. This indicates that individuals who indicate a usual mode of transit are less likely to use transit as an actual mode on a given day.

Usual mode questions typically refer to the conditions for the prior week. Thus, the respondent is

TABLE 2 Comparisons of Various Survey Estimates of Public Transportation Mode Share

Year	Census journey-to-work, usual mode	American Community Survey, work trips, usual mode	American Housing Survey, work trips, usual mode	NHTS work trips, survey day, actual mode	NHTS work trips, usual mode	NPTS/ NHTS all trips	NHTS 2001 all trips, adjusted
1969						3.40	
1970	8.90						
1977						2.70	
1980	6.40						
1983						2.70	
1990	5.30			4.60		2.20	
1995				3.56	5.11	1.81	
1997			4.58				
1999			4.90				
2000	4.70	5.07					
2001		5.10	4.72	3.67	5.05	1.56	1.76
2002		5.00					

TABLE 3 Omnibus Household Survey Transit Use Results

Survey month	Transit share (%)	Sample size
February 2003	3.69	16
April 2003	1.89	10
June 2003	3.52	20
August 2003	3.36	21
October 2003	3.21	15

answering in the context of a specific period of time that may have included multiple work trips and multiple modes. Presumably, someone who travels on a given mode more than half the time would indicate that as the usual mode. It is not uncommon, for example, for a transit traveler to commute by transit four days per week and then take an auto on Fridays to facilitate an evening event or early work departure. Similarly, a regular auto traveler may choose to use transit on a given day due to auto unavailability or other factors. The usual mode range of categories also includes a "work at home" choice; thus, this deflates the shares for the other categories slightly as this category is now included in the denominator in the share calculation. For actual mode questions, work at home is not a choice; thus the shares for all other options are proportionally slightly larger.

Table 4 presents an analysis of the usual and actual travel mode for work trips from the 1995 and 2001 surveys. This table confirms an interesting phenomenon. Auto usual mode travelers are far more likely to be strongly loyal to the auto mode, with very modest use of transit for their actual trips (0.1% transit use for actual trips in 2001), whereas usual transit mode travelers used auto modes for 18.4% of their actual trips in 2001. These data confirm the behavior that is required to produce the differences between the usual and actual mode shares observed in NHTS data. To further verify this phenomenon, we analyzed 1995 data, also presented in table 4.

One can apply some algebraic calculations to derive the required mode loyalty for auto and transit travelers for the reported differences in usual and actual mode relationships to be valid. For these conditions to be true required less than 4% of usual mode auto travelers to use transit on a given trip. As the actual transit use by usual auto travelers declines there is an opportunity for greater auto use by transit travelers. An equation can be defined to describe the conditions that would be required to produce any given combination of transit usual mode and actual mode shares.

TABLE 4 Work Travel Usual vs. Actual Mode Choice Percentages

	Usual mode		Actual mode on travel day					
Year	Mode	Share	Drive alone	Carpool	Transit	Walk	Bike	Other nonreport
1995	Private vehicle	92.2	81.8	15.3	0.3	0.5	0.1	2.0
	Transit	**4.7**	11.5	10.8	65.6	7.4	0.2	4.5
	Walk	2.6	13.5	9.0	3.2	50.8	0.3	23.3
	Bike	0.5	9.4	11.9	0.3	4.9	68.7	4.9
	Share	100.0	75.6	14.8	**3.6**	2.4	0.5	3.1
2001	Drive alone	83.2	89.8	9.4	0.1	0.4	0.1	0.2
	Carpool	9.1	21.7	75.7	1.0	1.3	0.1	0.2
	Transit	5.0	8.4	10.0	69.3	8.5	0.4	3.3
	Walk	2.3	10.3	9.0	2.7	77.3	0.2	0.5
	Bike	0.4	8.1	10.1	1.4	7.7	72.7	0.1
	Share	100.0	77.4	15.4	**3.7**	2.7	0.4	0.4

Concept source: Analysis by University of South Florida, Center for Urban Transportation Research, based on N. McGuckin, Work, Automobility, and Commuting (Chapter 4), *Travel Patterns of People of Color, Final Report,* prepared by Battelle (Washington, DC: U.S. Department of Transportation, Federal Highway Administration, June 30, 2000). Data based on January 2004 release of NHTS data.

These data, while logical when analyzed, are contrary to some perceptions of the impacts of reliance on the usual mode question in many travel surveys. The usual mode question has troubled some policy analysts because of the lack of certainty it creates in the data, and because there is the expectation that over time the loyalty to any given mode is lessened as more choices are available. It is more common—due to higher auto ownership/availability, more working spouses and flexible work arrangements, more prevalent alternatives to driving (work at home, transit, car/vanpool)—to presume that travel arrangements may be becoming more diverse with individuals choosing different modes in response to specific activity plans for the day. Thus, there have been some concerns that the usual mode measure might be underestimating transit use as well as use of other modes like walk, bike, and shared ride. However, for transit, the data do not bear out this perception. In fact, the usual mode question strategy appears to overstate the actual share of workers commuting on transit in any given day.

As the data in table 4 suggest, actual travel day behavior can vary significantly from usual mode. In general, for individuals whose usual mode is transit, less than 70% of them used transit on the actual day. Usual transit users frequently share a ride, walk, or use a single-occupant auto. The 2001 survey suggests that usual transit users were slightly more loyal to transit on the actual day than was the case in 1995. In the case of auto, the data suggest that of those with auto travel as a usual mode, over 97 percent used auto on the specific travel day.

These shares are consistent with those that would make the reported usual mode and actual mode transit shares mathematically correct. Usual mode auto travelers seldom use transit for their actual trip, with travel on transit being only a fraction of 1% of actual trips. Usual transit travelers have the least loyalty to their mode of all the usual categories, whereas drive alone auto usual mode individuals have the greatest mode loyalty.

Between 1995 and 2001, the differences between usual mode and actual mode grew significantly indicating that non-usual transit users were less likely to use transit. This increased loyalty to all modes perhaps runs counter to perceptions of a commuting force that is using a range of travel options. A host of factors, including the share of households with more than one vehicle and the tight time constraints for workers in the strong economy period leading up to 2001, might be supporting the trend to greater mode loyalty. It is important to remember that a significant share of the auto commuters do not have walk or shared ride options available to them and may not have transit access at one or both ends of their work trip, thus they use autos 100% of the time.

It is difficult to draw many conclusions from these data on the degree of captivity of transit travelers or the degree of options open to auto travelers. However, it does make both mathematical and logical sense when reviewed in the context of the observed travel behavior of the public. While occasional use of transit may be growing for work travel as the work force grows, there is no evidence that the share of occasional transit use by usual auto travelers is growing. This analysis is restricted to work travel and should not be generalized to other trip purposes.

TRANSIT MODE SHARE AND DATA QUALITY

One of our objectives in compiling various measures of transit use was to take into account that individual measures of transit use or mode share are each subject to various limitations associated with the way they are collected. As noted, each source has limitations including sample sizes and response rates as well as the inherent sensitivity to response bias that is of particular concern to transit researchers.

Transit, by its very nature, suffers from response bias in many travel surveys. Transit travelers are more inclined to have language or literacy problems, be reluctant to disclose sensitive information, be less likely to have telephones available for phone surveys, or otherwise be at risk of being underrepresented in survey data-collection efforts. Nonetheless, within each data source the quality of the data appears to be improving over time and longitudinal comparisons provide insight into overall aggregate national trends. While one would appropriately use greater caution when working with smaller subsets of these databases, the application of ridership and mode share trend data for national transit policy is supported by the collective set of available data. There is no evidence or speculation that response bias has changed over time; thus, no basis exists for discounting the value of the longitudinal survey data as an indicator of trends.

Within the transit community there are a number of initiatives that are likely to improve the data on ridership levels over time. APTA and the Federal Transit Administration are collaborating to produce a single national estimate of annual transit ridership that will reduce the confusion of having multiple national count measures. NTD data are continuing to improve with more agencies using farebox count data rather than sampling as the basis for ridership estimates, and swipe or scan passes improve the count information. The emerging implementation of Automated Passenger Counters offers the prospect of more reliable ridership and trip length counts. The initiative to implement monthly NTD data collection also offers the prospect of greater quality control for NTD data.

Survey design and sampling methods continue to improve; however, resource constraints and the intractability of literacy, language, and disclosure fears impeding response rates will continue to challenge survey methods to provide highly confident measures of transit use, particularly for sensitive subgroups. Larger samples offer the greatest chance of improving confidence in survey data, particularly for geographic or other more narrowly defined market segments. Careful sample design and more aggressive nonresponse followup can increase the response rates and minimize bias.

Various initiatives underway such as the American Community Survey, which is planned to provide year-round samples and more experienced professional data-gathering staff, may improve data quality for work trip commute questions. Beyond data-gathering initiatives already discussed, there is certainly room for a richer understanding of how transit is used as a component of a multimodal trip and in developing a better understanding of the relationship between linked and unlinked trips at the national level.

A greater understanding of the alternative mode for transit trips would provide more insight into the transportation impacts of transit trips. However, simply understanding the data differences in existing data sources and presenting the variations is an important first step in using transit mode share data for various policy deliberations. At the national level, the collective body of existing data provides a sound basis for having stronger insight into mode share trends. While the data are not perfect, their shortcomings should not dissuade its use for policy deliberations nor discredit the messages that can be gleaned from a multiple source review of mode share trends.

INTERPRETATION OF THE MODE SHARE TREND DATA

The following observations can be discerned from the body of data on transit use and mode share:

- The evidence on transit use trends across sources is consistent with declines in unlinked trips in the early 1990s followed by strong ridership growth through 2001, at which point ridership began declining.
- No body of data exists on the industrywide changes in the relationship between linked and unlinked transit trip making (the ratio of unlinked to linked trips). However, the evolution of more transfer-friendly fare media—such as all day passes and the expansion of rail systems that can produce higher total boardings (as some one seat bus trips now become a feeder bus and rail trip)—may be increasing the ratio of unlinked to linked trips. For example, fare structure changes in New York City, where an all day pass was instituted, contributed to the growth in ridership because individuals no longer had to pay for boarding each subsequent vehicle for multiple vehicle trips. However, the trend in public transit PMT is clear and tracks with the trend in trips, as the average trip length has remained relatively constant according to NTD measures.
- All the data sources appear to confirm the decline in mode share for both work and nonwork trips through 1995.
- All the survey data sources appear to confirm the stable to slight upward trend in work trip mode share from 1995 to 2001 (unfortunately count

data do not provide trip distribution information to confirm these trends). The census data in 1990 and 2000 bridged the trough in transit mode share and do not reflect the turn in trend in the mid-1990s.

The most challenging discontinuity among the various data items is that the NHTS overall mode share trend from 1995 to 2001 does not appear to support the calculations of the ridership count data sources. The PMT-based measures of mode share showed an increase in the share of trips on transit between 1995 and 2001. Had that been confirmed by the NHTS, the NHTS mode share number would have been approximately 1.95% rather than 1.76%. It is not possible at this point to explain the differences in share. All the data sources, including the count data, are subject to a variety of uncertainties. For example, the significant differences between NTD and APTA data for a given property and for the country as a whole are uncomfortably large (Chu 2004).

Count data are likely to become more reliable and higher due to electronic fareboxes and automated ticket vending and, thus, part of the recent trends in ridership growth may not be actual increases in transit use. Others have speculated that the shift to pass-based fare systems has resulted in unlinked trip numbers increasing faster than linked trips or passenger-mile numbers. Each measure of transit use has a slightly different definition and trip linking and trip length are not robustly determined. Among the possible explanations for the NPTS/NHTS trend from 1995 to 2001, which is inconsistent with industry data, are that the 1995 NPTS overstated transit use or that the NHTS survey method resulted in a noticeable undercount of transit ridership.[5] In spite of this inability to completely rationalize the various data sources, some clear conclusions can be drawn:

- Regardless of various refinements that may be identified over time, it is clear that transit has grown in total trip terms and has stabilized its overall mode share or perhaps changed modestly through 2001. The work trip share appears to have grown slightly in the late 1990s, but the duration of the growing mode share may have been quite limited. Also, national aggregate ridership count data do not include trip purpose data nor do they enable measurement of linked trips, thus complicating interpretation.
- It is equally clear that transit will need to reverse course from the most recent trends and continue to post meaningful year-over-year ridership gains if it is to play a larger role in meeting overall urban travel needs. While there is a heightened sensitivity to transit mode share as it fluctuates between growth and decline, the pace of change has moderated from the long-term historic trend of significant declines. It is also evident that the absolute level of transit use at the national aggregate level is modest in terms of trips and especially modest in terms of PMT.
- Nationally, transit mode share is not changing rapidly. One should also note, however, that the particular context in each community may deviate substantially from this national trend. The national mode share for transit does not provide a full picture of the contribution of transit to peak-period peak-direction travel in critical corridors in many of the larger urban areas in the United States, nor does it reflect the importance of transit to those who are dependent on or choose to use public transit services. Nonetheless, the overall role of transit as seen from the various measures of mode share is a relevant consideration in public policy and investment programming decisions.
- This paper does not speak to other trends that are occurring within the population of transit users. There is an acknowledged shift toward a greater share of transit being on the rail mode, there are more long trips as express and rail services penetrate ever more distant suburbs, and there is a growing trend for transit service to access major attractors such as airports, sports stadiums, and major retail complexes. Aggregate measures of ridership and mode share do not

[5] Additional perspective on this issue may be gained by reading "Counting Transit so that Transit Counts" (TransManagement 2004).

fully capture the nature of the mobility role that transit plays in various urban areas.

While there are opportunities for improvements in measuring transit use and mode share, and these improvements may be important to individual agency service design and planning activities, the strategy of synthesizing multiple measures of mode share provides a knowledge base that is useful in informing national-level policy deliberations and in identifying data needs and differences.

REFERENCES

Chu, X. 2004. Ridership Accuracy and Transit Formula Grants, Transportation Research Board Pre-Print Number 05-1537.

Cox, W. and R.D. Utt. 2002. Census Shows Commuters are Rejecting Transit, The Heritage Foundation, Executive Memorandum, Sept. 5.

Joint Center for Political Studies. 1985. Demographic Change and Recent Worktrip Transit Trends, Working Paper No. 2. Urban Mass Transportation Administration, Washington, DC.

Kain, J.F. and Z. Liu. 1999. Secrets of Success: Assessing the Large Increases in Transit Ridership Achieved by Houston and San Diego Transit Providers. *Transportation Research, Part A Policy and Practice* 33(7–8):601–624.

Mason, J.W. 1998. The Buses Don't Stop Here Anymore: Sick Transit and How to Fix It. *The American Prospect* 56–62, March.

Millar, W.W. 1999. Overview of Public Transit in America. *Business Briefing: Global Mass Transit Systems* 44–46.

Norman, J. 2003. Statewide Study of Public Transportation to Attract Non-Traditional Transit Riders in California. California Department of Transportation.

Pisarski, A.E. 2003. Some Thoughts on the Census: Transit Statistical Match-Up. *Transportation Quarterly* 57(3):11–16.

Project for Public Spaces, Inc. 1999. The Role of Transit Amenities and Vehicle Characteristics in Building Transit Ridership, TCRP Report 46.

Schmidt, S. 2001. Incentive Effects of Expanding Federal Mass Transit Formula Grants. *Journal of Policy Analysis and Management* 20(2):239–261.

Stanley, R. 1998. Continuing Examination of Successful Transit Ridership Initiatives. *TCRP Research Results Digest* 29.

Surface Transportation Policy Project (STPP). 2002. Transit Growing Faster than Driving. *Decoding Transportation Policy & Practice* 3, May 29.

Taylor, B. and P. Haas. 2002. Increasing Transit Ridership: Lessons from the Most Successful Transit Systems in the 1990s. Mineta Transportation Institute, San Jose State University.

Taylor, B.D. and W.S. McCullough. 1998. Lost Riders. *Access* 13:27–31.

TransManagement. 2004. Counting Transit so that Transit Counts, prepared for the American Public Transportation Association. May. Available at http://www.apta.com/research/info/online/documents/counting_transit.pdf.

Transportation Research Board (TRB). 2001. *Making Transit Work: Insights from Western Europe, Canada, and the United States*, Special Report 257. Washington, DC.

Urban Mobility Corp. 2002. Mass Transit Debate Continues. *Innovation Briefs* 13:6.

Urbitran Associates, Inc. 1999. Guidelines for Enhancing Suburban Mobility Using Public Transportation, TCRP Report 55.

Built Environment and Nonmotorized Travel: Evidence from Baltimore City Using the NHTS

FELIPE TARGA[1]
KELLY J. CLIFTON[2,*]

[1] Inter-American Development Bank
Finance and Infrastructure Division, RE1/FI1
1300 New York Avenue
Washington, DC 20577

[2] Assistant Professor
Urban Studies and Planning
National Center for Smart Growth Research and Education
Preinkert Fieldhouse, Suite 1112
University of Maryland-College Park
College Park, MD 20742

ABSTRACT

The relationship between land use and travel behavior is a topic of debate among researchers and practitioners seeking to find land-use planning interventions to manage travel demand. This paper presents an empirical analysis of the effects of several land-use, urban form, and neighborhood-level design attributes, as well as traveler attitudes/perceptions of the urban system, on the frequency of walking, and the share of walking trips relative to total trips. Using the 2001 National Household Travel Survey add-on for the Baltimore metropolitan region, the paper estimates Poisson regression models at the person-level for the number of walking trips and a linear regression model for the share of walking trips made during a single travel day. The results suggest that neighborhoods with higher densities, more diverse land-use mixes, better street connectivity, and better access to bus transit lines are associated with persons who walk more frequently and make more walking trips with respect to trips made by other modes. Among the built environment variables, street network connectivity had the largest elasticity with respect to frequency of walking. Potential limitations of the analytical approach, as well as the degree of

Email addresses:
*Corresponding author—kclifton@umd.edu
F. Targa—felipet@iadb.org

KEYWORDS: Built environment, urban form, land use, nonmotorized travel, trip generation, NHTS, Poisson regression.

generalization of the results and their policy implications, are discussed.

INTRODUCTION

In the last decade, transportation trends for American cities have been generally characterized by increasing automobile ownership and use. These trends, on the one hand, have originated or aggravated existing transportation problems in urban areas, such as traffic congestion, air quality, energy consumption, livability, and public health. On the other hand, these trends have also motivated planning policy initiatives aimed at reducing some of these negative impacts by managing the demand for travel. Among the set of transportation demand management policy measures recently proposed, one of the most controversial relies on the connection between transportation and land use.

Planning practitioners and researchers who advocate for neotraditional and transit-oriented development rely on a hypothesized relationship between land use and travel activity behavior. Recent initiatives along this line of thinking consider land-use policies as a means to ease congestion, improve air quality, curb automobile demand, and contribute to improved quality of life by making urban areas more livable. Although these benefits may be expected if the urban built environment becomes more accessible, success of these policies hinges on understanding and anticipating travelers' responses.

Over the last decade, a substantial body of research has been accumulated that focuses on empirically testing the effect of several measures of urban form and neighborhood-level design characteristics on travel demand. The widely dissimilar approaches to measuring urban form attributes, the use of different levels of aggregation for both land-use and travel data, and the focus on different transportation-related outcomes have resulted in mixed, and sometimes, insignificant evidence. The empirical analysis conducted in this paper contributes to the general understanding of the relationship between land use and travel behavior by testing the effects of several land-use and urban form characteristics, as well as traveler attitudes/perceptions of the urban system, on two travel-related outcomes measured at the individual level—frequency of walking trips and share of walking trips with respect to total all-mode trips.

CONCEPTUAL FRAMEWORK

From a theoretical perspective, it is unclear what the net effect on the intensity of travel should be if urban areas are made more accessible. For example, improvements in the transportation system are expected to decrease travel times and thus lower the cost of travel. If the cost of travel goes down, consumption of travel can be expected to go up. Some studies use this rationale to argue that in urban environments where destinations are close by or more accessible, the cost per trips will be lower and higher trip generation rates or vehicle-miles travel can be expected (Crane 1996).

This is just a theoretical expectation, however, because the actual effect will depend on the elasticities of the demand with respect to price and the availability and feasibility of alternative modes. Indeed, depending on contextual attributes of the urban and transportation systems, the cost of travel may not be lower. The rationale for this argument is that the cost of a trip is a function of time, which in turn is a function of distance and speed and other travel-related attributes, such as out-of-pocket expenses, safety, and comfort. In a compact, clustered, and mixed-use development, we might expect that origins and destinations would be closer to each other, but the effect on speeds will depend on other factors. Therefore, the total impact on travel times and travel costs will depend on whether or not speed impacts overcome distance effects.

Additionally, the net effect on travel consumption (i.e., making more or fewer trips) might not only depend on the elasticity of the demand, but also on the degree of mode substitution and regional accessibility aspects. This net effect could be also moderated by particular travel activity attributes, such as trip purpose or mode of travel, as well as individual and household sociodemographic characteristics. In general, if there are cases where differences in travel-related outcomes can be attributed to differences in the built environment, they might also depend on the elasticity of the demand for travel with respect to person- and trip-related attributes.

The conceptual structure developed in this paper focuses on the empirical examination of observed travel outcomes, in a single point of the time (i.e., cross-sectional design), for households located in neighborhoods with different built environment attributes. Although the analytical framework allows for statistical association of these effects, it also has some limitations for inference about causality and net substitution effects. On the one hand, self-selected travelers (i.e., those who decided to locate in dense, mixed-use, and accessible neighborhoods as a result of their desire to live in these urban environments in order to walk more frequently) might confound the ability to make causality inferences. On the other hand, statistical associations between observed travel outcomes and households' built environment characteristics do not allow for formal inference about net travel substitution effects (e.g., more walking trips are substituting for car trips). These two aspects are crucial for understanding the implications of any land-use policy aiming to leverage the demand for auto travel and promoting more competitive and sustainable modes of travel.

These critical issues could be addressed with a more comprehensive study design, particularly with longitudinal or panel data structures, which could contain more detailed data (e.g., the travel decision-making process on consecutive days of travel, including short-term travel activity behavior, and long-term decisions, such as auto ownership and residential location choices). Based on the cross-sectional data available for this paper, the analytical framework estimates measures of statistical association between walking trip generation rates and households' built environment attributes, including land-use, urban form, and other neighborhood-level design characteristics. In addition to trip generation rates, statistical associations are also estimated for the share of walking trips with respect to all trips. If more accessible neighborhoods are associated with households where travelers walk more frequently, the set of associations based on modal share can provide additional insights about the possible trip substitution effects.

Finally, this paper advances the limited knowledge of built environment effects on nonmotorized and nonwork travel. Likewise, instead of using only a handful of built environment attributes, this study captures the association effects of a full complement of land-use and urban form measures, including disaggregate measures and composite indexes. In addition to objective or direct measures of the built environment attributes, the analytical framework also captures the association effects of attitudes and perceptions of travelers with respect to the urban system. Recent studies of the relationship between land use and travel behavior (Targa and Clifton 2004) have recognized that travel-related choices are expected to not depend exclusively on objective measures of the transportation system or the land-use characteristics but also on the perceived subjective attributes of the system.

LITERATURE REVIEW

Over the last decade, researchers have focused on empirically testing the effect of several measures of land-use, urban form, and neighborhood-level characteristics on travel behavior or travel-related outcomes (Badoe and Miller (2000) and Ewing and Cervero (2001) provide a detailed review of these studies). Overall, results from the most disaggregated and carefully controlled studies suggest that effects on trip generation rates depend mainly on household socioeconomic characteristics and that travel demand is inelastic with respect to accessibility (Ewing and Cervero 2001). Likewise, one common finding that comes from these studies is that the built environment has a greater impact on trip lengths than on trip frequencies. Nonetheless, some studies have also shown that urban environments with higher densities, a mix of land uses, and grid-style street configurations are associated with higher frequencies of walking/biking and other nonwork-based trips (Handy 1993, 1995, 1996; Friedman et al. 1994; Cervero and Gorham 1995; Kulkarni et al. 1995; and Cervero and Radisch 1996). Studies focusing on mode choice have found that this decision depends as much on built environment attributes as on socioeconomic characteristics. The association effects of built environment attributes with other travel-related outcomes, such as vehicle-miles traveled, have been documented as small but statistically significant.

Within the existing empirical studies, questions remain about the degree of trip substitution effects among different modes of travel and issues of self-selectivity (e.g., people who prefer walking/biking choose to live in built environments that facilitate that behavior as opposed to the urban form influencing their behaviors). Few studies have provided formal evidence of the underlying direction of causality, and among these studies, the results are mixed. Using cross-sectional data and controlling for preferences and attitudes, some studies have found that observed associations between travel behavior and neighborhood characteristics are largely explained by the self-selection of residents with certain attitudes (Bagley and Mokhtarian 2002), while others have not found such an impact after accounting for attitudes (Schwanen and Mokhtarian 2005). A recent study found that characteristics of the built environment influence walking behavior after accounting for a preference for walking-friendly neighborhoods (Cao et al. In press).

DATA DESCRIPTION

The primary data source for this study is the 2001 National Household Travel Survey (NHTS), in particular, the additional 3,446 households surveyed from June 2001 through July 2002 in the Baltimore metropolitan region. Households were randomly selected for participation in the Baltimore add-on sample. The survey was gathered through computer-assisted telephone interviews. In order to be consistent with the national data, the 2001 NHTS add-on survey was conducted following basically the same definitions and procedures of the 2001 NHTS national sample.

Land-use and urban form/design attributes used in the empirical examination were computed from several archived sources, such as census and county TIGER-enhanced files for the year 2000. Household locations were geocoded based on the respondent-provided closer location place of residence. Using geographic information systems (GIS), land-use, urban form, and other neighborhood-level design characteristics were assigned to each household record based on its geographic location. Most of these measures were operationalized consistently with previous efforts focusing on the characterization of built environment attributes (Galster et al. 2001; Song and Knaap In press). GIS and the increasing availability of land-use and transportation data in electronic format aided in the production of these secondary data. Census 2000 sociodemographic information was also obtained for the area of study. The geographic area of analysis consisted of the city of Baltimore, including 1,539 surveyed households (figure 1) or 2,934 persons with reported travel-day data.

Among the 2,934 travelers, 2,061 (70.25%) did not make any walking trips during the reported travel day, and 580 travelers (19.77%) reported one or two walking trips. Figure 2 depicts the spatial distribution of the frequency of walking trips on the reported day of travel. Analyzing the trip purpose variable from the trip-level data (not shown here), we confirmed that the majority of these walking trips (91.5%) were generated from nonwork-related activities (38.48% were home-related). In terms of modal share, walking trips accounted for 100% of the total trips on the reported day of travel for 10.87% of the surveyed travelers, while the average modal share for walking trips was 19.24%.

In addition, to control for traditional socioeconomic and demographic characteristics and trip-related attributes, the analytical framework developed in this paper uses attitudinal and perceptual data as proxies for sociopsychological factors influencing travel activity behavior. Perceptual data include attitudes toward traffic accidents, highway congestion, the presence of drunk drivers on the road, lack of sidewalks and walkways, and the price of gasoline. The existence of a medical condition that impedes the mobility of the respondent is also expected to influence travel behavior by limiting driving, the use of transit, or reducing the amount of travel made.

The set of explanatory variables of interest consists of urban form, neighborhood design, and land-use attributes associated with the geographic location of each traveler's household. Although several variables were constructed using GIS-based data, household unit density at the census block-level, street connectivity (measured as the perimeter of the census block), the diversity of land-use mix at the census block group-level (measured as an 0 to 1 index indicating the degree of land-use mixing), and distance to the nearest bus transit stop were the variables finally selected. The selection of explanatory variables was based on statistical (e.g., the most

FIGURE 1 Surveyed Households in the 2001 NHTS Baltimore Add-On Sample

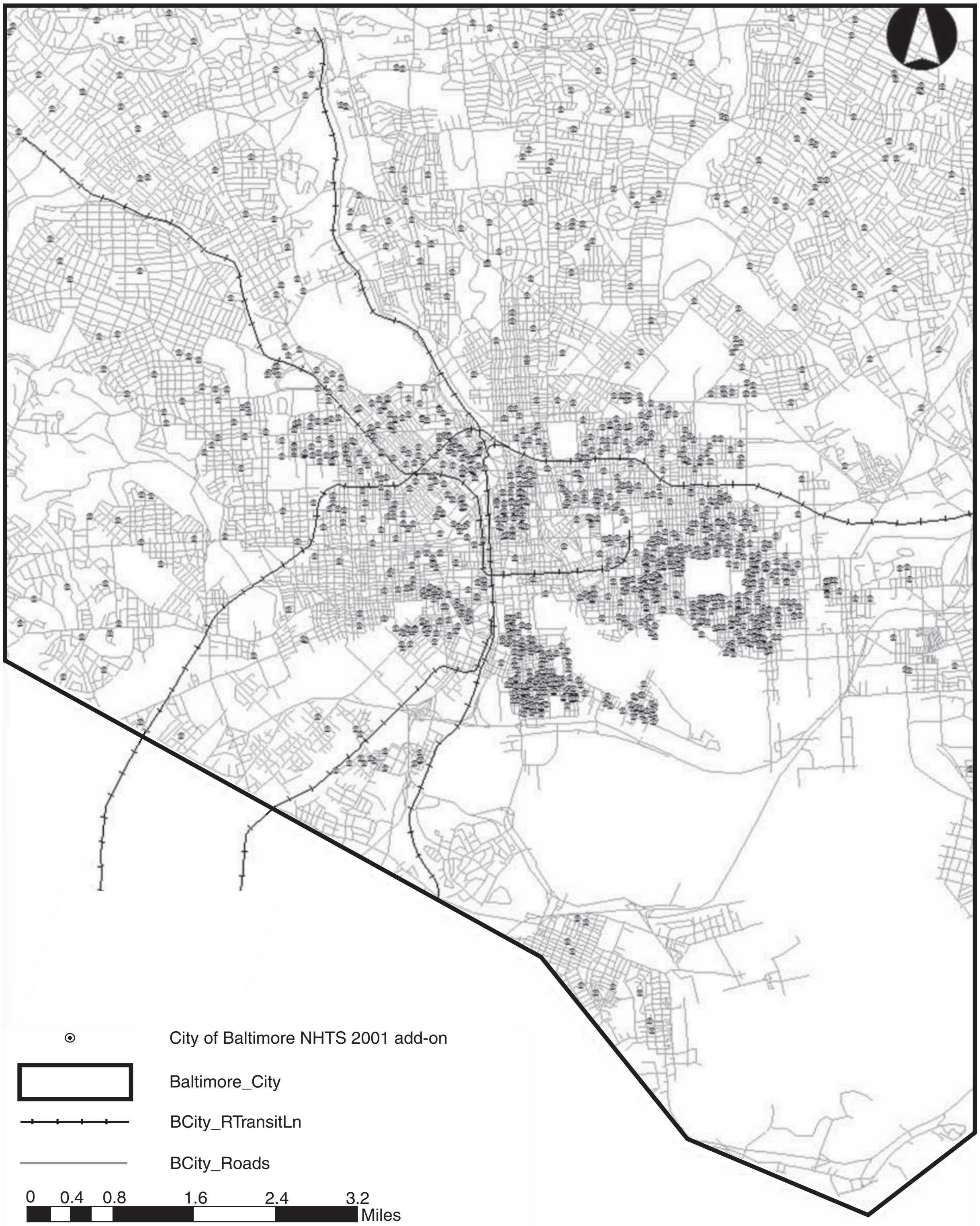
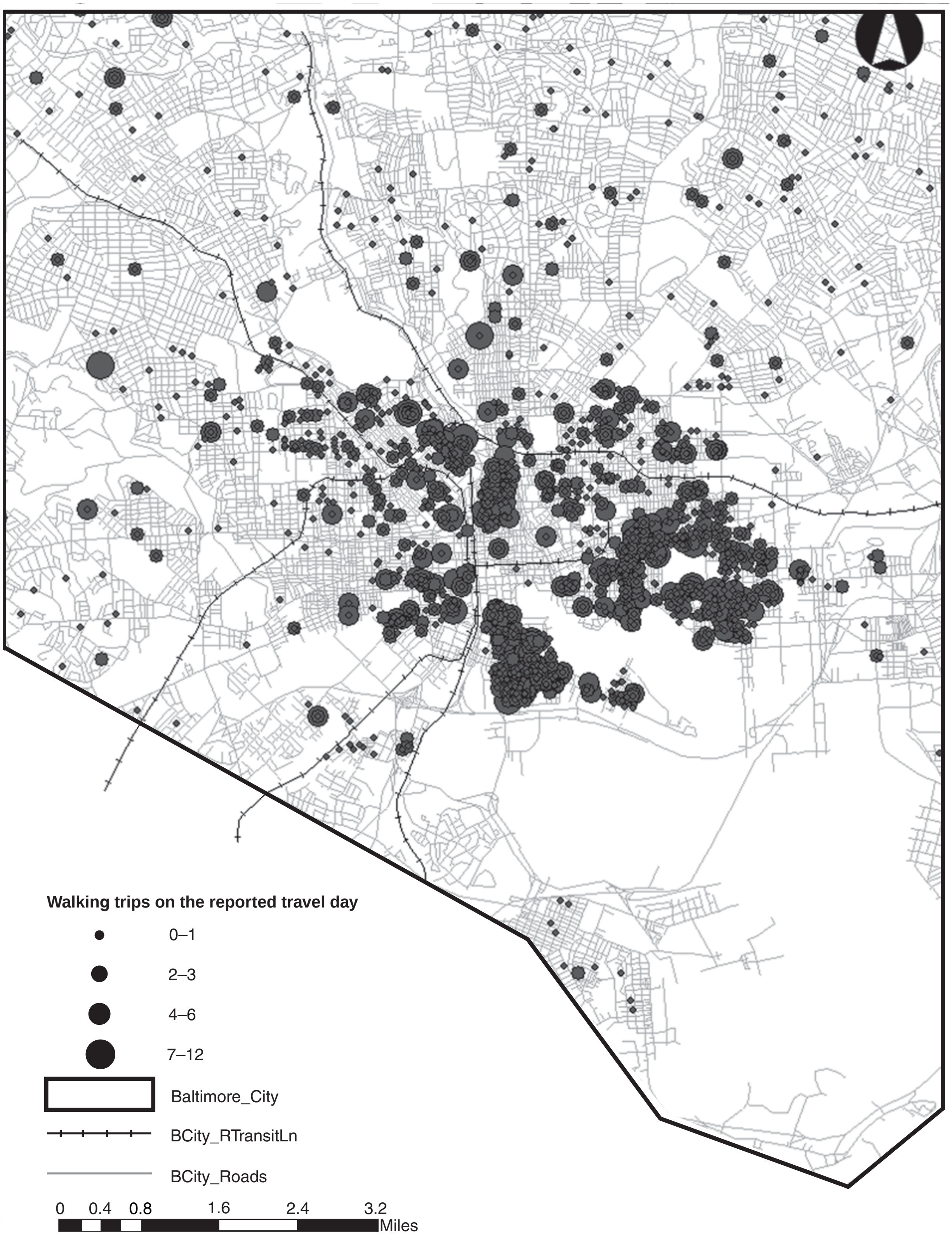


FIGURE 2 Frequency of Walking Trips on the Reported Day of Travel
2001 NHTS Baltimore Add-On, City of Baltimore



parsimonious model specification) and study-specific considerations (e.g., different levels of spatial aggregations will impact the effects of some land-use or urban form attributes). Neighborhood sociodemographic characteristics were also obtained at the census block-level for 2000.

The possibility exists that some land-use variables could be correlated with household variables, or that some built environment attributes may be correlated with specific socioeconomic variables. However, a correlation analysis (not shown here) confirmed that all pair correlations were low, except for density and street connectivity (p = 0.51), which were part of the same set of built environment attributes. Table 1 presents summary statistics for the set of dependent and explanatory variables for the area of analysis (i.e., Baltimore City).

Conceptually, we expected that neighborhoods with higher densities, fine land-use mixes, better street connectivity, and generally better access to transit would be associated with persons making more walking trips and with a higher walking modal share. Because shopping trips and other nonwork-based trips tend to be more elastic with respect to accessibility and are more likely to be done by non-motorized modes than work trips, we expected differences in urban form and design attributes to be more influential for these trips.

METHODOLOGY

Walking trip generation rates and modal share proportions were calculated at the person-level for all household members with reported travel-day data (a 24-hour period). Given the count-type nature of the data for the number of walking trips, the methodological approach consisted, initially, of specifying and estimating a Poisson regression model. In a Poisson model specification, a random variable indicates the number of events (i.e., walking trips) during an interval of time (i.e., reported travel day). In the regression model, the number of events y has a Poisson distribution with a conditional mean that depends on household or travelers' characteristics, trip characteristics, and land-use/urban form attributes according to the following structural model:

$$\mu_i = E(y_i|x_i) = \exp(\boldsymbol{x}_i \boldsymbol{\beta})$$

where x_i is a row vector with observations of the explanatory variables for each person, and $\boldsymbol{\beta}$ is a column vector of estimated coefficients associated with each explanatory variable. This structural model is estimated by means of maximum likelihood (ML) estimation techniques. Asymptotic tests of the coefficient estimates and calculation of marginal effects are used to evaluate the statistical significance and relative magnitude of the effects of land-use/urban form measures on frequency of walking.

McFadden's likelihood ratio index (McFadden 1973) and adjusted McFadden's R^2 for the number of parameters (Ben-Akiva and Lerman 1985) are used as scalar measures of fit for the Poisson models. These measures are the most popular approximations to the coefficient of determination R^2 in linear regression models. In particular, the log-likelihood of the model without regressors is thought of as the total sum of squares, while the log-likelihood of the model with regressors is thought of as the residual sum of squares.

The second step in the methodological approach consisted of specifying and estimating a linear regression model by means of the ordinary least squares (OLS) for the walking modal share variable. This regression model uses the same set of explanatory variables specified in the Poisson models and follows the structural model below:

$$y_i = c + \boldsymbol{x}_i \boldsymbol{\beta} + u_i$$

MODEL ESTIMATION

This section presents the estimation results for the models discussed in the preceding section. The coefficients of the explanatory variables included in the Poisson model specification are estimated by means of ML and represent the relative effect of the associated variable on the frequency of walking. Coefficients are estimated without expansion factors or analysis weights commonly used to avoid bias in the statistical analysis. Particular attention is devoted to the estimates of the built environment attributes, the primary interest of this paper.

TABLE 1 Data Description and Summary Statistics

Variable name	Variable label/response category description	Valid *N*	Mean	Std. Dev.	Min.	Max.
Dependent variables						
Walking trips	Number of walking trips on the surveyed day	2,934	0.80	1.48	0	12
Share walking	% of walking trips with respect to all trips made on the survey day	2,934	0.19	0.34	0	1
Household (HH) characteristics						
Vehicles in HH	Number of vehicles in household per household member	2,934	0.53	0.47	0	4.5
Bikes in HH	Number of full-size bicycles per household member	2,932	0.25	0.41	0	4
Type HU	Type of housing unit (HU)					
Detached	Detached single house	2,931	0.13	0.34	0	1
Duplex-triplex	Duplex, triplex	2,931	0.02	0.15	0	1
Row/townhouse	Row house, townhouse	2,931	0.62	0.49	0	1
Apart/condo	Apartment, condominium	2,931	0.22	0.41	0	1
Dorm	Dorm room, fraternity or sorority house	2,931	0.00	0.03	0	1
Semi	Semi-attached/semi-detached house	2,931	0.00	0.04	0	1
Boat	Boat	2,931	0.00	0.05	0	1
Owned	Housing unit (= 1 owned, = 0 rented)	2,925	0.62	0.49	0	1
Income HH	Household income (= 1 $30K or less, = 0 $30K and more)	2,934	0.27	0.45	0	1
Individual characteristics						
Age	Age (years)	2,891	41.26	22.61	0	96
Female	Gender (= 1 female, = 0 male)	2,934	0.57	0.50	0	1
Driver	Driver status (= 1 driver, = 0 nondriver)	2,933	0.58	0.49	0	1
Status	Working/school status last week					
Working	Working	2,933	0.44	0.50	0	1
Absent	Temporarily absent from a job or business	2,933	0.03	0.16	0	1
Looking	Looking for work	2,933	0.02	0.14	0	1
Homemaker	A homemaker	2,933	0.04	0.21	0	1
Student	Going to school	2,933	0.05	0.22	0	1
Retired	Retired	2,933	0.21	0.40	0	1
Other	Doing something else	2,933	0.06	0.24	0	1
Full-time	Work status (= 1 full time, = 0 part time)	2,934	0.41	0.49	0	1
Occupation	Occupation category					
Sales/service	Sales or service	2,934	0.13	0.34	0	1
Clerical/admin	Clerical or administrative support	2,934	0.07	0.25	0	1
Manuf/const/maint	Manufacturing, construction, maintenance	2,934	0.05	0.22	0	1
Prof/manag/tech	Professional, managerial, or technical	2,934	0.24	0.42	0	1
Transport	Transportation/machine operator	2,934	0.00	0.05	0	1
Military	Military	2,934	0.00	0.03	0	1
Other	Police/firefighter/corrections officer	2,934	0.00	0.03	0	1
Licensed	Drive licensed vehicle as part of work (= 1 yes, = 0 no)	2,934	0.07	0.25	0	1
Distance-to-work	One-way distance to work (miles)	2,905	4.45	10.59	0	200
Walk-exercising	Number of outside (exercising) walk trips in past week	2,919	0.24	0.66	0	7
Bike-exercising	Number of outside (exercising) bike trips in past week	2,928	0.00	0.08	0	3
Med condition	Medical condition makes travel out of home difficult (= 1 yes, = 0 no)	2,930	0.12	0.33	0	1
Graduate	Highest grade of school completed (= 1 graduate, = 0 other)	2,923	0.15	0.36	0	1

(continues on next page)

TABLE 1 Data Description and Summary Statistics *(Continued)*

Variable name	Variable label/response category description	Valid *N*	Mean	Std. Dev.	Min.	Max.
Attitudes/perceptions						
Traffic accidents (TA)	Worry about traffic accident					
TA-No	1—Not a problem	2,922	0.05	0.21	0	1
TA-Little	2—A little problem	2,922	0.03	0.17	0	1
TA-Somewhat	3—Somewhat of a problem	2,922	0.03	0.17	0	1
TA-Very	4—Very much of a problem	2,922	0.01	0.10	0	1
TA-Severe	5—A severe problem	2,922	0.02	0.14	0	1
Highway congestion	Worry about highway congestion (HC)					
HC-No	1—Not a problem	2,860	0.16	0.36	0	1
HC-Little	2—A little problem	2,860	0.09	0.29	0	1
HC-Somewhat	3—Somewhat of a problem	2,860	0.13	0.33	0	1
HC-Very	4—Very much of a problem	2,860	0.08	0.28	0	1
HC-Severe	5—A severe problem	2,860	0.10	0.30	0	1
Drunk drivers (DD)	Worry about drunk drivers					
DD-No	1—Not a problem	2,913	0.05	0.21	0	1
DD-Little	2—A little problem	2,913	0.02	0.14	0	1
DD-Somewhat	3—Somewhat of a problem	2,913	0.01	0.12	0	1
DD-Very	4—Very much of a problem	2,913	0.01	0.12	0	1
DD-Severe	5—A severe problem	2,913	0.04	0.21	0	1
Price of gasoline (PG)	Worry about price of gasoline					
PG-No	1—Not a problem	2,817	0.17	0.38	0	1
PG-Little	2—A little problem	2,817	0.09	0.29	0	1
PG-Somewhat	3—Somewhat of a problem	2,817	0.11	0.31	0	1
PG-Very	4—Very much of a problem	2,817	0.06	0.23	0	1
PG-Severe	5—A severe problem	2,817	0.12	0.32	0	1
Walkways/sidewalks	Worry about poor walkways or sidewalks (WS)					
WS-No	1—Not a problem	2,922	0.09	0.28	0	1
WS-Little	2—A little problem	2,922	0.02	0.14	0	1
WS-Somewhat	3—Somewhat of a problem	2,922	0.01	0.11	0	1
WS-Very	4—Very much of a problem	2,922	0.01	0.07	0	1
WS-Severe	5—A severe problem	2,922	0.01	0.09	0	1
Urban form, neighborhood design, and land-use attributes						
Ln(density)	Household density at the census block-level *ln*(1,000 HH units/mile^2)	2,857	15.21	14.07	0.00	121.85
Connectivity	Street connectivity (census block's perimeter in miles)	2,934	0.34	0.26	0.07	3.01
Diversity	Land-use mix diversity index at the census block-level	2,932	0.19	0.20	0.00	0.75
Accessibility	Transit accessibility (distance in miles to the nearest bus transit stop)	2,934	0.08	0.06	0.05	0.35
Neighborhood sociodemographics						
Age in N	Median age of population at the census block-level	2,934	35.37	11.29	0	77.80
Race in N	Proportion of white population at the census block-level	2,863	0.45	0.39	0	1.00

The Poisson model was estimated for three different specifications (table 2). Model 1 includes only traditional household and person socioeconomic characteristics. Model 2 includes all variables used in model 1, along with attitudinal and perceptual data of the urban and transportation system. Model 3 includes all variables used in model 2, as well as all the built environment attributes and the neighborhood sociodemographic characteristics. The OLS model (model 4) for walking modal share is

TABLE 2 Estimated Poisson Models for Number of Walking Trips and Estimated Ordinary Least Squares (OLS) Model for Share of Walking Trips

	Poisson models (number of trips)						OLS (share of trips)	
	Model 1		Model 2		Model 3		Model 4	
Variable name	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value
Household (HH) characteristics								
Vehicles in HH	–0.313***	–4.88	–0.240***	–3.77	–0.297***	–4.39	–0.078***	–4.45
Bikes in HH	0.362***	8.19	0.320***	7.11	0.285***	6.16	0.061***	3.85
Type housing unit								
Apart/condo	0.052	1.00	0.022	0.41	0.116**	2.01	0.034**	1.97
Dorm	1.200***	3.66	0.897***	2.61	1.641***	4.62	0.563**	2.47
Income HH	0.086*	1.74	0.095*	1.90	0.143***	2.80	0.032**	2.24
Individual characteristics								
Age	–0.002*	–1.87	–0.005***	–3.73	–0.004***	–3.56	–0.001***	–3.55
Female	–0.052	–1.18	–0.072	–1.61	–0.072	–1.62	–0.017	–1.30
Driver	–0.216***	–3.42	–0.188***	–2.97	–0.280***	–4.38	–0.060***	–3.34
Status								
Absent	0.287**	2.26	0.278**	2.18	0.324**	2.53	0.077*	1.92
Looking	0.351***	2.82	0.274**	2.20	0.225*	1.80	0.070	1.61
Full-time	–0.256***	–3.73	–0.308***	–4.44	–0.311***	–4.41	–0.037*	–1.90
Occupation								
Sales/service	–0.028	–0.32	–0.089	–1.03	–0.114	–1.30	–0.027	–1.20
Manuf/const/maint	–0.069	–0.54	–0.144	–1.12	–0.261**	–1.98	–0.046	–1.42
Prof/manag/tech	0.500***	6.37	0.464***	5.90	0.349***	4.39	0.040*	1.72
Licensed	–0.539***	–4.48	–0.507***	–4.20	–0.443***	–3.66	–0.053**	–2.07
Distance-to-work	–0.016***	–4.85	–0.016***	–5.04	–0.016***	–5.02	–0.003***	–4.28
Walk-exercising	0.406***	20.63	0.398***	19.77	0.359***	17.44	0.083***	8.49
Med condition	–0.584***	–5.91	–0.617***	–6.19	–0.662***	–6.63	–0.098***	–4.38
Graduate	0.300***	4.63	0.263***	4.02	0.185***	2.82	0.014	0.71
Attitudes/perceptions								
Traffic accidents (TA)								
TA-No			–0.345***	–2.66	–0.341***	–2.62	–0.036	–1.21
Highway congestion (HC)								
HC-Very			0.186**	2.26	0.168**	2.03	–0.001	–0.05
HC-Severe			0.392***	5.25	0.389***	5.19	0.051**	2.29
Drunk drivers (DD)								
DD-Severe			0.275***	2.99	0.268***	2.91	0.008	0.26
Price of gasoline (PG)								
PG-No			0.348***	6.03	0.389***	6.73	0.053***	2.90
PG-Severe			–0.295***	–3.36	–0.204**	–2.31	–0.020	–0.97
Walkways/sidewalks (WS)								
WS-Little			0.255*	1.91	0.230*	1.70	0.041	0.98
WS-Somewhat			0.432***	2.56	0.549***	3.23	0.039	0.70
Urban form, neighborhood design, and land-use attributes								
Ln(density)					0.041	1.32	0.011	1.25
Connectivity					–0.934***	–5.62	–0.089***	–2.78
Diversity					0.439***	4.07	0.063**	1.97
Accessibility					–1.030***	–2.64	–0.117	–1.15

(continues on next page)

TABLE 2 Estimated Poisson Models for Number of Walking Trips and Estimated Ordinary Least Squares (OLS) Model for Share of Walking Trips *(Continued)*

	Poisson models (number of trips)						OLS (share of trips)	
	Model 1		Model 2		Model 3		Model 4	
Variable name	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value	Coeff.	*t*-value
Neighborhood sociodemographics								
Age in N					–0.005**	–1.99	–0.001	–1.33
Race in N					0.731***	11.01	0.092***	5.17
Constant	–0.075	–1.26	–0.067	–1.12	–0.260	–0.79	0.224**	2.51
Valid N =	2,630		2,630		2,630		2,630	
Log-likelihood intercept only:	–3,974.97		–3,974.97		–3,974.97			
Log-likelihood full model:	–3,564.04		–3,510.16		–3,381.89			
McFadden's R^2: (R^2 for OLS)	0.103		0.117		0.149		0.146	
McFadden's adjusted R^2: (Adjusted R^2 for OLS)	0.098		0.110		0.141		0.135	
Likelihood-ratio (LR) test with respect to model 1 (p-value):			0.001		0.001			
Likelihood-ratio (LR) test with respect to model 2 (p-value):					0.001			

KEY: ***, **, and * denote coefficient significantly different from zero at the 1%, 5%, and 10% level of significance (two-tail test), respectively.

estimated with all of the explanatory variables used in the Poisson model 3.

Table 2 summarizes the corresponding coefficient estimates, t statistics, and the statistical significance test for each estimated coefficient. All models were statistically significant at the 99% confidence level ($p < 0.001$ for the χ^2 test). The model specification with traditional explanatory variables for trip generation rates (model 1) helps to explain some of the variability of frequency of walking compared with a model without regressors (McFadden's adjusted R^2 = 0.098). Among household characteristics, lower number of vehicles and higher number of bicycles per household member, college dorm home type, and lower household income were characteristics associated with a higher frequency of walking, as expected. Traveler characteristics associated with a higher frequency of walking included young, nonlicensed driver, temporarily absent from a job or looking for work, full-time workers, professional or managerial occupation category if working, healthy, graduate-level education, and people who frequently walk for exercise and have their work location closer to home.

Adding attitudinal variables to the model specification (McFadden's adjusted R^2 = 0.110) increased the statistical explanatory power of the model with respect to model 1 (likelihood ratio test; $\chi^2_{(8)}$ = 107.7 and $p < 0.001$). Among attitudinal variables, the individual estimated coefficients suggest that people who are more concerned with traffic accidents, highway congestion, and drunk drivers are likely to walk more frequently than people less concerned with these urban system characteristics. Interestingly, people who drive frequently and express more concern about the price of gasoline were less likely to walk. Those who indicated that sidewalk conditions presented "a little" and "somewhat" of a problem tended to walk more frequently.

A particularly notable finding of our analysis is the statistically significant association between built environment attributes and the frequency of walking. Comparing the overall performance of model 3 (McFadden's adjusted R^2 = 0.141), the explanatory power of the model increases statistically with respect to model 2 (likelihood ratio test; $\chi^2_{(6)}$ = 256.5 and $p < 0.001$) and with respect to model 1 (likelihood ratio test; $\chi^2_{(14)}$ = 364.3 and $p < 0.001$). The same set of explanatory variables explain

13.5% (adjusted R^2 = 0.135) of the variability of the proportion of walking trips with respect to all trips made on the surveyed day of travel (model 4). Overall, these results show that both attitudinal variables and built environment attributes increase the statistical explanatory power of the models, and consequently, help to better explain the variance of the dependent variables (i.e., frequency of walking and walking modal share). However, this study's primary focus is the relative effect that specific land-use, urban form, and neighborhood-level design variables have on walking trip generation rates and on the share of walking trips.

In particular, people living in denser urban settings, measured as the number of household units per square mile in the corresponding household census block, tend to walk more frequently on the surveyed day of travel, all else being equal. We hypothesize that the sign of the *Density* coefficient is positive and statistically different from zero based on a one-tail test (p = 0.093). The marginal effect of *Density* (table 3) suggests that an increase of 1% in the number of household units per square mile (within the census block where the traveler's household is located) is associated with an increase of the expected number walking trips of 0.026, all else being equal. This translates into an elasticity of 0.033, evaluated at the mean value of the walking trip generation rate.[1] In other words, a 1% increase in the number of household units per unit of area is associated with a 0.033% increase in the expected number of walking trips on a given day. This elasticity is even lower than the average density elasticity of vehicle-miles traveled (0.05) estimated in previous studies (Ewing and Cervero 2001).

Likewise, people living in neighborhoods with higher street connectivity or with more grid-like street networks, measured as the perimeter of the corresponding household census block, are likely to walk more frequently, as reported on the surveyed day of travel. The marginal effect of the street connectivity variable (table 3) suggests that a one mile decrease in the perimeter of the corresponding census block (i.e., more connected street networks) is associated with an increase of the expected number of walking trips of 0.587, all else being equal. Evaluated at the mean of the walking trip generation rate, this translates into an elasticity of –0.258.[2] This means that a 1% decrease in street network connectivity (i.e., length of a census block perimeter) is associated with a 0.258% increase in the expected number of walking trips made on the reported day of travel, all else being equal.

Figure 3 depicts the probability of making one or more walking trips on the reported day of travel as the length of the census block perimeter varies from 0 to 3 miles, holding the rest of the variables at their means. Moreover, the coefficient of street connectivity in the walking share model (model 4) suggests that the same one mile decrease in street network connectivity is associated with an increase of 8.9% in the proportion of walking trips with respect to all other trips made by other modes, including car, on the reported day of travel.

The degree of land-use mix was captured by an index of land-use mix diversity ranging from 0 to 1 (Song and Knapp In press). If the land use in the census block group associated with the traveler's household is dedicated exclusively to a single use, the diversity index variable takes a value of 0. Conversely, a value of 1 indicates perfect mixing of the land uses considered in this study (i.e., residential, commercial, industrial, institutional, and open urban space). Evaluated at the mean of the walking trip generation rate, the marginal effect of this land-use mix index (table 3) translates into an elasticity of 0.065.

Figure 4 depicts the probability of making one or more walking trips on the reported travel day as the land-use mix diversity index varies from 0 to 0.75, holding the rest of the variables at their means. The slope of the fitted line in figure 4 shows the low elasticity value for the land-use mix index. Likewise, the

[1] Because *Density* was transformed with a logarithmic function, a change in the transformed variable is associated with a 1% change in the units of the untransformed variable. Moreover, the elasticity is calculated at the mean of the walking trip generation rate; an increase of 0.026 trips is equivalent to an increase of 3.3% with respect to the mean value (0.80 trips).

[2] A one mile decrease with respect to the mean of the perimeter of the census blocks is equivalent to a 296% decrease, and an increase of 0.587 trips is equivalent to an increase of 76.4% with respect to the mean value (0.80 trips). This translates into an elasticity of –0.258.

TABLE 3 Marginal Effects for Predicted Probabilities of Poisson Model 3

Variable name	Poisson model 3: One-unit change around the mean[1]	Marginal change[1]
Household (HH) characteristics		
Vehicles in HH	–0.187	–0.187
Bikes in HH	0.180	0.179
Type housing unit		
Apart/condo	0.073	0.073
Dorm	1.151	1.031
Income HH	0.090	0.090
Individual characteristics		
Age	–0.003	–0.003
Female	–0.045	–0.045
Driver	–0.177	–0.176
Status		
Absent	0.205	0.204
Looking	0.142	0.141
Full-time	–0.197	–0.196
Occupation		
Sales/service	–0.072	–0.072
Manuf/const/maint	–0.164	–0.164
Prof/manag/tech	0.221	0.220
Licensed	–0.281	–0.279
Distance-to-work	–0.010	–0.010
Walk-exercising	0.227	0.226
Med condition	–0.424	–0.416
Graduate	0.116	0.116
Attitudes/perceptions		
Traffic accidents (TA)		
TA-No	–0.215	–0.214
Highway congestion (HC)		
HC-Very	0.105	0.105
HC-Severe	0.246	0.244
Drunk drivers (DD)		
DD-Severe	0.169	0.168
Price of gasoline (PG)		
PG-No	0.246	0.245
PG-Severe	–0.128	–0.128
Walkways/sidewalks (WS)		
WS-Little	0.145	0.144
WS-Somewhat	0.349	0.345
Urban form, neighborhood design, and land-use attributes		
Ln(density)	0.026	0.026
Connectivity	–0.608	–0.587
Diversity	0.278	0.276
Accessibility	–0.676	–0.647
Neighborhood sociodemographics		
Age in N	–0.003	–0.003
Race in N	0.469	0.459

[1] Binary (0–1) dummy indicator variables correspond to a discrete change as the variable changes from 0 to 1.

FIGURE 3 Predicted Probabilities for Frequency of Walking Trips to Changes in Street Network Connectivity

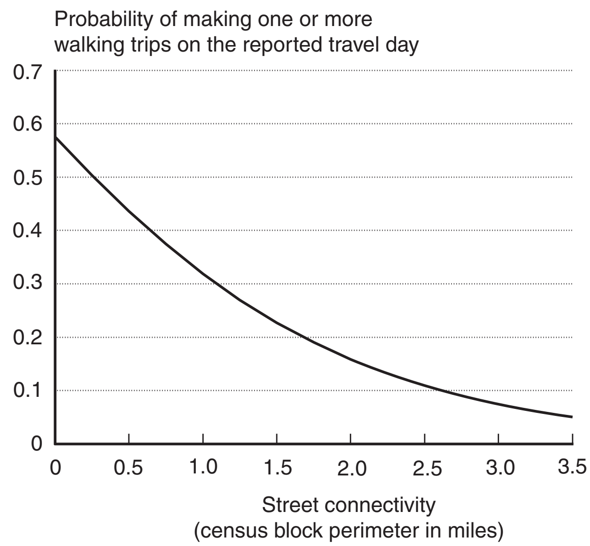
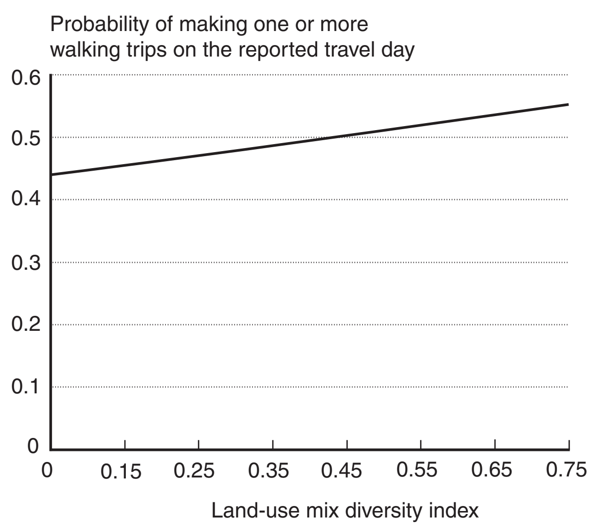


FIGURE 4 Predicted Probabilities for Frequency of Walking Trips to Changes in Land-Use Mix Diversity



coefficient of the land-use mix index in the walking share model (model 4) suggests the same significant, but marginal, association of this index for the proportion of walking trips with respect to all trips made on the reported day of travel.

Access to bus transit lines is also statistically associated with higher walking frequency, as expected. In particular, the marginal effect suggests that people living one mile closer to a bus stop are expected to make 0.647 more walking trips on the reported day of travel. Evaluated at the mean of the walking trip generation rate, this effect translates into an elasticity of –0.070.

The last set of estimated coefficients suggests that people living in neighborhoods with a lower median age and higher white population proportion are likely to walk more frequently as reported on the surveyed day of travel. Only neighborhoods with a greater proportion of white residents are associated with a larger walking share with respect to all trips in the surveyed day of travel.

CONCLUSIONS AND LIMITATIONS

The conceptual structure and the empirical results presented in this paper advance the understanding of the relationship between land use and travel behavior. In particular, our findings contribute to the general understanding of this relationship by testing the effects of several land-use and urban form characteristics, as well as traveler attitudes and perceptions of the urban system, on two nonmotorized travel-related outcomes. Based on a cross-sectional study design, the paper estimated measures of statistical association between walking trip generation rates and households' built environment attributes, including land use, urban form, and other neighborhood-level design characteristics. In addition to trip generation rates, statistical associations were also estimated for the share of walking trips with respect to total travel.

Using the 2001 NHTS for the City of Baltimore (2,630 travelers), the paper estimated three Poisson regression models at the person-level for the number of walking trips during a single surveyed day of travel and a linear regression model for the share of walking trips made on the same travel day. The results suggest that neighborhoods with higher densities, fine land-use mixes (i.e., more diverse), better street connectivity, and generally better access to bus transit lines were associated with persons who walk more frequently and have a higher proportion of walking trips with respect to all trips. These results were expected given the theoretical elasticity of nonmotorized travel with respect to accessibility.

Results from the model for walking share with respect to total travel suggest that more accessible neighborhoods are not only associated statistically with households where travelers walk more frequently, but also with households where the proportion of walking trips is higher on the same day of travel. These results provide some insights into possible trip substitution effects in these neighborhoods but are restricted to inference limitations discussed later in this section.

The 2001 NHTS Baltimore add-on survey used random selection, and previous analysis of this dataset has shown that the sample is representative of the population (Battelle and Morpace 2002). However, caution should be taken when trying to transfer the results here to different locations. Indeed, the degree of generalization of the results and the general external validity of the empirical findings is limited to the context of travel and urban setting characteristics in the geographic area of study (i.e., Baltimore City).

Nonetheless, two critical issues could not be addressed comprehensively given the limitations of the study design and data availability. In particular, the analytical framework allowed for statistical association between land-use and travel behavior effects, but it had some limitations for inference on causality and net travel substitution effects. Without longitudinal or panel data structures containing more detailed information (e.g., the travel decision-making process on consecutive days of travel, including short-term travel activity behavior and long-term decisions, such as auto ownership and residential location choices), we were unable to formally evaluate possible confounded effects under conditions of self-selected travelers (e.g., those located in dense, mix-used, accessible neighborhoods as a result of their desire to live in those urban environments).

One of the complications of using cross-sectional data is that the model specification cannot capture the endogenous processes typically found in travel decisionmaking. In general, the presence of endogeneity in the model estimation might yield inconsistent and biased estimates of the relationships. However, few studies have provided formal empirical evidence of the underlying direction of causality, and among these studies the results are mixed. While some studies found that observed associations between travel behavior and neighborhood characteristics are largely explained by the self-selection of residents with certain attitudes, others have not found such an impact after accounting for attitudes toward travel and location preferences. Future research will benefit from more detailed travel data, particularly longer periods of observed or surveyed travel (i.e., panel or longitudinal studies).

Despite the potential limitations of the analytical approach, the results of this paper are highly relevant for transportation planning practitioners and researchers and improve our understanding of the relationship between land use and travel behavior. Ultimately, the empirical evidence provided in this paper is expected to contribute to the growing body of literature focusing on the interaction between land use and travel demand.

ACKNOWLEDGMENT

The opinions expressed in this paper are those of the authors and do not necessarily reflect the official position of the Inter-American Development Bank.

REFERENCES

Badoe, D.A. and E.J. Miller. 2000. Transportation-Land-Use Interaction: Empirical Findings in North America and Their Implications for Modeling. *Transportation Research Part D: Transport and Environment* 5(4):235–263.

Bagley, M.N. and P.L. Mokhtarian. 2002. The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach. *Annals of Regional Science* 36:279–297.

Battelle and Morpace. 2002. 2001 National Household Travel Survey Add-On Program: Baltimore Regional Transportation Board, final report and data codebook. October.

Ben-Akiva, M.E. and S.R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Cao, X., S. Handy, and P.L. Mokhtarian. In press. The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior. *Transportation*.

Cervero, R. and R. Gorham. 1995. Commuting in Transit Versus Automobile Neighborhoods. *Journal of the American Planning Association* 61:210–225.

Cervero, R. and C. Radisch. 1996. Travel Choices in Pedestrian Versus Automobile Oriented Neighborhoods. *Transport Policy* 3:127–141.

Crane, R. 1996. Cars and Drivers in the New Suburbs: Linking Access to Travel in Neotraditional Planning. *Journal of the American Planning Association* 62:51–65, Winter.

Ewing, R. and R. Cervero. 2001. Travel and the Built Environment. *Transportation Research Record* 1780:87–114.

Friedman, B., S.P. Gordon, and J.B. Peers. 1994. Effect of Neotraditional Neighborhood Design on Travel Characteristics. *Transportation Research Record* 1466:63–70.

Galster, G., R. Hanson, M.R. Ratcliffe, H. Wolman, S. Coleman, and J. Freihage. 2001. Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept. *Housing Policy Debate* 12(4):681–717.

Handy, S. 1993. Regional Versus Local Accessibility: Implications for Non-Work Travel. *Transportation Research Record* 1400:58–66.

______. 1995. Understanding the Link Between Urban Form and Travel Behavior, presented at the 74th Annual Meetings of the Transportation Research Board, Washington, DC, 1995.

______. 1996. Urban Form and Pedestrian Choices: Study of Austin Neighborhoods. *Transportation Research Record* 1552:135–144.

Kulkarni, A., R. Wang, and M.G. McNally. 1995. Variation of Travel Behavior in Alternative Network and Land Use Structures. *ITE 1995 Compendium of Technical Papers*. Washington, DC: ITE, pp. 372–375.

McFadden, D. 1973. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers of Econometrics*. Edited by P. Zarembka. New York, NY: Academic Press.

Schwanen, T. and P.L. Mokhtarian. 2005. What Affects Commute Mode Choice: Neighborhood Physical Structure or Preferences Toward Neighborhoods? *Journal of Transport Geography* 13:83–99.

Song, Y. and G.J. Knaap. In press. Measuring Urban Form: Is Portland Winning the War on Sprawl? *Journal of the American Planning Association*.

Targa, F. and K.J. Clifton. 2004. Integrating Social and Psychological Processes into the Land-Use Travel Behavior Research Agenda: Theories, Concepts and Empirical Study Design, presented at the 7th International Conference on Travel Survey Methods, Los Sueños, Costa Rica, Aug. 1–6, 2004.

Using Bayesian Updating to Enhance 2001 NHTS Kentucky Sample Data for Travel Demand Modeling

BING MEI[1,*]
THOMAS A. COONEY[2]
NIELS R. BOSTROM[3]

[1] Institute for Transportation Research and Education
North Carolina State University
909 Capability Drive, Suite 3600
Raleigh, NC 27606

[2] Pima Association of Governments
177 N. Church Ave, #405
Tucson, AZ 85701

[3] Wilbur Smith Associates
465 East High Street, Suite 100
Lexington, KY 40507

Email addresses:
*Corresponding author—bmei@ncsu.edu
T.A. Cooney—tCooney@pagnet.org
N.R. Bostrom—rbostrom@wilbursmith.com

ABSTRACT

This paper investigates the utility of the 2001 National Household Travel Survey Kentucky standard and add-on samples for statewide, rural county, and small urban area travel demand modeling. The weaknesses of the Kentucky standard sample for deriving trip rates and average trip lengths are identified, which include greater uncertainty caused by a small sample size and suspiciously low trip rates for urban clusters (urban areas with less than 50,000 population). We show that the Kentucky add-on sample can be used to enhance the Kentucky standard sample for developing trip rates and average trip lengths. Combining the two samples using Bayesian updating resulted in improved trip rates and average trip lengths.

INTRODUCTION

The objective of this research was to evaluate the utility of the 2001 National Household Travel Survey (NHTS) Kentucky samples, including both standard and add-on, for rural county, small urban area, and statewide travel demand modeling in Kentucky. Specifically, trip rates by trip purpose and by area type derived from the samples were examined by comparison with other samples constructed based

KEYWORDS: National Household Travel Survey, surveys, Kentucky, trip generation, trip distribution, Bayes' theorem, data transferability, travel demand model, transportation planning.

on the 2001 NHTS national sample. Average trip lengths were also analyzed in the same manner.

The 2001 NHTS was conducted as an update to and integration of the earlier Nationwide Personal Transportation Survey, which focused on short daily household trips, and the American Travel Survey, which focused on long trips. Approximately 26,000 households were surveyed nationwide for the NHTS. The NHTS survey data include characteristics of households, people, and vehicles, as well as detailed information on daily and long-distance travel for all purposes and by all modes.[1]

The NHTS is designed to collect data from a nationally representative sample of households in order to provide statistically accurate travel estimates at the national level. Sample data in the NHTS are not intended to be adequate for statewide or area-specific estimates. As a result, the Kentucky Transportation Cabinet (KYTC) participated in the NHTS add-on program initiated by the U.S. Department of Transportation to obtain more household travel data within Kentucky. Under the add-on program, an additional 1,154 households were surveyed in four counties in Kentucky (KYTC 2002). The primary purpose of conducting the Kentucky add-on survey was to achieve a larger sample size suitable for revising travel demand model parameters for rural county, small urban area, and statewide modeling.

DATA

2001 NHTS Kentucky Standard Sample

The Kentucky households in the 2001 NHTS national sample were selected to form a separate sample, which is called the Kentucky *standard sample* in this paper to distinguish it from the Kentucky *add-on sample*. Because the NHTS was designed to collect data from a nationally representative sample of households to provide statistically valid estimates at the national level, the number of households in each state is relatively small. There were only 390 Kentucky households in the national sample, of which 338 had completed trip reports for all household members. These 338 households made 2,785 motorized trips on their assigned travel days, including both weekdays and weekends. Of these, 378 (13.6%) were home-based work trips, 1,546 (55.5%) were home-based other trips, and 861 (30.9%) were nonhome-based trips. Due to the inclusion of weekend trips, home-based work trips account for a lower percentage than that commonly reported for weekday travel demand modeling.

As a fairly common modeling practice, different trip rates are used for different area types. Since the census area-type classification of urbanized area, urban cluster, and rural area has been incorporated into the NHTS dataset, this study employed these area types. The U.S. Census Bureau (USDOC 2000) defines an *urban cluster* as a densely settled area that has a population of 2,500 to 49,999, while an *urbanized area* is defined as a densely settled area that has a population of at least 50,000. Both urban clusters and urbanized areas generally consist of a geographic core of block groups or blocks that have a population density of at least 1,000 people per square mile, adjacent block groups and blocks with at least 500 people per square mile, and less densely settled blocks that form enclaves or indentations, or are used to connect discontiguous areas with qualifying densities. *Rural* consists of all territory, population, and housing units located outside of urbanized areas and urban clusters (USDOC 2000).

Following census definitions, 140 of the 338 Kentucky standard sample households were in urbanized areas, 63 in urban clusters, and 135 in rural areas. These relatively small sample sizes, especially for the urban clusters, raise concerns about the reliability of the survey results for estimating Kentucky-specific trip rates by area type and purpose. Trip rates by area type and trip purpose derived from this sample are presented later with a detailed set of tables.

2001 NHTS Kentucky Add-On Sample

For the Kentucky add-on sample, an additional 1,154 households were randomly selected and surveyed between June 2001 and June 2002 in Carter, Edmonson, Pulaski, and Scott Counties in Kentucky. KYTC classified these four counties as typical non-urbanized counties. Figure 1 shows the location of these counties, while table 1 summarizes the dis-

[1] For more information, see the 2001 NHTS website: http://nhts.ornl.gov/2001/index.shtml.

FIGURE 1 Kentucky Add-On Sampe County Locations

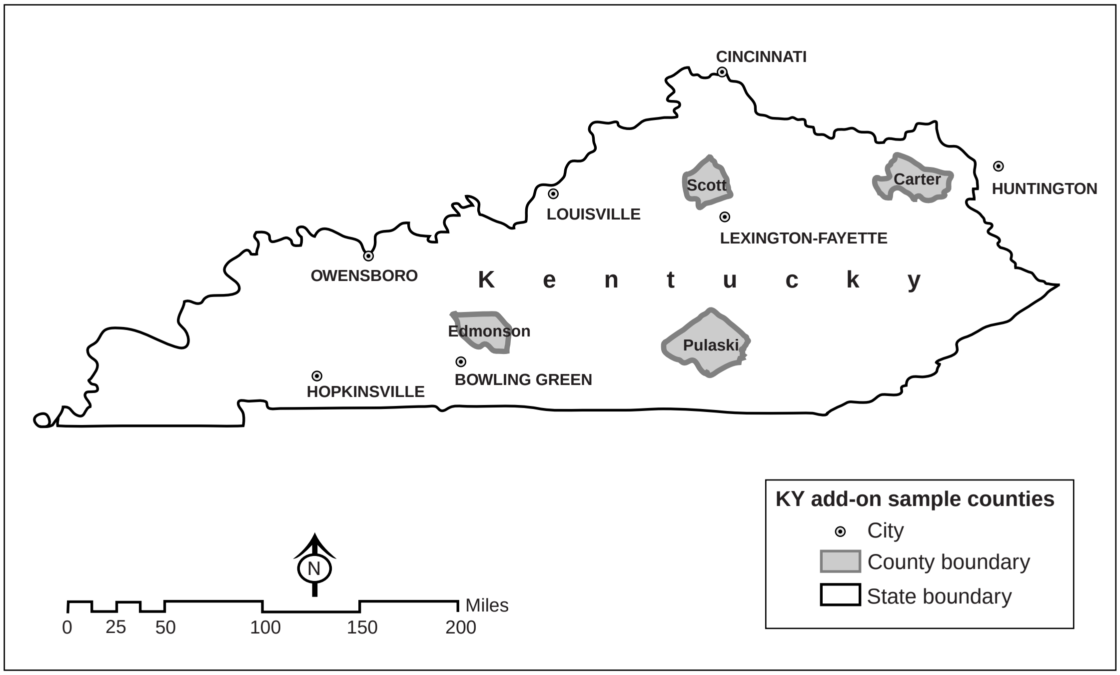


TABLE 1 Kentucky Add-On Sample Allocation

County	Population (Census 2000)	Number of sample households: Urban cluster	Rural area	Overall	Percentage of total
Carter County	26,889	34	116	150	13.0
Edmonson County	11,644	0	165	165	14.3
Pulaski County	56,217	199	224	423	36.7
Scott County	33,061	102	314	416	36.0
Total	**127,811**	**335**	**819**	**1,154**	**100.0**

Source: 2001 NHTS Kentucky add-on sample.

tribution of add-on sample households by area type. Daily household trip information was collected in the sample. Of the 1,154 households, 335 were located in urban clusters, 819 in rural areas, and none were located in urbanized areas. All 2,828 persons in these households reported their travel day trips, which included 9,710 motorized person-trips in total on both weekdays and weekends. Broken down by trip purpose, there were 1,249 (12.9%) home-based work trips, 5,347 (55.1%) home-based other trips, and 3,114 (32%) nonhome-based trips.

Other Datasets

At the time the NHTS survey was conducted, Kentucky was part of the East South Central Census Division and had an overall population of approximately 4.1 million, with 7 metropolitan statistical areas (MSAs) completely or partially located within the state. Two of the 7 MSAs had a population over 1 million and fell into the MSA size category of 1 million to 3 million population. Populations of the others ranged from 100,000 to 500,000. There were no MSAs with a population of 3 million or more.

To check the reasonableness of the trip rates and average trip lengths derived from the Kentucky standard and add-on samples, several other datasets with larger sample sizes were constructed by selecting relevant households from the 2001 NHTS national sample. These datasets are collectively called *non-Kentucky samples* in this paper and were constructed as follows:

- **Households nationwide with the exclusion of those in MSAs with a population of 3 million or more.** We refer to this sample as the *national sample*. A total of 15,443 households that made 141,769 daily motorized trips fell into this category.
- **Households in the East South Central Census Division excluding the state of Kentucky.** This division included Alabama, Mississippi, Tennessee, and Kentucky. No MSAs had a population of 3 million or more in this area. A total of 902 households that made 8,348 daily motorized trips fell into this category. We refer to this sample as the *East South Central sample*.
- **Households in selected states surrounding Kentucky, excluding those in MSAs with 3 million residents or more.** The selected states included Tennessee, Missouri, Illinois, Indiana, and Ohio. We refer to this sample as the *surrounding-states sample*. The resulting dataset consisted of 2,904 households reporting a total of 27,638 motorized trips.
- **Households in states with similar socioeconomic characteristics (in terms of household annual income and household size) to Kentucky, excluding those in MSAs of 3 million residents or more.** This sample is referred to as the *similar-SE-states sample* and consisted of 2,781 households reporting 25,838 motorized trips.

To select states for the similar-SE-states sample, distributions of household annual income and household size of candidate states were compared with those of Kentucky. Only those with similar distributions were included in the sample, which finally consisted of Alabama, Arkansas, Iowa, Louisiana, Missouri, Mississippi, Oklahoma, South Carolina, and Tennessee. Collectively, the household income and size distributions of the similar-SE-states sample were relatively close to those of the Kentucky standard sample.

ANALYSIS

KY Standard Sample Trip Rates and Average Trip Lengths

Trip Rates

The NHTS dataset includes weights to expand the sample data to the U.S. population. For this study, household person-trip rates were developed with both unweighted and weighted survey data. Only motorized trips were included; bicycle, walk, and other nonmotorized trips were excluded from the data. Both weekday and weekend trips were included. To reduce the effect of smaller sample sizes on the accuracy of estimates, trip rates assessed by statistical tests were only classified by trip purpose (home-based work, home-based other, and nonhome-based) and by area type (urbanized area, urban cluster, and rural area). More detailed classifications of trip rates (e.g., by household size, number of vehicles owned, and/or household income) were not attempted in this study.

The statistical comparison of trip rates from different data sources was made using the common *t*-statistic of the difference of two means. Trip rates for each of the samples are displayed in tables 2 through 6.

Table 2 shows that the Kentucky standard sample produced the lowest all-purpose household trip rate (8.53 weighted, 8.24 unweighted). The *t*-tests indicate a significant difference exists between the Kentucky standard sample rate and those of the national, the East South Central, the surrounding-states, and the similar-SE-states samples at the 0.05 level of significance. By area type, Kentucky standard sample rates are also lower than these rates, especially in the urban cluster category where 5.73 weighted and 6.16 unweighted trips were generated on average. These rates are significantly different from all other non-Kentucky rates at the 0.05 level of significance. Because there were only 63 households in this category, the Kentucky urban cluster rates may not be statistically reliable.

Trip rates by trip purpose and by area type are shown in tables 3 through 6. As seen from the tables, the Kentucky standard sample rates agree with the non-Kentucky sample values relatively well for the

TABLE 2 All-Purpose Average Household Person-Trip Rates

	Sample	Urbanized area Mean	Urbanized area Sample size	Urban cluster Mean	Urban cluster Sample size	Rural area Mean	Rural area Sample size	All areas Mean	All areas Sample size
Unweighted	National sample	9.21	7,988	9.00*	2,541	9.22	4,914	9.18*	15,443
	East South Central sample	9.79	391	8.98*	126	8.80	385	9.26*	902
	Surrounding-states sample	9.41	1,503	9.32*	481	9.80*	920	9.52*	2,904
	Similar-SE-states sample	9.68	1,272	9.15*	465	8.88	1,044	9.29*	2,781
	KY standard sample	**9.15**	**140**	**6.16**	**63**	**8.27**	**135**	**8.24**	**338**
	KY add-on sample	—	—	8.81*	335	8.25	819	8.41	1,154
Weighted	National sample	9.55	7,988	9.15*	2,541	9.64	4,914	9.51*	15,443
	East South Central sample	10.04	391	9.35*	126	9.67	385	9.79*	902
	Surrounding-states sample	9.46	1,503	9.37*	481	10.20*	920	9.68*	2,904
	Similar-SE-states sample	9.88	1,272	9.43*	465	9.53	1,044	9.67*	2,781
	KY standard sample	**9.44**	**140**	**5.73**	**63**	**8.89**	**135**	**8.53**	**338**
	KY add-on sample	—	—	9.31*	335	8.42	819	8.68	1,154

Key: * indicates the value is significantly different from the Kentucky standard sample value at the 0.05 level of significance.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

TABLE 3 Average Household Person Trip Rates by Purpose (All Area Types)

Sample	HBW Unweighted	HBW Weighted	HBO Unweighted	HBO Weighted	NHB Unweighted	NHB Weighted
National sample	1.13	1.20	5.05*	5.19	3.00*	3.12*
East South Central sample	1.04	1.13	5.08	5.33*	3.13*	3.32*
Surrounding-states sample	1.20	1.25	5.14*	5.18	3.18*	3.25*
Similar-SE-states sample	1.09	1.17	5.03	5.19	3.17*	3.30*
KY standard sample	**1.12**	**1.19**	**4.57**	**4.69**	**2.55**	**2.65**
KY add-on sample	1.08	1.15	4.63	4.74	2.70	2.79

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; * indicates the value is significantly different from the Kentucky standard sample value at the 0.05 level of significance.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

home-based work purpose for all area types; no significant difference was observed at the 0.05 level of significance. For home-based other trips, the Kentucky standard sample rates are also close to the others in urbanized areas and rural areas; again, no significant difference was observed at the 0.05 level of significance. However, large differences exist for the urban clusters, where the Kentucky standard sample rates are lower than non-Kentucky rates by approximately 30% to 40% and are significantly different from all of them at the 0.05 level of significance. More significant differences exist for the nonhome-based trips. Except for the urban cluster category, the Kentucky standard sample nonhome-based trip rate is also much lower than and significantly different from all non-Kentucky rates, both

TABLE 4 Average Household Person Trip Rates by Purpose (Urbanized Area)

	HBW		HBO		NHB	
Sample	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
National sample	1.12	1.19	5.14	5.29	2.95	3.07
East South Central sample	1.14	1.17	5.29	5.37	3.36	3.51
Surrounding-states sample	1.19	1.22	5.13	5.14	3.09	3.11
Similar-SE-states sample	1.14	1.20	5.22	5.27	3.32	3.41
KY standard sample	**1.20**	**1.25**	**4.99**	**5.10**	**2.96**	**3.09**

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; * indicates the value is significantly different from the Kentucky standard sample value at the 0.05 level of significance.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

TABLE 5 Average Household Person Trip Rates by Purpose (Urban Cluster)

	HBW		HBO		NHB	
Sample	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
National sample	1.11	1.16	5.05*	5.11*	2.84*	2.88*
East South Central sample	0.91	1.01	5.03*	5.15*	3.04*	3.20*
Surrounding-states sample	1.28	1.35	5.15*	5.11*	2.89*	2.91*
Similar-SE-states sample	1.10	1.20	4.96*	5.06*	3.10*	3.18*
KY standard sample	**1.05**	**1.00**	**3.56**	**3.33**	**1.56**	**1.40**
KY add-on sample	1.14	1.24	5.06*	5.37*	2.61*	2.70*

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; * indicates the value is significantly different from the Kentucky standard sample value at the 0.05 level of significance.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

TABLE 6 Average Household Person Trip Rates by Purpose (Rural Area)

	HBW		HBO		NHB	
Sample	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
National sample	1.15	1.22	4.90	5.09	3.18*	3.33
East South Central sample	0.99	1.13	4.88	5.36	2.93	3.18
Surrounding-states sample	1.19	1.25	5.13	5.29	3.49*	3.66*
Similar-SE-states sample	1.03	1.14	4.82	5.16	3.02	3.24
KY standard sample	**1.07**	**1.22**	**4.62**	**4.90**	**2.58**	**2.77**
KY add-on sample	1.06	1.11	4.46	4.49	2.73	2.83

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; * indicates the value is significantly different from the Kentucky standard sample value at the 0.05 level of significance.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

weighted and unweighted, for all area types. Again, this overall significant difference is probably due to the low rates in the urban cluster category.

Average Trip Lengths

Average motorized person-trip lengths by trip purpose were calculated for the national sample, the similar-SE-state sample, and the Kentucky standard sample and included both weekday and weekend trips. Table 7 presents the average trip lengths from the three datasets. It was found that the Kentucky standard sample produced longer trips than the national and similar-SE-states samples for all trip purposes. The *t*-test shows that the home-based other and nonhome-based trip lengths from the Kentucky standard sample are significantly different from those in the national and similar-SE-states samples at the 0.05 level of significance. This is not

TABLE 7 Average Trip Lengths by Trip Purpose

	HBW		HBO		NHB	
Sample	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**	**Unweighted**	**Weighted**
National sample	21.1	21.5	17.3*	17.7*	18.6*	18.6*
Similar-SE-states sample	21.0	21.7	17.1*	17.6*	18.3*	18.8*
KY standard sample	**23.7**	**24.2**	**20.1**	**20.3**	**20.6**	**20.3**
KY add-on sample	25.6	25.1	21.3	20.4	18.7	18.0

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; * indicates the value is significantly different from the Kentucky standard sample value at the 0.05 level of significance.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

surprising, because Kentucky has large rural areas where trip lengths tend to be longer than those in nonrural areas. However, due to the small sample size, the average trip lengths from the Kentucky standard sample had larger standard errors of the mean (discussed in greater detail later).

Kentucky Add-On Sample Trip Rates and Average Trip Lengths

Trip Rates

As introduced above, the Kentucky add-on sample was randomly collected from four counties in the state. Two of the four counties (Carter and Scott) are in MSAs with 300,000 and 500,000 million populations, respectively. Although the sample is not statewide and lacks households in urbanized areas, it still partially reflects the socioeconomic and travel characteristics of Kentucky residents. This, along with a larger sample size, makes the add-on data an appealing source of additional information. Trip rates by county, as well as in total, developed from the add-on sample are shown in table 8.

All-county trip rates are also presented in tables 2 through 6 along with rates from the other samples. As can be seen in table 2, the urban cluster all-purpose trip rates from the add-on sample are much higher than those from the Kentucky standard sample: 8.81 vs. 6.16 for unweighted and 9.31 vs. 5.73 for weighted. In rural areas, the rates from the two samples are close to each other. Compared with the all-purpose trip rates from the non-Kentucky samples, the Kentucky add-on sample rates agree relatively well in urban clusters. However, they are slightly lower in rural areas and more similar to the Kentucky standard sample rates.

Broken down by trip purpose, add-on sample home-based work rates are in good agreement with rates from all other samples. The home-based other trip rates from the add-on sample match well in urban clusters but are slightly lower in rural areas. The add-on sample nonhome-based trip rates are all lower than non-Kentucky samples but higher than the Kentucky standard sample, especially in urban clusters (tables 5 and 6).

Based on the above observations, the Kentucky add-on sample overall appears to provide more reasonable information than does the Kentucky standard sample for urban clusters and rural areas.

Average Trip Lengths

Average trip lengths by county and trip purpose from the Kentucky add-on sample are shown in table 9. Edmonson County produced the longest trips for all trip purposes and Pulaski County produced the shortest trips on average. This travel pattern was found to be consistent with area development patterns. Edmonson County is the most rural of the four counties and Pulaski the most urban. However, they all produced longer trips than the areas included in the national and similar-SE-states samples, as shown in table 7. They even produced longer home-based work and home-based other trips than the Kentucky standard sample. Because statewide models, as well as county and small urban area models, are typically used for studying travel demand and transportation systems in rural areas, the add-on sample provides useful data for developing those models.

TABLE 8 Add-On Sample Average Person Trip Rates by Trip Purpose by Area Type

County		HBW		HBO		NHB		All purposes	
		Urban cluster	Rural area	Urban cluster	Rural area	Urban cluster	Rural area	Urban cluster	Rural area
Unweighted	Carter County	1.26	1.03	5.32	4.52	2.56	2.27	9.15	7.81
	Edmonson County	—	1.00	—	4.50	—	2.47	—	7.97
	Pulaski County	1.05	0.88	4.85	4.62	2.62	2.84	8.53	8.34
	Scott County	1.26	1.23	5.37	4.30	2.61	2.97	9.25	8.50
	All counties	1.14	1.06	5.06	4.46	2.61	2.73	8.81	8.25
Weighted	Carter County	1.25	1.11	5.38	4.92	2.65	2.51	9.28	8.54
	Edmonson County	—	0.99	—	4.33	—	2.52	—	7.84
	Pulaski County	1.13	1.00	5.26	4.91	2.69	3.15	9.08	9.06
	Scott County	1.47	1.25	5.58	4.10	2.73	2.87	9.78	8.22
	All counties	1.24	1.11	5.37	4.49	2.70	2.82	9.31	8.42

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based.
Source: Calculations based on data from 2001 NHTS Kentucky add-on sample.

TABLE 9 Add-On Sample Average Trip Lengths by Trip Purpose (in minutes)

County		HBW	HBO	NHB	All purposes
Unweighted	Carter County	26.8	20.0	15.5	19.6
	Edmonson County	33.8	22.6	26.1	25.1
	Pulaski County	21.4	19.1	17.2	18.7
	Scott County	25.9	23.6	18.7	22.3
	All counties	25.6	21.3	18.7	21.0
Weighted	Carter County	28.5	19.8	15.8	19.8
	Edmonson County	34.5	22.6	25.9	25.2
	Pulaski County	21.8	19.0	16.8	18.6
	Scott County	25.2	23.3	19.2	22.2
	All counties	25.1	20.4	18.0	20.3

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based.
Source: Calculations based on data from 2001 NHTS Kentucky add-on sample.

DATA TRANSFER AND BAYESIAN UPDATING

Based on the above analysis, the advantages and disadvantages of both the Kentucky standard sample and the add-on sample can be summarized as follows:

- The Kentucky standard sample is representative of the entire state, but due to its small sample size large uncertainty exists with the trip rates and average trip lengths derived from the sample.
- The Kentucky add-on sample produces statistically more reliable trip rates and average trip lengths, but due to the way the data were collected the sample does not represent the whole state.

Therefore, data updating techniques were considered to integrate the two data sources into a new set of data, which takes advantage of the Kentucky standard sample in reflecting travel characteristics statewide and the Kentucky add-on sample in providing greater certainty with data values derived from it.

Literature Review

While the transferability of travel demand model parameters, especially those of discrete choice models, has been studied extensively for years (Atherton and Ben-Akiva 1976; Badoe and Miller 1995), it appears that the transferability of transportation planning data has not been much investigated, even though transportation professionals have been using transferred travel data for many years in many places. One of the most typical examples may be the application of national averages of trip rates, trip length distributions, etc., as default values (NCHRP 1998) in the development of travel demand models for small to medium-sized urban areas, where funding was not available for collecting local-specific travel data. In this case, an underlying assumption that people may not be aware of is that those national average values are assumed to be transferable to the study area. If those data are applied to the study area without any adjustments, the transfer is considered a *full* transfer. However, in order to produce reasonable model validation statistics, a common practice for model developers is to adjust the initially adopted national data based on "professional judgment." With adjustments, the data transfer is considered a *partial* transfer. The most critical aspect of a partial transfer is the transfer methodology.

Wilmot and Stopher (2001) conducted research on the transferability of transportation planning data. They stated that disaggregate data at the trip level are very context-specific and therefore intrinsically untransferable, but aggregate data that express the general travel behavior of individuals collectively have a much better chance of being transferable. In their study, they used Bayesian updating to update national averages of trip rates, trip length frequency distributions, and mode shares with local data derived from a small travel survey conducted in Baton Rouge, Louisiana. The updated values and the national averages were then compared with the data values derived from the 1995 Nationwide Personal Transportation Survey Baton Rouge add-on sample, which is much larger in size than the updating sample.

Wilmot and Stopher (2001) found that: 1) transportation planning data (e.g., trip rates, mode shares, and trip length distributions) at certain aggregate levels can be transferred from multiple sources and combined into a single set of updated data; and 2) the data created by updating the transfer data with a small sample of local data were found to be improved over the *fully* transferred data, which were the national averages in their study. They also found that Bayesian updating appears to be a feasible method for data transfer and updating. With Bayesian updating, the influence of all contributory data sources is incorporated into the newly created data. Thus, our study used Bayesian updating to combine the Kentucky standard sample with the Kentucky add-on sample to develop a set of new trip rates and average trip lengths.

Theory of Bayesian Updating

The method of Bayesian updating is based on Bayes' Theorem, which has been widely used for statistical inference (Berry and Lindgren 1990). It starts with prior information and a measure of certainty regarding the prior information. When new sample data are available they are incorporated with the prior into a new answer, which is also called the posterior. With more sample data, the uncertainty regarding the new answer diminishes and the following answers improve.

Since both the prior and the updating sample are normally distributed and the variance is known, the mean and variance of the posterior can be expressed as a function of the mean and variance of the prior and the updating data in the functional form as shown in the equations below (Atherton and Ben-Akiva 1976). This functional form is known as a normal-normal conjugate prior. The posterior produced with this function is also normally distributed.

$$\theta_{updated} = \frac{\dfrac{\theta_{prior}}{\sigma^2_{prior}} + \dfrac{\theta_{updating}}{\sigma^2_{updating}}}{\dfrac{1}{\sigma^2_{prior}} + \dfrac{1}{\sigma^2_{updating}}} \text{ and } \sigma^2_{updated} = \frac{1}{\dfrac{1}{\sigma^2_{prior}} + \dfrac{1}{\sigma^2_{updating}}}$$

where

θ_i is the mean of data item *i;* and

σ_i^2 is the variance of the mean of data item *i.*

Subscription meaning: *updated* represents the updated dataset; *prior* represents the prior dataset; and *updating* represents the dataset used for updating the prior dataset. As a note, since

TABLE 10 Comparison of Updated and Observed Trip Rates for Urban Clusters

	Sample	HBW Mean	HBW SE of mean	HBO Mean	HBO SE of mean	NHB Mean	NHB SE of mean
Unweighted	KY add-on sample	1.14	0.088	5.06	0.260	2.61	0.204
	KY standard sample	1.05	0.174	3.56	0.430	1.56	0.272
	Updated data	**1.12**	**0.078**	**4.66**	**0.222**	**2.23**	**0.163**
	Similar-SE-states sample	1.10	0.074	4.96	0.238	3.10	0.171
	Surrounding-states sample	1.28	0.077	5.15	0.228	2.89	0.165
	East South Central sample	0.91	0.127	5.03	0.459	3.04	0.315
	National sample	1.11	0.032	5.05	0.099	2.84	0.078
Weighted	KY add-on sample	1.24	0.099	5.37	0.284	2.70	0.207
	KY standard sample	1.00	0.170	3.33	0.410	1.40	0.255
	Updated data	**1.18**	**0.085**	**4.71**	**0.233**	**2.18**	**0.160**
	Similar-SE-states sample	1.20	0.086	5.06	0.264	3.18	0.187
	Surrounding-states sample	1.35	0.087	5.11	0.234	2.91	0.174
	East South Central sample	1.01	0.148	5.15	0.520	3.20	0.378
	National sample	1.16	0.035	5.11	0.107	2.88	0.083

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; SE = standard error.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

transfer data serve as the prior in data updating, the terms "prior" and "transfer data" are used interchangeably in this paper.

In the above equations, data item values from the data sources are weighted by the inverse of their variance to achieve a value for the updated data item. This feature is appealing, because data values with greater certainty (i.e., with smaller variance) contribute more to the estimate of the updated data item than those with less certainty (i.e., with larger variance).

When the prior data are reliable, a relatively small sample can be used for updating. However, in the cases where the prior data are not very reliable, a relatively large updating sample is more likely needed. In both cases, the variance of the posterior data will always be less than that of both the prior and the updating sample.

Bayesian updating has been studied and used in the past to update the parameters of travel demand models and the method has been reported to perform well (Atherton and Ben-Akiva 1976). Wilmot and Stopher's study (2001) appears to be the first one that applied this technique in a data transferability study. They reported that updating of transfer data with local information using Bayesian updating seems to improve transfer data consistently.

Bayesian Updating for Kentucky Samples

This study adopted the Bayesian updating method tested by Wilmot and Stopher (2001) and emphasized the strength of the method in combining contributory datasets. The Kentucky standard sample data was utilized as the prior and the add-on sample data as the updating data in the process. The two datasets were combined and a new improved dataset was produced in which the advantages of one input dataset compensated for the disadvantages of the other input dataset.

Updated Trip Rates

We combined the Kentucky standard sample with the Kentucky add-on sample using Bayesian updating, and the updated trip rates were compared with the trip rates from the non-Kentucky samples to investigate the improvement. The updated trip rates are shown in tables 10 and 11 along with the rates from the Kentucky standard sample, the Kentucky add-on sample, and the non-Kentucky samples. These tables show that the updated trip rates are improved overall. For instance, compared with the

TABLE 11 Comparison of Updated and Observed Trip Rates for Rural Areas

		HBW		HBO		NHB	
	Sample	**Mean**	**SE of mean**	**Mean**	**SE of mean**	**Mean**	**SE of mean**
Unweighted	KY add-on sample	1.06	0.050	4.46	0.140	2.73	0.137
	KY standard sample	1.07	0.125	4.62	0.348	2.58	0.259
	Updated data	**1.06**	**0.047**	**4.48**	**0.130**	**2.70**	**0.121**
	Similar-SE-states sample	1.03	0.045	4.82	0.148	3.02	0.124
	Surrounding-states sample	1.19	0.056	5.13	0.152	3.49	0.143
	East South Central sample	0.99	0.067	4.88	0.234	2.93	0.193
	National sample	1.15	0.023	4.90	0.066	3.18	0.058
Weighted	KY add-on sample	1.11	0.054	4.49	0.146	2.83	0.148
	KY standard sample	1.22	0.144	4.90	0.375	2.77	0.301
	Updated data	**1.13**	**0.051**	**4.54**	**0.136**	**2.81**	**0.133**
	Similar-SE-states sample	1.14	0.052	5.16	0.169	3.24	0.139
	Surrounding-states sample	1.25	0.060	5.29	0.164	3.66	0.160
	East South Central sample	1.13	0.085	5.36	0.279	3.18	0.220
	National sample	1.22	0.025	5.09	0.073	3.33	0.064

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; SE = standard error.
Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

trip rates from the similar-SE-states sample, in 10 of the 12 mean cells, updated trip rates are closer than the Kentucky standard sample rates. In particular, the values for urban clusters are improved substantially and look more reasonable. Only two rates appear to deviate more, though slightly. From the data uncertainty perspective, all updated trip rates have much lower standard errors of the mean than the rates from the Kentucky standard sample, which indicates that uncertainty in the updated trip rates has been substantially reduced.

Updated Average Trip Lengths

The major issue identified earlier concerning the average trip lengths from the Kentucky standard sample was the relatively large standard errors of the means, which indicate less confidence with the means. To achieve lower standard errors, the Kentucky standard sample average trip lengths were combined with the Kentucky add-on sample values using Bayesian updating. Table 12 presents the updated average trip lengths along with their standard errors. As stated earlier, the standard error (or variance) of the posterior data produced by the Bayesian updating process is always less than that of both the prior and the updating sample. Therefore, as expected, for unweighted average trip lengths, the standard error of the mean decreased from 1.692 to 0.422 for home-based work trips, from 0.690 to 0.289 for home-based other trips, and from 0.685 to 0.388 for nonhome-based trips. Similar patterns were also observed for the weighted trip lengths. While the process reduced uncertainty in the average trip lengths, the values of the updated average trip lengths stayed close to both the Kentucky standard sample and the add-on sample values but are longer than the national averages and the similar-SE-states sample values. This may reflect the actual trip length characteristics of Kentucky.

CONCLUSIONS

Based on the above analysis, the following conclusions are drawn:

1. The 2001 NHTS Kentucky standard sample is partially useable for travel demand modeling, although its sample size is not large and it produces unreasonably low trip rates for urban clusters.
2. The Kentucky add-on sample produces statistically more reliable trip rates and average trip

TABLE 12 Comparison of Updated and Observed Average Trip Lengths

	Sample	HBW Mean	HBW SE of mean	HBO Mean	HBO SE of mean	NHB Mean	NHB SE of mean
Unweighted	KY add-on sample	25.6	0.604	21.3	0.295	18.7	0.403
	KY standard sample	23.7	1.692	20.1	0.690	20.6	0.685
	Updated data	**25.4**	**0.422**	**21.1**	**0.289**	**19.2**	**0.388**
	Similar-SE-states sample	21.0	0.473	17.1	0.188	18.3	0.285
	National sample	21.1	0.194	17.3	0.082	18.6	0.120
Weighted	KY add-on sample	25.1	0.610	20.4	0.303	18.0	0.411
	KY standard sample	24.2	1.705	20.3	0.711	20.3	0.702
	Updated data	**25.0**	**0.428**	**20.4**	**0.297**	**18.6**	**0.396**
	Similar-SE-states sample	21.7	0.492	17.6	0.197	18.8	0.308
	National sample	21.5	0.199	17.7	0.088	18.6	0.125

Key: HBW = home-based work; HBO = home-based other; NHB = nonhome-based; SE = standard error.

Sources: Calculations based on data from 2001 NHTS national sample and Kentucky add-on sample.

lengths although it is only partially representative of Kentucky.

3. The Kentucky add-on sample can be used to enhance significantly the NHTS Kentucky standard sample for statewide, rural county, and small urban area travel demand modeling. The trip rates generated by combining the Kentucky add-on sample rates with the Kentucky standard sample rates using Bayesian updating showed substantial and consistent improvements. The Bayesian updating process also improved average trip lengths by reducing uncertainty in the data.
4. The findings of this study support the conclusions drawn by Wilmot and Stopher (2001) that transportation planning data can be improved by combining two or more reasonable datasets into one, and Bayesian updating seems to be a feasible method to combine or update transportation planning data.

NOTE

At the time this study was conducted, Mr. Mei and Mr. Cooney were with Wilbur Smith Associates and Mr. Bostrom was with the Kentucky Transportation Cabinet.

REFERENCES

Atherton, T.J. and M.E. Ben-Akiva. 1976. Transferability and Updating of Disaggregate Travel Demand Models. *Transportation Research Record* 610:12–18.

Badoe, D.A. and E.J. Miller. 1995. Analysis of the Temporal Transferability of Disaggregate Work Trip Mode Choice Models. *Transportation Research Record* 1493:1–11.

Berry, D.A. and B.W. Lindgren. 1990. *Statistics: Theory and Methods*. Pacific Grove, CA: Brooks/Cole Publishing Co.

Kentucky Transportation Cabinet (KYTC). 2002. 2001 NHTS Add-On Program: Final Report and Data Codebook. October.

National Cooperative Highway Research Program (NCHRP). 1998. *Travel Estimation Techniques for Urban Planning,* NCHRP Report 365. Washington, DC: National Academy Press.

U.S. Department of Commerce (USDOC), U.S. Census Bureau. 2000. Geographic Terms and Concepts. Available at http://www.census.gov.

Wilmot, C.G. and P.R. Stopher. 2001. Transferability of Transportation Planning Data. *Transportation Research Record* 1768:36–43.

Using National Data to Simulate Metropolitan Area Household Travel Data

PETER R. STOPHER[1,*]
STEPHEN P. GREAVES[1]
MIN XU[2]

[1] Institute of Transport & Logistics Studies (C37)
The University of Sydney, NSW 2006
Sydney, Australia

[2] Transport & Population Data Center (TPDC)
New South Wales Department of Planning
GPO Box 39, NSW 2001
Sydney, Australia

ABSTRACT

This paper describes the overall approach to simulating household travel survey data, and provides an overview of the results from three metropolitan areas in the United States and two in Australia using the 1995 Nationwide Personal Transportation Survey data. The applications we use demonstrate the benefits of the approach that could save substantial amounts of money on data collection. We identify the need for improvement of the approach and propose a new procedure for simulating trip tours and their characteristics, instead of trips. The paper concludes by providing the preliminary findings based on tours as the unit of analysis using the 2001 National Household Travel Survey.

INTRODUCTION

Household travel surveys (HTSs) are increasingly expensive to undertake (Schofer 2002). In addition, problems with response rates, misreporting of travel, and the increasing difficulty of conducting computer-assisted telephone interview (CATI) surveys make it likely that the HTS as we know it will have to change significantly in the future. For modeling purposes, sample sizes of 3,000 households and upwards are required irrespective of the size of

E-mail addresses:
* Corresponding author—peters@itls.usyd.edu.au
S.P. Greaves—stepheng@itls.usyd.edu.au
M. Xu—min.xu@dipnr.nsw.gov.au

KEYWORDS: Household travel survey data, Classification and Regression Trees (C&RT), Monte Carlo simulation, microsimulation, Bayesian updating, trip tours.

the region, putting an adequate sample size beyond the reach of many urban areas.

Concurrent with the increasing expense and difficulties of obtaining HTS data is the demand for greater spatial coverage of the sample within a region to support micro-level planning along transport corridors and within subareas. Finally, there is greater interest in obtaining statewide HTS data. This may entail surveys in both small urban and rural areas, which are potentially more expensive to conduct than the standard metropolitan HTS. Survey sample sizes for rural and small urban areas are likely to be relatively small, although the desire may still be to produce models for each such geographic grouping.

In an effort to find an alternative to the large-scale HTS, recent research has aimed at developing a method to simulate HTS data (Greaves 2000; Greaves and Stopher 2000; Stopher et al. 2003). The method uses distributions of travel characteristics obtained from a nationwide sample that are updated to a specific locality using a small local update sample and Bayesian updating with subjective priors. A Monte Carlo simulation of specific travel attributes is performed, namely, the number of trips by purpose, and, for each simulated trip, the main mode of travel, the time of departure, and the trip duration. All of the U.S. work reported to date (covering Baton Rouge, Dallas, and Salt Lake City) uses the 1995 Nationwide Personal Transportation Survey (NPTS) as the source for the distributions. The U.S. distributions have also been used to simulate HTS data for Adelaide and Sydney in the Australian cases because of the lack of a recent nationwide travel survey (Stopher et al. 2003; Pointer et al. 2004).

Overall, the results so far have demonstrated the method capable of creating HTS data that are a reasonable approximation to observed data in a variety of urban settings. This said, several needed enhancements to the methods have also been identified, chief among them is replacing the trip-based method with a tour-based simulation method. Following a synopsis of the progress to date, this paper provides the rationale for this latest development, together with our initial thoughts on how one should classify tours for the purposes of simulation and, ultimately, how to simulate the tours using the 2001 Nationwide Household Travel Survey.

Clarifying the Role of Simulated HTS Data

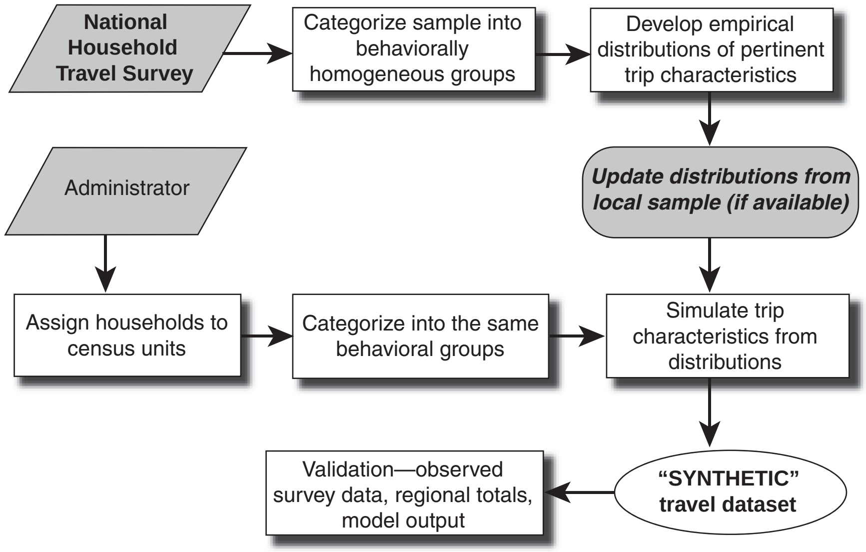
While the rationale for simulating HTS data is apparent, the role those data should play within the travel forecasting landscape has proven more contentious. At the heart of this debate is whether it is appropriate to use simulated data as an input to another modeling procedure that may be based on different underlying assumptions from those used to generate the data. If the various elements (e.g., trips, destination, modes, routes) are modeled separately, as is predominantly the case in practice, then it is arguably an appropriate application. However, problems could arise when the simulated HTS data are used to estimate models based on interdependent relationships, which typify most disaggregate modeling applications. For instance, problems could exist with the mode-choice step of current procedures, because this is typically done at a disaggregate level. One other area where problems could occur is within the simulated data because of the built-in assumptions as to city size and public transport service, albeit averages generated from a nationwide sample. Nevertheless, it is possible that there are inherent assumptions that may even run contrary to those used in an aggregate model of trip distribution, for example. However, if simulated data can be generated that are themselves a function of such things as city size and transport networks, then this objection would largely evaporate.

In light of this issue and the well-voiced concerns with conventional models (whose structures have been somewhat dictated by data restrictions anyway), the possibility that a simulation-based approach can generate data for large samples and even entire populations suggests that a more fruitful direction could be in the use of national data to estimate travel in a local region.

OVERVIEW OF THE SIMULATION APPROACH

The method proceeds through the steps depicted in figure 1. Initially, the NPTS data were classified using the Classification and Regression Tree (C&RT) method (Breiman et al. 1984) into behav-

FIGURE 1 Overview of the Approach



iorally homogeneous groupings based on the pertinent trip attributes of interest: trip rates by purpose, mode, departure times, and trip durations in minutes. In the delineation of the categories, while attempts were made to build characteristics of the metropolitan area and transport supply directly into the categories used in the simulation, these added little to the explanation and therefore were included indirectly through the local data updating procedures described in a later section.

Table 1 presents the household classifications for predicting home-work, home-school, and home-other trip rates used in the simulation (Greaves and Stopher 2000). The generalized linear modeling results indicate both the statistical significance of the groupings (F-statistic) and the proportion of variance explained by the schemes (R^2). The remaining trip attributes use categories based on household demographics and the prior simulated attribute. This is indicated for mode in table 2 for selected trip purposes.

Having created the categories, we next developed the distributions for each category. For trip purpose, the distributions represent the relative frequency that zero, one, two, or more trips would be produced by the household for that purpose (illustrated in figure 2 as a cumulative relative frequency graph using the example of home-work trips). In the case of mode, the distributions represent the relative frequency of taking each of the five modes indicated in table 2 (an example is shown for home-work modal trips in figure 3).

It should be noted that a considerable amount of data are required to construct reliable distributions. In using the 1995 NPTS data, we removed certain records based on missing data and proxy reporting, so that the final number of household records available for the C&RT analysis, the development of frequency distributions, and the Monte Carlo simulation was 30,400. Our simulation used random numbers that were treated as probabilities. Each probability was then read from the cumulative distribution and the value corresponding to it picked from the distribution and assigned to the household or trip of concern.

Next a sample of households was drawn (not micro-simulated) using real households from census data. In our U.S. applications, we used the 5% Public Use Micro-Data Sample (PUMS), providing unit records for "long-form" households in the decennial census. In the Australian work, the Australian Bureau of Statistics produces the Household Sample File (HSF), which contains full unit records for 1% of the households in the five-year Australian census. In both cases, to protect the confidentiality of the records, the geographic location of the household was given only at a large geographic area level, so that the actual location of the household, with respect to the transport system, was not known. In the case of the PUMS data, households were given

TABLE 1 Categories Used for Selected Trip Purpose Distributions

Trip purpose	Categorization scheme	Mean	Standard deviation	GLM results
Home-work	0 workers	0	0	$F = 3{,}228$
	1 worker, 0–1 vehicles	1.29	1.05	$df = 9$
	1 worker, 2+ vehicles	1.45	1.09	$R^2 = 0.489$
	2 workers, 0 children (0–4), 0 children (5–17)	2.78	1.56	
	2 workers, 0 children (0–4), 1+ children (5–17)	2.56	1.56	
	2 workers, 1+ children (0–4), 0 children (5–17)	2.14	1.40	
	2 workers, 1+ children (0–4), 1+ children (5–17)	2.32	1.39	
	3 workers, 0 children (5–17)	4.12	2.05	
	3 workers, 1+ children (5–17)	3.75	1.94	
	4 + workers	5.56	2.41	
Home-school	0 children (5–17)	0	0	$F = 14{,}039$
	1 children (5–17)	1.30	0.90	$df = 4$
	2 children (5–17)	2.73	1.50	$R^2 = 0.704$
	3 children (5–17)	4.16	2.17	
	4+ children (5–17)	5.46	3.02	
Home-other	1 person, 0 workers, 0 vehicles	1.06	1.36	$F = 871$
	1 person, 0 workers, 1+ vehicles	1.92	1.87	$df = 15$
	1 person, 1 worker	1.11	1.42	$R^2 = 0.300$
	2 persons, 0–1 workers, 0 vehicles	1.90	2.03	
	2 persons, 0 workers, 1+ vehicles	3.68	3.01	
	2 persons, 1 worker, 1+ vehicles	2.93	2.66	
	2 persons, 2 workers	2.09	2.16	
	3 persons, 1–2 children (0–4)	2.59	2.18	
	3 persons, 0 children (5–17), 0 children (0–4)	3.70	3.25	
	3 persons, 1–2 children (5–17), 0 children (0–4)	4.67	3.50	
	4 persons, 0–1 children (5–17), 0–1 children (0–4)	5.08	3.49	
	4 persons, 0–1 children (5–17), 2–3 children (0–4)	2.88	2.38	
	4 persons, 2+ children (5–17)	6.83	4.81	
	5+ persons, 0–1 children (5–17)	5.63	4.21	
	5+ persons, 2 children (5–17)	7.39	5.09	
	5+ persons, 3+ children (5–17)	9.16	6.29	

Key: GLM = generalized linear modeling; F = F-statistic to test the null hypothesis of equal means; df = degrees of freedom; R^2 = proportion of variance explained by the schemes (within group sum-of-squares/total sum-of-squares).

Note: Trip rates shown are for those categories in the NPTS sample.

weights in the data, because they were not sampled uniformly into the PUMS. In using the PUMS data, prior to sampling, each household was replicated the number of times appropriate for its weight. In the HSF data, households were sampled on a strictly random basis, so that each household represented 100 households.

Next, a sample design was chosen. In many HTSs, the sample will be stratified geographically, and then by household size and number of vehicles in the household. The first stratification is usually a disproportionate sample, to ensure that there is either a statistically or politically adequate sample in each geographic subregion of the study area (e.g., a

TABLE 2 Categorization Scheme for Travel Mode Simulation (in percent)

Trip purpose	Mode categories	POV driver	POV passenger	Bus	School bus	Bike or walk
Home-work	0 vehicles	12	34	37	0	16
	1 vehicle, 1 worker	91	7	1	0	2
	1 vehicle, 2+ workers	65	25	5	0	4
	2+ vehicles, 1–2 workers	95	4	0	0	1
	2+ vehicles, 3+ workers	87	9	1	0	2
Home-school	0 vehicles	0	12	4	59	24
	1 vehicle, 1–3 persons	0	45	3	47	5
	1 vehicle, 4+ persons	2	26	4	52	16
	2 vehicles, 1–2 persons	50	0	50	0	0
	2 vehicles, 3 persons	2	50	0	39	9
	2 vehicles, 4 persons	1	37	1	54	7
	2 vehicles, 5+ persons	0	44	2	50	4
	3+ vehicles, 1–3 persons	21	32	0	47	0
	3+ vehicles, 4+ persons	11	37	2	41	9
Home-other	0 vehicles	16	50	14	2	18
	1+ vehicle, 1 person (18+), 0 children (5–17)	88	8	0	0	3
	1+ vehicle, 2+ persons (18+), 0 children (5–17)	73	23	0	0	3
	1 vehicle, 1 child (5–17)	48	44	0	2	5
	2+ vehicles, 1 child (5–17)	62	33	0	1	4
	1 vehicle, 2+ children (5–17)	38	48	1	2	11
	2+ vehicles, 2+ children (5–17)	49	42	0	1	7

Key: POV = privately operated vehicle.

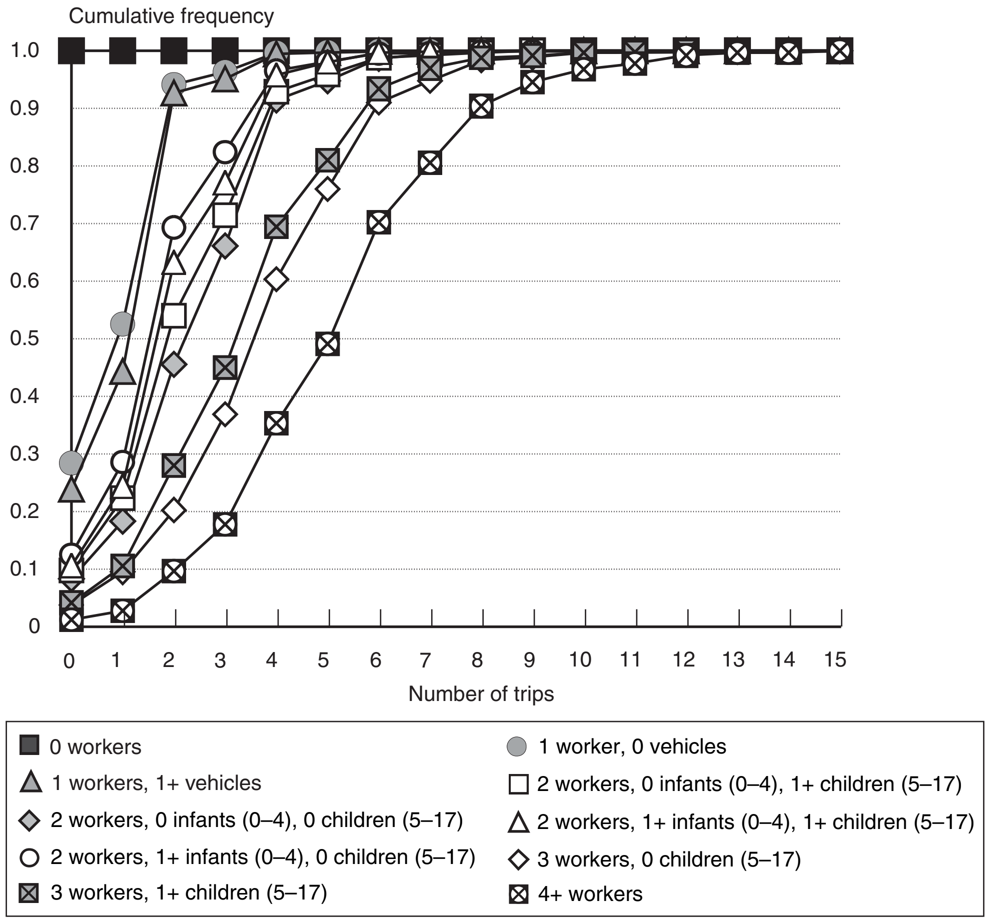
Note: Percentages shown are the proportion of trips within each purpose/category using that particular mode.

sufficient sample in each county for a multi-county region). To reduce sampling costs, the household size and number of vehicles stratification is usually designed as a proportionate sample. Such a sampling scheme can be used, provided that the census subdivisions used to record the PUMS or HSF data correspond roughly to the geographic subregions for sampling.

In this case, the unit record data were first grouped into the appropriate geographic subdivisions for each of which a sampling goal was established, so that there was a known distribution of households required by cell of a household size by vehicle availability matrix. Households were then randomly sampled from the unit records. The sampling was undertaken without replacement, because each household unit record appeared multiple times in the region, based on its weight. This does mean, however, that some household unit records may have been used more than once in a sample, to represent the appropriate stratified random sample.

Once households were sampled, we had complete sociodemographic data on the household, allowing us to determine the specific group to which the household belonged for each travel characteristic to be simulated. This permitted us to draw the travel characteristics from the appropriate distributions for each travel characteristic and to simulate an entire day's worth of travel for the household. Initially, we simulated the number of trips by each trip purpose made by the household. From this point on, the simulation related to each individual trip that had been simulated for the household, for which mode, time of departure, and trip duration were each simulated, conditional on the preceding travel characteristics.

FIGURE 2 Distributions for Simulating Home-Work Trips



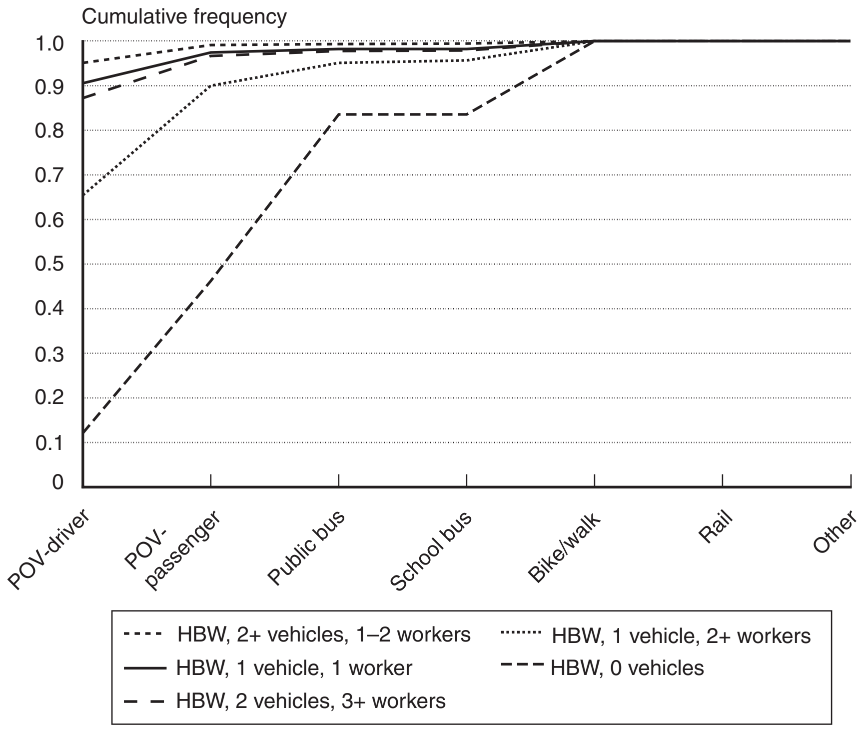
In standard Monte Carlo simulation, it is customary to make many hundreds of drawings from the distributions for each characteristic and then to average the results. However, in our case, because we were usually simulating thousands of households and tens of thousands of trips, we found it unnecessary to use repetitive drawings, especially when the results were to be used in an aggregate manner. In effect, the thousands of households for which travel was to be simulated and the tens of thousands of trips for which characteristics were to be simulated approximate the normal Monte Carlo procedure of multiple drawings.

Bayesian Updating

In the early simulations for Baton Rouge, Dallas-Fort Worth, and Salt Lake City, we found that while trip rates were reproduced reasonably well based on a purely demographic categorization of the population, not surprisingly, mode shares and trip lengths were not. To address this, we then included measures of transportation supply, urban area size, density, and other contextual measures in the categorization known to correlate with travel. However, despite repeated attempts, this approach produced only marginal improvements in the simulation results.

As a consequence, we explored alternative methods in which we assumed that while a full HTS of 3,000 or more households may be out of the budget of the area of concern, a small HTS of, say, 300 to 750 households might be feasible to adjust/update the information coming from the national survey. The original rationale for this approach was taken from the literature on travel-behavior model transferability between regions, which consistently demonstrates that the quality of the transfer is substantially improved if local data (particularly from a small sample of households) are available to update model parameters (Atherton and Ben-Akiva 1976; Badoe and Miller 1995). In our case, we

FIGURE 3 Distributions for Simulating Mode of Travel for Home-Work Trips



Key: HBW = home-based work; POV = privately owned vehicle.

applied similar logic to update the travel data, in this case by updating the NPTS probability distributions that drove the simulation.

For this research, we used Bayesian updating with subjective priors, a procedure used in model updating (Koppelman et al. 1985). Under this procedure, an unknown parameter θ is related to its prior distribution and the likelihood function of the local data by the probability expression:

$$\begin{pmatrix}\text{Posterior probability}\\ \text{of } \theta \text{ given the local data}\end{pmatrix} \propto \begin{pmatrix}\text{prior probability}\\ \text{of } \theta\end{pmatrix} * \begin{pmatrix}\text{likelihood function of}\\ \text{the local data given } \theta\end{pmatrix}$$

The critical issue with using Bayesian updating is to define the prior distribution of θ. The most widely used approach is to assume θ is normally distributed with mean θ_t and variance σ_t. Similarly, the sampling distribution of the local data is assumed to be normally distributed with mean θ_s and variance σ_s. This assumption (conjugate prior) enables data from the two sources to be combined to produce a posterior distribution that is also normally distributed with parameters θ_p and variance σ_p that are calculated as follows:

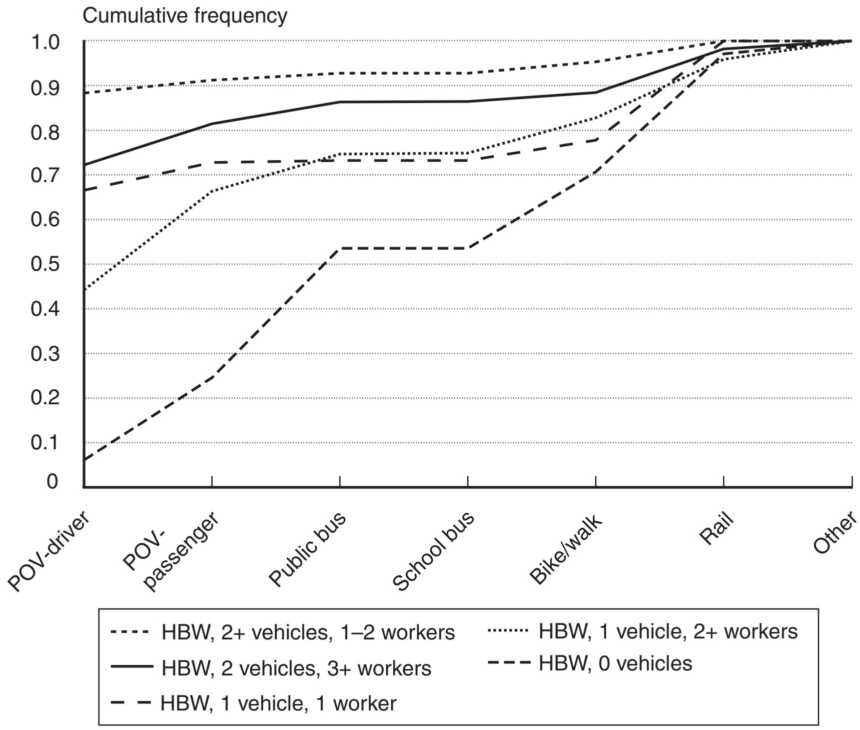
$$\theta_p = [\theta_t / \sigma_t^2 + \theta_s / \sigma_s^2] / [1/\sigma_t^2 + 1/\sigma_s^2] \qquad (1)$$

$$\sigma_p^2 = [1/\sigma_t^2 + 1/\sigma_s^2] \qquad (2)$$

Equation 1 shows that θ_p was derived from the prior and local samples, which had effectively been weighted by the inverse of their respective variances. These weights can be altered manually if they are deemed inappropriate. As a practical matter, the discrepancy in sample size between the update sample and the NPTS sample meant that without some manual adjustment of weights no effects were observed. However, it is clearly important that this manual adjustment is done based on sound reasoning.

The method was applied by updating the frequencies at each appropriate value of the distributions to be used in the simulation to produce modified distributions. Given that each interval was treated as a proportion, an estimate was needed for the standard error of the share (this is actually analogous to the standard deviation of the sampling distribution of a sample proportion). This can be derived from the following expression, although it must be noted that this requires five or more estimates for the assumption of normality to hold. This was problematic given the size of the update sample and the level of disaggregation used in the categorization schemes.

FIGURE 4 Distributions for Simulating Mode Using an Updated Sample of 300 Households (Sydney)



Key: HBW = home-based work; POV = privately owned vehicle.

$$std.error = \sqrt{\frac{p(1-p)}{n}}$$

where the sample proportion x/n is substituted for p, x = share, and n = sample size.

The impact of Bayesian updating of the distributions is illustrated by the case of simulating the mode of travel for home-work trips in Sydney, Australia (Pointer et al. 2004). In this case, 300 Sydney households were drawn randomly and their information used to update the NPTS distributions (shown in figure 3) to those shown in figure 4. The impact of the updating is apparent. In particular, the large increase in rail travel and the decline in car driver trips were consistent with what we would expect. The evidence from this updating experiment suggests that a local sample of about 500 households might be an optimal balance between the costs of surveying additional households and accuracy of prediction.

APPLICATIONS

Summary of Previous Applications

In work reported to date, this procedure has concentrated on applications for demonstration purposes and has been restricted to running the simulations to replicate an actual HTS. Initial work was performed using Baton Rouge as a test case. We selected Baton Rouge because its household travel survey was conducted immediately following the NPTS and used the exact same survey methods and materials. This meant there would be no intervening problems of methodological difference in data collection between the survey used for comparison and the source of the simulation. Results from this test were encouraging, showing generally only small numeric differences between the simulations and the actual HTS (Greaves and Stopher 2000). Subsequent work with Baton Rouge data included initial tests of Bayesian updating, which improved most of the data fits to the original household travel survey (Greaves 2001).

Since the original Baton Rouge work, the methods have been tested in four regions with markedly different demographic, transport, and urban form characteristics: Dallas-Fort Worth, Salt Lake City, Adelaide, and Sydney. In the case of the two Australian cities, the original NPTS distributions served as the source of the simulated travel data and the HSF was used as the source of the local demographic data. In all cases, while there were numerical differences in most of the travel characteristics produced from the simulations, aggregate totals of trips by purpose, by mode, by time of day, and by trip duration were generally within 2% to 5% of the actual results of the household travel surveys (Stopher et al. 2001; Stopher et al. 2003). In all cases, Bayesian updating produced marked improvements in these figures. In the Sydney application, we also explored the effects of sample size requirements for the Bayesian updating, concluding a sample of about 500 households represented a minimum desired number in this particular case (Pointer et al. 2004).

Potential Uses of the Simulations

Overall, the results have been encouraging. Using a Monte Carlo simulation of travel characteristics appears to be viable to replace or augment standard household travel surveys, especially if Bayesian updating is undertaken with a small local sample (about 500 households) for updating purposes. In the United States, it has been suggested that simulated data could be used to create a database for smaller metropolitan areas that lack sufficient resources to undertake a full HTS (i.e., a sample of 3,000 or more households). In place of this, such areas can undertake a small sample of about 500 households and then undertake simulation of a sample as large as may be desired, up to and including the entire metropolitan population.

In Australia, somewhat different applications have been suggested. First, it can be used as a means to increase the sample available for subregional and corridor planning, where a small sample may be available from a metropolitan HTS that can be used for Bayesian updating. Second, it can be used to augment the sample sizes for extending metropolitan surveys into the nonmetropolitan areas of states or to full statewide surveys. In this context, a state may be subdivided into regions and a small sample drawn for each, which is then extended by simulation to provide a much larger sample for each major area of the state.

The simulated data represent a much cheaper option than a full HTS. In general, the creation of a simulated dataset of almost any size, with Bayesian updating, would cost about $20,000 to $30,000. A small sample of 500 households, using current average costs for a CATI survey, is likely to cost about $100,000, giving a total cost of the simulated dataset of less than $130,000. This compares favorably to the cost of a full survey of 3,000 or more households, which could cost in excess of $500,000.

In all of the above applications, the data could be used just like actual HTS data, that is, as an input to model estimation or calibration. However, as suggested in the early part of this paper, another alternative is to apply the data directly in forecasting. This could be done with a second simulation, in which the future characteristics of households are simulated from the present, using an available microsimulation procedure (Chung and Goulias 1997) and then re-running the simulation of the travel characteristics with the new distribution of household demographics.

Limitations of the Current Simulation Methods

One limitation of the simulation procedure is that the lack of transport system characteristics will show little or no difference in travel patterns, irrespective of changes made to the transportation system. That is, total numbers of trips by purpose, by mode, by time of day, and by trip length would not vary as changes are made to the system, in contrast to travel patterns revealed by trip distribution and assignment models, where the actual destinations and routes chosen may change. However, modeling trip distribution and assignment is generally not possible as a result of the lack of geographic specificity in the simulation.

Another problem that arises from current simulations is that disaggregation of the simulated data will reveal potential inconsistencies in the data. In this procedure, trips are simulated independently of each other, even at the household level, which results in two undesirable properties. The first applies to households and individuals making no trips. The simulations always underestimate the number of households reporting no travel on the travel day compared with the actual HTS. This occurs because of the independence in the generation of trip numbers by purpose. The probabilities associated with zero trips for any purpose are generally quite small. The simulated probability of a household making no trips is the product of the individual probabilities of zero trips for all trip purposes, which will be very low. In reality, there is interdependence among the trips and also in a household generating no trips on a given day. For example, a household with workers, none of whom make a work trip on a given day, may be much more likely to be a zero-trip household. This interdependence is not present in the trip-based simulations.

The work-around for this problem in the applications described in the previous subsection was to estimate values for trip-making households only and compare them, and then to estimate the number of zero trip-making households separately, based on the actual HTS. This works when there is an actual HTS but becomes problematic when the simulation is performed as a substitute for an HTS.

The second inconsistency arises from the independence of the simulations of both trips by purpose and the characteristics of each trip. The following scenario could occur in a simulation. A household is simulated to have three home-based work (HBW) trips, one home-based school (HBSch) trip, one home-based shopping (HBSh) trip, and one nonhome-based (NHB) trip. The household has two workers (both adults) and one school age child. The first problem is that some members of the household may be unable to return home, because there are not sufficient return trips to get everyone home by the end of the day. For example, we could hypothesize that one person went to work, came home, and did nothing else (two HBW trips). A second person could have gone to work, then gone from work to shop, and then returned home (one HBW trip, one NHB trip and one HBSh trip). The child in the household then went to school, but there is no trip left to get the child home again. In addition, the mode of one HBW trip might be car driver, one might be car passenger, and one might be bus, while the HBSch trip is by walk, the HBSh trip is by bicycle, and the NHB trip is by bus. This makes little sense, because the car is not driven home by a household member, and the bicycle also either never returns home or is picked up at the shopping location. Furthermore, times of day may not match our inferences here—all the HBW trips and the HBSh trip may have a morning time of departure, while the HBSch and NHB trips may have an afternoon departure time. There is also no reason for the durations of these trips to resemble one another. Another problem would be a household that is simulated as making one trip only, which might even be a NHB trip.

In aggregate analysis, none of the above matters, because the trip totals by purpose, mode, time of day, and duration will provide reasonable descriptions for a population. However, these aspects of independence in the trip simulation make it impossible to disaggregate the data below a certain point and also preclude the simulation of the geographic location of trip ends. Therefore, to be able to introduce the geographic aspects of trip making into the simulated data, it is clearly necessary to change the simulation procedure so that disaggregation is possible and produces sensible results. In addition, dependence among simulated trips will also permit better simulation of the number of nonmobile households on a given day. How dependence on the transport service can be introduced is more problematic and is not directly solved by the steps that may be necessary to introduce trip dependence. However, these issues are explored more in the next section of this paper.

SIMULATING TOURS

Our initial thinking on how to resolve these problems was to simulate first whether a household would make trips or not, then the (nonzero) number of trips, and then the purposes of those trips. However, while this would deal with two of the problems, namely the correct number of zero-trip households and having an appropriate number of trips for a household, it would not take care of the problem of simulating illogical or unrealistic combinations of trip purposes, departure times, or trip durations. On further consideration, it seemed most productive to move away from simulating trips to simulating tours instead.

We define a tour, similar to the Adler and Ben-Akiva (1979) definition of it in their early work, as a "set of consecutive trip links that begin and end at an individual's home." We chose to depart from this by using a set of trip links that may begin and end at

TABLE 3 O'Fallon and Sullivan's Classification of Tours

Tour	Tour description	Sequence
1	Simple work	h – w – h
2	Multipart work	h – w – (w) – w – h
3	Composite to work	h – nw/e – (nw/e/w) – w – h
4	Composite from work	h – w – (nw/e/w) – nw/e – h
5	Composite to and from work	h – nw/e – (nw/e/w) – w – (nw/e/w) – nw/e – h
6	Composite at work	h – w – (nw/w/e) – nw/e – (nw/w/e) – w – h
7	Simple/multipart education	h – e – (e) – h
8	Composite education and nonwork	h – nw – e – h **and** h – (nw) – nw – h
9	Simple nonwork/non-education	h – nw/ne –h
10	Multipart nonwork/non-education	h – nw/ne – nw/ne – (nw/ne) – h

Key: h = home; w = work; e = education; ne = non-education; nw = nonwork.

TABLE 4 Proposed Tour Classification for Simulation

Tour	Tour description	Sequence
1	Simple home-work	h – w – h
2	Simple education	h – e – h
3	Simple nonwork/non-education	h – nw/ne – h
4	Simple work-based	w – nw/e/w – w
5	Complex home-work	h – w – w – h **and** h – nw/e – (nw/e/w) – w – h **and** h – w – (nw/e/w) – nw/e – h **and** h – nw/e – (nw/e/w) – w – (nw/e/w) – nw/e – h
6	Complex education	h – nw/ne – (nw/e) – e – h **and** h – e – (nw/e) – nw/ne – h **and** h – nw/ne – (nw/e) – e – (nw/e) – nw/ne – h
7	Complex nonwork/non-education	h – nw/ne – nw/ne – (nw/ne) – h
8	Complex work-based	w – nw/e/w – nw/e/w – (nw/e/w) – w

Key: h = home; w = work; e = education; ne = non-education; nw = nonwork.

home, defining a home-based tour, or that may begin and end at work, defining a work-based tour. This conforms a little more closely to a definition suggested by Axhausen (2000) that a tour is any sequence of trip links that begin and end at the same location.

Home-based or work-based tours can also be either simple or complex. A simple tour involves only one other activity and is accomplished with two trips, for example, a trip from home to work, followed by one from work to home, or a trip from home to shop and a trip from shop to home. A complex tour is any tour involving multiple stops, such as a trip from home to school to work to shop to school to home. Both work-based tours and home-based tours can be simple or complex.

The literature describes many possible ways to classify tours. We have followed a scheme proposed by Strathman and Dueker (1995) and modified by O'Fallon and Sullivan (2004). Table 3 presents O'Fallon and Sullivan's classification where stops shown in parentheses may occur zero, one, or more times. We modified this further to separate out the simple from complex tours. Also, at this stage, it is not clear that the simulation would call for as many classifications into "to work," "from work," etc. Our preliminary modification of this is shown in table 4. The same interpretation of the stops in parentheses applies in table 4 as in table 3.

The next issue was how to simulate the tours. There are two approaches, referred to in the literature as *sequential* and *simultaneous/holistic* approaches.

Sequential approaches involve the incremental generation of each trip or activity in the tour, based on previous elements of the tour (Kitamura et al. 1997). Simultaneous approaches involve the initial generation of the entire tour/pattern; such an approach is used in the TRANSIMS framework, where activity tours are generated for each household (Vaughn et al. 1997). Evidence and intuition suggest entire tours may provide a more tangible method by which to categorize households, which is an essential component of the methodology we are developing (Kulkarni and McNally 2001).

A cautionary note on the simultaneous approach is that, in defining categories of tours, complex and rare chains must not be excluded, because this will result in an underestimation of the number of trips. For simple tours, this is not a problem. However, for complex tours, which may involve (based on our preliminary analysis of the U.S. National Household Travel Survey (NHTS) reported below) up to 15 stops on a tour, this is an issue. We, therefore, took a two-tiered approach in which we first simulated whether the tour was simple or complex. Then, for each complex tour, we simulated the number and subsequent purpose of each stop on the tour. In essence, this combined elements of both the simultaneous and sequential approaches.

USING THE NHTS AS A SOURCE OF TOURS

We processed the NHTS data into 47,648 tours, starting from 134,400 trips. In comparison, O'Fallon and Sullivan (2004) reduced 124,089 trips to 37,565 tours from the 1997/1998 New Zealand HTS. This represents a reduction to 30.3%, which is very similar to the 35.5% reduction that we achieved.

From the classification into tours, 22 tours were from home to home with no intervening stops. These were presumably for such activities as exercise, walking the dog, or simply where something was forgotten and the trip was abandoned for a return to home. There were 81 tours that included missing purposes. These required additional processing to be usable, but were included in the unclassified group of trips in the following analysis. There were also eight trip chains that did not represent a tour, because they did not start from either home or work and did not end at the place where they started. These may also be erroneous records and will be checked further.

TABLE 5 Count of Tours from NHTS by Simplified Types

Tour description	Count	Percent
Simple home-work	5,700	12.0
Simple education	4,495	9.4
Simple shop	4,301	9.0
Simple social-recreational	5,494	11.5
Simple personal business	1,901	4.0
Simple meal	1,888	4.0
Simple medical/dental	763	1.6
Simple serve passenger	2,783	5.8
Simple other	10	0.0
Complex home-work	7,072	14.8
Complex home-education	2,103	4.4
Complex home-other	9,546	20.0
Unclassified	1,599	3.4
Total	47,655	100.0

We found that the maximum number of intermediate stops in a tour was 14. There were only six tours that had this many stops. The typology of table 4 requires considerable further manipulation of the data to determine how many tours of each type exist in the NHTS data. This manipulation has not yet been done, including splitting out the work-based tours. However, to give an idea of the profile of tours in the NHTS data, table 5 presents a basic count. It is interesting to note, from this preliminary analysis, that simple tours comprise a total of 59.7% of all home-based tours. Of the complex tours, just under one-third involve an initial stop at work or school.

CONCLUSIONS

Previous research, using the 1995 NPTS data and PUMS or the HSF, has shown that simulating household travel characteristics using a Monte Carlo simulation, especially with Bayesian updating with subjective priors, produces reasonable approximations to actual travel characteristics obtained from HTSs. This suggests that this simulation method may be highly productive for generating HTS data, especially under circumstances where budgets do not permit collecting a normal sample size or where needs exist for much larger than usual samples.

This paper presents some shortcomings of the original approach, particularly issues relating to the independent simulation of trips and their characteristics. These appear to be susceptible to mitigation by changing from a trip-based to a tour-based simulation. In light of that, we have also shown that the 2001 NHTS appears suitable as a source of data for a Monte Carlo simulation of household tours. In addition, preliminary analysis of the tour-based data indicates that the NHTS data contain a sufficient number of tours for the distributions, while there is scope for exploring some alternative typologies of tours that may be more useful for simulation. The tour-based approach is also a necessary step for including a geographic simulation of the stop locations. The geographic simulation is essential if the resulting trips are to be loaded on a network, in order to investigate impacts on transport infrastructure, etc.

The NPTS data originally, and now the NHTS data, represent an invaluable resource for this work. The transfer of the distributions to Australia has worked better than expected. Given that there are still relatively few nations that undertake a nationwide travel survey on a periodic or continuing basis, the U.S. nationwide surveys have made an important contribution to this field of research.

REFERENCES

Adler, T. and M. Ben-Akiva. 1979. A Theoretical and Empirical Model of Trip Chaining Behaviour. *Transportation Research, Part B* 13:243–257.

Atherton, T.J. and M.E. Ben-Akiva. 1976. Transferability and Updating of Disaggregate Travel Demand Models. *Transportation Research Record* 794:1–8.

Axhausen, K. 2000. Definition of Movement and Activity for Transport Modelling. *Handbooks in Transport: Transport Modelling.* Edited by D. Hensher and K. Button. Oxford, England: Elsevier.

Badoe, D.A. and E.J. Miller. 1995. Comparison of Alternative Methods for Updating Disaggregate Logit Mode Choice Models. *Transportation Research Record* 1493:90–100.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. *Classification and Regression Trees.* Belmont, CA: Wadsworth International Group.

Chung, J. and K.G. Goulias. 1997. Travel Demand Forecasting Using Microsimulation: Initial Results from a Case Study in Pennsylvania, *Proceedings of the 76th Annual Meetings of the Transportation Research Board,* Washington, DC.

Greaves, S.P. 2000. Simulating Household Travel Survey Data in Metropolitan Areas. Unpublished Ph.D. dissertation, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge.

______. 2001. Local Sample Updates for Synthetic Household Travel Survey Data, *Proceedings of the 23rd Conference of Australian Institutes of Transport Research,* CD-ROM, Monash University, Australia, December 2001.

Greaves, S.P. and P.R. Stopher. 2000. Creating a Simulated Household Travel/Activity Survey: Rationale and Feasibility Analysis. *Transportation Research Record* 1706:82–91.

Kitamura, R., C. Chen, and R.M. Pendyala. 1997. Generation of Synthetic Daily Activity-Travel Patterns, *Proceedings of the 76th Annual Meetings of the Transportation Research Board,* Washington, DC.

Koppelman, F.S., G.K. Kuah, and C.G. Wilmot. 1985. Transfer Model Updating with Disaggregate Data. *Transportation Research Record* 1037:102–107.

Kulkarni, A.A. and M.G. McNally. 2001. A Microsimulation of Daily Activity Patterns, *Proceedings of the 80th Annual Meetings of the Transportation Research Board,* Washington, DC, 2001.

O'Fallon, C. and C. Sullivan. 2004. Trip Chaining: Understanding How New Zealanders Link Their Travel, Research Report to Transfund New Zealand. September.

Pointer, G., P. Stopher, and P. Bullock. 2004. Monte Carlo Simulation of Sydney Household Travel Survey Data with Bayesian Updating Using Different Local Sample Sizes, *Proceedings of the 83rd Annual Meetings of the Transportation Research Board,* Washington, DC, 2004.

Schofer, J.L. 2002. National Household Travel Survey, letter report to the Committee to Review the Bureau of Transportation Statistics Survey Programs, Washington, DC, June 18.

Stopher, P.R., P. Bullock, and J. Rose. 2003. Simulating Household Travel Data in Australia: Adelaide Case Study. *Road and Transport Research* 12(3):29–44.

Stopher, P.R., S.P. Greaves, and P. Bullock. 2001. Synthesizing Household Travel Survey Data: Application to Two Urban Areas, *Proceedings of the 82nd Annual Meetings of the Transportation Research Board,* Washington, DC, 2003.

Strathman, J. and K. Dueker. 1995. Understanding Trip Chaining. *Special Reports on Trips and Vehicle Attributes,* 1990 Report Series. Washington, DC: U.S. Department of Transportation, Federal Highway Administration. February.

Vaughn, K.M., P. Speckman, and E.I. Pas. 1997. Generating Household Activity-Travel Patterns (HATPs) for Synthetic Populations, *Proceedings of the 76th Annual Meetings of the Transportation Research Board,* Washington, DC.

Travel Surveys: Methodological and Technology-Related Considerations

JOY SHARP[1,*]
ELAINE MURAKAMI[2]

[1] Bureau of Transportation Statistics
Research and Innovative Technology Administration
U.S. Department of Transportation
400 Seventh St, SW, Room 4432
Washington, DC 20590

[2] Federal Highway Administration
U.S. Department of Transportation
c/o FTA
915 Second Avenue, Room 3142
Seattle, WA 98174

ABSTRACT

This paper discusses considerations for the next series of personal travel surveys conducted by the Department of Transportation. After a brief discussion of the current National Household Travel Survey (NHTS) design, a broad range of methodological and design considerations are introduced—often in the context of other federal surveys and household travel survey experiences. This paper introduces topics such as whether the current design allows the NHTS to fulfill major objectives of the survey, the efficacy of simultaneous collection of daily and long-distance travel, considerations for improvement of data quality, the need to improve response rates, and the desire to maintain data on travel behavior trends.

BACKGROUND

Objective

The primary objective of this paper is to introduce and discuss survey methodology considerations for the next series of personal travel surveys conducted by the U.S. Department of Transportation (DOT). After a brief discussion of the current National Household Travel Survey (NHTS) design and issues, a broad range of methodological and design consid-

Email addresses:
Corresponding author—joy.sharp@dot.gov
E. Murakami—elaine.murakami@fhwa.dot.gov

KEYWORDS: Travel surveys, survey methodology, data quality, emerging technologies.

erations are introduced, often in the context of other federal surveys and household travel survey experiences. In addition, the following questions are posed that must be carefully considered in designing the next series of personal travel surveys.

- Does the current design allow the NHTS to fulfill major objectives of the survey?
- Can daily and long-distance travel effectively be captured in the same survey effort?
- What changes should be considered to improve the overall quality of the NHTS data?
- What are the most important improvements?
- How might new technologies be incorporated into the design?
- What methods to improve response rates have proved effective in other surveys?
- How can these changes be implemented, yet allow for the ability to monitor travel behavior trends?

History and Current Methodology

The 2001 NHTS combined a daily travel survey, the Nationwide Personal Transportation Survey (NPTS), and a long-distance travel survey, the American Travel Survey (ATS). Both predecessor surveys were last conducted in 1995. The goal in combining the two surveys was to build a more comprehensive picture of household travel while reducing the cost and respondent burden.

The 2001 NHTS primarily employed the 1995 NPTS design with an expanded and more detailed long-distance travel section (i.e., trips of 50 miles or more) added at the end of the interview. The design consisted of a cross-sectional, random-digit dial (RDD) sample of approximately 26,000 households and 60,000 persons nationally, with additional samples in nine states and metropolitan areas.[1] All interviews were conducted via telephone using a two-stage data-collection design. Interviews were conducted over a 14-month period, March 2001 to May 2002, to capture travel throughout the year.

Sampled households (with matched addresses) first received an advance letter with a $5 incentive, followed by a telephone screener interview to collect basic household information, and finally an extended telephone interview to collect trip detail from all household members on their assigned *travel day* and *travel period*. The travel day was preassigned for each household to ensure equal representation among days of the week and across the entire year. The travel period for long-distance travel was defined as the four-week period prior to and including the travel day. Attempts were made to collect travel information on all persons in the household. In order to be considered a completed or useable household interview, interviews had to be obtained from at least 50% of all household adults. Proxy interviews were required for all children under 14 years of age and were allowed, only in very limited situations, for adult household members.

Issues and Constraints

One of the greatest challenges of any statistical survey is producing high-quality, useful data with a limited budget and resources. This will likely be an even greater challenge with the next series of personal travel surveys. In 2002, DOT commissioned the Transportation Research Board and the Committee on National Statistics to review and evaluate the NHTS. The group suggested several improvements that will need to be carefully addressed in the next survey (TRB 2003). In addition, federal statistical surveys are obligated to adhere to the policy and guidelines of external stakeholders, most notably Congress and the Office of Management and Budget (OMB).[2]

In June 2004, OMB in conjunction with the Federal Committee on Statistical Methodology, drafted a revised series of standards and guidelines for all

[1] Five states (Hawaii, Kentucky, New York, Texas, and Wisconsin) and four metropolitan areas (Baltimore, MD; Des Moines, IA; Lancaster, PA; and Oahu, HI) purchased an additional sample for their areas through the NHTS "add-on" program.

[2] The Paperwork Reduction Act of 1995 requires federal agency requests submitted to OMB, ". . . to use effective and efficient statistical methodology appropriate to the purpose for which information is to be collected and directs OMB to develop and oversee the implementation of government-wide policies, principles, standards, and guidelines concerning statistical collection procedures and methods."

federal surveys.[3] Finally, the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA),[4] enacted in December 2002, mandates more stringent procedures pertaining to the collection, protection, and release of federal survey data. This legislation has significant implications for the accessibility of NHTS data and can impact what data are collected.

With the exception of modest changes,[5] the NHTS design remained largely consistent for the collection of daily travel with that of the 1995 NPTS. Thus, the repeated design helped to preserve daily travel trend analysis over time. The long-distance travel component, however, underwent significant changes in definition, content, and methodology as compared with the 1995 ATS. Some key issues faced in conducting the 2001 NHTS and some important considerations for the next series of passenger travel survey(s) follow.

Sample size and methodology significantly changed for the long-distance component. The number of sampled households was reduced from almost 70,000 (1995 ATS) to 26,000 (2001 NHTS). The long-distance trip definition was also revised to include trips of 50 miles or more away from home (as compared with 100 miles in the previous survey). Long-distance trips were collected once for a four-week reference period, as compared with four waves of interviews over a one-year period. These changes resulted in a sample of far fewer long-distance trips, diminished ability to track long-distance travel trends, and difficulty in producing annual and seasonal long-distance travel estimates. In addition, the smaller sample size all but eliminated the ability to produce lower level geographic estimates and analyze travel flows.

Response rate to this multistage telephone survey was 41%. One of the biggest challenges for the 2001 NHTS was obtaining a high response rate, primarily meant to reduce the impact of nonresponse bias likely in surveys with lower response rates. In spite of many efforts—use of incentives, refusal aversion training for interviewers, refusal conversion—the 2001 survey achieved a household response rate of 41%. Given the complexities of the survey and the difficulty of achieving high response rates in an RDD design, the response rate was considerably lower than what is commonly expected in a federal statistical survey and the survey was only reluctantly approved by OMB.

Technology with positional information should be considered in future surveys. Research in the last 10 years suggests that using global positioning systems and, perhaps, cellular phone technology can be effective tools in capturing trips that are often missed using self-reported methods. Expenses incurred for incorporating these technologies have dropped dramatically as hardware costs decline, particularly in light of the Federal Communications Commission directives for positional accuracy for emergency calls (FCC E911). While many benefits may be gained, the decision must also consider the potential tradeoffs in cost, data quality, and statistical reliability.

Consideration of alternative sample design and data-collection methodology. The next series of passenger travel surveys should consider different sampling strategies and data-collection methodologies to address concerns related to coverage, nonresponse, timeliness, and operational efficiency.

SURVEY DESIGN CONSIDERATIONS

Alternative Survey Designs

Survey designs can generally be classified into two broad categories on the basis of whether they obtain repeated measurements on the sample of units over time: panel surveys do and cross-sectional surveys do not. In the United States, most travel surveys rely on one-time cross-sectional designs to collect information on travel consumption and behavior (Tourangeau et al. 1997). The NPTS/NHTS series can be most accurately described as a repeated cross-sectional design, since essentially the same survey is repeated over time with different samples using very similar survey questions and procedures.[6] Given the

[3] Office of Management and Budget, memorandum to the Interagency Council on Statistical Methodology, June 4, 2004.

[4] See http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:pub1347.107.pdf.

[5] See appendix A for a more comprehensive list of changes in survey design.

[6] Modest changes were implemented between data collections to improve overall data quality. These revisions made analyzing trends across years more difficult. See appendix A for a more detailed description of changes.

significant changes in long-distance travel data collection in the 2001 NHTS, the 1995 ATS is more accurately defined as a single, cross-sectional survey. (Although respondents were interviewed four different times, similar to a panel survey, this design was used to pool estimates over a one year period and not analyze the change within households between interviews.) Table 1 describes four common travel survey designs, along with a brief description of advantages and disadvantages, and provides examples of each.

Other countries have transformed many large cross-sectional and panel surveys into continuously repeated surveys for many of the same reasons currently given for transitioning the decennial census "long form" into the continuously collected American Community Survey. Primary reasons include: 1) to flatten the budget for all years, rather than having high peaks during periodic data collection; 2) to retain staff expertise, both in field implementation, data processing, and data analysis; and 3) to improve estimates over time (i.e., national-level data are more timely while retaining the ability to make small area estimates over time). Some examples of national household travel surveys that are conducted continuously include:

- United Kingdom Travel Survey—ranges from 3,500 to 10,000 households per year;
- Household Travel Survey (HTS) for the Greater Metropolitan Region of Sydney, Australia—about 5,000 households per year; and
- German Mobility Panel (MOP), which is a rotating panel.

However, as smaller samples are collected each year, multiple years of data are required for reporting subpopulations, including geographic subregions or subgroups based on sociodemographic or economic characteristics, such as a specific age or income group.

Sample Design

Sample Frame

For the purpose of making objective statistical inference, the sample must be selected using probability methods, that is, where everyone in the target population has a known, non-zero probability of selection (Kish 1965). Likewise, the sample frame utilized in a national probability sample must be complete, accurate, and up-to-date to ensure adequate representation of the larger target population. For telephone surveys, problems with completeness of frame include the growing number of persons who have only a cell phone and who are not included due to costs incurred by potential respondents for incoming as well as outgoing calls, and the small number of households without telephones. For address-based surveys, the completeness of the address list must be evaluated. For example, the Census Bureau has a "Master Address File" that is updated with a U.S. Postal Service Delivery Sequence File. Omissions of certain groups in the sample frame can introduce coverage bias because of the exclusion of these groups.

To the extent that the nontelephone households differ from telephone households in their travel behaviors, coverage bias makes the results less representative of the U.S. population. According to the 2000 Census,[7] the number of households without telephone service was estimated to be 2.4%. While the number of households without telephone service has decreased in the last decade, the number with only a cell phone is rising. A recent analysis using the February 2004 Current Population Survey supplement estimated the number of cell-phone only households to be as high as 6% (Tuckel and O'Neill 2004). In addition, these households were found to be disproportionately single-person, central city, and renters.

Address frames are subject to errors of omission as well. In the pre-test of the 2001 NHTS, it was estimated that the address frame contained addresses for 90% to 95% of households (USDOT 2001).

As a result of continued problems with population coverage (and low response rates,) the future of RDD-only design remains in question. A number of other sample frame options exist, but each introduces other disadvantages related to survey costs, data-collection methodology, response rates, and estimation. Table 2 summarizes the sample frames currently being used by other national surveys, along with some associated benefits and limitations.

[7] See http://www.census.gov/hhes/www/housing/census/historic/phone.html.

TABLE 1 Comparison of Four Designs for Passenger Travel Surveys

	Single cross-sectional	Repeated cross-sectional	Longitudinal panel	Rotating panel
Description	A single sample of households or individuals complete survey during a single period of time	Different samples of households or individuals complete survey for multiple periods of time	Same sample of households or individuals complete survey for multiple periods of time	Same sample of households or individuals complete survey for multiple, specified time periods. Sample gradually, occasionally replaced with new sample (cross between repeated cross-sectional and longitudinal designs)
Advantages	• Provides snapshot of behavior for given time period • Typically provides more representative sample of population of interest • Eliminates potential response bias due to respondent "conditioning" resulting from participating multiple times	• Provides snapshot of behavior for given time periods • Typically provides more representative sample of population of interest • Eliminates potential response bias due to respondent "conditioning" resulting from participating multiple times • Allows for comparisons of population between field periods (assuming similar survey conditions)	• Allows for analysis of change in behavior of same units due to changes in environment and other factors (cause and effect analysis) • Cost and resource efficiencies in subsequent waves from building off of previous interviews	• Allows analysis of change in behavior of same units due to changes in environment and other factors (cause and effect analysis) • Allows for long-term analysis of population change (longer than the duration of a longitudinal study) • Cost and resource efficiencies from building off of previous interviews
Disadvantages	• Does not permit analysis of changes over time; cannot track trends • Operationally, higher costs often incurred to initiate a project; costs and resources not evenly distributed and maintained • Potential "telescoping" effects	• Limited change can be implemented across enumerations to allow for population trend analysis • Does not allow for analyzing behavior and change among same sample units • If not continuously conducted or with small time lag between enumerations, cost and resources not evenly distributed and maintained • Potential "telescoping" effects	• High respondent burden and panel attrition • Response bias due to respondent conditioning • High cost of respondent tracking (e.g., following "movers") • More complicated weighting and estimation • Potential "seam" effects	Though often less than nonrotating panel, still subject to: • High respondent burden and panel attrition • Response bias due to respondent conditioning • High cost of respondent tracking (e.g., following "movers") • Even more complicated weighting and estimation • Potential "seam" effects
Examples	• The American Travel Survey (1995)	• National Household Travel Survey • American Community Survey • UK National Travel Survey • Sydney Household Travel Survey	• Dutch National Mobility Panel • Puget Sound Transportation Panel	• German Mobility Panel • Current Population Survey

TABLE 2 Alternative Sample Frame Designs

	Area frame	List frame	Retired sample (area or list)	Dual frames
Description	Frame consists of smaller geographical units (e.g., county or county clusters) of the country; involves a multi-stage approach to selecting sample units, the first being selected areas	A frame compiled from a listing of the target population unit (typically people, households, or establishments)	Sample frame constructed or borrowed from rotated out or "retired" frames (typically area or list sample frames)	Utilization of more than one frame (e.g., random-digit dial and area)
Advantages	• Often provides more complete coverage of the target population • Efficiencies in (in-person) interviewing because of clustered design	• Often provides more complete coverage of the target population (assuming lists are up-to-date) • List often contains auxiliary information that can be used for sampling or data-collection purposes	• More cost-effective sample frame, because it has already been listed/compiled • Provides already collected unit information that can be used for sampling and data-collection purposes	• Can provide for more cost-effective survey design • Can help to address coverage issues
Disadvantages	• Data-collection costs are typically much more expensive, because sampled households cannot always be interviewed by less expensive data-collection modes, e.g., by mail or telephone. • Precision of resulting survey estimates is often negatively impacted because of clustered (areas) sample design	• Cost and resource limitations sometimes make this option unaffordable • Appropriate, accurate lists do not always exist (or are not accessible)	• Effect of respondent burden and fatigue resulting from prior survey participation • Coverage can become an issue if the list is not updated or new growth in the area is not accounted for	• Adds complexity to sampling, weighting and estimation • Can increase overall data-collection costs (especially when combining with area frame)
Examples	• Current Population Survey (Bureau of Labor Statistics—BLS)	• American Community Survey (Census Bureau)	• 1995 American Travel Survey • 2003 American Time Use Survey (BLS)	• National Survey of America's Families (NSAF) (Urban Institute)

Other sample frame alternatives are currently being researched, yet they pose many coverage issues and operational difficulties for household travel surveys, even as a dual-frame or mixed-mode approach. Internet surveys, for example, offer limited household coverage, because less than half (41.5% from the 2000 Current Population Survey[8]) of U.S. households have access to the internet. Offered as another alternative data-collection mode, they appear to have more merit, but results to date have been mixed with frequent reports of low response and other data quality issues. In addition, the use of cellular telephone sampling frames is also limited, because no comprehensive list of cellular telephone numbers exists.

Sample Size

The number of units to sample is another key decision in the survey design. In choosing the required sample size, these factors should be considered:

- the desired level of precision for the survey estimates (including needed levels of geography and for subgroups);

[8] See http://www.census.gov/prod/2001pubs/p.23-207.pdf.

- the sample design and resulting design effect; and
- factors likely to impact achieved sample size—nonresponse, eligibility rates, and attrition (for panel designs).

With these factors in mind, the 2001 NHTS was designed to achieve interviews from approximately 25,000 households and 60,000 individuals within these households. This resulted in the collection of approximately 250,000 daily trips and 45,000 long-distance trips. This sample size allows for reliable estimation of many national-level trip characteristics for both daily and long-distance travel, but affords very limited estimation for lower levels of geography.[9]

Long-distance trip analysis was greatly affected by the sample size. The 2001 NHTS captured less than 5% of comparable trips captured by the 1995 ATS (i.e., noncommuting trips of 100 miles one way), resulting in the inability to make state estimates or analyze flows between states and major metropolitan areas. In addition, only limited analysis can be performed for rarer transportation modes (e.g., trips by bus and train) and for certain subpopulations (e.g., elderly travelers). The sample size would need to be increased substantially in order to ensure reliability of estimates at lower levels and for rarer groups, as well as to measure flows; however, an increase would be very costly.

Subgroups of Interest

As mentioned above, the NHTS sample size posed limitations for specific analyses due to the small size of the geographic and demographic subgroups. The 2001 NHTS national design included one stratum for metropolitan areas (MSAs) with rail service, but otherwise did not oversample specific, rarer groups. Historically, the NHTS sample design and size has limited the analysis of several important transportation modes and groups. A description of a few of these follows.

Transit users. Personal vehicle use dominates passenger travel and accounts for nearly 9 out of 10 trips in this country. Analysis of lesser used but important transportation modes, such as transit, is limited due to both the small sample and, in some cases, geographic sensitivity. Analysis of transit behavior has relied heavily on decennial census information on the share of transit for "usual mode to work" and the NHTS for transit share for all trips, regardless of purpose. These results are often cited by Congress in decisions related to transit investments. The add-on program in the NHTS that allows states and metropolitan planning organizations to purchase additional samples provides a disproportionate number of cases in certain geographical areas. For example, the New York state add-on sample in the 2001 NHTS also resulted in an unintended benefit. The New York metropolitan area has nearly 40% of the U.S. transit market based on the Federal Transit Administration National Transit Database. The addition of the New York add-on sample allowed for a unique analysis of transit behavior in an area representing the largest share of transit use. The added sample, reweighted for national estimation, also provided modest gains in precision for transit estimates at the national level.

Households without vehicles. Data from the 2001 NHTS showed that approximately 8% of households were without a vehicle for regular use. In the United States, households without vehicles are thought to include two main groups: 1) people who live in high-density urban neighborhoods (e.g., Manhattan, downtown Chicago, San Francisco, or Washington, DC) with good transit and taxi accessibility, and 2) recent immigrants who have not acquired a car (Murakami 2003; Pisarski 1996). Those in the second group may be less likely to participate in a national travel survey for several reasons: language barriers, potential distrust of government activities, and different concerns about privacy and confidentiality.

Race and ethnic groups. African Americans have been less likely to respond to travel surveys than white households. Little documentation is available on participation rates by Hispanic households in transportation surveys. However, larger households are more likely to fail to complete travel surveys as

[9] Reliable aggregated estimates can be made for the nine census regions (and divisions based on metropolitan area size and the presence of rail). An exception to limited geographical analysis is the selected add-on areas; nine states and metropolitan areas purchased additional samples for their areas to obtain a more reliable estimation.

the burden of reporting is greater, and this finding may affect participation by Hispanic households (Murakami and Watterson 1992; DRCOG and Parsons 2000; Nustats 2003, 2004a, and 2004b).

Contrino and Liss (2004) conducted research on nonresponders in three regional surveys (Atlanta, Phoenix, and Ohio) and found that, in all three areas, minority and low income groups were more likely to be nonrespondents. As a result, Contrino and Liss recommend oversampling for low responding and special interest population groups and using targeted approaches for recruitment and retrieval, as well as the use of post-stratification weights to adjust for low participation. Oversampling strategies, if effectively employed, can be a valuable mechanism for producing more reliable estimates for specific subgroup analysis. However, assuming the overall sample size remains constant, it can also result in a loss of precision for national estimates. In addition, while oversampling may reduce the variance of these estimates, it does not necessarily reduce the potential nonresponse bias.

Household vs. Within-Household Sampling

Many transportation analysts and modelers require data from all members of households for use in their analysis and models in order to capture joint decisions and household interactions. Consequently, attempts were made in the NHTS to interview all household members, and only households where at least half the household members were interviewed were considered complete or useable. Approximately 85% of NHTS useable households resulted in a complete enumeration of the household members. These households were provided with an additional set of weights to allow researchers the choice of using only those households with complete enumeration. Complete enumeration of a household is a challenging task and has negative impacts on the response rate. In the last several years, transportation researchers (Erhardt 2000) have begun to investigate whether changing the sampling unit to a person, rather than a household would improve response while maintaining the ability to simulate travel for a household.

DATA-COLLECTION METHODOLOGY

Mode of Data Collection

Another key decision in travel surveys is the mode of data collection. Similar to sample design, each choice of data-collection mode has inherent advantages and disadvantages. Selection of an appropriate mode requires careful consideration of many factors, not the least of which is coverage of the target population. While the method of data collection might be largely dictated by the population coverage and sample frame, other common determinants include survey costs, response rates, and data quality issues. Mode selection can also be influenced by the complexity and length of the survey and timeliness needs.

Table 3 provides a summary of four popular data-collection modes: in-person, telephone, mail, and internet, along with their features. As this table illustrates, in-person data collection typically yields the most complete coverage, achieves the highest response rate, and produces the best quality data. Not surprisingly, in-person interviews are also the most expensive of the four modes. For this reason, telephone and mail modes are more commonly used despite well-recognized tradeoffs in data quality. Telephone interviews have been the most commonly used collection method in the United States over the past 20 years, as field costs for personal visits increased to prohibitive levels and other obstacles to personal interviews have arisen (e.g., personal security and gated communities).

The 2001 NHTS used a telephone data-collection methodology. Telephone interviews are often preferred over mail-back methods for travel/activity surveys as they allow for more probing for complete reporting of trips. In addition, because the NHTS captures travel by all household members, telephone retrieval allows for correction and validation of travel among household members.

Mixed-mode approaches are commonly used to strike a balance between survey costs and data-quality issues (most often response and coverage). A commonly used approach, taken in the American Community Survey (ACS), is to employ the least costly mode for initial contact, followed by a more costly mode for nonresponse followup, such as using a mail survey with telephone nonresponse or a

TABLE 3 Description and Comparison of Data-Collection Methodologies

	In-person	Telephone	Mail	Internet
Description	Interviewer travels to respondent's home or office and administers questions in face-to-face interview	Interviewer contacts respondent and administers questions over the telephone	Questionnaire mailed to respondent and returned by mail, or data are retrieved by telephone	Respondent completes survey on web
Coverage	Most complete	Omits nontelephone households	Similar to in-person depending on how the addresses were obtained	Only households with internet connection or access to internet
Response rate	Highest of all modes	Intermediate	Among the lowest	Among the lowest
Data quality	Highest of all modes	Intermediate	Lowest of all modes	Intermediate; mixed results
Cost	Most expensive (this often leads to geographically clustered sample cases and a reduction in the effective sample size)	Intermediate	Among least expensive	Among least expensive; has a high startup cost compared with data-collection cost

telephone survey with in-person nonresponse followup. In dual-frame designs aimed at improving coverage, different modes are often required to capture the sampled units from each frame. While a mixed-mode approach can offer an effective mechanism for improving response and coverage, it also potentially introduces bias resulting from mode effects (i.e., a difference in responses due entirely to the method of data collection). Therefore, it is important to first evaluate the tradeoff of improved coverage and response with potential response error (bias) before deciding on a mixed-mode methodology. Recognizing the greater likelihood that future surveys, including the NHTS, will need to allow for multiple modes of data retrieval, appropriate research on modal influences on travel behavior data collections will be needed.

Electronic or computerized data-collection options are also commonly used for the entirety of the interviews or as an alternative methodology. The NHTS used a computer-assisted telephone interview (CATI) with an additional option for reporting specific information via the internet. Use of CATI was especially beneficial for the NHTS, allowing for trip rostering and capture of trips made by multiple household members. Therefore, trips already capture during preceding interviews could be verified by a subsequent household member instead of captured anew. Computer-assisted designs require more upfront planning and increased time to implement compared with paper-and-pencil surveys, but result in faster access to data and higher quality control as the need for data entry from paper forms is eliminated. In the future, the use of CATI, whether by telephone or in-person, presents opportunities for better integration of geographic information and other location-based data.

Nonresponse Minimizing Techniques

Nonresponse is not just an issue for the NHTS. In June 2004, Robert Groves presented an overview of the status of current household nonresponse to the Committee on National Statistics (Groves 2004). He found declining response rates in attributes that match those of the NHTS (i.e., surveys conducted in the developed world, one-time surveys compared with longitudinal surveys, and telephone surveys compared with in-person surveys). He listed techniques to reduce nonresponse, including prepaid incentives, cash incentives, more followup calls, and longer data-collection periods.

Many efforts were implemented in the 2001 NHTS design to achieve as high a response rate as possible. For example, respondents were sent incentives prior to contact. Interviewers were provided with special refusal aversion training, and refusal conversion efforts were attempted with survey nonresponders. The following are some factors thought to contribute to nonresponse in the NHTS:

- multistage telephone data collection (RDD sample design) with no nonresponse followup;
- short data-collection window (interviews were allowed up to six days);
- attempts to enumerate all household members—strict requirements for interviews (at least 50% of adult household members had to be interviewed before a case was considered complete);
- limited reference period of travel;
- limited proxy allowance; and
- interviewer assignments.

Stages of Contact to Complete Interview

Most cross-sectional household travel surveys utilize a multistage approach for interviewing households about their travel. As with the NHTS, advance letters introducing the survey were sent first, followed by a telephone contact to conduct a basic household-level screener interview. Respondents were mailed a travel diary, with information retrieved by another telephone interview. In any multistage approach, nonresponse can occur at each stage and compounds the overall nonresponse rate. In the 2001 NHTS, the recruitment rate was 58.2%, and the subsequent completion of the extended survey was 70.8%. The composite response rate was 41.2% (USDOT 2004).

Nonresponse Followup Studies

The 2001 NHTS did not include a nonresponse followup survey. Traditionally, low response rates have been suspected of resulting in more biased results. However, more recent research (Groves 2004) also cites examples of surveys with high nonresponse rates and low bias, and interestingly, some attempts to reduce nonresponse that resulted in greater bias.

Nonresponse followup (NRFU) surveys have not yet been incorporated as standard practice in activity and travel behavior surveys. Some exceptions include work done by Richardson (2003) in Australia and a small test funded by the Federal Highway Administration (FHWA) in Denver in the late 1990s. The Victorian Activity and Travel Survey in Melbourne, Australia, conducted in-home interviews with a sample of nonrespondents to the main mail-back survey. This study indicated that nonrespondents to mail-back surveys were more like early respondents than late respondents in daily trip rates.

In the Denver Regional Council of Governments project, a brief mail-out/mail-back survey was conducted for nonrespondents to an RDD telephone survey (for those where an address could be found). Small cash incentives were found to double the response rate to the NRFU survey. They did not find statistically significant differences in household trip rates between the households who completed the full survey compared with those in the "quick refusal" and "noncontact" households. Therefore, the hypothesis that these nonrespondents to the telephone survey led to underreporting of trips was not supported.

For the ACS, the U.S. Census Bureau found that response rates varied widely, with particularly low mail-back responses in neighborhoods that were predominantly Native American (17%), Hispanic (34%), and African American (35%) (USDOC 2002) . The original plan for the ACS nonresponse followup, which was tested in their pilot, was a one in three field followup. The response rate to the field followup has been uniformly very high (between 92% and 95%). The Census Bureau now plans to implement differential nonresponse followup, with higher followup rates in areas with low mail-back returns (USDOC 2004).

Data-Collection Window

Current NHTS methodology requires that an interview be completed from a respondent within six days of the assigned travel day. For respondents who neglect to complete a diary, recall errors are felt to be much higher after six days, especially for daily travel. The limited six-day window also eliminates the confusion of referencing the particular travel day that was assigned (e.g., Tuesday this week as opposed Tuesday last week). Although the six-day data-collection period appears to help in reducing response problems, it could also potentially contribute to nonresponse, especially if attempts are made to interview all household members within this relatively small window. In addition, potential bias may also be introduced in the capture of long-distance travel. Respondents who travel often and are away from home for longer periods of time are more likely not to respond.

Reference Period

Daily travel. As previously mentioned, the reference period for the travel day in the NHTS is a pre-assigned one-day period (from 4:00 a.m. on the travel day through 4:00 a.m. the following day). Other national travel surveys, such as the U.K. National Travel Survey and the German Mobility Panel use a seven-day diary. These surveys allow for examination of travel variability over a longer period. For example, a respondent may not go grocery shopping each day but only once a week. Similarly, a respondent may ride transit only two days a week. These longer reference periods, however, are more burdensome, typically achieve very low participation rates, and may result in fewer trips reported each day as the survey period continues. The Dutch National Mobility Panel found significant "trip reporting fatigue" in a seven-day diary (Golob and Meurs 1986).

Cost efficiency might suggest that a smaller sample with a larger reference period should be considered in order to continue generating similar numbers of trips overall. However, moving from a larger sample size with a one-day reference period to a smaller sample with a longer reference period would create additional estimation issues for lower geographic levels and subgroups. Although we might have the same number of trips, the effective sample size would be lower given the increased correlation between trip reports. Given current criticisms, reducing the household sample size—thus requiring more aggregation on characteristics and even more limited analytic potential—would not likely be perceived as an improvement.

Long-distance travel. The NHTS reference period for long-distance travel was the four-week period before and including the travel day. Therefore, if a respondent's travel day was July 30, their assigned travel period would be July 3 to July 30. This brings into question the respondent's ability to accurately recall trips for this period, and telescoping effects are potentially introduced (i.e., they might be reporting trips taken outside the travel period, e.g. on July 1 to 2). Due to the rotating nature of the travel period, it further introduces difficulties in producing seasonal and annual estimates of long-distance travel. Introducing longer, more salient reference periods, however, can also be problematic. For example, given that the respondent is interviewed only once, it is unlikely that he or she would be able to accurately recall all trips for one year, not to mention the burden of this request.

In the 1995 ATS design, respondents were interviewed quarterly over a one-year period. This methodology allowed for bounding and dependent recall of previous trips, thus reducing telescoping effects. As is common in panel designs, however, time-in-panel or conditioning effects were also evidenced by the declining trip rates in later waves of interviewing.

Proxy Allowance and Effects

Only limited proxy reporting for adult household members was allowed in the NHTS, resulting in approximately 80% of the interviews being conducted with the respondent. Self-reports are preferred in travel surveys due to diminished accuracy and completeness in trip reporting often experienced when proxy reporting is allowed. In one travel survey conducted in Toronto, researchers found that home-based discretionary and nonhome-based trips were underreported by proxy, with gender a related factor (Badoe and Steuart 2002). Bose and Giesbrecht (2004) found in the 2001 NHTS that average trip rates for persons interviewed by proxy were much lower than those who reported for themselves. The average daily trip rate was 4.5 for self-reports as compared with 3.7 for proxy reports. Proxy reports were more likely in the NHTS if the respondent was male, a nondriver, had less education, was away from home on the travel day, or had a disability that affects travel. Proxy reports also tended to have fewer daily, long-distance, walk, and bike trips, and transit usage.

Interviewer Assignments

Broeg and Ampt's (1983) continuing work assigns "caseloads" to individual interviewers or "motivators." Their methods rely primarily on mail-out/mail-back techniques, with telephone calls used for reminders and for queries when responses are missing or other problems exist. Because the 2001 NHTS used telephone retrieval, it was nearly impossible to assign cases to an individual as call-backs were scheduled over many different hours and different days of the week. A small test was completed in 2002 in the Washington, DC, region and a small team was assigned a caseload. The survey period was very compressed, and thus results were inconclusive but seemed positive (Freedman and Machado 2003). In debriefing the interviewers, the interviewers felt more confident and comfortable when making subsequent phone calls. Some respondents said they wanted the first caller to call them back, not someone else on the team. One of the drawbacks was that the current scheduling software was not optimized for team assignments.

QUESTIONNAIRE DESIGN

Improvements in questionnaire design should be made to assist the respondent in providing complete and accurate information that the analyst is attempting to collect. The main issue for travel surveys is to ensure that all trips are reported, otherwise, they become a serious problem of item nonresponse. Techniques that have been used to improve reporting include:

- clear definitions of what travel is to be captured,
- different diary designs, and
- different approaches to capture trips, activities, and time use.

Daily trips. A daily trip in the NHTS was defined as each time the respondent went from one address to another. One of the greatest difficulties in travel surveys is to capture short stops, because people may neglect to report them for several reasons:

- reporting burden—every stop generates a series of questions that make phone retrieval time longer;
- they are considered incidental and, therefore, the respondent assumes that the researcher is not interested in knowing about them; and
- easy to forget (e.g., pick up milk, cigarettes).

Long-distance trips. In contrast, people in the United States do not have a good estimate of distance, so questions about trips of over 50 miles in length are often overreported (in spite of interviewer aids such as maps). Respondents will often report trips that are closer to home than the long-distance definition. In the processing of both the 1995 ATS and the 2001 NHTS, 20% to 25% of long-distance trips were later excluded after the calculated route distance illustrated that these trips were under the specified mileage required (i.e., 100 miles for the 1995 ATS and 50 miles for the 2001 NHTS). Current issues facing the next collection of long-distance data include trip length criteria, the amount and type of detail most important, and how to define and collect trips or journeys with multiple stops and/or side trips.

Diaries and Recall Aids

Diary formats have evolved primarily through focus group testing. Some of the questions on the visual appearance and format of travel/activity diaries include:

- Should answers be open-ended or fixed? If fixed, check boxes have been generally applied. Some diaries ask people to read a code list and to enter the code in each box.
- Should the diary primarily be a "memory jogger" and include only some of the questions about each trip, or should it include all questions that will be asked in the CATI retrieval? NHTS is more like a memory jogger, because some questions are asked on the phone that are not included on the diary form.
- If respondents will not read instructions on how to complete the diary, what can we do to help them complete it to meet our needs? The 2001 NHTS added a pictogram showing "activities" and "trips" as an example.

In addition, NHTS respondents were also mailed a map delineating a 50-mile radius from their home location. Although the map was somewhat misleading, it was thought to have served as an effective memory jogger.

A couple of pilot tests have been conducted (Bachu et al. 2001; Stopher et al. 2004) using passive global positioning systems (GPS) and then supplying a map to respondents to use as a recall device. They found that people were able to recount their trips, even two weeks later by looking at the printed maps showing their GPS-recorded travel. This approach relied on the map-reading ability of respondents. In the Australian pilot (Stopher et al. 2004), respondents were also given the option of looking at a tabular description of each stop showing street names, arrival times, and travel times.

Trip-Based vs. Activity-Based (Time-Use) Surveys

During the last couple of decades, travel behavior researchers have focused increasingly on activity-based travel survey approaches. (One of the first known uses of an activity diary for travel research was in Belgium in 1986–1987.) This is due in part to the desire to understand travel in the context of daily activities and allows analysts to bring this context into travel analysis and modeling. Activity approaches allow transportation researchers to examine the activities and relationships that generate the need for travel (Harvey 2003). Traditional trip-based travel surveys, such as the NHTS, enumerate all trips taken by persons during a specified time period, followed by the collection of trip detail that typically includes items such as origin and destination, time, purpose, and mode. Activity-based travel surveys, on the other hand, collect all activities undertaken by the respondent in the given time period. Trips are captured as just another activity. Much of the trip detail is not asked directly but is inherent to the activity diary structure and can be derived. Harvey's review of approximately 10 activity surveys showed that travel accounts for approximately 19% of reported activities.

Time-use surveys recently conducted in Europe found a lower proportion of persons who were "immobile," that is, not traveling on a given day, compared with a travel survey. For the French, 8% were immobile in the time-use survey compared with 17% in a travel survey (Armoogum et al. 2004). One hypothesis is that the reporting of "no trips" in a travel survey is that the answer is given as a "soft refusal." Now that the American Time Use Survey data are available, it is important for transportation analysts to do a similar comparison.

While some preliminary comparisons between trip-based and activity-based surveys have been performed, additional research is still needed. Important measures—trip rates and trip frequency distributions—should be analyzed across survey types while controlling for survey conditions. While some research has shown that activity surveys produce better data quality, it is still unclear what tradeoffs there may be between quality and cost (Pendyala 2003).

NEWER TECHNOLOGY

GPS

Over the last several years, several passenger travel surveys have introduced multiple approaches for integrating GPS into travel surveys. Most commonly these have included vehicle-based passive surveys, person-based passive surveys, and vehicle-based interactive surveys. Original benefits were expected to include reduction of missing (unreported) trips, improved accuracy of travel distance and time, routing and speed data previously unobtainable, and ability to capture longer periods of travel.

Between 2001 and 2004, several regional household travel/activity surveys incorporated a GPS component as a subsample to their household dairy (Wolf 2004). The ability to capture unreported trips (item nonresponse) has ranged widely from 20% to 80%. Zmud and Wolf (2003) found that unreported trips were most likely to be trips of less than 10 minutes. Household characteristics leading to less complete reporting include having three or more vehicles, three or more workers, an annual income below $50,000, and persons younger than 25 years.

One of the most exciting passive GPS studies is the Commute Atlanta project (an FHWA value pricing pilot) with 365 days of 1-second GPS data for over 450 household vehicles. This long period of data collection allows for examination of the variability of travel and a better understanding of long-distance trips made by private vehicles. A "sunset clause," in which the data must be destroyed within six months of the end of the project, is one major handicap of this project.

How a GPS component could be incorporated into a national survey raises many questions, because completeness of responses in self-reported diaries, compared with GPS-recorded information, may be linked to demographic characteristics (e.g., English-language capability and education) and metropolitan characteristics (e.g., population size, density, and transportation network complexity).

Cell Phones

Several tests have been conducted to trace personal movements using cellular/mobile phones. Some advantages of mobile phones relative to a GPS system are that they function underground and inside buildings more often, and the density of cellular base stations/towers is higher in the densest urban areas. Also, the market penetration of cellular phones is very high, so the cost of equipment is low.

In Germany, Wermuth et al. (2003) tested tracking of cellular phones for a long-distance survey. Recently, Kracht (2004), also in Germany, began testing cellular phone use for tracking daily personal travel, especially as many phones already have the capability of recording and storing the position (cell) over time. The positional accuracy afforded by cellular phones is not as good as using GPS, so while gross measurements of distance and travel time are achievable, specific routes or travel modes are less likely to be determined without greater respondent interface.

Web Utilities

Using the internet as a response method is becoming more robust; however, it has more often been applied to shorter origin/destination travel surveys, for example, after license plate capture with a mail-out postcard and responses allowed either by mail-back or the internet. One test using only internet responses for household travel diaries, completed by the Resource Systems Group (2002), showed higher interest among older men with higher incomes and young men. In a regional test, the ability to incorporate a pre-geocoded electronic yellow pages was an advantage for selecting destinations. However, in the 2002 National Transportation Availability and Use Survey conducted by the Bureau of Transportation Statistics, only 3 out of over 5,000 respondents chose to complete the survey using the web option.

FUTURE CONSIDERATIONS

Preserving Trends

Obviously, any changes made to the design of the future NHTS surveys has an obvious impact on the ability to monitor travel behavior trends. Any change will need to be carefully weighed and tradeoffs made between needed improvements and the continued ability to track trends over time. Although minimal changes were made to the capture of daily travel, certain changes were introduced that obscured the ability to detect what was an actual change in travel behavior versus a change due solely to methodological or definitional differences. For example, additional probes were added to the 2001 NHTS to better capture more incidental types of trips thought to be underreported, such as walk and bike trips. As a result, the number of walking trips reported increased significantly from 1995 to 2001. Due to this change, it is not possible to discern what was a real change in walking behavior as compared with the improved capture of walking trips. As previously described, the substantial changes in the collection of long-distance travel data severely limited the ability for longer distance trend analysis.

CONCLUSIONS

Despite methodological hurdles, data from surveys on personal travel in the United States are a valuable commodity. Over time, the community of data users has grown and the application of these data in numerous studies has increased exponentially, especially as data accessibility has increased (ORNL 2004). Understanding who is traveling, how much they are traveling, and why they are traveling assists decisions on transportation investments and the potential implications of these investments on different communities.

Due to the disparate objectives and uses of the daily and long-distance travel components, current plans call for the next series of travel surveys to be separated again into two different data-collection efforts. Continued outreach and participation with the data user community to better understand data needs and anticipate emerging needs will be necessary for their successful design. Future surveys could greatly benefit from shared methodological improvements, including the following areas identified as priorities:

- **Improved response rates.** Given new requirements for response rates in federal surveys and issues related to nonresponse bias, additional efforts and methodologies will need to be employed in the next surveys to elevate response rates to more acceptable levels. Particular attention also needs to be given to those subpopulations with disproportionately higher response rates and that are of particular interest to transportation researchers and planners because of differential travel needs and behavior (e.g., African American and Hispanic households, and younger, more mobile persons). This may call for tests to adjust recruitment mechanisms to further promote interest and increase credibility among these subpopulations.
- **Nonresponse followup study and bias assessment.** A concurrent NRFU study is needed to improve response rates, monitor and assess potential bias, and potentially adjust final estimates.
- **Modified sample design/frame.** Due to increasing coverage problems caused by nontelephone households and cell-phone-only households, the future of RDD samples as a sole sample frame is not a likely option for travel surveys conducted by DOT. Alternative frames or dual-frame approaches will need to be further explored and implemented. Since nonresponse is also impacted by the sample frame, any new sample design should also weigh the potential for increasing response rates as well.

The next long-distance survey, in particular, will need careful consideration of its design. The disparate designs employed for the 1995 and 2001 surveys resulted in inconsistent definitions and trip characteristics, limited data utility, and diminished ability to monitor longer travel behavior. The next long-distance travel survey needs to be developed with a sustainable design, so that data users can come to expect a more useful and consistent product that allows for monitoring of long-distance travel behavior trends. In addition, the design needs to focus on two critical areas:

1. **Larger sample size.** One of the most critical needs identified for long-distance travel data is a sample size sufficient for analysis of smaller geographic analysis and flows between states and larger metropolitan areas.
2. **Research into appropriate reference periods.** Estimation and reliability were affected by the four-week reference period for long-distance travel in the NHTS. The next survey needs to ensure better capture of trips throughout the year, and further research is needed to ensure the appropriate reference period(s), especially considering the length and type of long-distance trip.

The NHTS/NPTS series, as a whole, has provided the daily travel data user community with a fairly consistent, useful product for nearly 30 years. However, further improvements are still needed to expand on its utility and better respond to changes in transportation, the population in general, and technology. Below are a few areas considered priorities in the next daily travel survey:

- **Improved coverage of rarer, important modes of travel.** Personal vehicle travel dominates daily travel and rarer modes, such as transit, are often not captured well enough to allow for reliable estimates to be made on its use and characteristics. Sampling strategies and data-collection methodologies need to be considered that would allow for better estimation of these modes without unduly compromising other survey estimates.
- **Reduce trip underreporting.** It is still unclear how accurately or comprehensively the NHTS captures U.S. daily travel. Research using other survey designs and technologies suggests that the NHTS may still be missing trips, especially those involving nonhome-based, more incidental travel. Incorporating GPS technology, at least as a subsample, will allow analysts to more fully assess and validate the comprehensiveness of trip capture, the accuracy of trip characteristics (e.g., time and distance), and may also allow for adjustments of travel estimates. In addition, other external data sources are currently becoming

available (e.g., the 2003 American Time Use Survey), which will allow for further comparison and validation of trip estimates and may also suggest areas where additional methodological improvements can be made.

REFERENCES

Armoogum, J., K. Axhausen, J.-P. Hubert, and J.L. Madre. 2004. Immobility and Mobility Seen Through Trip-Based vs. Time-Use Surveys, presented at the 7th International Conference on Travel Survey Methods, Playa Herradurra, Costa Rica, August 2004.

Bachu, P.K., T. Dudala, and S.M. Kothuri. 2001. Prompted Recall in Global Positioning System Survey, Proof of Concept Study. *Transportation Research Record* 1768: 106–113.

Badoe, D.A. and G.N. Steuart. 2002. Impact of Interviewing by Proxy in Travel Survey Conducted by Telephone. *Journal of Advanced Transportation* 6(1):43–62, Winter.

Bose, J. and L. Giesbrecht. 2004. Impact of Proxy Reporting in the 2001 National Household Travel Survey, presented at the Joint Statistical Meetings, Toronto, Canada, August 2004.

Broeg, W. and E. Ampt. 1983. State of the Art in the Collection of Travel Behavior Data: Workshop on Data Needs and Collection. *Transportation Research Board Special Report 201*. Washington DC: Transportation Research Board. 48–62

Contrino, H. and S. Liss. 2004. Sources and Impacts of Nonresponse in Household Travel Surveys: Three Case Studies, presented at the 7th International Conference on Travel Survey Methods, Playa Herradurra, Costa Rica, August.

Denver Regional Council of Governments (DRCOG) and Parsons Transportation Group. 2000. Travel Behavior Inventory: Decribing and Reaching Nonresponding Populations, report for the Federal Highway Administration, U.S. Department of Transportation. June.

Erhardt, G.D. 2000. Validation of Trip Production Rates for Synthesized Households, draft paper, Cornell University. January.

Freedman, M. and J. Machado. 2003. Pilot Study of Modified CATI Interview Technique, draft proposed for NCHRP Project 08-37, Standardized Procedures for Household Travel Surveys. July.

Golob, T.F. and H. Meurs. 1986. Biases in Response Over Time in a Seven-Day Travel Diary. *Transportation* 13:163–181.

Groves, R.M. 2004. Household Survey Nonresponse: Research Relevant to Statistical Policies, presented at the 94th Meeting of the Committee on National Statistics: Household Survey Nonresponse: What Do We Know? What Can We Do? Washington, DC, June 18, 2004.

Harvey, A.S. 2003. Time-Space Diaries: Merging Traditions. *Transport Survey Quality and Innovation*. Edited by P. Stopher and P. Jones. Amsterdam, The Netherlands: Pergamon Press.

Kish, L.F. 1965. *Survey Sampling*. New York, NY: John Wiley and Sons, Ltd.

Kracht, M. 2004. Tracking and Interviewing Individuals with GPS and GSM Technology on Mobile Electronic Devices, presented at the 7th International Conference on Travel Survey Methods, Playa Herradurra, Costa Rica, August 2004.

Murakami, E. 2003. Households without Vehicles, 2000. *CTPP Status Report*, January. Available at http://www.fhwa.dot.gov/ctpp/status.htm, as of October 2004.

Murakami, E. and W.T. Watterson. 1992. The Puget Sound Transportation Panel after Two Waves. *Transportation* 19:141–158.

NuStats. 2003. BCD-COG Household Travel Behavior Survey: Final Report for the Berkeley-Charels-Dorchester Council of Governments. September.

______. 2004a. Goldsboro Urbanized Area Travel Survey: Final Report for the City of Goldsboro. January.

______. 2004b. Kansas City Regional Household Travel Survey: Final Report for the Mid-American Regional Council. June.

Oak Ridge National Laboratory (ORNL). 2004. Citations and Uses of NPTS/NHTS Data. Available at http://nhts.ornl.gov/2001/pub/NHTS_Citations.pdf, as of Sept. 27, 2005.

Pendyala, R. 2003. Quality and Innovation in Time Use and Activity Surveys. *Transport Survey Quality and Innovation*. Edited by P.R. Stopher and P. Jones. Amsterdam, The Netherlands: Pergamon Press.

Pisarski, A. 1996. *Commuting in America II*. Lansdowne, VA: Eno Transportation Foundation.

Resource Systems Group. 2002. SBIR Phase II Final Report: Computer-Based Intelligent Travel Survey System, Report for the Federal Highway Administration. October. Available at http://www.fhwa.dot.gov/ohim/trb/sbir/sbir.htm, as of Oct. 18, 2004.

Richardson, A.J. 2003. Behavioral Mechanisms of Nonresponse in Mail-Back Travel Surveys. *Transportation Research Record* 1855:191–199.

Stopher, P., A. Collins, and P. Bullock. 2004. GPS Surveys and the Internet, presented at the 27th Australasian Transport Research Forum, Adelaide, Australia, Sept. 29–Oct. 1, 2004.

Transportation Research Board (TRB). 2003. *Measuring Personal Travel and Goods Movement: A Review of the Bureau of Transportation Statistics' Surveys*, Special Report 277. Washington, DC: National Research Council.

Tourangeau, R., M. Zimowski, and R. Ghadialy. 1997. *Introduction to Panel Surveys in Transportation Studies*, Travel Model Improvement Program, DOT-T-98-3. Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Tuckel, P. and H. O'Neill. 2004. Ownership and Usage Patterns of Cell Phones: 2000–2004, presented at the Annual Meeting of American Association for Public Opinion Research, Phoenix, Arizona, May 13–16, 2003.

U.S. Department of Commerce (USDOC), U.S. Census Bureau. 2002. Meeting 21st Century Demographic Data Needs: Implementing the American Community Survey—Report 2: Demonstrating Survey Quality. Washington, DC. May. Available at http://www.census.gov/acs/www/AdvMeth/Reports.htm, as of August 2004.

______. 2004. *Meeting 21st Century Demographic Data Needs: Implementing the American Community Survey—Report 6: The 2001–2002 Operational Feasibility Report of the American Community Survey.* Washington, DC. 2004. Available at http://www.census.gov/acs/www/AdvMeth/Reports.htm, as of August 2004.

U.S. Department of Transportation (USDOT), Federal Highway Administration. 2001. Issues in the 2001 Address vs. List-Assisted Design, unpublished document.

______. 2004. *National Household Travel Survey User's Guide, January 2004 (Version 3) (National Sample with Add-Ons),* p. 4-3. Available at http://nhts.ornl.gov/2001/usersguide/index.shtml, as of December 2005.

Wermuth, M., C. Sommer, and M. Kreitz. 2003. Impact of New Technologies in Travel Surveys. *Transport Survey Quality and Innovation*. Edited by P. Stopher and P. Jones. Amsterdam, The Netherlands: Pergamon Press.

Wolf, J. 2004. Application of New Technologies in Travel Surveys, presented at the 7th International Conference on Travel Survey Methods, Playa Herradurra, Costa Rica, August 2004.

Zmud, J. and J. Wolf. 2003. Identifying the Correlates of Trip Misreporting: Results from the California Statewide Household Travel Survey GPS Study, presented at the International Conference on Travel Behaviour Research, Lucerne, Switzerland, 2003.

APPENDIX A

The following is a list of improvements and changes in the 2001 NHTS as compared with the 1995 NPTS:

1. travel taken by persons younger than five years old are enumerated;
2. more emphasis on walk and bike trips by prompting specifically for these trips;
3. for respondents not traveling on the designated travel day, information on the most recent trip was collected;
4. for respondents not traveling during the designated travel period (the 28-day period when data on long-distance trips were collected), information on the most recent long-distance trip was collected;
5. information on access and egress to the transit station was explicitly collected;
6. the travel period was a 28-day period in 2001 but a 2-week period in 1995; and
7. long-distance trips in 2001 were those with the farthest destination 50 miles away from home while the criterion in 1995 was 75 miles.

In addition, a number of questions were added to (or data elements later derived from) the 2001 NHTS to cover emerging trends pertinent to personal travel behavior.

At the household level:

- cell phone ownership;
- number of phone lines owned and how they were used (voice, fax, modem); and
- vehicle fuel consumption and annual fuel costs.

At the person level:

- internet access and frequency and location of use;
- travel disability and its effect on mobility;
- primary activities during the "last week;"
- explicit coding of multiple jobs;
- broad categories of occupation;
- immigrant status; and
- frequency of walk and bike trips during the week prior to the interview day.

At the individual daily trip level:

- more detail on trip purpose; and
- access and egress modes to transit stations.

At the individual long-distance trip level:

- access and egress modes to airport, train station, etc.;
- overnight stops and purpose of stops; and
- all modes used at the final destination.

REVIEWERS for 2005

A

Mohamed Abdel-Aty	University of Central Florida, Orlando, FL
William Anderson	Boston University, Boston, MA
Lisa Aultman-Hall	University of Connecticut, Storrs, CT
Kay Axhausen	Institute for Transport Planning, Zurich, Switzerland

B

David Banks	Duke University, Durham, NC
Jonaki Bose	Research and Innovative Technology Administration, USDOT, Washington, DC
George Bresson	Université Panthéon-Assas Paris II, Paris, France
Stacey Bricka	NuStats, Austin, TX
Martijn Brons	Free University of Amsterdam, Netherlands
Michael Bronzini	George Mason University, Fairfax, VA
John Bukowski	Federal Highway Administration, USDOT, Washington, DC
Joe Button	Texas A&M University System, College Station, TX

C

Amrut Champaneri	Research and Innovative Technology Administration, USDOT, Washington, DC
Dan Citrenbaum	Federal Aviation Administration, USDOT, Washington, DC
Kelly Clifton	University of Maryland, College Park, MD
Timothy Coburn	Abilene Christian University, Abilene, TX
Michael Cohen	Research and Innovative Technology Administration, USDOT, Washington, DC
Steve Cohen	Bureau of Labor Statistics, U.S. Department of Labor, Washington, DC
Demetra Collia	Research and Innovative Technology Administration, USDOT, Washington, DC
Matthew Coogan	New England Transportation Institute, White River Junction, VT

D

Gary DeLorme	Federal Transit Administration, USDOT, Washington, DC
Raymond Dezzani	University of Idaho, Moscow, ID
Jennifer Dill	Portland State University, Portland, OR
Kieran Donaghy	University of Illinois at Urbana-Champaign, IL

F

Stacey Falzarano	Resource Systems Group, Inc., White River Junction, VT
Ellis Feldman	Federal Aviation Administration, USDOT, Washington, DC
Erik Ferguson	University of California, Irvine, CA
Gary Feuerberg	Research and Innovative Technology Administration, USDOT, Washington, DC

G

Byron Gajewski	University of Kansas Medical Center, Kansas City, KS
Marie Gantz	RTI International, Research Triangle Park, NC
Nicholas Garber	University of Virginia, Charlottesville, VA
Lee Giesbrecht	Research and Innovative Technology Administration, USDOT, Washington, DC
Genevieve Giuliano	University of Southern California, Los Angeles, CA
Arturo Gonzalez	University College, Dublin, Ireland
Thea Graham	Federal Aviation Administration, USDOT, Washington, DC
Stephen Greaves	University of Sydney, Australia
Michael Greenwald	University of Wisconsin, Milwaukee, WI
Richard Gruberg	Federal Highway Administration, USDOT, Washington, DC

H

Susan Handy	University of California, Davis, CA
Mark Hansen	University of California, Berkeley, CA
David Hensher	University of Sydney, Australia
Patricia Hu	Oak Ridge National Laboratory, Knoxville, TN
Ho-Ling Hwang	Oak Ridge National Laboratory, Knoxville, TN

K

Edward Kashuba	Federal Aviation Administration, USDOT, Washington, DC

L

Partha Lahiri	University of Maryland, College Park, MD
David Levinson	University of Minnesota, Minneapolis, MN
Susan Liss	Federal Highway Administration, USDOT, Washington, DC
Todd Litman	Victoria Transport Policy Institute, Victoria, Canada
Timothy Lomax	Texas A&M University System, College Station, TX
Dom Lord	Texas A&M University System, College Station, TX

M

Randy Machemehl	University of Texas at Austin, Austin, TX
Fred Mannering	Purdue University, West Lafayette, IN
Puring McDonald	Research and Innovative Technology Administration, USDOT, Washington, DC
Shaw-Pin Miaou	Texas A&M University, College Station, TX
Patricia Mokhtarian	University of California, Davis, CA
Greg Moyer	Athens-Clarke County Planning Department, Athens, GA
Elaine Murakami	Federal Aviation Administration, USDOT, Seattle, WA

N

Peter Nijkamp	Free University, Amsterdam, Netherlands

O

Keith Ord	Georgetown University, Washington, DC

P

Ram Pendyala	University of South Florida, Tampa, FL
Steven Polzin	University of South Florida, Tampa, FL
Pilar Poncela	Universidad Autónoma de Madrid, Spain

R

Robert Raeside	Napier University, Edinburgh, Scotland
Luis Rizzi	Pontificia Universidad Catolica de Chile, Chile
Matthew Roorda	University of Toronto, Canada
Neil Russell	Research and Innovative Technology Administration, USDOT, Washington, DC

S

Jerome Sacks	National Institute of Statistical Sciences, Research Triangle Park, NC
Miriam Scaglione	University of Applied Sciences Valais, Sierre, Switzerland
Yoram Shiftan	Technion-Israel Institute of Technology, Technion City, Israel
Eleanor Singer	University of Michigan, Ann Arbor, MI
Kumares Sinha	Purdue University, West Lafayette, IN
Brian Sloboda	Research and Innovative Technology Administration, USDOT, Washington, DC
Frank Southworth	Oak Ridge National Laboratory, Knoxville, TN
Herman Stekler	George Washington University, Washington, DC
Peter Stopher	University of Sydney, Australia
Dazhi Sun	University of Illinois at Urbana-Champaign, IL

T

Piyushimita Thakuriah	University of Illinois, Chicago, IL
Darren Timothy	Federal Highway Administration, USDOT, Washington, DC
Tomar Toledo	Massachusetts Institute of Technology, Cambridge, MA
Victor Torres	Bureau of Labor Statistics, U.S. Department of Labor, Washington, DC
Martin Treiber	Dresden University of Technology, Dresden, Germany

W

Martin Wachs	RAND Corporation, Santa Monica, CA
Simon Washington	Arizona State University, Tempe, AZ
John V. Wells	Office of the Secretary of Transportation, USDOT, Washington, DC
Fred Williams	Federal Transit Administration, USDOT, Washington, DC

Y

Vincent Yao	Institute for Economic Advancement, University of Arkansas, Little Rock, AR
Theodore Younglove	University of California, Riverside, CA
Chunyan Yu	University of British Columbia, Vancouver, Canada

Z

Johanna Zmud	NuStats, Austin, TX
Aleš Žnidaric	Slovenian National Building and Civil Engineering Institute, Ljubljana, Slovenia

INDEX FOR VOLUME 8*

* A complete index of all volumes of the journal is available online at www.bts.gov/jts.

N

O

P

R

S

T

U.S. Department of Transportation
Research and Innovative Technology Administration
Bureau of Transportation Statistics

JOURNAL OF TRANSPORTATION AND STATISTICS

Volume 8 Number 3, 2005
ISSN 1094-8848

CONTENTS