DESIGN AND FEASIBILITY STUDY: CONNECTICUT TRANSPORTATION PLANNING DATA

October 2008

Lisa Aultman-Hall
Stephanie Mather
Eric Jackson
Hyeon-Shic Shin
John Ivan

JHR 08-315 Project 05-7

This research was sponsored by the Joint Highway Research Advisory Council (JHRAC) of the University of Connecticut and the Connecticut Department of Transportation and was performed through the Connecticut Transportation Institute of the University of Connecticut.

Technical Report Documentation Page

1 Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JHR 08-315	N/A	

4. Title and Subtitle	5. Report Date
Design and Feasibility Study: Connecticut Transportation Planning Data	October 2008
	6. Performing Organization Code N/A
7. Authors Lisa Aultman-Hall, Stephanie Mather, Eric Jackson, Hyeon-Shic Shin, John Ivan	8. Performing Organization Report No. JHR 08-315
9. Performing Organization Name and Address University of Connecticut Connecticut Transportation Institute Storrs, CT 06269	10. Work Unit No (TRAIS) N/A 11. Contract or Grant No. N/A
12. Sponsoring Agency name and Address Connecticut Department of Transportation 280 West Street Rocky Hill, CT 06067-0207	13. Type of Report and Period Covered FINAL 14. Sponsoring Agency Code N/A

15. Supplementary Notes

This study was conducted under the Connecticut Cooperative Highway Research Program (CCHRP, http://www.cti.uconn.edu/chwrp/index.php).

16. Abstract

Transportation decision-making requires many types of data and ultimately the quality of a policy decision is dependant on the quality of that data. The specific objective of this research was to prioritize the household travel data needs of Connecticut transportation planners and to evaluate the available data collection options. These options included; transferability of national travel surveys, internet based travel surveys and Connecticut based phone surveys. A pilot web-base survey was conducted. The process was evaluated and results were compared to data transferred from national surveys. These preliminary results indicate that the web survey bias was limited for the rate data most critical for Connecticut planners. Recommendations regarding a plan to routinely collect, maintain and distribute the transportation planning data are made. A final decision on data collection needs and methodology is ultimately within the domain of the Connecticut Department of Transportation (ConnDOT).

17. Key Words	18. Distribution Statement
Travel Survey Methods NHTS Transferability Travel Data Collection Internet Travel Survey	No restrictions. This document is available to the public through the National Technical Information Service Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	71	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^{o}F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^{o}C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^{o}C	Celsius	1.8C+32	Fahrenheit	^{o}F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 Introduction

Transportation decision-making requires many types of data and ultimately the quality of a policy decision is dependant on the quality of that data. Many transportation planning activities, including demand forecasting, highway safety analysis and mobile emissions modeling, require household travel data. Household travel data consist of travel patterns (trip purpose, frequency, mode choice and possibly route), household characteristics and individual personal attributes. Unfortunately, with budget limitations, scarce resources limit our ability to collect relevant data. Yet at the same time, infrastructure decisions and maintenance of our aging system require careful planning. Large databases that are representative of household and individual travel patterns are not only expensive to collect, but also to maintain and update. Indeed, the inability to fund the collection of up-to-date travel behavior data for calibrating four-step transportation planning models in particular forces many transportation planning agencies to rely instead on average national data or older local survey data that no long applies to current travel conditions within the state. Data collection is further restricted by increasing concerns over individual privacy, a public outcry against telephone surveys, and attention to safety which effectively ended the practice of on-road OD surveys. While each of these concerns is merited, it is little consolation to the planner who wants to base recommendations on sound, comprehensive and recent data.

Even when traditional phone and mail-back surveys are conducted, it is difficult to obtain trip rate and length information due to trip tour (chaining) behavior and participant recall errors, including underreporting. The real-time use of hand-held computers by travelers has been proposed as one technology for more accurate and comprehensive data collection. Moreover, new Global Positioning System (GPS) and Intelligent Transportation Systems (ITS) technologies offer the potential to cost-effectively collect up-to-date household travel behavior data without burdening the individual traveler. GPS receivers can collect these trip length and route data automatically for multiple trips and multiple days without direct imposition on survey participants. Another possible source of information is new video technology which could be used for automated origin destination (OD) collection along transportation corridors. The wide-spread use of ITS has resulted in large quantities of operational data, such as traffic counts, that could be adapted for use as planning data. For example, a recent Connecticut Cooperative Highway Research Program project used truck volume counts to estimate a synthetic OD for truck trip generation in the state-wide model (Aultman-Hall et al. 2004). At a national workshop hosted in 1998 by ITS America, the US DOT and other agencies, the uses of ITS data in highway and transit planning, operations and freight planning were identified (ITS America, 1998). There was at that time a need to plan for aggregation of the vast amount of raw data generated by ITS. The priority data elements available from ITS were considered to be traffic volume, speed and travel time, origin/destination information (from Smart cards and similar devices), vehicle occupancy and density.

These examples illustrate the potential for new technologies to cost-effectively improve the ability for transportation agencies to collect up-to-date household travel behavior information. Approximately three years ago, the state DOT in Connecticut approached researchers at the University of Connecticut to jointly pursue routine household travel data collection for state planning purposes. Moreover, it was hoped that the same dataset might serve the needs of the Metropolitan Planning Organizations (MPOs) and other planning groups within the state. Several research questions formed the basis of this resultant project. First, it was necessary to identify the most critical data needs of planners in Connecticut State and MPO agencies. This was determined based on two large roundtable sessions hosted early in the project. Second, the methods to collect and maintain travel data were to be evaluated for technical feasibility, statistical validity and cost effectiveness. The following methods to obtain recent travel data for Connecticut were included in the project: transferring national data from the NHTS survey; conducting a phone survey; use of ITS data; and use of a web-based survey.

The specific objective of this research was to prioritize the household travel data needs of Connecticut transportation planners and to evaluate the available options, including new-technology driven options such as the Internet, to collect these data. Recommendations regarding a plan to routinely collect, maintain and distribute the transportation planning data are made. A final decision is ultimately within the domain of the Connecticut Department of Transportation (ConnDOT).

Following the background section, this report contains a summary of the project's roundtables, results of the analysis of transferring national data to Connecticut; and the results of a pilot web-based household travel survey which was conducted during the project.

2.0 Background

The relevant background for this project can be categorized into the five topics that are discussed in this section of the report: 1) the context and approaches to state-wide transportation planning; 2) data needs for state-wide planning; 3) traditional data collection techniques; 4) existing sources of data; and finally 5) the potential for the use of new technology to collect transportation planning data. Much of the information presented in this section was presented at the roundtables described in section3.0 before the planners were asked for input on the project work plan.

2.1 The Context and Approaches to State-wide Transportation Planning

Stopher and Metcalf (1996) define two reasons for collection of household travel data: understanding the ways that the transportation system is currently being used; and forecasting future demands on the system. While urban areas, including MPOs, were the traditional study areas for household travel surveys and demand models, Horowitz (1999) points to several reasons why a state would be interested in forecasting state-wide or rural travel. First, the overall assessment of the adequacy of the state-wide transportation network and the programming of projects requires overall state-wide forecasts. Second, urban forecasts must be supplemented with intercity travel in order to produce state-wide indicators such as air quality merits or other total travel measures that are collected for national policy. Finally, TEA21 mandates that several issues be addressed in state-wide transportation plans and Horowitz (1999) suggests many of these issues can be facilitated by a good multimodal travel forecasting model. Connecticut is one of the states that maintains such a state-wide travel demand model.

A travel demand forecasting model usually consists of a series of sequential mathematical models aimed at estimating future traffic demand by network segment by mode. Use of a forecasting model offers several advantages over the more traditional use of traffic growth rates. Travel demand is inherently spatial and necessarily linked to land use patterns. Travel is a derived demand and as such traffic volume growth is more complex than linear growth rates suggest. Demand models can account for the causal impact of changes in land use and infrastructure, as well as the human factors which affect mode and route choice. While demand models are not necessarily perfect, in today's policy environment, accounting for the complex interaction of land use and transportation is essential. Particularly since ISTEA, the interest in state-wide travel forecasting models has grown. In addition, to assisting with infrastructure decisions, these models can provide critical information for air quality models, traffic simulation and public policy debates. Moreover, given congestion effects combined with larger and larger traffic volumes, the simple extrapolation of past traffic growth produces unrealistic future traffic volume estimates.

Over the last 15 years, the opportunities to build better state-wide travel forecasting models have also grown. The proliferation of inexpensive computing power means that larger study areas and networks can more easily be modeled. Further, the widespread use of Geographic Information Systems (GIS), and the associated network databases, aid model development and allow planners to display meaningful graphics to the public and decision makers. The modeling methodologies in transportation planning have advanced significantly. These improved models include stochastic user equilibrium traffic assignment, and probabilistic mode choice models. However, the state of the practice has not, in most cases, kept pace with the state of the art in demand forecasting.

Significant challenges exist for all demand forecasting models, including state models. First, data collection is an expensive endeavor and few jurisdictions have the funds needed to properly populate their model calibration datasets with updated trip rates, route or destination choices. Second, changes in the way people travel have challenged the four-stage methodology itself. For example, trip tours (chaining) and increasing levels of non-work discretionary travel do not fit well in common model methodologies. Third, freight traffic volumes have become more significant over the last two decades. Collecting freight data is challenging due to shipper confidentiality issues. Incorporating freight and multi-modal freight networks has created challenges. Many states are turning to modeling commodity flows as an alternative to simply modeling truck volumes. But the distances and mode choice factors for freight differ from those of passenger travel making inclusion in a common model difficult. A fourth challenge for state-wide models is the integration with metropolitan regional models which were often developed first and have a different scale and focus. This challenge goes beyond the overlaying of networks and boundaries, and includes the different nature of travel being modeled. For example, bicycling and walking might be included in the regional model, but not at the state level. Transit systems vary significantly throughout a state. Regional freight trips, such as distribution or trash collection, do not appear at the state level, where interstate commerce and the need for modal substitution dominate the focus.

In summary, state demand forecasting models are considered by some as necessary because they offer policy makers analytical techniques which connect policy sensitive land use and infrastructure variables to the levels of future traffic demand. Computer power and modeling approaches have advanced opportunities for these models, yet challenges related to data and model scope remain.

2.2 Data Needs for State-wide Planning

Ultimately, the data needed for state-wide planning depends on the analyses that will be undertaken, which are in turn dictated by the policies being contemplated for the transportation system within the state. The focus of statewide planning differs from that of corridor, project level or metropolitan area planning. Federal transportation legislation requires states undertake certain planning efforts but now allow flexibility in how planning is undertaken. It has been suggested that states consider the following factors in their planning efforts: energy use, border crossings, connectivity between metropolitan areas, efficiency of existing facilities, traffic congestion management and commercial vehicles (Horowitz 1999).

Connecticut is a small state in terms of physical size (4800 square miles) with a moderate population (3.5 million). Two major interstate corridors traverse the state and the congestion on I-95 where the southwest corner of the state meets greater New York city often dominates traffic concerns. Together, the state owned and operated Metro North Railroad into Grand Central terminal in New York, as well as regular Amtrak service along several routes in the state, represent a significant amount of non-automobile travel. This makes mode choice of interest in state-wide planning and also means that the New York MPO collects data in southwest Connecticut for their planning model. In many ways, Connecticut lies at a “crossroads” within New England and a significant volume of out-of-state traffic passes through Connecticut without having an origin or destination within the state. This section outlines the ideal set of data required for a comprehensive state-wide demand forecasting model, as well as the minimum data required to update an existing model.

Ideally, the following information would be available for all trips undertaken by all individuals: purpose, time of departure, origin location, mode, route, travel time (by segment), and destination location. Location is more recently captured with precision potentially latitude and longitude but often nearest road intersections. Ideally trips on both weekdays and weekends would be documented as well as both discretionary and non-discretionary travel. The demographic characteristics of the individual (age, gender) and their household type (number and age of people, workers, vehicles, and drivers) are critical data because they represent the predictor variables for model development. They allow for transferability of the model for estimation of travel in different locations and times (usually the future). Comprehensive travel surveys collect this full range of data.

The minimum data required to update a model consists of trip rate (by person or household), mode split and trip length distributions. Planners at the roundtables described in the next section (2.3) indicated trip rate and mean trip length as the two most critical data elements needed for their work. The number of trips made by a household or an individual in a period of time (usually an hour or a day), is an essential measure in travel behavior research and transportation demand modeling. This measure is widely used in planning for trip generation, emissions modeling, as well as other transportation related evaluations. Trip lengths allow for calibration of the trip distribution portion of the models. Conventional survey methods, such as mail and phone surveys that collect travel information based on participant recall, are limited even with respect to collection of these most basic quantities, especially in capturing short trips or trip chains. Short trips are frequently omitted by survey participants especially when data collection lasts for multiple days and participants are required to record trips at the end of a day or at the end of the entire data collection period.

2.3 Traditional Data Collection Techniques

A synthesis of methods for household travel surveys was conducted in 1996 by NCHRP (Stopher and Metcalf 1996). Stopher and Metcalf identified 55 agencies across the country that had conducted a household travel survey in the late 1980s and early 1990s. They found that agencies were typically motivated by aging data and collected data only for weekdays from approximately 2500 households each. The data collected typically included a description of the household, the persons in the household and the vehicles available to the household, as well as the trips and/or activities undertaken by each person. The data were usually collected for a specific 24-hour period using an interview process. Interviews were face-to-face, telephone or mail-back questionnaire. The University of Connecticut is particularly well-qualified to perform a telephone-based survey given the on-campus Survey Research Center and therefore the costs to perform data collection this way were estimated within the project. However, response burden and ensuring a representative sample have been increasing problematic in survey research and this option is no longer recommended for those starting new data programs.

Other traditional data collection methods are presented by Stopher and Metcalf (1996). These include workplace surveys, on-board transit surveys, cordon surveys and screenline counts. The authors also point to the need for supply-side data for planning, which includes information on the current status and future projections for land use, highway and transit facilities, and traffic conditions. These supply side data items are beyond the scope of this project and the focus is instead on collecting data on travel patterns especially the trip rate and length data requested by the stakeholder planners invinted to roundtables described later in section 3.0.

2.4 Existing Sources of Travel Data

The United States Department of Transportation (USDOT) Bureau of Transportation Statistics (BTS) collects and maintains several databases that can provide useful data on overall passenger travel for state-wide planning efforts. Many of these datasets are collected in conjunction with the US Census Bureau and some of them pre-date the establishment of the BTS itself.

In 1995, the first American Travel Survey was conducted by BTS. A description of the survey and certain data tables are available on the agency's web site. The objective of the survey was to collect information about households, persons and trips of over 100 miles in length. A total of 80,000 households were surveyed at quarterly intervals throughout the year. Aggregate data can be downloaded from the Internet, but there are some limitations on the use of the actual data itself. The Nationwide Personal Transportation Survey (NPTS) was conducted in 1969, 1977, 1983, 1990, and 1995. This survey collected information for personal trips for <u>all</u> (of any length) modes of travel and all purposes. In 1995, 42,000 households were surveyed and in some years MPOs or other agencies could pay a fee for an over-sample in their area. In 1990, the state of Connecticut requested such an over-sample and these data have been used by Dr. Ivan, co-PI on this research team, to investigate peak period trip generation (Ivan and Jha 1997). This project will consider whether this Connecticut-specific database can be updated and used for current state-wide planning efforts. In 2001, the National Household Travel Survey, a combined version of the Nationwide Personal Transportation Survey and the American Travel Survey was conducted for the first time and the more aggregate data tables are available from the BTS.

The FHWA is about to embark on the 2008 version of the NHTS. Field testing started in January 2008 and first data are to be collected in March. As the survey was not comprehensively funded in SAFETEA-LU (2005), data will only be collected in the 17 states and metropolitan areas which have purchased "add-on" samples. A total of 100,000 households will be sampled. Connecticut will be in the random national sample but not among the add-ons. Nearby, both Vermont and New York are included in this add-on project. Several new data items have been added since the 2001 survey including: school travel, home deliveries of internet shopping, interstate use and tolls paid.

The BTS also provides a Census Transportation Planning Package (CTPP) that contains a collection of summary tables that have been generated from both the census short and long forms. The BTS describes this dataset as a set of tables that contains information organized by urban area and state agencies about population and household characteristics, worker characteristics and characteristics of the Journey-to-Work (JTW). This product was produced for both 1990 and 2000.

These national databases have the advantage of being thoroughly designed and large enough to be statistically valid at the national level. However, as survey costs increase, sample size has decreased, making it difficult to disaggregate the data spatially for a given area and still protect confidentiality. This necessitates questioning the transferability of data and how national average data can be used for state level planning. Wilmot and Stopher (2001) studied the transferability of aggregate transportation planning data for use in demand forecasting models. Their analysis included trip rates, mode shares and trip length distributions. By comparing a model updated from the national data to results from an actual local survey, they were able to conclude that a panel of 500 households surveyed on an annual basis and used to update the transferred data represented a cost-effective solution for an urban transportation planning model. They further argued that this approach would be more cost effective than undertaking periodic full regional travel surveys. Mei et al. (2005) found similar findings in their comparison of the transferability of NHTS data for the state of Kentucky. While this study was on-going, work at the University of Chicago Illinois (Zhang and Mohammadian 2008a and 2008b, Mohammadian and Zhang 2007) on transferability of the NHTS 2001 data was also conducted using similar techniques (Bayesian updating and cluster analysis). The UIC team also used neural networks and simulation. All of their results suggest strong promise for use of transferred data.

In 2007, the Oakridge National Lab created a web-based transferability data program using the NHTS 2001 data. In essence the program estimates TAZ or Census tract level trip rates or lengths (as requested by planners in this project) based on the data collected in the national 2001 survey. This tool was released after the transferability analysis described in section 4.0 of this report.

2.5 Use of New Technology to Collect Transportation Planning Data

Today, with the availability of Global Positioning System receivers to capture vehicle location, Geographic Information Systems (GIS) for travel route data storage and analysis, and inexpensive computer memory, it has become realistic to use GPS for travel data collection. GPS-collected data is high in quality compared to the data obtained through conventional surveys, especially after May 2000 when the Department of National Defense removed Selective Availability[1] (S/A). Ochieng (2002) has shown that the accuracy level of GPS is now satisfactory for route data collection.

GPS has numerous advantages over the traditional data collection methods, such as travel diaries or telephone interviews. First, GPS devices collect data automatically over a large geographic area at the same time trips are made thus avoiding some errors caused during trip recollection, such as survey participants' inaccurate travel time or distance estimation, an inability to recall or describe exact routes, or the omission of short trips. Second, the automatic data recording ability of GPS devices lowers participant burden and therefore increases the survey recruiting success rate and decreases the rate of participants' withdrawal midway through the study. For example, Murakami and Wagner (1999) addressed trip reporting fatigue in traditional surveys versus GPS surveys. In a seven day trip diary survey in the Netherlands, by day 7, only 30% of the respondents were still participating while only 50% of them completed the diary. In contrast, in the GPS-based project described by Murakami and Wagner, the recruiting success rate was 67% and only two of the 100 households declined to participate after reviewing the informed consent forms. Murakami and Wagner (1999) also concluded that the ability of GPS to capture multiple days and record routes and speed is better than retrospective surveys. Wolf et al. (2001) compared the number of trips collected by GPS to those collected in a traditional computer-assisted telephone interview (CATI) in a California study and concluded that GPS is significantly efficient in capturing trips in that GPS can capture 29.2% more trips than the equivalent CATI-reported method. The advantage of GPS is especially evident when "heavy travel households" (households that have at least one vehicle with more than 10 GPS trips recorded on one travel day) were considered. Yalamanchiliv et al. (1999) found that GPS-based data furnished more than twice as many multi-stop chains than the recall data.

There are generally two types of in-vehicle GPS configuration. One is a GPS device used together with Personal Data Assistant (PDA) where participants are required to input information for each trip, for example trip purpose and the names of the driver and passengers. The other is a passive GPS that requires no intervention from participants and collects data automatically. This can serve the need to minimize participant burden while at the same time collecting as complete a dataset as possible. The GPS and PDA combination has the advantage of collecting more complete trip data, but is more expensive and involves high participant burden. The passive GPS system is easier for participants, but requires post-processing because all trips are stored in a single continuous data stream. Although the data processing is challenging, the completeness of the data and the lower cost and burden of passive GPS merits its use and the specific development and evaluation of methodologies to post-process the data. Therefore, the focus of recent work at the University of Connecticut has been to develop post-processing techniques using in-vehicle logs for validation of large passive in-vehicle GPS datasets (Du et al. 2004). Stopher et al. (2003) also acknowledge that GPS data collection can be complicated and that technical difficulties also caused problems. These

[1] With SA, the navigation accuracy was artificially degraded to the level of 100 m horizontally and 156 m vertically.

advantages and challenges of GPS were again evaluated by the researchers, advisors and roundtable participants in this project. Ultimately, the cost for GPS data collection combined with challenges resulted in a decision not to pilot travel data collection using this method in this project.

The Internet offers the potential to collect transportation planning data. The representativeness of this type of data has not been completely defined. However, a study by Abdelwahab and Abdel-Aty (2003) compared the quality and completeness of data obtained in a mail-back survey to that collected via the Internet and found the latter to be much better. Bricka et al. (2003) addressed the concern that highly mobile households are missed in phone-based surveys (they are out traveling when phone calls are made), by using an Internet-based system for a subset of households. Their conclusions suggest that Internet-based data collection captures information that might otherwise be excluded. During roundtables in this the biased sampling problems of Internet based recruiting and travel data collection were of concern. This project ultimately focused on conducting a pilot web survey for Connecticut and DOT stakeholders indicating sampling bias might be overcome in on-going data collection by networking through public libraries, schools and other community center. Stakeholders in this study also noted an Internet survey advantage might include having participants select origins and destinations on an interactive map. This technique has been piloted by others (Resources Systems Group 1999) but was not tested in this project.

The widespread use of Intelligent Transportation Systems (ITS) has resulted in a proliferation of video cameras throughout transportation networks. The primary intent of the cameras is often surveillance such as the detection or confirmation of incidents. But coupled with expert system software, some cameras in some locations, although not necessarily on the Connecticut state system at this time, are routinely used for traffic volume counts and speed measurements. They have been programmed to estimate occupancy (a surrogate measure for traffic density) which makes them a potential replacement for inductive loop detectors depending on site conditions. Microwave and infrared detectors have also replaced inductive loop detectors in ITS applications, but may not offer as many sources of data for planning as the video cameras. Connecticut has a series of video cameras on their freeway system and posts this information in real time to the Internet. Recently, ConnDOT also started operation of microwave detection systems and posts these data to the Internet as well.

Automatic Vehicle Identification (AVI) in more than one location of the transportation system allows for the estimation of segment travel times as well as the origin and destination (or sub-origin and destination) of travelers (Antoniou et al. 2004). Current research is addressing where within the transportation network identification devices need to be placed to produce the most useful information (Chen et al. 2004). The re-identification of vehicles can be made through video technology (vehicle recognition or license plate reading) or through a unique signature such as that on toll transponders. Both the travel time and route information is useful to the planners and for calibration of the demand models used in forecasting. These data can also be used in the analysis of road tolls or congestion pricing (Porter et al. 2004). Note that some researchers (Tawfik et al. 2004 and Oh et al. 2004) have conducted experiments with reidentification using inductive loop detectors. Use of inductive loop detectors is promising as they are so widely distributed throughout both the arterial and collector roadways of our transportation network. Coifman and Yang (2004) go one step further and suggest that although ground-based sensors can provide rich datasets they lack spatial coverage thus limiting their use and application. They suggest that "High-resolution imagery remotely sensed from satellite or airborne platforms is an attractive alternative that can potentially supplement and enhance the existing traffic monitoring programs with a spatially rich dataset." Ultimately, ITS-based data were not tested in this project because it was not able to provide the type of planning data identified as most needed by the roundtable stakeholders.

Travel is a derived demand and many transportation planning efforts have turned to the modeling of activity patterns in an effort to predict travel needs as an output. Collecting activity by individual and household type might be considered even more complex than collecting travel data. Use of advanced technology has been suggested for activity data generation. Greaves and Stopher (2000) for example have proposed those household activities and the resultant travel patterns could be simulated based on existing national databases as an alternative to expensive data collection. The conclusions from this and later papers (Stopher et al. 2003 and Stopher and Greaves 2007) suggest these approaches including data fusion holds promise. Ideally however, the most basic household travel parameters such as trip rates, length and household demographics (including the spatial distribution of household type) would be state specific

3.0 Summary of Stakeholder and Roundtable Input

The overall objective of stakeholder input was to assess data needs for transportation planning and to refine the research objectives of the project. In June 2005, this project team traveled to Washington D.C. to meet with representatives of the FHWA and BTS to discuss travel planning data and assess the future plans of national database managers for future studies. In person meetings were held with six individuals and one conference call was undertaken. Both sets of meetings indicated a positive utility for NHTS add-on data. Problems previously encountered by the state of Connecticut in using the data, in particular geocoding, were expected to be resolved for the 2007 survey. Ultimately, the NHTS will not be conducted until 2008. The price of $175 per completed household and a minimum sample size in the state of 1500 was considered potentially cost prohibitive for CT. Later in the fall, a phone interview was conducted with the state of Massachusetts in which they indicated use of the national NHTS and other products but no intention to collect their own traditional household travel survey data.

A technical advisory committee was formed for the project in Summer 2005. The committee consisted of 3 ConnDOT planners and two MPO planners. In early fall this group met and considered the following questions:

a. What data do planners in CT need most?
b. How should we routinely collect and maintain these data?
c. What pilot data collection technique should be evaluated in this research project:
    i. Conduct traditional household travel survey by phone
    ii. Collect CT data with new technologies
    iii. Use federal data (NHTS, "transferability", ATS)
    iv. Buy federal NHTS add-on
d.Who should be invited to the project's roundtables

As a result of input from the technical advisory committee, urban and rural planning agencies were invited to participate in the half day roundtable in September 2005. A total of 24 attendees participated including representatives from state DOT, urban and rural CT planning agencies, state police and New York Metropolitan Transportation Council (NYMTC). The main planning data concerns and suggestions raised by participants are shown in Figure 1. There was strong support for the state model but a desire to have more refined zones, better geo-coding of trips ends, and model compatibility with regional models. Planners expressed interest in having data on non-work trips and special large trip generators. Suggestions for innovative sampling methods were made.

- Air Quality conformity is an important use and requires accuracy
- Approaches to data collection in adjacent areas should influence our data collection
- In addition to data need, models approaches need serious updating
- Better capture of chained trips and interaction within the household requires attention
- Limited bike and pedestrian trip information is problematic
- The geo-coding and accuracy of the origins and destinations is important
- Off-peak, non-work (non-journey to work) data is lacking\
- Data for special generators such as casinos, entertainment, racetracks, etc will require attention
- Redevelopment of certain areas of the state has affect travel and data is needed
- Buses and freight are not currently captured and should be
- There is a challenge to incorporate future travel demand into modeling when we lack even present day data
- Crash data was suggested as a source for vehicle occupancy
- Statewide model is too aggregate and the cities and towns need more refined information including non-major roads but the high quality state model is an important framework for local models
- State model zones are too large and need improvement for dealing with new generators- especially in outlying areas and/or in developing areas
- Time of day modeling may be needed
- Forecasts needed (not currently given in State model)
- Mode choice is particularly difficult to include
- DOT lacks the resources or staff to update the model for all needs, this includes workforce development challenges
- Need for weekend data increasing holiday/recreation
- There is a desire to ultimately link regional models and the state model with the ability to “window in” on project areas
- Planning agencies rely on state to be the comprehensive data source
- The data need is for travel behavior data for calibrating the models; the land use and demographic data for input is currently stronger
- Concerns about phone surveys and GPS surveys missing people were expressed
- Use CT income tax return to collect extra data was suggested
- It was suggested to intercept people at places they wait; DMV or during jury duty

Figure 1: Roundtable Input (September 2005)

Ideally the planners would like to have data for special generators; off-peak, non-work trips: projections for future; more detailed network (regional overlay); more zones in suburban areas; and more temporal detail. The conclusion of the roundtable was threefold. First, the state planning model is necessary as the backbone and link for regional planning efforts, including regional models. Second, stakeholders look to the state DOT as the leader for travel data collection and maintenance. Third, certain agencies and location outside of Connecticut were recommended for the research team to look for innovative data collection methods.

In June 2006, roundtable stakeholders were again invited to meet for a half day with the project team. Between roundtables, the research team collected travel surveys from 9 state and local agencies beyond Connecticut and interviewed individuals by phone. The following general conclusions were reached:

- Mail back and phone surveys are still most commonly used
- GPS pilots are frequent but have significant challenges
- New methods are being sought widely but concensus for the next generation of data collection does not exist
- Need to change from trip-based to activity-based data collection is commonly recognized as essential

Table 1 indicates the case studies considered and a summary of the methods used, variables collected and costs (where known). Note the continued prevalence of phone surveys (CATI), the increasing costs over time and the overall low response rates. During interviews other agencies indicated the same concerns expressed by the Connecticut planners including response rates, non-work and non-auto trips, recall bias, geo-coding and the cost of maintaining current data.

Table 1: Brief Summary of Household Surveys

Survey Reviewed	Agency / Region	Year	Type	Sample Size	Percent Household Sampled	Weighted Response Rate	Incentives	Cost
Transportation Tomorrow	Greater Toronto Area, Canada	2001	CATI	Approximately 150,000 HHs sampled	5%			$15.56 (Canadian) p HH (2001); $18.65 p HH (1996)
2001 National Household Travel Survey	United States	2001	CATI	69,817 usable / 106,598 HHs		38.90%	$5-10 / HH	
2001 Atlanta Household Travel Survey	Atlanta Regional Commission	2001	CATI	8.069 usable / 12,184 Has		30.40%		
2000-2003 Ohio Statewide Household Travel Survey	Nine Small MPOs in Ohio	2001 - 2003	CATI	16,112 useable / 22,413 HHs		30%		
Bay Area Travel Survey (BATS)	Metropolitan Transportation Commission, CA	2000	CATI	18,068 hhs usable (14,563 MTC has; 503 1990 panel; 1995 BART users add-on; 1007 non-BART users add-on)			$5 per diary mail-out	
CATS 1990 Household Travel Survey	Chicago Area Transportation Study (CATS)	1990	Mail-back	19,314 usable / 79,346 mailed				
Kansas City Regional HH Travel Survey	Mid-America Regional Council	2004	CATI GPS Mail-back	3,049 of 4,001 recruited 228 / 3049 29 / 228		37%		
Metro HH Activity and Travel Behavior Survey	Portland Metropolitan Service District (Metro), Oregon	1994 - 1995	CATI / Stated-preference	4451 HH completed		33%		Total $ 603,180 ($13 HH)
1997 Regional Travel - Household Interview Survey (RT-HIS)	New York Metropolitan Transportation Council	1997-1998	CATI	11,264 HHs completed / 14,441 HHs recruited		26%	Two types of incentives (Pilot Study) - $1 Cash or Airline ticket drawing	
Connecticut Statewide HH Travel Survey	Connecticut	1976 - 1977	Mail-back	6784 usable returns	3%	24.50%		$1.28 for each usabl response / $1.17 fo each returned questionnaire (only material costs are included

1. Boston Case Study still in development

Appendix A contains a summary table of the typical household travel variables measured in the collection of surveys examined. This list was presented to the participants of the second roundtable and also used as the basis for cost estimates for a phone survey discussed later in the report. Participants were asked to consider the following options for updating Connecticut household travel data: transfer of national and NYMTC data; use of GPS; use of an Internet survey; use of ITS or vehicle enforcement data. A demonstration of the web-based survey software purchased for the project was conducted. Preliminary results of the categories of household for data transfer were presented. Fully supported by the ConnDOT reps, the roundtable participants recommended the data collection pilot in this project consist of ONLY a web survey instead of piloting two different survey techniques. This decision allowed for increased sample size. A GPS-based PDA was not recommended as a second pilot effort. Part of the motivation for this decision was the intention that a large enough sample of web data might be collected to be useful for planning purposes beyond pilot evaluation. The raw tabulated data has already been provided to ConnDOT and is available from CTI. Furthermore, UConn researchers had significant experience with GPS for travel data collection and the limitations and advantages had been documented (Appendix B)

4.0 Transferability of National Travel Surveys

In response to questions which arose at the roundtable meetings, transferability of the National Household Travel Survey to Connecticut was investigated. The main objective of this part of the work was to evaluate the feasibility of developing new, accurate household trip rates and trip lengths for Connecticut by combining existing national and New York (including two Connecticut counties) household travel survey data from the past 16 years. This section will summarize her thesis and outline how the national survey dataset can be transferred and made to be representative of travel in Connecticut.

4.1 Data Description

Three of the datasets used for this analysis were from the household travel survey conducted approximately every five years by the Federal Highway Administration (FHWA). These datasets include the Connecticut add-on[2] household survey subset from the 1990 National Personal Travel Survey (NPTS), the Connecticut subset from the 1995 NPTS, and the Connecticut and New England subset from the 2001 National Household Travel Survey (NHTS). The second source is the New York Metropolitan Transportation Council (NYMTC), which conducted their RHTS survey in 1997-98. The NYMTC is the MPO for the New York City, New York area and this survey included two Connecticut counties which have access to the Metro-North Railroad (MNRR) line that travels into New York City. The trip variables extracted from these datasets included: household trip rate, trip purpose, trip length, travel mode and household socioeconomic data. Recall the roundtable participants indicated that trip rate and length were the two most important variables for which updated values were needed. Therefore the other variables were used

[2] NPTS/NHTS surveys are conducted nationwide and states can purchase add-ons which increase the number of households included in that particular state.

to compile the final household trip rate and trip length datasets used for analysis in this project.

Two datasets for analysis were tabulated from these multiple datasets, one by household and one by individual trip. For this study, a trip is defined as a person-trip, thus the trip rate for a household is the sum of person-trips made by all household members. The final aggregate household and trip datasets contain 4,343 households and 35,201 trips. Table 2 breaks down these observations by data source. Table 2 suggests that more trips per household were taken in 2001 than in 1990, this indicates there could be a temporal increase in travel rates that should be accounted for in the analysis.

Table 2: Number of Observations in Each Subset

Data Source	Total Number of Households	Total Number of Trips	Mean Trips Per Household
NPTS 1990	2,266	15,088	6.7
NPTS 1995	225	2,257	10.0
NYMTC 1997	430	3,477	8.1
NHTS 2001	1,422	14,379	10.1
Total	4,343	35,201	8.1

4.2 Distribution of Socioeconomic Variables

The following section explores only a small subset of the distributions of socioeconomic and trip variables by subset that may bias results or suggest a temporal pattern. A detailed description of all variables considered relevant can be found in the original thesis. Socioeconomic variables are used to weigh the travel information in order to “transfer” it for use in another zone or area.

The distribution of the number of vehicles per household is shown in Figure 2. Each dataset has about 6% of households with no vehicle. Most households (46%) own two vehicles. Less than 4% of households in each dataset own five or more vehicles. The distribution of vehicle count varies little between data sources.

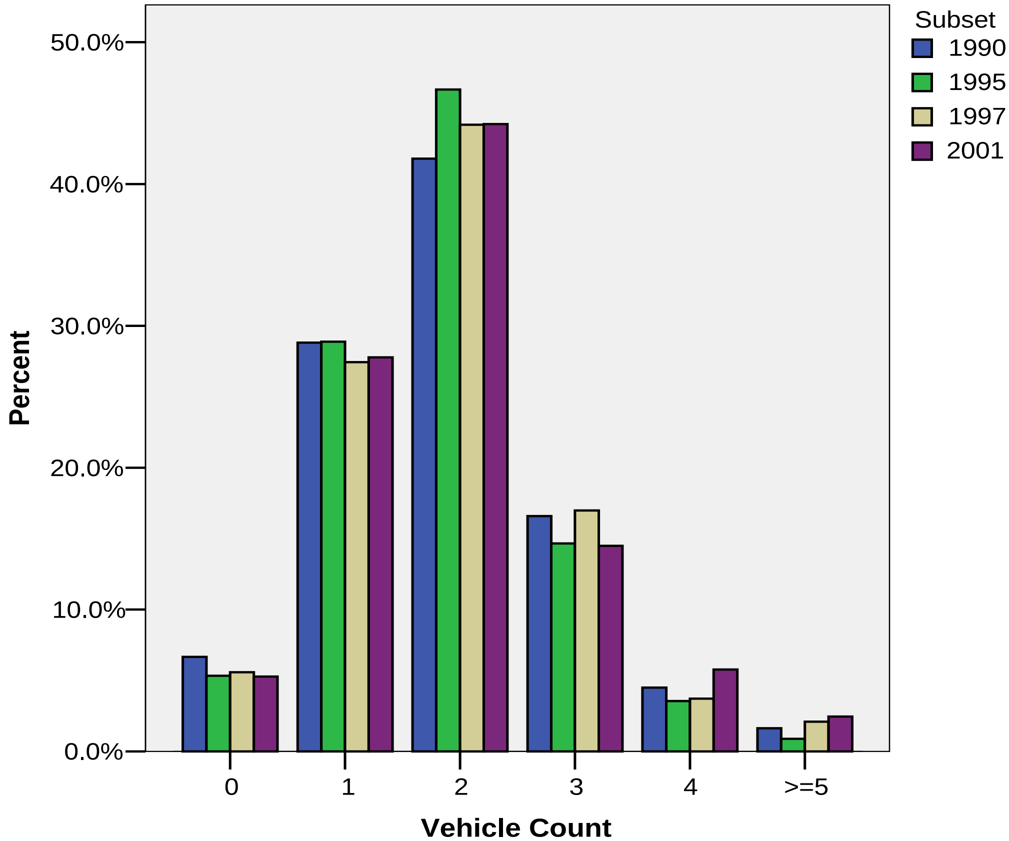


Figure 2: Vehicle Count Distribution by Source

Figure 3 indicates the distribution of household size for each dataset. A total of 17-20% of households in the 1990, 1995, and 2001 surveys have only one person, whereas about 28% of households from 1997 NYMTC data for New Haven and Fairfield counties in CT have one-person households. All subsets have mostly two-person households. The most significant difference between subsets in household size distribution is the difference in one-person households, with 1997 being about 8-10% higher than the other subsets. This finding suggests that perhaps the household demographic characteristics in New Haven and Fairfield counties differ from the rest of Connecticut and New England.

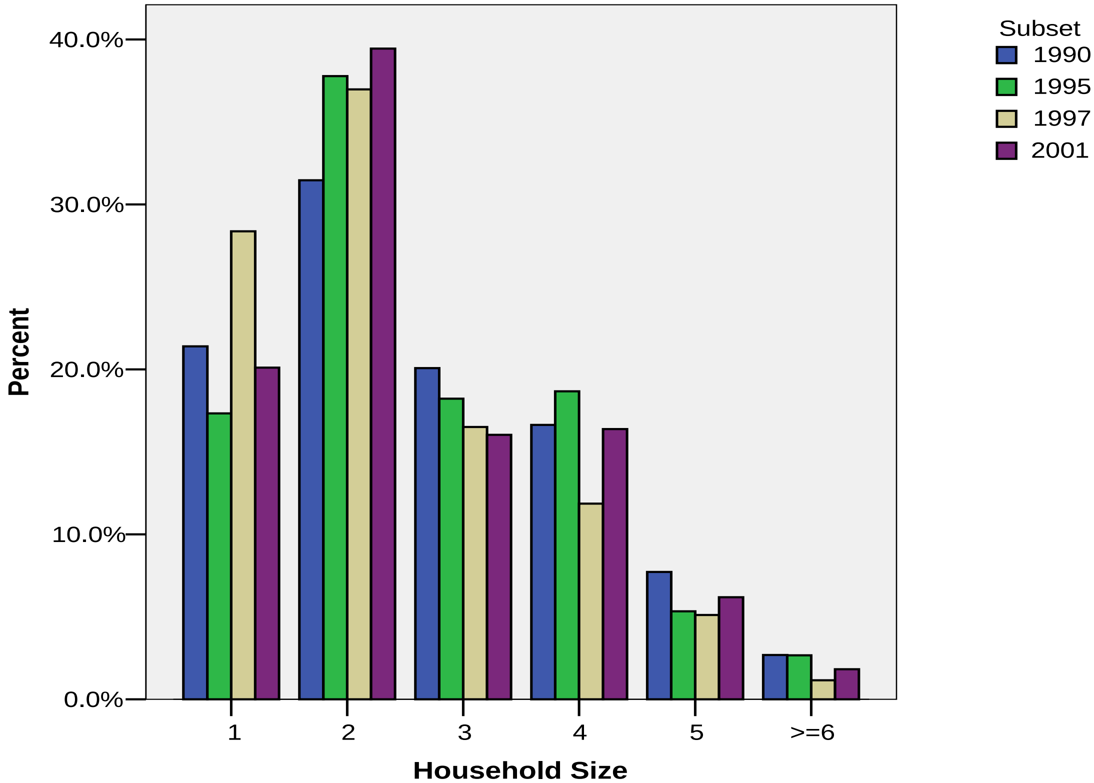


Figure 3: Household Size Distribution by Source

Figure 4 presents the distribution of workers by household for each dataset. The 1990 NPTS dataset contains the most households with no workers at about 26%. The 1995 and 2001 datasets both have about 21% of their households with no workers, while 1997 has the lowest percentage at about 18%. Most households in all datasets have one or two-workers. The 1990 and 1997 subsets both have more one-worker households with 39% and 43% respectively. The 1995 and 2001 subsets both have more two-worker households, with 42% and 39% of households reporting two-workers, respectively. The difference in one-worker and two-worker households may be a cause for differences in trip rates amongst subsets, especially work trips.

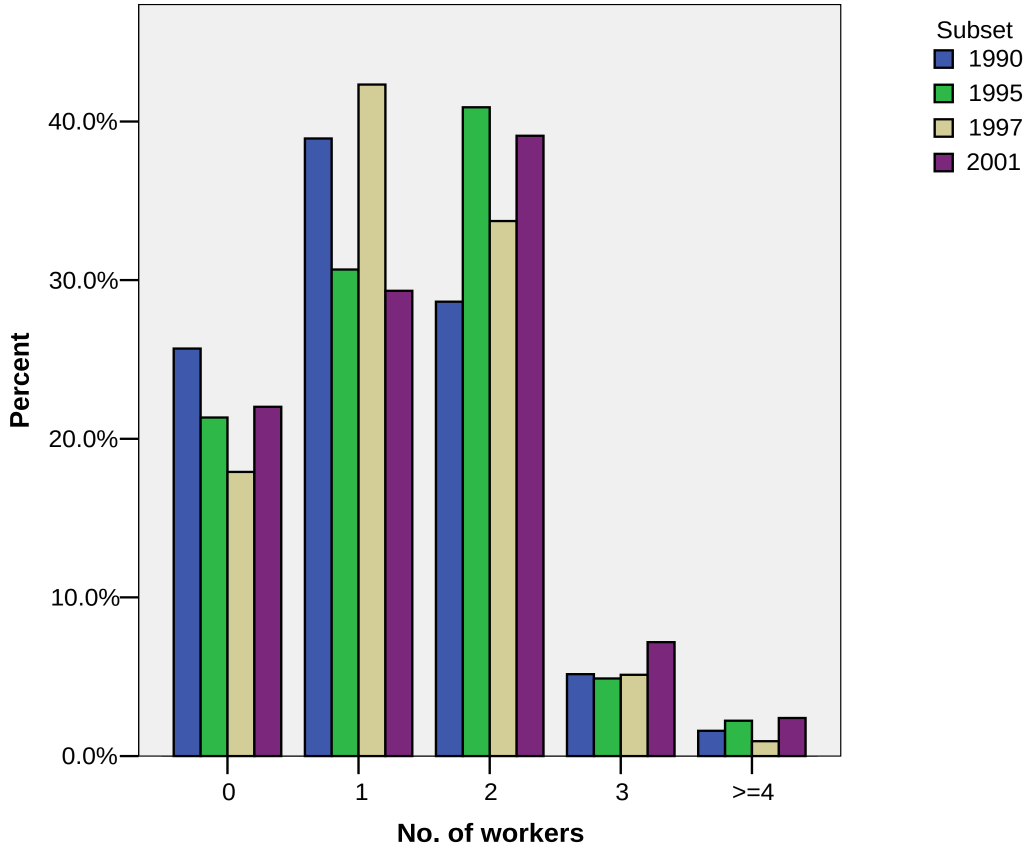


Figure 4: Number of Workers Distribution by Source

Overall, household descriptive statistics indicate each data source is similar with respect to the number of vehicles. The distributions for number of workers, number of drivers, household size and household type show small differences among subsets that may be a result of different travel behavior between subsets, but may simply be a function of sampling error.

4.3 Description of Travel Patterns in the Final Dataset

It is also important to explore the travel data variables before conducting analysis to evaluate if any patterns present in the travel data may impact the results. Furthermore, any temporal patterns in travel behavior may also affect study results.

Many studies use trip purpose to stratify the data and identified the variables that influence travel behavior. Figure 5 shows the distribution of person-trips by purpose for each dataset. All subsets have approximately the same distribution of trips by purpose. About 5-10% of trips are home-based work trips, 25-30% are home-based non-work trips, and about 65% are non-home based trips. As shown, the NPTS/NHTS subsets are the only two subsets that allowed households to respond without identifying a trip purpose. As a result, 11 trips for 1990, 41 trips from 1995 and 32 trips from 2001 subsets reported trips without identifying a trip purpose.

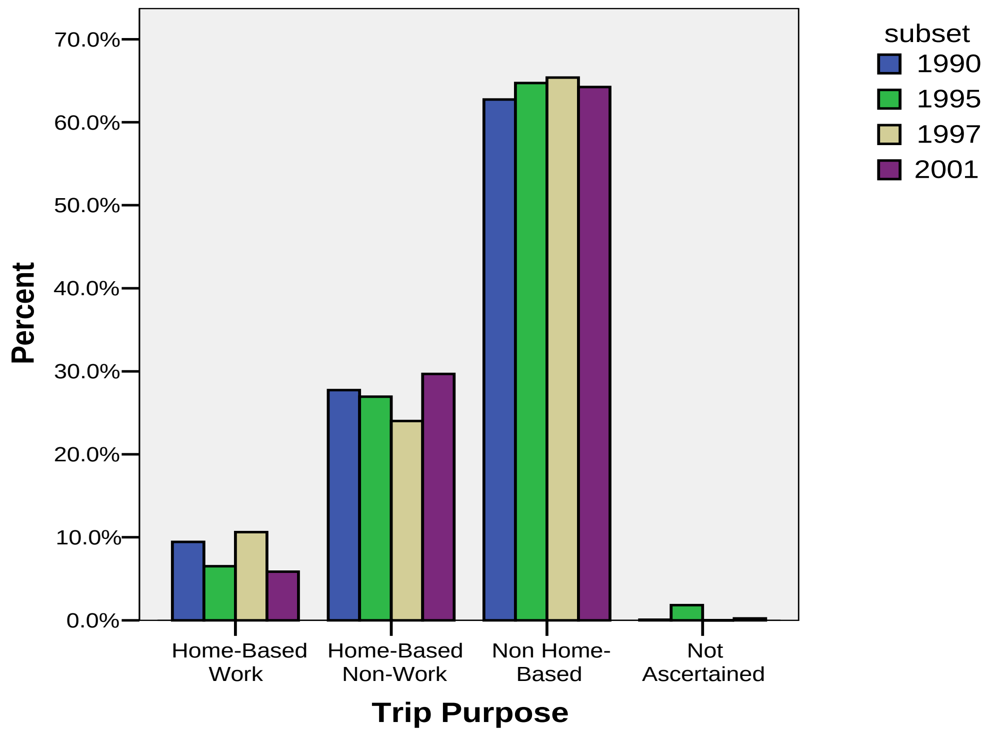


Figure 5: Trip Purpose Distribution by Source

Table 3 illustrates the mean trip rates per household for each data source. The household trip rates range between 6-10 trips per day. The two highest average household trip rate values correspond to the datasets with the highest standard deviation. It appears that if standard deviation were taken into account as a weight in a linear regression model, that there may be a temporal trend present since the 1995 dataset would have lower weights than 1990 and 1997 subsets because it has a larger standard deviation whereas in an unweighted linear regression model each trip rate would have equal weight. Table 3 also indicates the percent of households with zero trips per day. The 1990 dataset has the highest percent of households with zero trips at 13.8%, and 2001 has the lowest with 4.9%. The highest percent of households correspond to the lowest trip rates for each dataset.

Table 3: Household Person-Trip Rate Descriptive Statistics

Data Source		Mean Trip Rate (trips/day)	Standard Deviation – Trip Rate	% Households with zero trips/day
1990 NPTS	(CT Add-on)	6.7	6.0	13.8
1995 NPTS	(CT only)	10.0	7.7	5.8
1997 NYMTC	(NH/FFLD counties only)	8.1	6.6	9.1
2001 NHTS	(CT/New England)	10.1	8.0	4.9

The mean and standard deviation of trip length, along with the number of missing trip lengths are presented in Table 4. For each dataset, a trip is defined as one-way travel from one location to another. As a result, traveling from home to work to the grocery store, and then back home represents three separate trips with three different purposes. The mean trip length is similar for the four datasets (approximately 9-10 miles). The standard deviation for 1990 and 2001 are higher than the other two subsets. The 1997 subset has the least variance, although it is also the subset with the most missing trip lengths at about 6%. Although this high percentage of missing trip lengths, there are still more trip lengths reported in this subset than in the 1995 subset which has nearly twice the standard deviation of the 1995 set. As a result, it is clear that the 1997 subset is the most homogeneous of the four subsets. One reason for a high standard deviation and a higher mean trip length are outlier trips with very long trip lengths. However, the high trip lengths do not appear to be a source of inflated standard deviation, as the 1997 NYMTC data has the highest percentage of trips over 100 miles and also has the lowest standard deviation. Instead, this may reflect the more uniform urban form in the 1997 dataset study area compared to the state wide or regional data. There does not seem to be any temporal patterns present based on Table 4.

Table 4: Trip Length Descriptive Statistics

Data Source		Mean Trip Length	Std Deviation – Trip Length	% of Total Trips Reported with missing lengths (# of trips)	% of Total Trips with lengths over 100 miles (# of Trips)
1990 NPTS	(CT Add-on)	10.1	48.5	1.9 (287)	0.7 (104)
1995 NPTS	(CT only)	9.7	36.1	1.4 (32)	0.8 (18)
1997 NYMTC	(NH/FFLD counties only)	10.4	19.6	6.2 (215)	1.5 (53)
2001 NHTS	(CT/New England)	9.4	48.0	1.6 (232)	0.7 (101)

Figure 6 indicates the mean trip length distributions by data source for home-based work trips. As shown, most home-based work trip lengths are between 0-<5 miles (30-35%). Similar to the results shown in Table 4, the 1997 NYMTC dataset has the highest percentage (16%) of trip lengths of 30 miles and more. This result suggests that many households in the New Haven and Fairfield counties might have a worker that commutes into New York City. There are no differences among data source distributions apparent from Figure 6 that would suggest temporal changes in home-based work trip lengths.

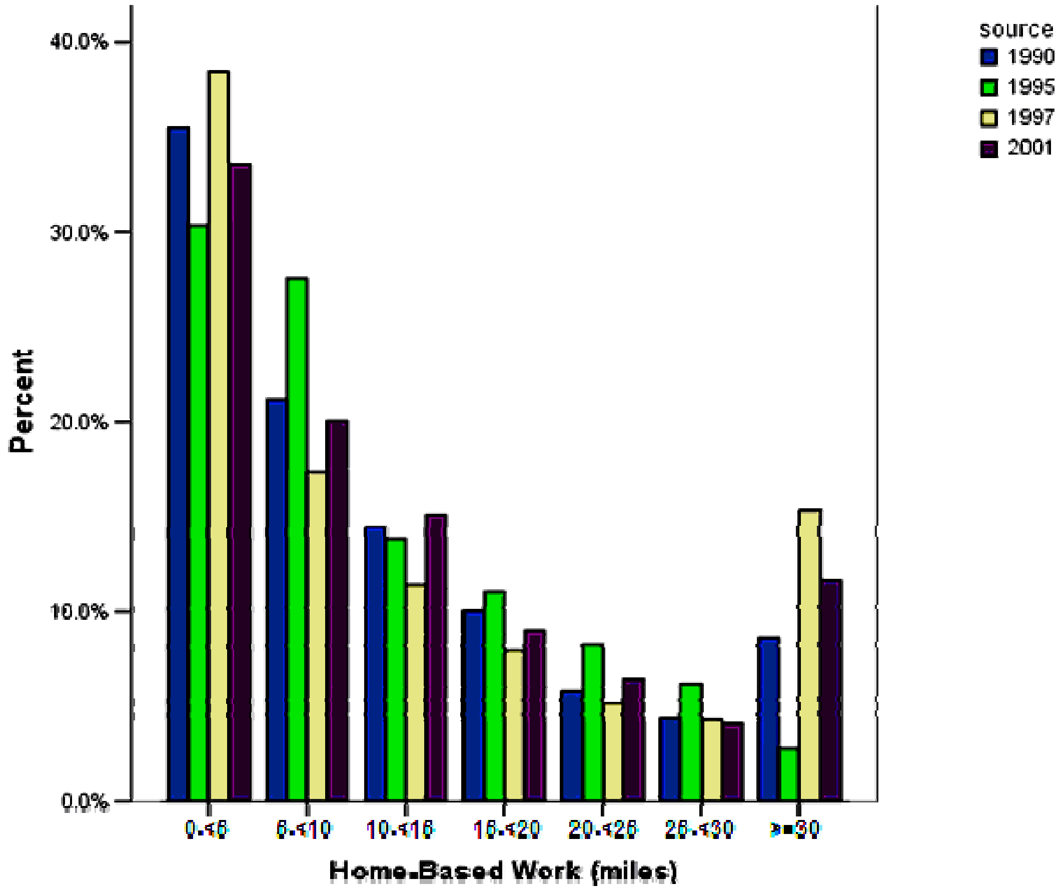


Figure 6: Home-Based Work Trip Length Distributions

Figure 7 presents the trip length distributions by data source for home-based non-work trips. Most home-based non-work trip lengths are between 0-<5 miles. This category includes the majority of reported home-based non-work trip lengths. The distributions decrease sharply as trip lengths increase for each data source. There are no differences among data source distributions apparent from Figure 7 that would suggest temporal changes in home-based non-work trip lengths.

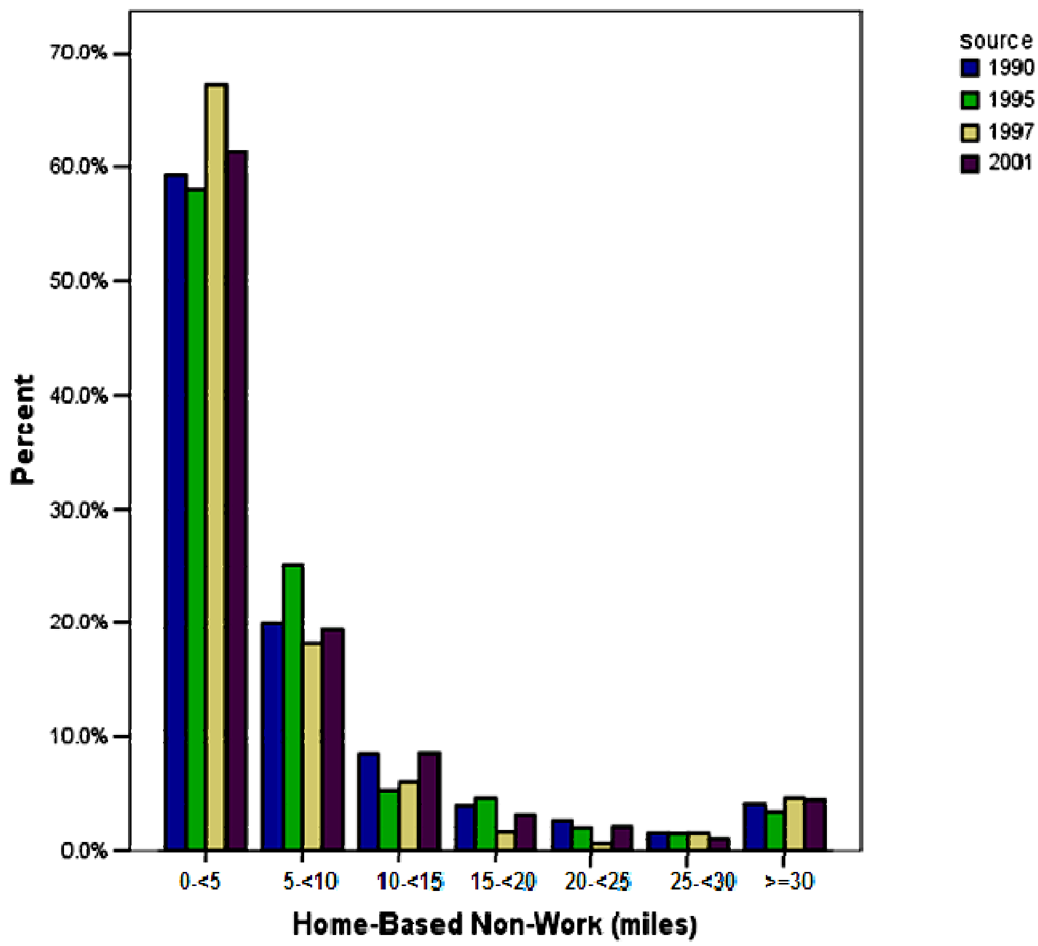


Figure 7: Home-Based Non-Work Trip Length Distributions

Similar to the results for home-based non-work trip lengths, the majority of reported non home-based trip lengths are between 0-<5 miles as shown in Figure 8 below. The distributions in each category for each data source decrease as the trip lengths increase. Once again, the 1997 NYMTC data has the most trips over 30 miles in lengths, accounting for about 10% of reported non home-based trips. There are no differences among data source distributions apparent from Figure 8 that would suggest temporal changes in non home-based trip lengths.

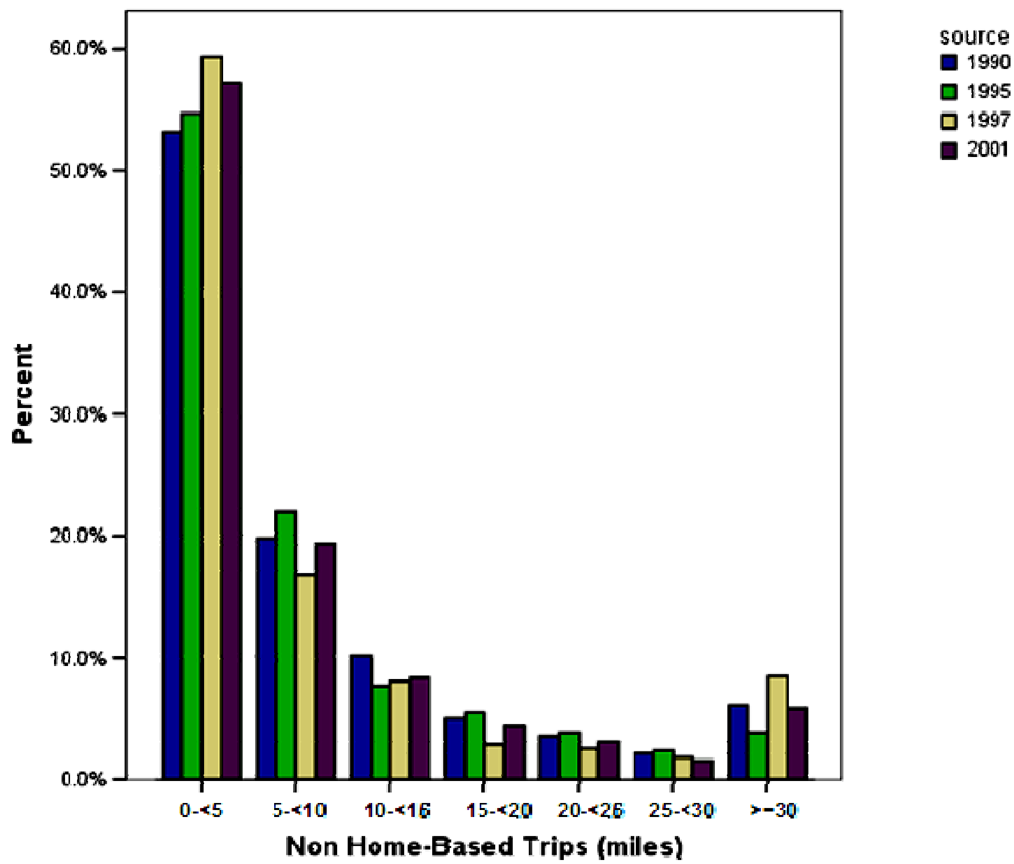


Figure 8: Non Home-Based Trip Length Distributions

4.4 Final Transferability Dataset Limitations

There are many limitations to this transferability study due to the datasets. First, the number of stratifications that can be included when conducting classification & regression trees is limited by the small sample size. This study includes a total sample size of only 4,343 households and 35,201 trips. In order for a stratification group to have a reliable trip rate or trip length predictor, the sample size in the stratified sample should be at least n=30 (Ortuzar et al. 2004).

Another limitation is the variable classification schemes presented in the final dataset. The classification scheme selected for a variable such as income will only be as precise in the final dataset as the most restrictive classification scheme in the subsets. As a result, if a variable is classified based on eight different categories, it is more accurate than a classification scheme based on four categories for that same variable. However, the final dataset is limited to only four categories.

There are also some geographic limitations. The 1990 household locations are identified by zip code rather than census tract. The use of zip code to identify locations is less accurate than the use of census tracts. Census tracts are smaller in area and higher in frequency than zip codes which create a more precise trip origin and destination locations. This impacts the travel distances calculated from the datasets since distances are calculated from the centroid to centroid of the zip code or census tract.

4.5 Transferability Analysis and Results

Linear regression was used to evaluate temporal patterns. The response variable in each linear regression model equation is either trip length or trip rate, and the year of the recorded day trip record minus 1990 is input as the only predictor, so that 1990 is set as year zero. Linear regression is performed in two ways. The first method is a weighted linear regression that uses the aggregated mean trip rate/length for each year and the trip rate/length variance as the weight. This method will result in a more homogeneous model with less variability that will be weighted based on trip rate/length variance within a given year instead of giving more weight to the bigger datasets, which would happen if all disaggregate data was used for estimation. On the other hand, this method may mask some of the real trends in the dataset, since the dependent variables are just the mean of the whole datasets by year. As a result, the second method of linear regression uses the disaggregate data for all years.

Three different linear regression models were estimated for both trip rate and trip length. The first model contains all the NPTS/NHTS, and NYMTC data. The second model includes data sets that contain households from only Connecticut households. This model does not include the NYMTC subset since it does not include households from all over Connecticut. This model is used to evaluate whether there are any temporal trends within the state of Connecticut, by excluding the New England dataset added to the 2001 NHTS data. The third and final model contains all the data from the New Haven and Fairfield counties in Connecticut, since this is the only location which is included in all four subsets. These three different model types are referred to as Model 1, Model 2, and Model 3, respectively, and each is estimated with the aggregated means by subset and with the disaggregated data. Each model is also estimated separately by trip purpose to see if there are temporal trends for some trip purposes that are not present for other trip purposes.

4.6 Temporal Analysis: Trip Rate Results

The results from the three models for the weighted linear regression model suggest that home-based work trip rates should have no adjustments made for temporal patterns, as each of them show insignificant regression models for this purpose. The home-based non-work models suggest that an adjustment for temporal patterns should be made of 0.1 trips per year, as both significant models for models 2 and 3 both show this result and the statistically insignificant home-based non-work model for model 1 shows this same value as well. The non home-based work trip models suggest that an adjustment of 0.2 trips per year should be made, which was significant in models 1 and 3, and is statistically insignificant at 95% confidence for model 2 for this same adjustment. The results of this analysis suggest that there are temporal patterns in the trip rates and adjustments were made to the datasets to account for temporal patterns.

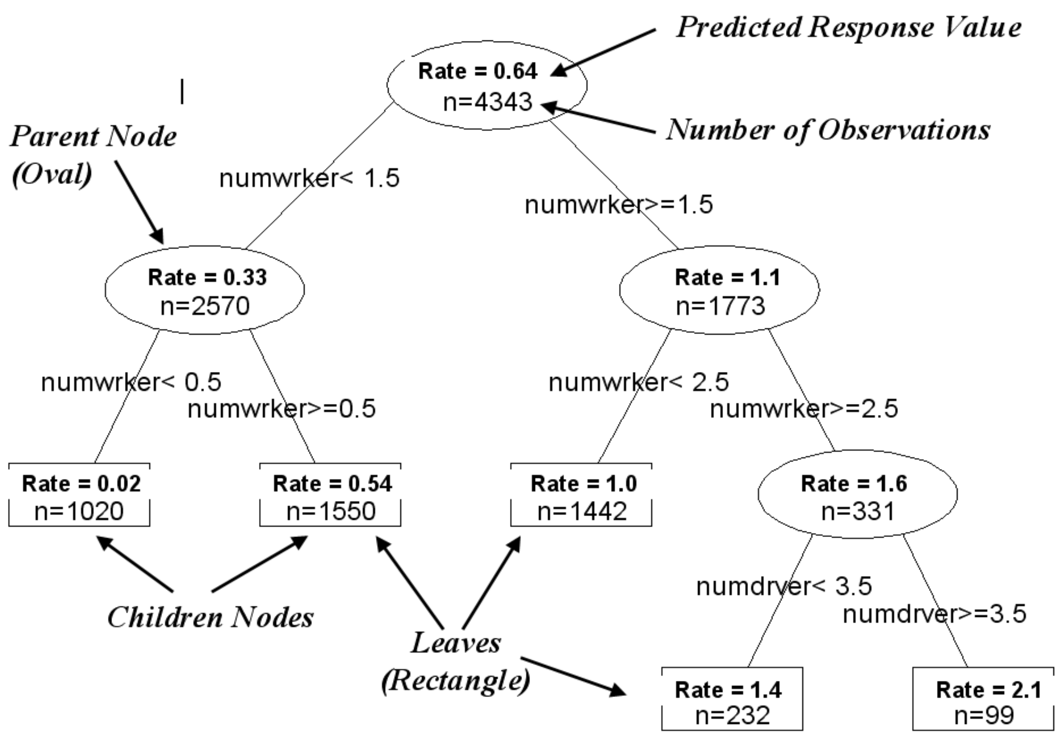
4.7 Temporal Analysis: Trip Length

Using the same methods as in the previous section, regression models imply there were no temporal patterns in trip length distributions. For all the models developed the R-square values are very poor and do not exceed an R-square value of 0.011. This indicated that the variability in trip length is not explained by the year in which the data were collected.

4.8 Classification Scheme for Transferability Analysis

The classification scheme selected is determined using the C&RT method of clustering. Since the response variable (trip rate or trip length) is a continuous variable, a regression tree is created to stratify the sample. Figure 9 shows an example regression tree output and labels the various components of the tree structure. The parent nodes, which are ovals on the tree, are the nodes that are split upon that result in two children nodes. Regression trees only result in binary splits, but one variable can be used for splitting multiple times if it is found to be significant. This was the case with the regression tree result in Figure 9, where the number of workers was split several times. The ultimate predicted response value is found in the leaves of the tree, which are the rectangles in Figure 9. As shown, the regression tree results in several if-then statements for each case that ultimately lead to a predicted response value. For instance, the leaf that is furthest left in Figure 9 gives a predicted value of 0.02 that consists of 1020 observations and is the predicted value for households with zero workers.

Figure 9: Mechanisms of a Regression Tree Output (Labels are in italics)



The regression tree stratifies the sample by exhaustively evaluating every cluster and ultimately stratifying based on the importance of a split and by minimizing deviance within the child nodes and maximizing the deviance between the child and parent nodes. Theoretically, a regression tree could be grown large enough to output every single observed response value in the leaves and thus create a tree with zero deviance and an R-square of 1 if limiting parameters, such as the minimum deviance allowed between splits and the number of observations allowed in each node, are not set.

4.9 Trip Rate Classification Results

The first regression tree results presented are for the home-based work categories. This model shows a standardized R-square value of 0.34, indicating that this cross-classification scheme has 34% more explained variance than the home-based work trip rate tree model with no categories. As shown, as the number of workers in a household increases, so do the number of home-based work trips made per day. The category with no workers has a mean trip rate of 0.0 home-based work trips per day, which logically makes sense as no work-related trips would be made in a household with no workers. The next four categories have slightly fewer trips made than the number of workers in the household. These numbers appear to be fairly low, as one would expect that the number of home-based work trips would be the same as the number of workers in the family. Thus, households with two-workers would be expected to have two home-based work trips. However, the data from the NHTS/NPTS sources included in these estimates are from both weekday and weekend trips, which may lower the trip rates since most people do not work on weekends. The NYMTC data source, on the other hand, only includes data for weekday trips, so it is expected that the NYMTC estimates may be higher than the NHTS/NPTS subsets.

The regression tree developed for the home-based non-work trip rates resulted in a standardized R-square value of 0.28. The five categories selected are based on the socioeconomic variables for household size, number of drivers, and number of children under the age of 18 years. As might be expected, the trip rate increases as the household size increases. When the household size is greater than or equal to 4 people, the number of drivers influences the home-based non-work trip rate as well. Additionally, when there are 2 or more drivers, the number of children in a household is important. This is reasonable since if children are in a household there will be more school trips and trips for other activities. In addition, the number of drivers may influence trip rate since more drivers will offer opportunities for more vehicle-trips, which is the main source of mobility for Connecticut households.

Finally, the tree-based categories for the non-home based trip rates are included in Figure 10. The tree model has a standardized R-square value of 0.26, and uses the same variables for household size and number of drivers that were selected for the home-based non-work trip rate classification, but does not include the number of children in a household as was used for home-based non-work trip rates. There are four categories based on these variables. The category for household size of more than three people with less than or equal to one driver is also a category that was used to stratify the home-based non work trip rate. However, that the mean trips rates are different between the home-based non-work and non home-based trip rate results. Each data source shows an increase in mean trip rate as household size and the number of drivers increase. All datasets have a reasonable number of observations for each category with the exception of the 1995 NPTS and 1997 NYMTC datasets in the category for household size of less than three people with less than two drivers. The standard deviations are fairly similar for each category with the exception of the latter two categories. Similar to the results for home-based non-work trips, the third category for the 1995 NTPS data gives an inflated mean trip rate due to the low numbers of observations (n=3). In this same category the NYMTC data only includes one observation and thus has a standard deviation of zero. In the last category, the 1995 NPTS and 2001 NHTS standard deviations are slightly higher. These high standard deviations will lend to lower weights in the final Bayesian updated trip rate value for these datasets.

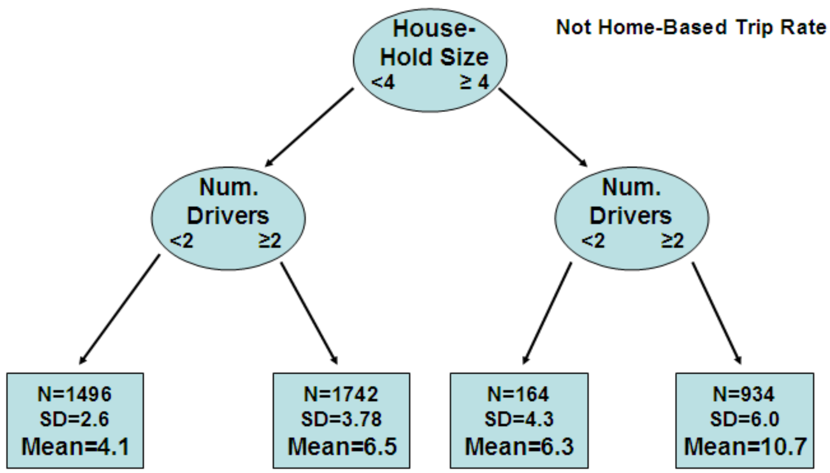


Figure 10: Non-Home Based Classification Scheme Results

4.10 Trip Length Classification Results

Unlike the trip rate results by trip purpose, the trip length regression tree results were not as accurate, as the error resulting from the complexity parameter tables was much higher for the trip length C&RT, and showed little decrease in error as more categories were added. The resulting R-squared values for each trip purpose were less than 0.1. This suggests that trip length is not easily explained by the predictor variables available in this study.

The results found in the previous two sections suggest that trip rate can be explained by the socioeconomic variables input into the model, whereas the socioeconomic and geographic variables were not sufficient at explaining trip length. As a result, no classification schemes were selected for trip lengths and Bayesian updating will be done by trip purpose for each data source to produce final trip length values.

4.11 Bayesian Updating

The main objective of the thesis was to combine several data source in a defensible way to have an updated household trip rate and trip length estimation for use in Connecticut. The final trip rate and trip lengths by trip purpose for each category were combined for each data source to get one final trip rate and trip length value by trip purpose for the categories. This is conducted using Bayesian updating which is typically conducted with only two datasets, consisting of the original trip generation dataset and a new small sample (Ortuzar et al. 2004). The original dataset is referred to as the prior information, and the new small sample is the new information. The parameters calculated in Bayesian updating are the mean trip rate or trip length (θ) and the population variance (σ^2). The parameters included in the Bayesian equation are sample size (n) and the sample variance (S^2) of these datasets. When σ^2 = S^2 /n, the prior parameters are $\sigma^2{}_1$ and θ_1 and the new sample parameters are a σ_s^2 and θ_s. As prescribed by Ortuzar et al. (2004), the equations for Bayesian updating are:

$$\theta_2 = \frac{1/\sigma^2{}_1 \times \theta_1 + 1/\sigma_s^2 \times \theta_s}{1/\sigma_1^2 + 1/\sigma_s^2} \tag{4-1}$$

and $$\sigma_2^2 = \sqrt{\frac{\sigma_1^2 \times \sigma_s^2}{\sigma_1^2 + \sigma_s^2}} \tag{4-2}$$

When substituting in $\sigma^2 = n/S^2$, these equations become

$$\theta_2 = \frac{n_1/S_1^2 \times \theta_1 + n_s/S_s^2 \times \theta_s}{n_1/S_1^2 + n_s/S_s^2} \tag{4-3}$$

and $$\sigma_2^2 = \sqrt{\frac{S_1^2 \times S_s^2}{n_s S_1^2 + n_1 S_s^2}} \tag{4-4}$$

Equation 4-3 shows that the updated mean trip rate or length value will be a weighted by the sample size and the variance of the prior and new datasets, so that the mean value for the dataset with the smaller variance and larger sample size will have more weight in the final mean value calculation.

In this study, however, there are four data sources instead of just two. As a result, the 1990 and 1995 NPTS datasets will be combined the same way as equations 4-3 and 4-4, as follows:

$$\theta_{90/95} = \frac{n_{1990}/S_{1990}^2 \times \theta_{1990} + n_{1995}/S_{1995}^2 \times \theta_{1995}}{n_{1990}/S_{1990}^2 + n_{1995}/S_{1995}^2} \tag{4-5}$$

$$\sigma_{90/95}^2 = \sqrt{\frac{S_{1990}^2 \times S_{1995}^2}{n_{1995} S_{1990}^2 + n_{1990} S_{1995}^2}} \tag{4-6}$$

The 1997 NYMTC dataset parameters are then combined with the new $\theta_{90/95}$ and $\sigma_{90/95}^2$ updated parameters using the following equations, as prescribed by Mahmassani et al. (1981):

$$\theta_{90/95/97} = \frac{\frac{1}{\sigma^{2}_{90/95}} \times \theta_{90/95} + \frac{n_{1997}}{S^{2}_{1997}} \times \theta_{1997}}{\frac{1}{\sigma^{2}_{90/95}} + \frac{n_{1997}}{S^{2}_{1997}}} \tag{4-7}$$

$$\sigma^{2}_{90/95/97=} \sqrt{\frac{\sigma^{2}_{90/95} \times S^{2}_{1997}}{S^{2}_{1997} + n_{1997}\sigma^{2}_{90/95}}} \tag{4-8}$$

These Bayesian mean values will then be compared to the 2001 NHTS mean values to evaluate goodness of fit.

4.12 Bayesian Updating Trip Rate Results

The estimated Bayesian mean and variance values are presented for each trip purpose. The final confidence intervals and standard deviations are calculated using the total number of households in each category for trip rates by trip purpose and the total number of trips in each category by trip purpose for trip length. Goodness of fit is evaluated by comparing the 95% confidence intervals of the mean values for the Bayesian updated categories for each trip rate and length by purpose to the 2001 NHTS Connecticut mean values. The 2001 NHTS data are used to evaluate goodness because they are the most recent actual Connecticut data that the Bayesian updating is representing.

For home-based trip rates the Bayesian variance and Bayesian standard error of the mean results are lower than the 2001 NHTS results, which is mainly due to the large sample size in the Bayesian results. The Bayesian results are statistically the same as the 2001 NHTS dataset results at 95% confidence for four of the five the categories. These results indicate that the Bayesian and 2001 NHTS results are both consistent with each other for these four categories.

The home-based non-work Bayesian results indicate the variance and standard error of the mean values are very low, which is mainly due to the high number of observations. Therefore, as household size, number of drivers and number of children increase, so do the mean trip rates in a household.

For the non home-based trip Bayesian results, as the household size and number of drivers increase, so does the trip rate. The variance and standard error of the mean results are very small, indicating that there is little variability in the predicted mean trip rate values. The 2001 NHTS results indicate mean trip rate increases as the household size and number of driver’s increases. The Bayesian mean trip rates are statistically the same as the 2001 NHTS results in three of the four categories. The last category for households with more than three people and more than one driver has a higher mean trip rate that is a statistically different mean value than the Bayesian updated mean value.

4.13 Bayesian Updating Trip Length Results

Finally, the trip length results are presented by trip purpose. The home-based work trips have the highest mean trip length, whereas the home-based non-work are the lowest trip lengths. The variance and standard error of the mean values are very low, indicating that there is little variability in the mean value. Similar to the Bayesian results, the highest mean trip length is for home-based work trips, whereas the lowest is for home-based non-work trip lengths. The results show that the 2001 NHTS mean trip lengths are statistically the same as the Bayesian updated mean trip lengths, indicating the Bayesian results are consistent with the 2001 NHTS data results.

Overall the 2001 actual trip rates are statistically the same as the updated Bayesian mean results for most categories for trip rates and for all trip purposes for the mean trip length results. These results show that the Bayesian results are satisfactory in predicting trip rate for most of the categories used and for all trip lengths by purpose. One possible reason for differences between the Bayesian and 2001 NHTS results in some trip rate categories may be due to the low numbers of observations in some categories, which may show mean trip rates that do not reflect reality due to the low sample sizes. Another possible reason may be due to the socioeconomic differences that were observed between the 1997 NYMTC data and the NPTS/NHTS datasets for household size and number of drivers. These differences may drive the Bayesian results and cause differences between the 2001 NHTS results and Bayesian results in specific trip rate categories, since these two variables were included in the categorical schemes for home-based non-work and non home-based trip rates.

4.14 Summary of Transferability Results

The analysis in this section demonstrates that the methodology of adjusting for temporal trends, categorizing data using socioeconomic and geographic data, and finally using Bayesian updating to combine the data sources is a feasible way of obtaining updated trip rate and trip length values for regional planners in Connecticut. Some differences were found for certain trip rate categories, which may be attributed to low sample size for the 2001 Connecticut NHTS trip rates, or due to socioeconomic differences between New Haven/Fairfield counties and the rest of Connecticut, which may result in trip rate differences between the Bayesian results and 2001 NHTS results in certain categories. With the exception of the NYMTC data source, the NPTS/NHTS data are available to all U.S. regional planners and proves to be a valuable data tool for travel demand modelers.

Temporal patterns were identified and adjusted for in the home-based non-work and non home-based trip rates, whereas the home-based work trip rates and all trip lengths by trip purposes did not show any indication of temporal patterns. This finding indicates that changes over time vary by trip purpose, which reaffirms the importance of stratifying travel data by trip purpose. If there are no changes over time, then there is no indication that collection of new travel data is really needed. This would imply that we could simply use socioeconomic data to stratify a sample and then take the weighted samples given the new demographic breakdowns in the new time or place. Although the changes found here were for only 2 of 3 purposes and indicated small temporal changes in trip rate, they were present suggesting a need to continue data collection at some level.

The classification schemes were identified using the regression tree models for trip rates and trip lengths by trip purpose. The resulting categories indicated that socioeconomic variables are sufficient in stratifying trip rates by trip purpose, and that the number of drivers, household size and number of children in a household are effective in cross-classifying home-based non-work trip rates, while number of drivers and household size are effective in cross-classifying non home-based trip rates. The tree model indicated that the number of workers alone was effective at predicting home-based work trip rates.

The trip lengths by trip purpose, on the other hand, are not as easy to explain. The models selected using regression trees were very complex and used both geographic and socioeconomic variables to attempt to explain the trip length variability. Even with the multiple categories selected using regression trees, the R-square value showed that there was little improvement in explained variability when compared to a model with no stratifications. As a result, no categories were selected to predict trip lengths, and trip lengths were stratified by trip purpose alone.

The trip length results reflect the complexity of modeling trip length. Suggestions for further investigations include looking into new methods of modeling trip length, as the methodology used in this study of cross-classifying the trip length was not effective in explaining trip length.

One question that this study raises is whether the previous trip taken in a chain of trips affects the next trip taken. Some MPOs, such as NYMTC, use trip chains in the trip distribution step of the four-step travel demand model to attempt to capture the affect the previous trip in a chain of trips has on the next trip. One question this method raises is whether producing trip chains in the trip distribution step would assist in explaining the non home-based trips, as there is a lot of variability in this trip purpose as is shown in this study. If trip chaining does explain more of the variability in trip length, it may be worthwhile for the Connecticut planners to look into this method of modeling trip distribution, as it currently only models trip length by trip purpose and does not recognize the influence of trips in a chain on a single trip.

There are several limitations identified in this analysis. One major limitation to this study was the sample size for each data source. This proved to be a weakness in stratifying the data, as the classification schemes selected produced categories that did not include the ideal number of households. However, the new NHTS surveys will be conducted in 2008 and will be a useful tool for Connecticut planners to validate or further update this trip rate and trip length results even if samples are not collected in Connecticut.

An additional challenge which should be addressed in further studies is the predictor variables included in the final dataset. One difficulty noted in this study is capturing the geographic factors that affect travel behavior in numerical or categorical variable form. It is unclear whether geographic factors affecting travel behavior are associated with household location, employment location, accessibility to highways and roadways, or to locations that are a source of activity. This study used the population density and distance from a household to a CBD as geographic variables to evaluate if travel is affected by a household location's proximity to possible sources of work and activities. Although these geographic factors are shown to not influence travel behavior, past studies in transferability and simulations suggest that geography and urban form affect our travel behavior. More complex geographic predictor variable could be developed given the widespread availability of GIS and GIS data layers.

Overall, the Bayesian updating method of combining data proved to be an effective way of determining trip rates and trip lengths by trip purpose. It may be worthwhile for the Connecticut DOT to purchase an add-on sample for the 2008 NHTS, as this new data in combination with the 1990 NPTS data would provide for a bigger sample size and possibly better results in Bayesian updating that may have shown statistical differences between the Bayesian and 1990 NPTS results, and would avoid low sample sizes that were present in this study. In addition, because temporal patterns were found and these same temporal patterns may not be constant over time, it is important to use the 2008 NHTS data to see what temporal patterns remain present.

5.0 Piloting an Internet-Based Travel Survey for Connecticut

The majority of the second year of this project was focused on developing and executing an Internet survey for travel data that is specific to Connecticut. One of the major benefits of this type of survey is that it could run continuously and if designed properly would require little maintenance and manipulation to collect a large amount of data. However there are limitations and bias in web-based surveys as well. This section will outline the development of the on-line travel survey as well as summarize the results including a comparison to the data sources used in section 4.0.

5.1 Design

The survey design and development began using the case studies and typical variables described in section 3.0 and Appendix A. Draft versions of the on-line survey were pre-tested by members of the project technical advisory committee and members of the transportation group at UConn. Dr. Shin led the design as well as the programming of the survey relatively inexpensive survey software package purchased for the project. While Dr. Shin coded the resulting survey questions into a software package SelectSurvey.NET the team realized the looping and repeating nature of a travel survey created challenges for both the software as well as the School of Engineering server where space to host the survey and its automatically generated database had been provided. The actual design and testing of the survey took much longer than initially expected, primarily due to the survey software. For example, every time Dr. Shin edited a question, it would take 10-20 minutes for the software to update that question and let him move to the next. For this reason we recommend the team develop and program their own website if this survey method is used again.

A copy of the survey questions can be found in Appendix C. The coding used for the actual online survey is over 350 pages long and is not included in this report. The survey loops for every person in the household and every trip. The survey is also branching in that the answers to certain questions result in the participant being guided to different questions. For example, one would only be asked the transit access mode question if one had indicated a trip was taken by transit.

Participants were asked to report all of their own travel for the previous day. Household questions included: address, household size, number of motor vehicles. Personal variables included age, gender, driver's license, worker/student status, as well as how often they personally bike, take transit or use a taxi. Home and school locations were queried. Individuals were asked how long they used the internet each day and about on-line shopping. Many of these questions were included with committee approval and intended to facilitate possible future research by ConnDOT for transportation planning. For each trip during the day the participant was asked to indicate the departure time, length in minutes, mode and destination location/type. The person was asked who they traveled with and whether or not they were the driver if auto was used as the mode. Transit access and egress mode was asked as well as if transit transfers were made. This information was repeated for each trip and then the participant was ask if they felt comfortable and wanted to respond to the survey for another household member.

5.2 Internet Survey Execution

After successful survey development and pilot testing the CT travel survey was made available to the public on April 23, 2007. The web-based survey remained active until May 29, 2007. Volunteers were solicited via email and person communication. Although, several large companies agree to email their employees to take the survey, many of the persons solicited to participate were close friends and family of the research team. It is likely these personal connections increased response rate but also willingness to complete the survey once started. The data collected during this period was non-random and potentially biased due to the necessity of an internet connection to compete the survey. The only runtime problem encountered was that at times too many people tried to access the site and the web page was "annoyingly" slow. This suggests the need to some stagger participation or use multiple or dedicated servers.

5.3 Database Description

The data exported from the survey was tabulated and the dataset was checked for errors. The data were evaluated for missing and erroneous data. In the period of just over a month the online survey was accessed 1146 times. However, each time the survey was accessed was not by a new, unique individual. Using the recorded IP address it appears many of these people would start the survey then quit and come back shortly after to start again perhaps due to the slow response time. Some participants came back and forth multiple times before actually completing the survey presumably due to slow server speed. However, quantifying the number of times this occurred is difficult to do in an automated manner. The main reason is that some respondents did not enter their street residence. Or even if they did enter their street residence, two households on the same street may have completed the survey. Thus making it difficult to determine if two people from the same household started or completed multiple surveys, or if the same person returned multiple times. While the automated collection of IP address aided in determining instances where a respondent returned multiple times to access the survey, there were still situations where a single person used multiple computers (i.e. home and work) to attempt to complete the survey. Therefore, an exact count of the number of times the survey was accessed by a returning respondent is difficult to quantify. Furthermore, it is likely that the volunteers were more tolerant than the general public would be due to their association with the researchers or the University. The possibility of repeated data collection is one advantage to web-based surveys and any future effort should ensure participant burden is minimized.

Figures 11 and 12 display the number of times the survey was attempted and is **not** the number of completed surveys on each day. Figure 11 shows that there were two large peaks in survey attempts, April 25th and May 8th, and a smaller peak on April 30th. The number of times the survey was accessed gradually decrease for the days following these peaks. These peak days correspond to days where mass emails were sent seeking participants. The survey was accessed the most on May 8th where 340 participants attempted to take the survey. Figure 12 contains a histogram of survey access by time of day. All days were aggregated into one histogram to show peak times at which the survey was taken. The time stamp from the collected data were rounded to the nearest 30 minute interval and then plotted above. From Figure 12 the peak number of surveys attempted took place from 9:45 am to 10:15 am. There was a second peak from 12:45 pm to 1:45 pm.

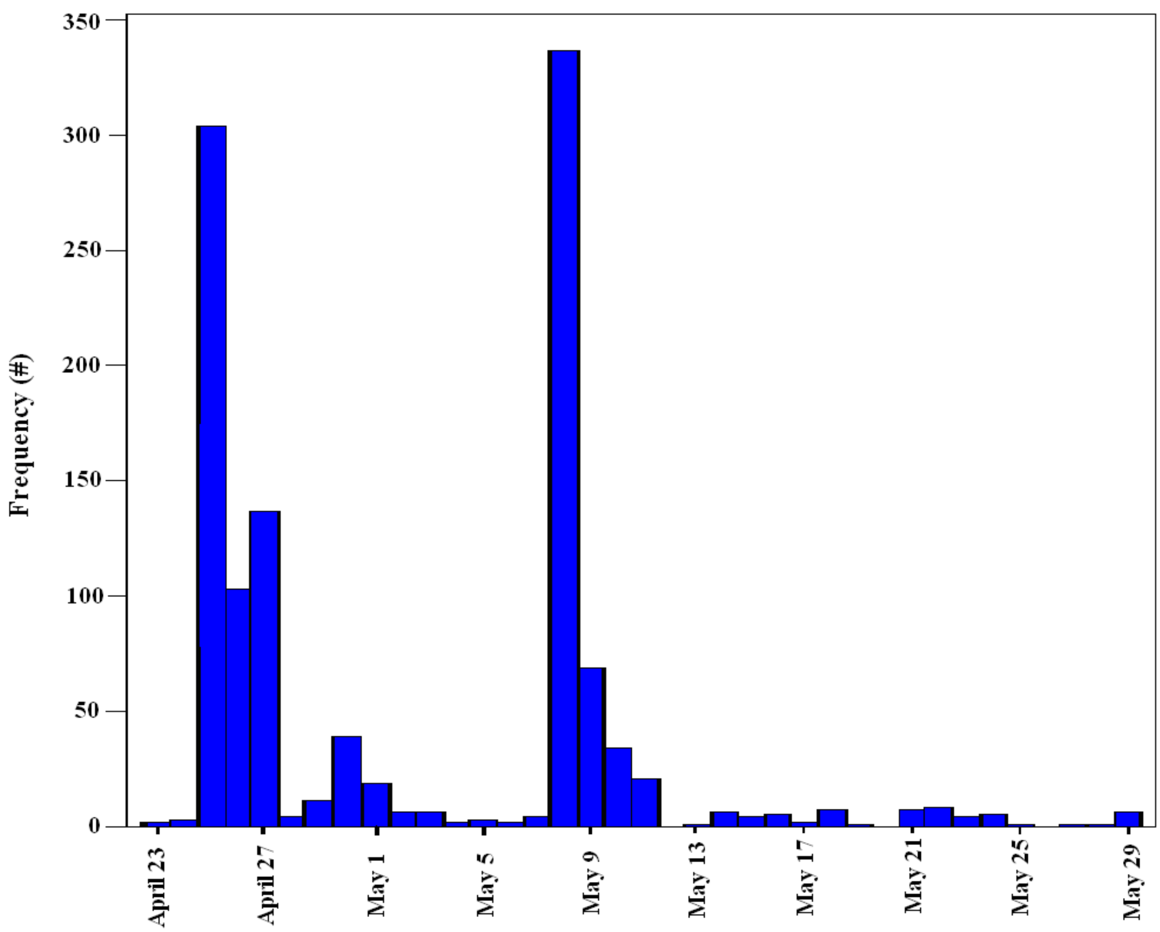


Figure 11: Number of Times the Survey was Accessed by Day (2006)

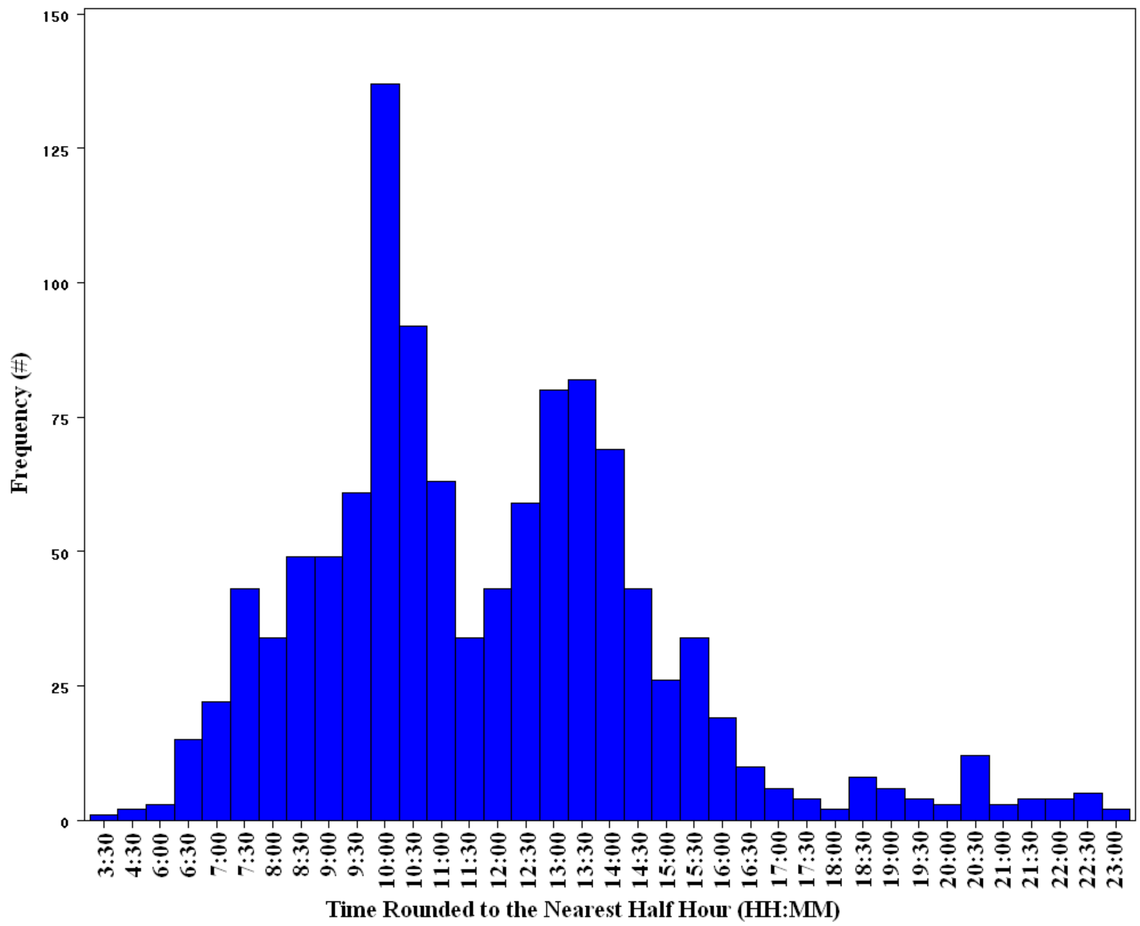


Figure 12: Survey Access by Time of Day

5.4 Personal Dataset Description

The data was divided into two tables for analysis: the person dataset and the trip dataset. For this research the term "Primary" corresponds to data reported by the respondent about themselves and about their travel. "Surrogate" corresponds to data reported by a respondent describing another member of the household and that person's travel. Often one member of the household reports for all members of the household in a travel survey however in this case the surrogate responses were voluntary. Table 5 presents the number of primary and surrogate person records for which data was collected. In total, travel was reported during the survey day for 993 individuals, but 51 individuals did not leave their home on the given day. A total of 475 surveys were started but not completed. A significant number of people reported travel for a surrogate in their household.

Table 5: Number of Primary and Surrogate Persons Surveyed

	Primary	Surrogate	Total
Did not complete person information	455	20	475
Completed personal information but did not leave home the previous day	22	29	51
Completed personal information and reported trips in the second portion of the survey	669	324	993
Total	1146	373	1519

Figure 13 displays the number of respondents by their reported home zip code. From this figure it is obvious that the majority of the survey responses were obtained from persons living in the central portion of the state. This bias in geography could have a significant impact on accuracy and quality of the data obtained. There were a limited number of surveys completed by persons residing in the southwestern portion of the state where congestion on I-95 is significant and many residents may travel greater distances to work in New York City. This non-random sample is not unexpected and is clearly related to the recruiting technique. Figures 14 and 15 indicate the sample had more women than men replying, and more working age individuals (likely due to recruiting via workplaces). But Figures 14 and 15 indicate a good range in the number of vehicles per household and household size was obtained in the survey.

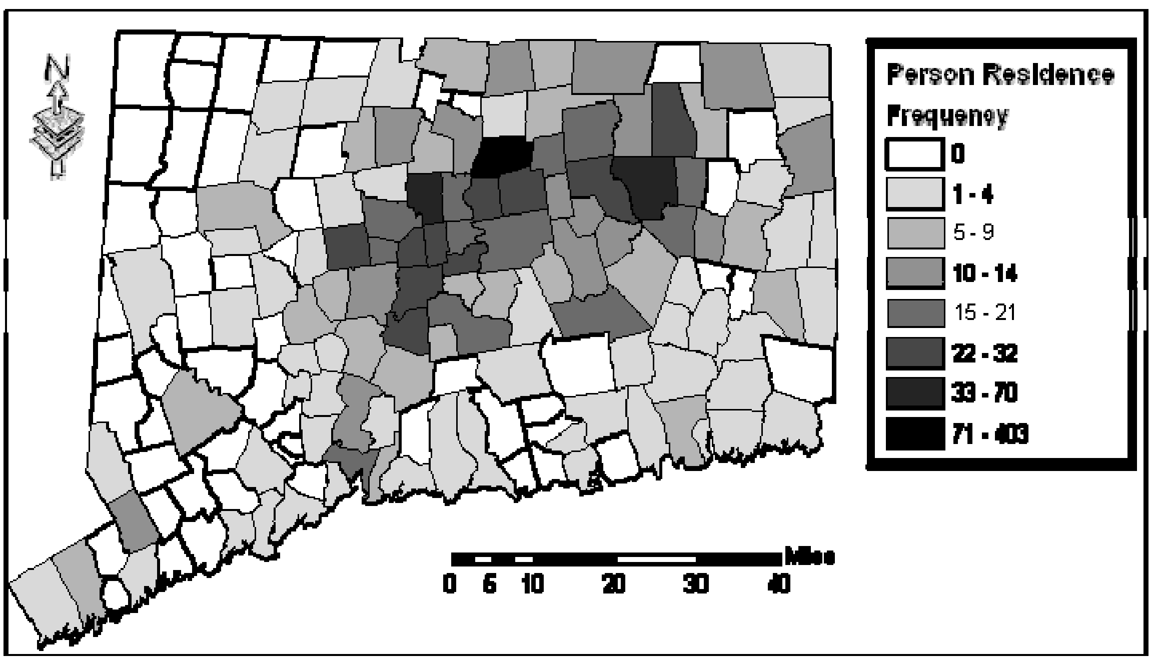


Figure 13: Respondent Home Locations

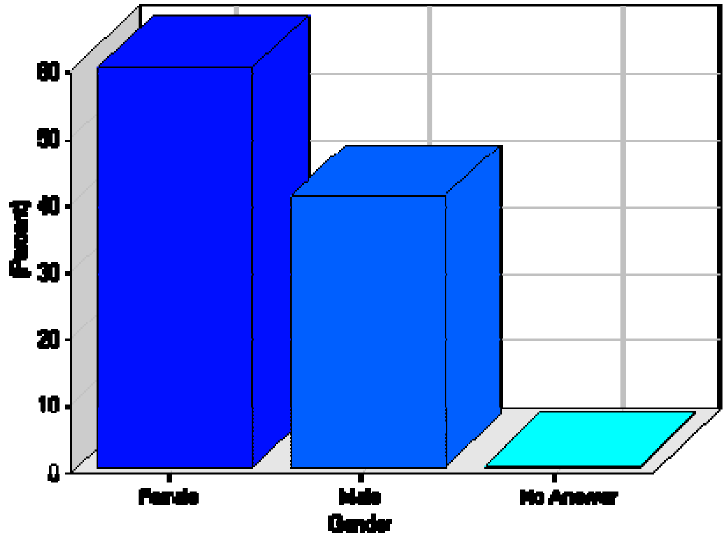


Figure 14: Gender Distribution

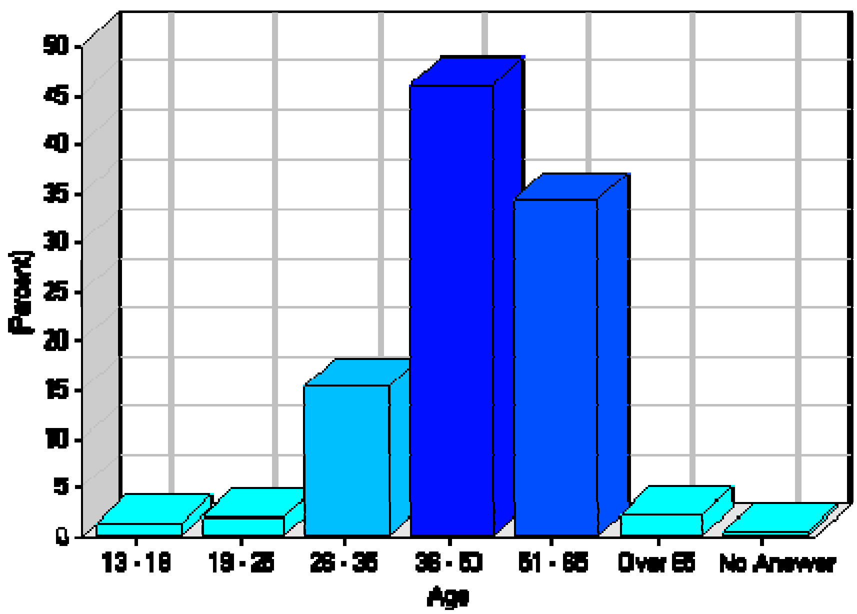


Figure 15: Age Distribution

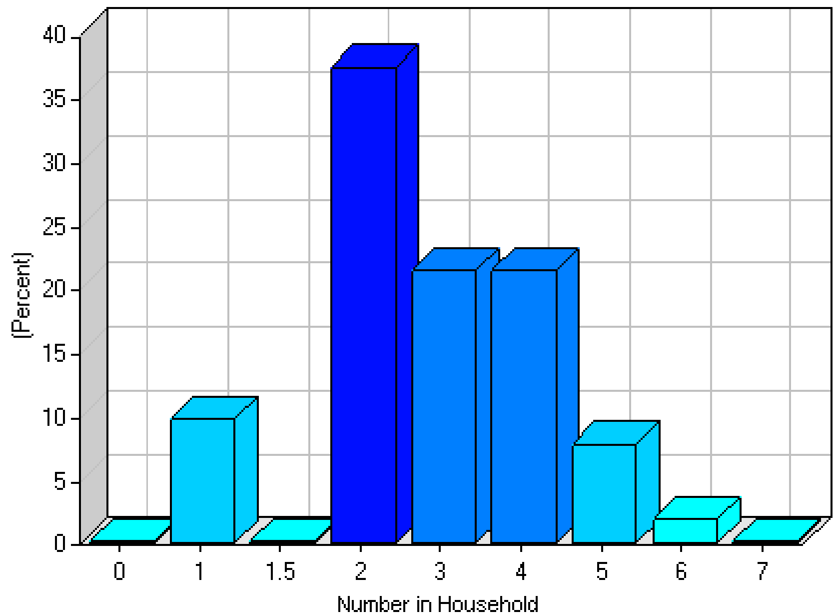


Figure 16: Number of Household Members

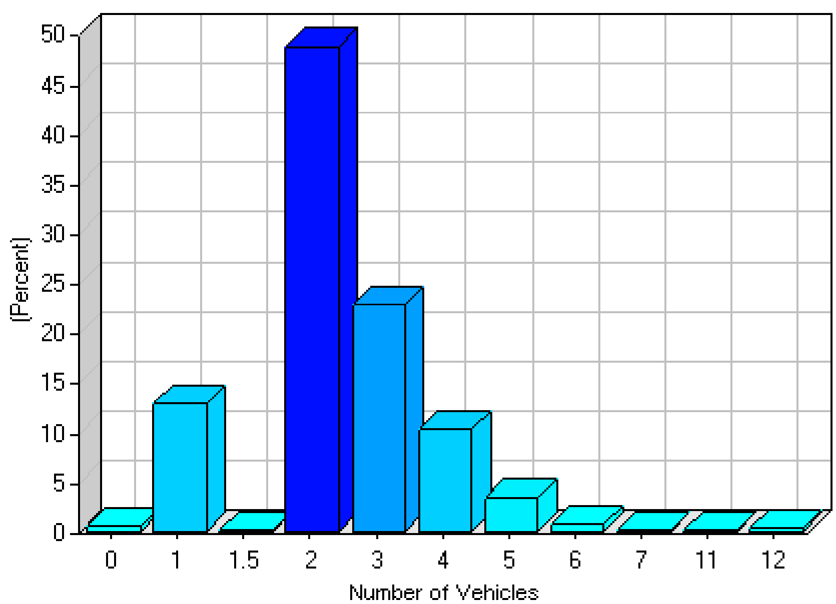


Figure 17: Vehicles in the Household

5.5 Trip Dataset Description

This section will describe the person trip data collected for 669 primary and 374 surrogate individuals. Of these a large portion reported only making one trip (Figure 18). This result is not logical since one trip would only correspond to trip to a destination (primarily the first trip of the day was to work). Therefore, the return trip home was not reported which is typical in diary recording as well. There are several possible explanations for this. The first is that respondents did not understand or consider last trip of the day (home) to be a trip or destination and therefore did not report that trip. A second possibility is that once the respondents realized they would have to recall all their trips for the previous day they opted to quickly quit the survey by stating this was their last trip of the day. Either of these two reasons lead to an under reporting of the number of trips made. Of the 1043 persons that reported making a trip in the trips dataset there were only 470 that reporting "Home" as the destination for their last trip of the day. Thus, 55% of people did not report "Home" as their last trip of the day.

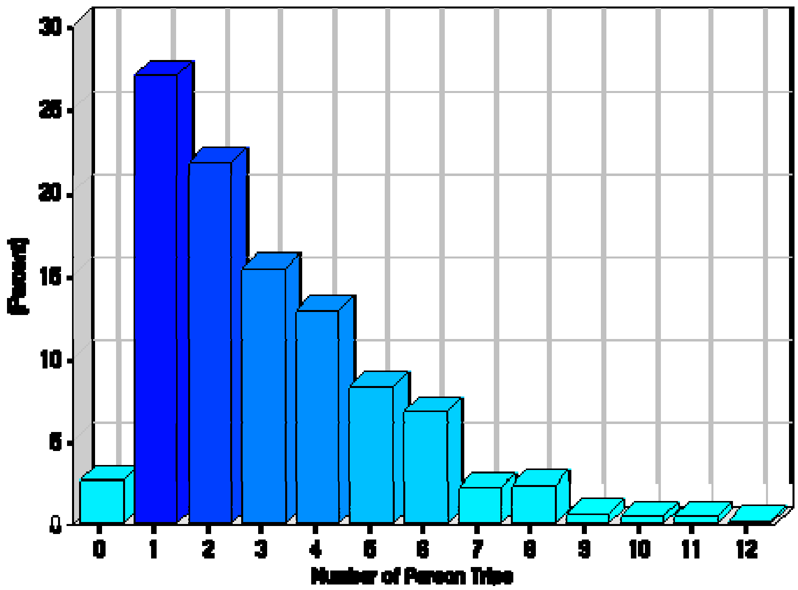


Figure 18: Number of Trips Made by Each Person

To account for persons not reporting their return "home" trip as an actual trip, the data from each person was reviewed to determine if their last trip of the day was made their "Home" location. This is a common response error and care should be taken to address this in future surveys but few robust methods of communicating trips as trip legs have been successful with the public. If this last trip was not a "Home" trip then an additional trip was added to their trip dataset. Trips were added by using their previously reported destination as their new origin and their residence data from the person dataset was used to specify their "Home" destination. These trips which were added during data tabulation were flagged in the trip dataset to mark which trips were added by the research team. Using the newly generated trip dataset a new histogram was created to display the number of trips made by each person (Figure 19). Comparing Figure 19 to Figure 18, persons making only two trips increased by less then 4% while the number of people making 4 trips increased by 13%. This suggests that by adding the home trip, for individuals that neglected to self report, impacted many respondents regardless of how many trips they originally reported. This could indicate that many respondents did not stop the survey early, but many people simply did not report their trip home at the end of the day.

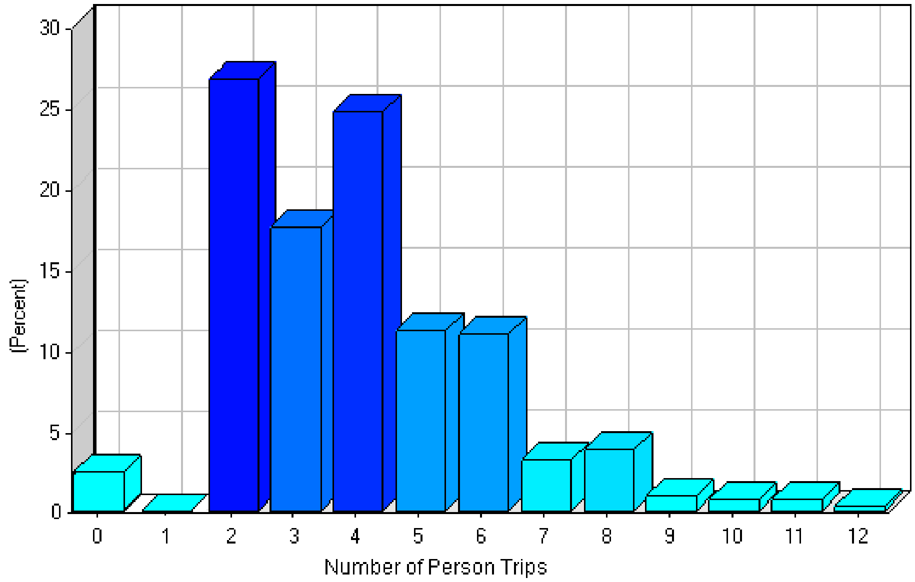


Figure 19: Number of Person Trips Made (corrected)

Once the trips were added to the dataset a trip rate per person could be calculated. Over the course of the survey, data were collected for a total of 1043 individuals and a total of 3483 person trips were reported (including the home trips added by the research team). This corresponds to an average of 3.3 trips per person. However, trip rates are more commonly reported as the number of trips a household make in a day, and not the number of daily trips made by an individual. Therefore, the person trip rates for individuals collected in the web-based survey needed to be converted or adjusted to allow for a comparison between the national datasets, add-on Connecticut dataset, and the regressions and Bayesian analysis presented in the previous section.

Given the nature of the data collected from the web-based survey there is not a defined method to convert individual trip rates to household rates. Simply by using the household ID in the web-based survey there were a total of 679 different households which responded to the survey and 3483 trips. This gives an average of 5.1 trips per household. However, survey respondent were not required to report the trips for every member of their household. Therefore, some household reported trips for up to three family members and some reported only their own trips. Even though there may have been other household members that made trips the previous day. To investigate the impact of surrogate reports on trip rates, trips were aggregated then trip rates were calculated based on the number of persons the respondent reported trips for.

When a respondent only reported trips for themselves the mean trip rate is only 3.6 trips per household (392 households made 417 trips). Respondents that reported trips for themselves and another household member have a mean trip rate of 6.7 trips per household, (248 households made 1681 trips) corresponding to 3.4 trips per person. Finally, when respondents reported trips for themselves and two other household members the average trip rate increased to 10.7 trips per household (36 households made 385 trips). This is an average of 3.6 trips per person. An average household trip rate of 10.7 is comparable to the trip rates calculated from the national dataset and the results of the regression and Bayesian methods in the previous section.

A second method was used to calculate household trip rates for evaluation of survey accuracy. This method calculated household trip rate by multiplying the reported number of people in each household by the mean number of trips calculated per person (3.6 trips per person). The result is a mean household trip rate of 10.1 trips per household and a median of 10.8 trips per household. These results for household trip rate are consistent with both the national study and the transferability study outlined in the previous section.

The analysis of a calculated household trip rate suggests that using the web-based travel study; mean trip rate information can be collected that is consistent with other methods of data collection. Variability by household type and location is not yet considered.

The trip reporting section of the survey was designed to allow the estimation of trip length in distance although participants were asked to report trip length in minutes. Respondents were asked to report each destination they traveled to in the previous day along with the city state and zip code of that destination. By using their previous destination zip code (now origin) and their current destination zip code a trip length was calculated based on the straight line distance between the centroids of each zip code. ArcGIS was used to calculate this “trip length” for each trip reported by the respondents.

If a trip origin and destination were located in the same zip code (intrazonal) ArcGIS would report a trip length of zero miles. Therefore, an alternate method was needed to estimate trip length. Intrazonal trip distances were calculated based on the area of the zip code. Approximating the shape of the zip code as a circle, a radius for each zip code area was calculated. The resulting radius was then used as the trip length for these intrazonal trips. Figure 20 contains a histogram of trip lengths based on the methods described above. Table 6 compares the data collected from the web-based survey to that obtained using the transferability techniques in the previous section. Figures 21 through 23 contain trip length distributions form the web-based survey and the data sources used above in section 4.0. The Figures and Table 6 suggest that the web-based survey has a mean trip length that is comparable to trip lengths obtained using other methods. The last two columns of the table (% missing length and % of trips over 100 miles) are also similar to the results using Bayesian updating and regression analysis. The only metric that is very different from the national dataset is the standard deviation column where most of the other surveys have a standard deviation 3 times that of the web-based survey. Although the 1997 NYMTC has a standard deviation that is in the same range as our web-based survey. The reason for this difference is not known but could be the focus of future analysis.

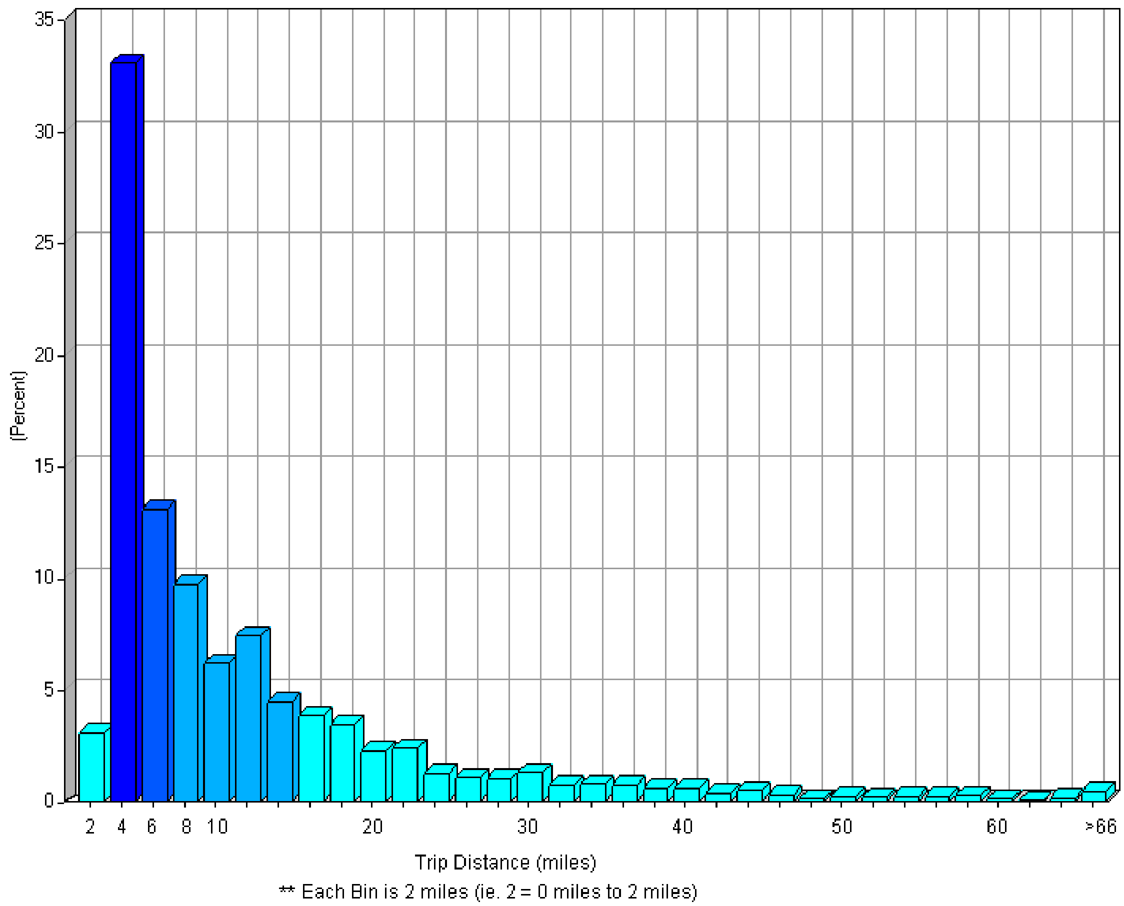


Figure 20: Histogram of Calculated Trip Distance

Table 6: Trip Length Comparisons

Data Source		Mean Trip Length (miles)	Std Deviation – Trip Length (miles)	% of Total Trips Reported with missing lengths (# of trips)	% of Total Trips with lengths over 100 miles (# of Trips)
1990 NPTS	(CT Add-on)	10.1	48.5	1.9 (287)	0.7 (104)
1995 NPTS	(CT only)	9.7	36.1	1.4 (32)	0.8 (18)
1997 NYMTC	(NH/FFLD counties only)	10.4	19.6	6.2 (215)	1.5 (53)
2001 NHTS	(CT/New England)	9.4	48	1.6 (232)	0.7 (101)
2007 Connecticut Web-based Pilot Survey		10.4	11.6	2.0 (70)	0.1 (4)

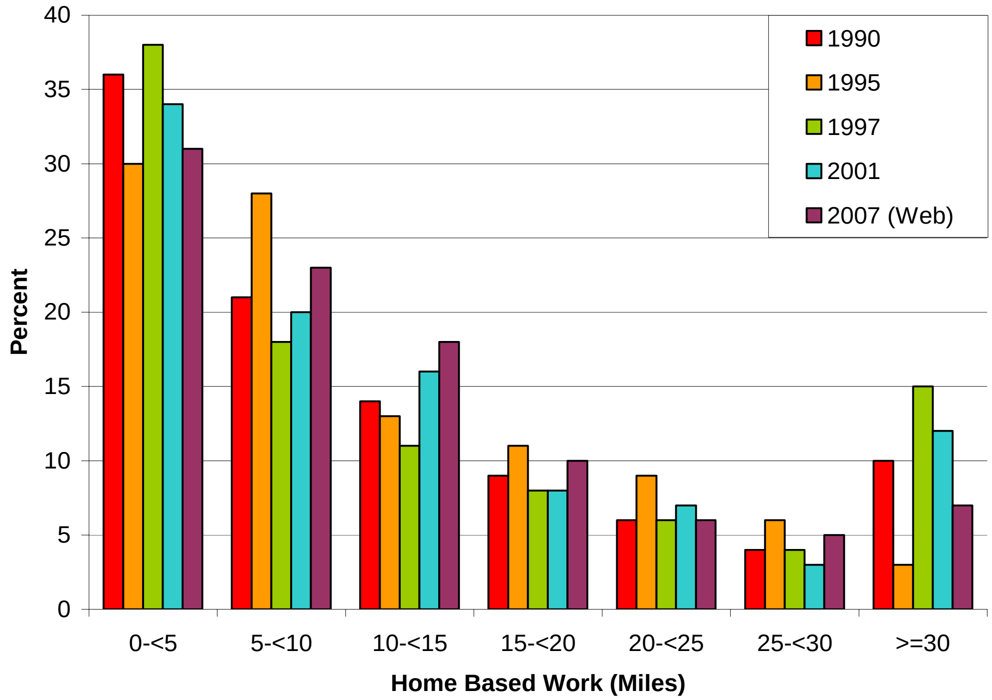


Figure 21: Histogram of Home-based Work Trip Length by Data Source

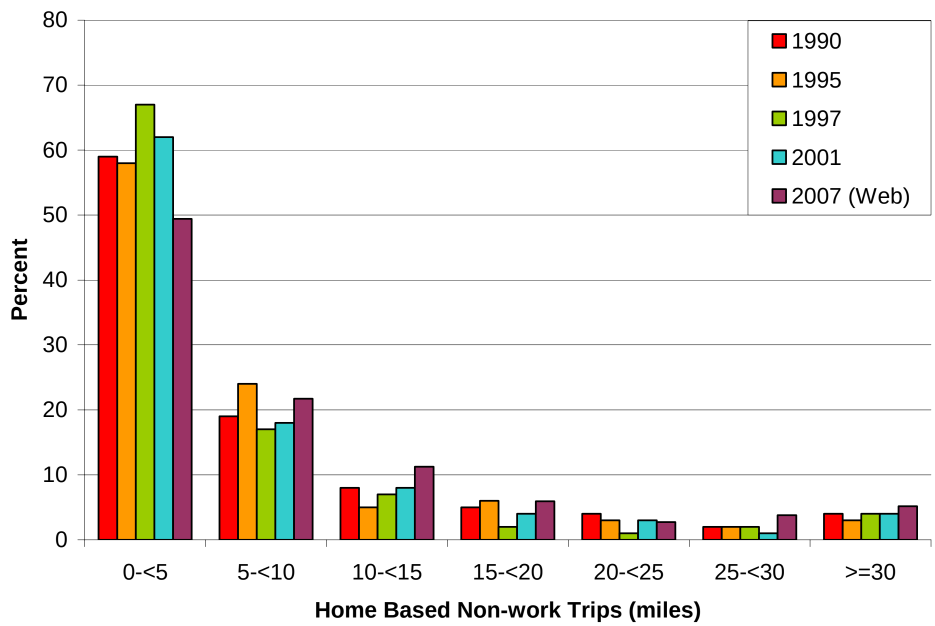


Figure 22: Histogram of Home-based Non-Work Trip Length by Data Source

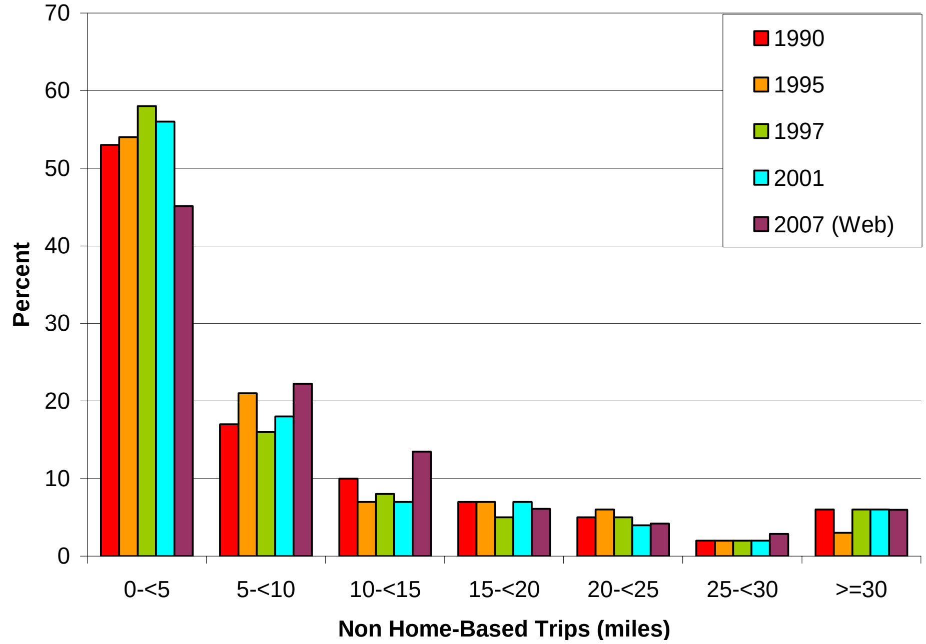


Figure 23: Histogram of Non-Home-based Trip Length by Data Source

While these preliminary results suggest that the web-based survey was able to collect data which were consistent with data collected using other means for the mean variables roundtable participants indicated they were most interested in, further research into the bias and consistency of web-based data is warranted. Moreover, this data collection method had many challenges that were time consuming and difficult to solve. If this method were to be used to collect data on a continuous basis a great deal of effort would be needed to design a survey which was efficient for both survey participants and prospective data users. The software used for this pilot study should not be used for a full scale data collection effort due to the numerous problems with survey design and database development. Although not possible within the resources of this project, evaluation of interactive maps for data collection might also be considered for future study. Overall evaluating of different Internet options might be ultimately very worthwhile, as once an instrument was in place it could be used continuous for years yielding the on-going data planners seek.

6.0 Cost Estimate for a Phone Survey

Using the questions developed for the web-based survey the Center for Survey Research and Analysis (CSRA) at the University of Connecticut was contacted and asked to provide price estimates to conduct the survey by phone. The phone survey could easily target Connecticut residents that are over the age of 18 years. Table 7 contains a cost estimate obtained in 2007 based on the number of completed surveys. Furthermore, for the pricing listed in Table 7, phone surveys would be limited to twenty minutes. The estimated cost does not include survey design but does include data tabulation and some summary statistics. This estimate includes CSRA identifying the sample, reviewing/editing the survey as needed, collecting the data using a Computer Aided Telephone Interviewing (CATI) program, and delivering a weighted dataset in SPSS format. The cost ranges from $32 to $39 per household. This is significantly less than the NHTS add-on cost of $175 per household but note that significantly less data would be collected in a 20 minute call and survey design costs would have to be added. Furthermore, this data would not be geo-coded and would become dated.

Table 7: Phone Survey Cost Estimates

Number of Completed Surveys	Projected Cost
500	$19,500
1000	$39,000
2000	$66,000
3000	$97,000

7.0 Conclusions and Recommendations

The priority needs for planning data were determined using focus group to be household trip rates and trip length distributions. Three viable options exist for the state of Connecticut to collect the type of basic household travel information desired by transportation planners. First, a telephone survey collecting the most basic information could be collected for relatively little cost. Survey design efforts could be minimized by adapting the web-based survey conducted for this project. IN addition to response burden and sampling issues, this survey method would not be on-going, would be limited in scope and could not provide some of the more advanced data options suggested by planners in this project's roundtables including geo-coding of trip ends. Second, a web-based survey could be adapted from the survey in this project and placed on-line for continuous data collection. We would recommend an interactive map be considered for geo-coded trip ends and better trip lengths. Careful attention to recruiting and sample bias would be required. This option would require a moderate level of resources for initial deployment but then become an on-going data source with relatively minimal effort. We recommend the team develop and program their own website in place of off the shelf software if this survey method is used again. Finally, the analysis of national and New York City datasets in this project re-confirms the transferability results found by others. Stratification by very basic household variables such as the number of workers and vehicles yields good mean trip rates and reasonable mean trip lengths. Weighting and transfer of the 2008 NHTS dataset which is about to be collected, might be the most cost effective way to obtain updated basic planning model data input for Connecticut. However, a well designed and managed web-based data source, while suffering from lack of complete random sampling, seems the best approach to create an on-going comprehensive planning, policy and research tool for travel data in Connecticut. Although the team originally proposed consideration of technology approaches to planning data collection including ITS and GPS, evaluation during this project indicated these methods could not provide the most key variables sought by transportation planners. Data mining these ITS and GPS data sources is valuable for other applications and advancing techniques could allow these sources as a complement to a household travel survey.

8.0 References

Abdelwahab, H. T. and Abdel-Aty, M., 2003. “Use of the Internet to Enhance Transportation Data Collection.” Institute of Transportation Engineers 2003 Annual Meeting and Exhibit, Seattle, Washington, Institute of Transportation Engineers.

Antoniou, Constantinos, Moshe Ben-Akiva, Haris N. Koutsopoulos. (2004) “Incorporating Automated Vehicle Identification Data into Origin-Destination Estimation”. 83rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Aultman-Hall, Lisa, Feng Guo, Chris O’Brien, Pat Padlo and Brian Hogge, *Incorporating Truck Flows into the State-wide Planning Traffic Model*. Final Research Report to the Connecticut Cooperative Highway Research Program Report, Draft, September 2004.

Chen, Anthony, Piya Chootinan, and Surachet Pravinvongvuth (2004). “A Multi-Objective Model for Locating Automatic Vehicle Identification Readers”, 83rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Coifman, B. and Y. Yang (2004). “Estimating Spatial Measures of Roadway Network Usage from Remotely Sensed Data.” 83rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Jianhe Du*, Lisa Aultman-Hall, Eric Jackson* and Wael ElDessouki. “Exploring the Models Needed to Convert GPS Point Data to Network-based Travel Route Data” Submitted February 2004 to *Transportation Research Part C*.

Greaves, S.P. and P.R. Stopher, Creating a Simulated Household Travel/Activity Survey B Rationale and Feasibility Analysis, *Transportation Research Record No. 1706*, 2000.

Horowitz, Alan. Guidebook on State-wide Travel Forecasting, Research Report to the Federal Highway Administration Office of the Intermodal and State-wide Programs, July 1999.

ITS America. *ITS as a Data Source*. Workshop Proceedings, Washington, D.C. January 1998.

Ivan, John and U. Jha. *Peak Period Trip Estimation Considering Level of Service and Socio-Economic Characteristics.* Joint Highway Research Council Report #97-258, 1997.

Mahmassani, H.S., and K.C. Sinha. (1981) Bayesian Updating of Trip Generation Parameters. *ASCE Journal of Transportation Engineering*, 107, 581-589.

Mei, Bing, N. R. Bostrom and T. Cooney “Using Bayesian Updating to Enhance 2001 NHTS Kentucky Sample Data for Travel Demand Modeling”, *Journal of Transportation and Statistics*, Vol. 8 No. 3, 2005, BTS.

Mohammadian, Abolfazi, and Yongping Zhang "Investigating the Transferability of National Household Travel Survey Data", Transportation Research Record 1993, 2007, pp67-79.

Murakami, E., and Wagner, D. P., 1999. "Can Using Global Positioning System (GPS) Improve Trip Reporting?". *Transportation Research C*, 7(3), 149-165.

Ochieng, W. Y., and Sauer, K., 2002. "Urban Road Transport Navigation: Performance of the Global Positioning System after Selective Availability". *Transportation Research C*, 10(3), 171-187.

Oh, Cheol , Stephen G. Ritchie, and Shin-Ting Jeng (2004). "Vehicle reidentification using heterogeneous detection systems" 83nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Ortuzar, J.D. and L.G. Willumson. (2004) *Modelling Transport*, 3rd Ed. John Wiley and Son, New York.

Porter, J. David, David S. Kim, Hector A. Vergara (2004). "Electronic Road Pricing Systems: Capabilities, and Existing and Needed Technology Solutions." 83nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Resource Systems Group. (1999) *Computer-Based Intelligent Travel Survey System: CASI/Internet Travel Diaries with Interactive Geocoding.* April.

Stopher, P.R. and Greaves, S.P. "Household travel surveys: Where are we going?" *Transportation Research Part A*, 41 (5), p.367-381, Jun 2007

Stopher, Peter. R., Bullock, P.J., and Horst, F.N.F., 2003. "Conducting a GPS Survey with a Time-Use Diary." 82nd Annual Meeting of the Transportation Research Board, Washington, D.C. Paper No. 03-2154.

Stopher, Peter. R. and Wilmot, Chester. G., 2001. "Development of a Prototype Time-Use-Diary and Application in Baton Rouge, Louisiana." 80th Annual Meeting of the Transportation Research Board, Washington, D.C. Paper No. 01-2351.

Stopher, Peter. R. and Metcalf, H.M.A., 1996. "Synthesis of Highway Practice. Methods for Household Travel Surveys." Washington, D.C., National Highway Research Program.

Tawfik, Ahmed Y., Baher Abdulhai, Aidong Peng, Seyed M. Tabib (2004). "Improving the Accuracy of Vehicle Signature Reidentification Using Decision Trees". 83nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Wolf, J., Loechl, M., Myers, J., and Arce, C., 2001. Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys. *International Conference on Transportation Survey Quality and Innovation*, Kruger Park, South Africa.

Yalamanchili, L., Pendyala, R. M., Prabaharan, N., and Chakravarthy, P., 1999. Analysis of Global Positioning System-based Data Collection Methods for Capturing Multi-Stop Trip-Chaining Behavior. *Transportation Research Record*, 1660, 58-65.

Zhang, Yongping, and Kouras Mohammadian, “Microsimulation of Household Travel Survey Data”, Presented at the January 2008 Annual Meeting of the Transportation Research Board, Washington D.C.

Zhang, Yongping, and Kouras Mohammadian, “Bayesian Updating of Transferred Household Travel Data Using Markov Chain Monte Carlo Simulation with Gibbs Sampler”, Presented at the January 2008 Annual Meeting of the Transportation Research Board, Washington D.C.

Appendix A: Travel Variables Collected in Case Studies

DATA TYPES	VARIABLE NAMES	Surveyed / Generated	Priority			Used in 1976/1977 CT HH Survey
			ESSENTIAL	IDEAL	NOT ESSENTIAL	
	HH id	X	X			
	Receipt of advance letter				X	
	HH location		X			X
	Coordinates of address			X		
	Phone number	X				
	Types of dwelling units			X		
	Owner / renter			X		
	Tenure at current residence				X	
	Ethnicity				X	X
	Bike usage last summer				X	
	Bike usage this summer			X		
	Travel date		X			
	Travel day of week		X			X
	GPS volunteer				X	
	Importance of transportation infrastructure components				X	
	Rating of transportation infrastructure components				X	
	HH size		X			X
	HH composition / life cycle indicator		X			
	Age of each HH member			X		
	Sex of each HH member			X		
	HH income			X		X
	Number of telephones				X	
	Relationship of each HH member to the HH respondent (Only relevant if only person is surveyed)			X		
	All individuals in HH are related?				X	
	Number of motorized vehicles		X			X
	Number of vehicles by types			X		
	Number borrowed HH vehicles				X	
	Driver status of each person		X			
	Work status of each person		X			
	Location of work		X			
	Student status for each person		X			
	Number of HH trips on trip day		X			

Appendix A (continued)

DATA TYPES	VARIABLE NAMES	Surveyed / Generated	Priority			Used in 1976/1977 CT HH Survey
			ESSENTIAL	IDEAL	NOT ESSENTIAL	
	HH id	X	X			
	Person id	X	X			
	Respondent status			X		
	Age		X			X
	Gender		X			X
	Mobility disability status		X			X
	Health condition				X	
	Education level			X		
	Driver status		X			X
	Employment status		X			X
	Occupation type			X		
	No work trip on travel day		X			
	Mode to work trip (Relevant only if no work trip is made on survey day)			X		
	Work parking cost			X		
	If worker: Employer location		X			
	If worker: typical work trip				X	
	If worker and driver: drive as part of work			X		
	Possession of a transit pass			X		
	Student status of person		X			
	School location		X			
	Number of trips made by the individual on trip day		X			
	Year		X			
	Body type		X			
	Number of cylinders			X		
	Fuel type used			X		
	Parking places				X	
	Parking cost				X	
	Time owned		X			
	Annual miles driven			X		
	Primary driver		X			
	Odometer readings			X		

Appendix A(continued)

DATA TYPES	VARIABLE NAMES	Surveyed / Generated	Priority			Used in 1976/1977 CT HH Survey
			ESSENTIAL	IDEAL	NOT ESSENTIAL	
	HH id	X	X			
	Person id	X	X			
	Trip id	X	X			
	Start time of each trip		X			X
	Arrival time at each destination		X			Did trip start at home?
	Trip duration		X			Did trip end at home?
	Activity duration (types)			X		
	Primary mode of the trip		X			
	Trip purpose (origin type)		X			Trip purpose
	Trip purpose (destination type)		X			
	Trip purpose (origin location)		X			X
	Trip purpose (Destination location)		X			X
	Mode of transportation		X			
	HH vehicle used		X			
	If HH vehicle used: which vehicle			X		If traveled by auto: Who was a driver?
	If someone else on trip: any HH members			X		If traveled by auto: number of people
	If HH members on trip: which HH members			X		If traveled by auto: parking costs
	If someone else on trip: any non-HH members			X		If traveled by auto: was transit available?
	If non-HH members on trip: how many			X		
	HH id	X	X			Distance to a public transit stop
	Person id	X	X			
	Trip id	X	X			
	Access mode		X			
	Egress mode		X			
	Access location		X			
	Egress location		X			
	Mode used		X			
	Number of transfers		X			
	Trip purpose				X	
	Mode				X	
	Travel time				X	
	Travel distance				X	
	Time of day when the trip started				X	
	Day of week when the trip started				X	
	If a private vehicle trip: vehicle occupancy; driver characteristics (age, sex, worker status, education, etc); vehicle attributes (make, model, year, annual mileage, etc)				X	
	If public transit: access and egress modes				X	
	Overnight stops, transportation mode and stop purpose				X	
	What was the activity?					
	Where did it take place?					
	When did activity start?					
	Did you have a vehicle available?					
	Parking costs, if any?					
	How long did it take?					
	Were you already there?					
	How did you get there?					
	Number in party					
	Start / end time					
	Bus trip information					

Appendix B: Summary of Connecticut Transportation Institute Previous Research with GPS Travel Data Collection

Between 2001 and 2005, Ms Jianhe Du completed her PhD at UConn under the direction of Dr. Aultman-Hall with funding through CTI. The overall objective of this research was to collect real world travel route data using Global Positioning System (GPS) receivers and to develop the models needed to use these data in route choice and other travel behavior research. To achieve the goal three specific analyses were conducted. First, a GIS model was developed to divide the data stream recorded by the in-vehicle GPS receivers into individual trips with the start and end point of the trip being specifically identified. Second, a spatial model was developed to change the typology of the routes (or trips) from representation as a series of points into a series of continuous network links. Automating this data processing will allow analysis of larger datasets for more generalizable results. Third, travel time on each road link in the whole network was estimated using the sparse sample of GPS travel data (256 vehicles each for 10 days spreading over the 18 month study period) as travel time probes. This model is necessary so that the complete and complex decisions faced by the drivers for each trip are known by researchers. The first two analyses are directly relevant to JHRAC 05-7 in that division of the GPS data into trips would be necessary to calculate both the trip rates and trip length of interest to the project's roundtable participants.

One specific unique aspect of this work was that data for calibration and evaluation of models were available. With this calibration data the UConn CTI study was able assess the accuracy of GPS trip rates which was not necessary possible in other studies. Dividing the GPS accurately requires a more complex algorithm than used by others and accuracy is not foolproof. The best trip dividing model correctly identified 94% of the trips. The accuracy level of the point-to-link data conversion model was 95%.

In this research, in-field travel route data were collected in Lexington, Kentucky March, 2002 to July, 2003 from 256 households. Each data collection cycle lasted for ten days, including 2 weekends and weekdays. This is a relatively large dataset compared to previous work, both in terms of the number of vehicles as well as the amount of time the GPS devices were left in the vehicles. This required significant labor both for field installation as well as pre and pst survey data collection. While the goal of this research was to automate the data processing needed for routing study, this method of data collection is still labor intensive.

Special features were built into the data collection to allow for the calibration and validation of the techniques proposed. This included an in-vehicle booklet for travel log data collection. Overall participants did not complete the booklets thoroughly and the GPS receivers had more complete information even given the post-processing necessary as described above.

Appendix C: Internet Survey Script

CONNECTICUT ONLINE HOUSHOLD TRAVEL SURVEY (2-1-07)

PERSON DATA SURVEY

INTRODUCTION

Are you willing to participate in this study?

→If Yes go to #1

→If No survey ends

HOUSEHOLD INFORMATION

1. Where is your home located? (City and Zip)
2. In what state is your home located?
3. How many people live in your household?
4. How many motorized vehicles does your household have?

INFORMATION ABOUT YOU

5. Gender
6. What age group do you belong to?
   - 0-5
   - 6-12
   - 13-18
   - 19-25
   - 26-35
   - 36-50
   - 51-65
   - >65
7. Do you have a driver's license?
8. Approximately how often do you use a bicycle?
   - Every day
   - More than THREE times a WEEK
   - TWICE a WEEK
   - ONCE a WEEK
   - TWICE a MONTH
   - ONCE a MONTH
   - Less than ONCE a MONTH
   - TWICE a YEAR
   - ONCE a YEAR
   - NEVER
9. Approximately how often do you use public transit?
   - Every day
   - More than THREE times a WEEK
   - TWICE a WEEK
   - ONCE a WEEK
   - TWICE a MONTH
   - ONCE a MONTH
   - Less than ONCE a MONTH
   - TWICE a YEAR
   - ONCE a YEAR
   - NEVER
10. Approximately how often do you use a taxi?
    - Every day
    - More than THREE times a WEEK
    - TWICE a WEEK
    - ONCE a WEEK
    - TWICE a MONTH
    - ONCE a MONTH
    - Less than ONCE a MONTH
    - TWICE a YEAR
    - ONCE a YEAR
    - NEVER
11. Do you work?

→ If Yes go to #12

→ If No go to #16

WORK INFORMATION

12. Where is your main work location? (city, Zip)
13. In what state is your workplace located?
14. On average how many days a week do you go to work?
15. On average how many days a week do you work at home?

---

16. Are you a student?

→ **If Yes** (Part-time Student, Full-time Student)

→ **If No go to #19**

17. Where is your school located? (city, Zip)
18. In what state is your school located?
19. Do you have an internet connection at home and / or work?

→ **If Yes**

→ **If No go to Trip Survey on next page**

INTERNET USE INFORMATION

20. On average how many hours a day do you use an internet at home and / or work?
21. How often do you shop online?
    - Every day
    - More than THREE times a WEEK
    - TWICE a WEEK
    - ONCE a WEEK
    - TWICE a MONTH
    - ONCE a MONTH
    - Less than ONCE a MONTH
    - TWICE a YEAR
    - ONCE a YEAR
    - NEVER
22. What items do you used to buy online?
    - Books / CDs / Movies
    - Clothing (for adults and children)
    - Electronics / Computers / Appliances / Camera
    - Grocery
    - Home furnishing (furniture, bed, etc)
    - Toys
    - Other

TRIP DATA SURVEY

INFORMATION ABOUT YOUR TRAVEL YESTERDAY

Did you leave your home YESTERDAY?

→ If No, go to "Thank you page and ask if the participant is comfortable answering question on behalf of other household members"

→ If Yes go to #23

[[[1st TRIP OF THE FIRST HOUSEHOLD MEMBER]]]

INFORMATION ABOUT YOUR TRAVEL YESTERDAY

23. At what time did you FIRST leave your home? (HH:MM)
24. Where did you go?
    - Home
    - Workplace
    - Pick-up location for passengers
    - Drop-off location for passengers
    - Restaurant
    - School
    - Stores for shopping
    - Medical office / facilities (not for work)
    - Park / recreational facilities
    - Personal business / errands
    - Private residence (not my home)
    - Religious / spiritual facilities
    - Airport
    - Other

    ****STORE ANSWER AS "DEST1"****
25. Where was "DEST1" located? (City, Zip)

26. In what state is the “DEST1” located?

27. How did you get to the “DEST1”?
    - Bike
    - Bus
    - Car (alone)
    - Car (with other passengers)
    - Train
    - Walk
    - Other

→IF MODE WAS TRAIN/BUS GO TO #31
→IF MODE WAS CAR ANSWER # 28 AND #29 THEN SKIP #40
→Others go to #40

28. How many other people were in the car?
29. Were you the driver?
30. Did someone in your household travel with you?
31. How did you get to the train station or bus stop?
32. How many other people were in the car?
33. Were you the driver?
34. Did someone in your household travel with you?
35. How many transfers did you make to get to the “DEST1”?
36. How did you get to the “DEST1” after getting off the train or bus?
37. How many other people were in the car?
38. Were you a driver?
39. Did someone in your household travel with you?
40. How long did it take to get to the “DEST1”?
    - Less than 10 minutes
    - 10-20 minutes
    - 20-30 minutes
    - 30-40 minutes
    - 40-50 minutes
    - 50-60 minutes
    - 60 minutes or more
41. Was it your last trip YESTERDAY?

→ If Yes the go to Question # 200
→ If No go to #42

42. When did you leave the “DEST1”? (HH:MM)
43. Where did you go NEXT?
    - Home
    - Workplace
    - Pick-up location for passengers
    - Drop-off location for passengers
    - Restaurant
    - School
    - Stores for shopping
    - Medical office / facilities (not for work)
    - Park / recreational facilities
    - Personal business / errands
    - Private residence (not my home)
    - Religious / spiritual facilities
    - Airport
    - Other

****STORE ANSWER AS “DEST2”****

[[[2ND TRIP OF THE FIRST HOUSEHOLD MEMBER]]]

44. Where was “DEST2” located? (City, Zip)
45. In what state is the “DEST2” located?
46. How did you get to the “DEST2”? (choose from list)
    - Bike
    - Bus
    - Car (alone)
    - Car (with other passengers)
    - Train

Walk
Other

→IF MODE WAS TRAIN/BUS GO TO #51
→IF MODE WAS CAR ANSWER # 47 AND #48 THEN SKIP #59
→Others go to #59

47. How many other people were in the car?
48. Were you the driver?
49. Did someone in your household travel with you?
50. How did you get to the train station or bus stop?
51. How many other people were in the car?
52. Were you the driver?
53. Did someone in your household travel with you?
54. How many transfers did you make to get to the "DEST2"?
55. How did you get to the "DEST2" after getting off the train or bus?
56. How many other people were in the car?
57. Were you a driver?
58. Did someone in your household travel with you?
59. How long did it take to get to the "DEST2"?
    Less than 10 minutes
    10-20 minutes
    20-30 minutes
    30-40 minutes
    40-50 minutes
    50-60 minutes
    60 minutes or more
60. Was it your last trip YESTERDAY?

SURVEY LOOPS BACK TO #42 UNTIL RESPONDENT ANSWERS THIS WAS THEIR LAST TRIP OF THE DAY. THE VARIABLE "DEST#" IS REPLACED IN EACH LOOP BUT STORED IN THE DATASET FOR ANALYSIS.
ONCE RESPONDENT ANSWERS THIS WAS THEIR LAST TRIP OF THE DAY, GO TO # 200.

200. Thank you for providing information about your travel. The next questions are about other household members. Do you feel confident that you can describe YESTERDAY's travel patterns for another member of your household?

→If Yes go to #310
→If No go to #999

[[[SECOND HOUSEHOLD Member Person INFO]]]

310. Relation to the participant
311. Gender of "2nd HH Member"
312. What age group does "2nd HH Member" belong to?
    0-5
    6-12
    13-18
    19-25
    26-35
    36-50
    51-65
    >65
313. Does "2nd HH Member" have a driver's license?
314. Approximately how often does "2nd HH Member" use a Bicycle?
    Every day
    More than THREE times a WEEK
    TWICE a WEEK
    ONCE a WEEK
    TWICE a MONTH
    ONCE a MONTH
    Less than ONCE a MONTH

TWICE a YEAR
ONCE a YEAR
NEVER

315. Approximately how often does “2nd HH Member” use Public Transit?
Every day
More than THREE times a WEEK
TWICE a WEEK
ONCE a WEEK
TWICE a MONTH
ONCE a MONTH
Less than ONCE a MONTH
TWICE a YEAR
ONCE a YEAR
NEVER

316. Approximately how often does “2nd HH Member” use a Taxi?
Every day
More than THREE times a WEEK
TWICE a WEEK
ONCE a WEEK
TWICE a MONTH
ONCE a MONTH
Less than ONCE a MONTH
TWICE a YEAR
ONCE a YEAR
NEVER

317. Does “2nd HH Member” work?

→ If Yes go to #318
→ If No go to # 322

WORK INFORMATION ABOUT THE SECOND HOUSEHOLD MEMBER

318. Where is the main work location? (city, Zip)
319. In what state is the workplace located?
320. On average how many days a week does “2nd person” go to work?
321. On average how many days a week does “2nd person” work at home?

SCHOOL INFORMATION ABOUT THE SECOND HOUSEHOLD MEMBER

322. Is “2nd person” a student?
→ if Yes (Part-time Student, Full-time Student) go to #323
→ if No go to #325
323. Where is the school located? (city, Zip)
324. In what state is the school located?

325. Does “2nd person” have INTERNET ACCESS at home and / or work?
→If Yes go to #326
→If No go to #329

INTERNET INFORMATION ABOUT THE SECOND HOUSEHOLD MEMBER

326. On average how many hours a day does “and person” use INTERNET at home and / or work?
327. How often does “and person” shop online?
Every day
More than THREE times a WEEK
TWICE a WEEK
ONCE a WEEK
TWICE a MONTH
ONCE a MONTH
Less than ONCE a MONTH
TWICE a YEAR
ONCE a YEAR
NEVER

328. What items does “2nd person” used to buy online?
Books / CDs / Movies
Clothing (for adults and children)
Electronics / Computers / Appliances / Camera
Grocery
Home furnishing (furniture, bed, etc)
Toys

Other

[[[TRAVEL SURVEY FOR 2nd HH MEMBER]]]

TRAVEL INFORMATION OF THE SECOND HOUSEHOLD MEMBER

329. Did "2nd HH member" leave your home YESTERDAY?
    → If Yes go to #330
    → If No go to #200

TRAVEL INFORMATION OF THE SECOND HOUSEHOLD MEMBER

330. At what time did "2nd person" FIRST leave your home? (HH:MM)
331. Where did "2nd person" go?
    Home
    Workplace
    Pick-up location for passengers
    Drop-off location for passengers
    Restaurant
    School
    Stores for shopping
    Medical office / facilities (not for work)
    Park / recreational facilities
    Personal business / errands
    Private residence (not my home)
    Religious / spiritual facilities
    Airport
    Other

STORE ANSWER AS "2_DEST1"

[First Trip for HH member 2]

332. Where is the "2_DEST1" located? (City, Zip)
333. In what state is the "2_DEST1" located?
334. How did "2nd HH Member" get to the "2_DEST1"?
    Bike
    Bus
    Car (alone)
    Car (with other passengers)
    Train
    Walk
    Other
    → if Car go to # 334 and #335 then skip to #346
    → if Bus/Train go to #337
    → Else go to #346
335. How many other people were in the car?
336. Was "2nd HH Member" the driver?
337. Did someone in your household travel with "2nd HH Member"?
338. How did "2nd HH Member" get to the train station or bus stop?
    Bike
    Bus
    Car (alone)
    Car (with other passengers)
    Train
    Walk
    Other
339. How many other people were in the car?
340. Was "2nd HH Member" the driver?
341. Did someone in your household travel with "2nd HH Member"?
342. How many transfers did "2nd HH Member" make to get to the "2_DEST1"?
343. How did "2nd HH Member" get to the "2_DEST1" after getting off the train or bus?
344. How many other people were in the car?
345. Was "2nd HH Member" a driver?
346. Did someone in your household travel with you?

347. How long did it take to get to the “2_DEST1”?
    - Less than 10 minutes
    - 10-20 minutes
    - 20-30 minutes
    - 30-40 minutes
    - 40-50 minutes
    - 50-60 minutes
    - 60 minutes or more
348. Was it the last trip of “2nd HH Member” YESTERDAY?
    → If Yes Go to #200
    → If No go to #349

[[[2nd trip of the 2nd person]]]

349. When did “2nd HH Member” leave “2_DEST2”? (HH:MM)
350. Where did “2nd HH Member” go NEXT?
    - Home
    - Workplace
    - Pick-up location for passengers
    - Drop-off location for passengers
    - Restaurant
    - School
    - Stores for shopping
    - Medical office / facilities (not for work)
    - Park / recreational facilities
    - Personal business / errands
    - Private residence (not my home)
    - Religious / spiritual facilities
    - Airport
    - Other

****STORE ANSWER AS “2_DEST2”****

[Second Trip For Household member 2]

351. Where is the “2_DEST1” located? (City, Zip)
352. In what state is the “2_DEST1” located?
    How did “2nd HH Member” get to the “2_DEST1”?
    - Bike
    - Bus
    - Car (alone)
    - Car (with other passengers)
    - Train
    - Walk
    - Other

    → if Car go to # 353 and #354 then skip to #365
    → if Bus/Train go to 338
    → Else go to #365
353. How many other people were in the car?
354. Was “2nd HH Member” the driver?
355. Did someone in your household travel with “2nd HH Member”?
356. How did “2nd HH Member” get to the train station or bus stop?
357. How many other people were in the car?
358. Was “2nd HH Member” the driver?
359. Did someone in your household travel with “2nd HH Member”?
360. How many transfers did “2nd HH Member” make to get to the “2_DEST1”?
361. How did “2nd HH Member” get to the “2_DEST1” after getting off the train or bus?
362. How many other people were in the car?
363. Was “2nd HH Member” a driver?
364. Did someone in your household travel with you?
365. How long did it take to get to the “2_DEST1”?
    - Less than 10 minutes
    - 10-20 minutes
    - 20-30 minutes
    - 30-40 minutes
    - 40-50 minutes
    - 50-60 minutes

60 minutes or more

366. Was it the last trip of “2nd HH Member” YESTERDAY?

SURVEY LOOPS BACK TO #349 UNTIL RESPONDENT ANSWERS THIS WAS “2ND HH MEMBERS” LAST TRIP OF THE DAY. THE VARIABLE “#DEST#” IS REPLACED IN EACH LOOP BUT STORED IN THE DATASET FOR ANALYSIS.

ONCE RESPONDENT ANSWERS THIS WAS THEIR LAST TRIP OF THE DAY. THEY ARE ASKED IF THEY WOULD LIKE TO REPORT TRIPS FOR ANOTHER HH MEMBER. IF YES THEY ARE RETURNED TO #310 TO FILL OUT MORE INFO ABOUT ANOTHER HH MEMBER. THIS LOOPS UNITL PERSON DOES NOT WANT TO REPORT ANY MORE TRIPS FOR ANY OTHER HOUSHOLD MEMBERS. THEN GO TO #999.

Page 877

999. Thank you for participation

Page 888

Evaluation Survey

Appendix D: Histograms of Select Travel Data Questions in Web Survey

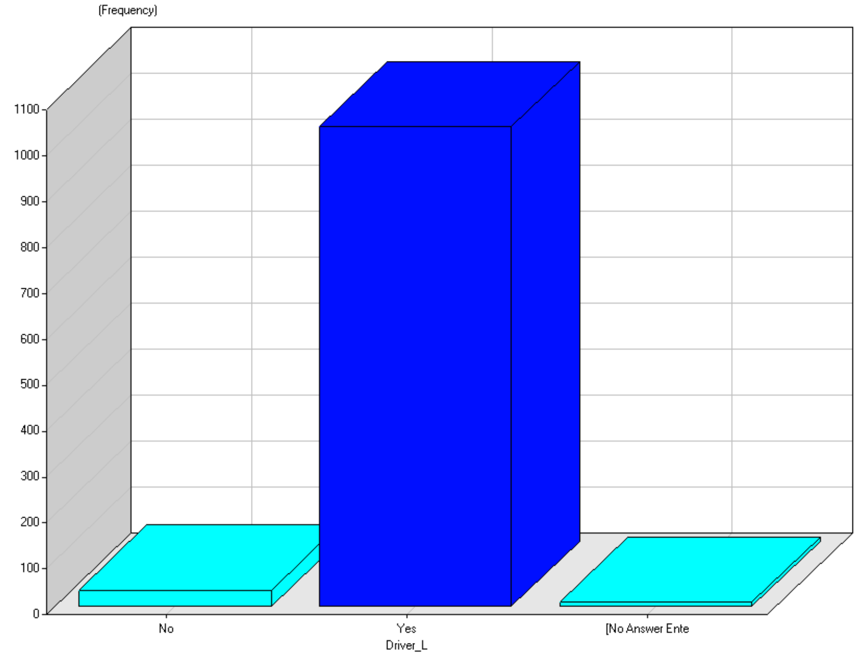


Figure C-1: Do You Have a Drivers License?

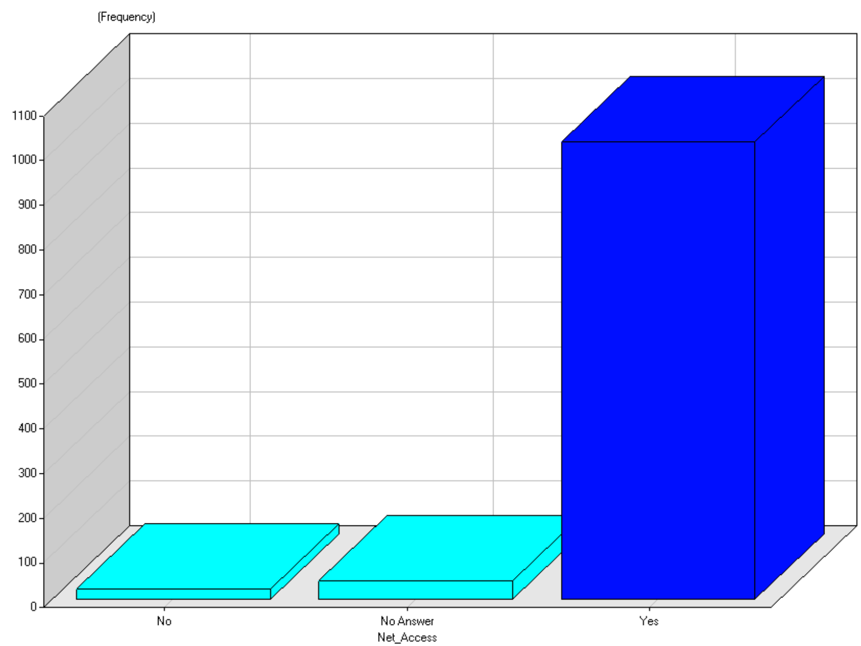


Figure C-2: Do You Have Internet Access at Home/Work?

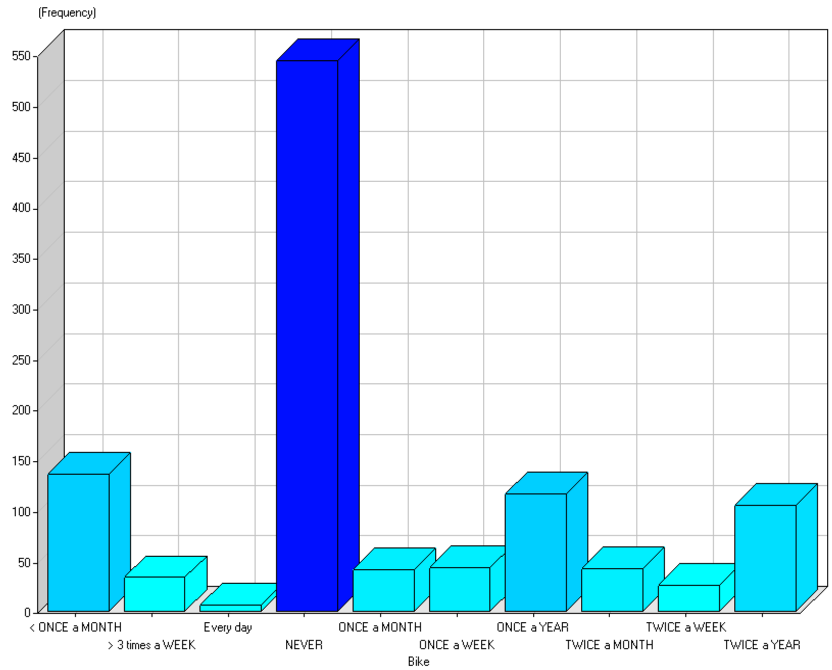


Figure C-3: How Often Do You Ride A Bike?

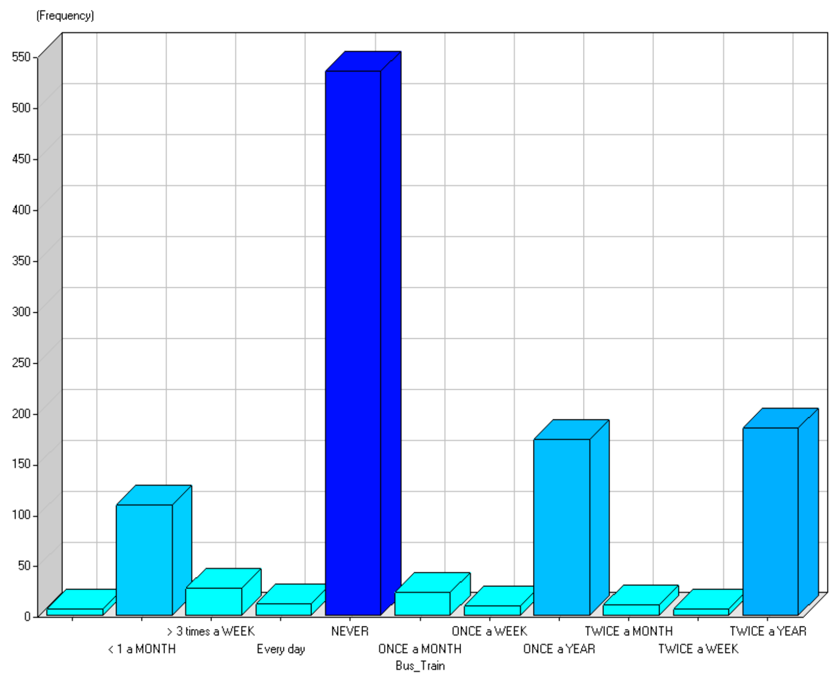


Figure C-4: How Often Do You Take A Bus Or Train?

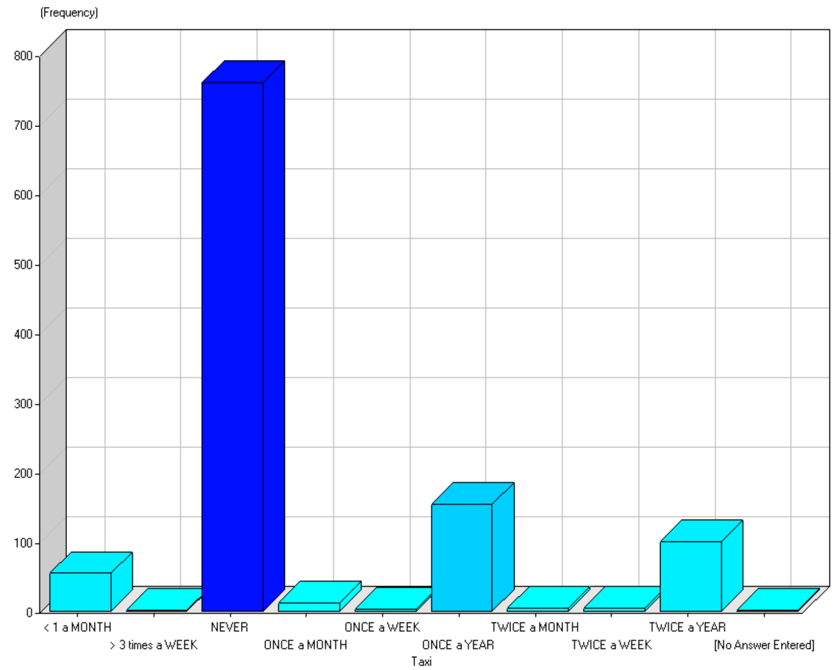


Figure C-5: How Often Do You Take A Taxi?

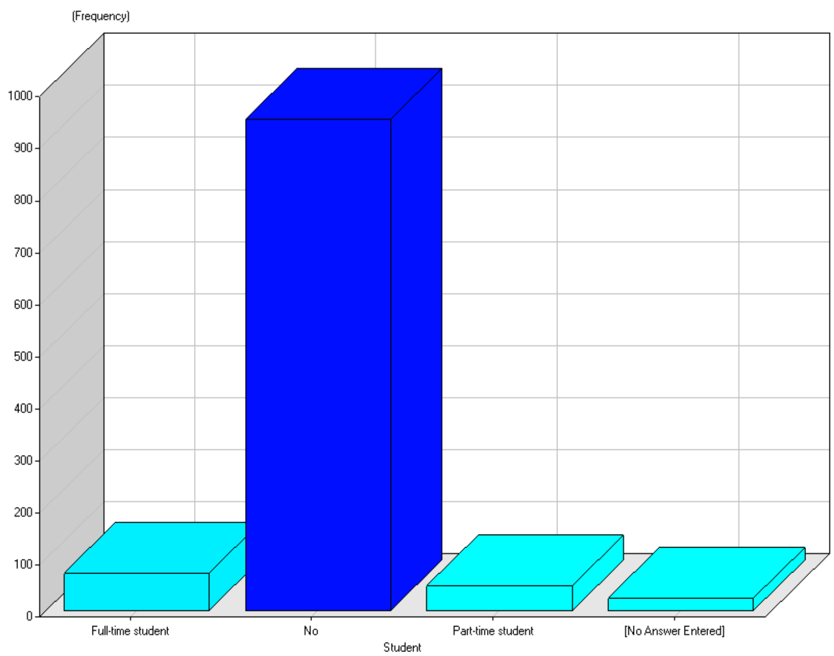


Figure C-6: Are you a Student?

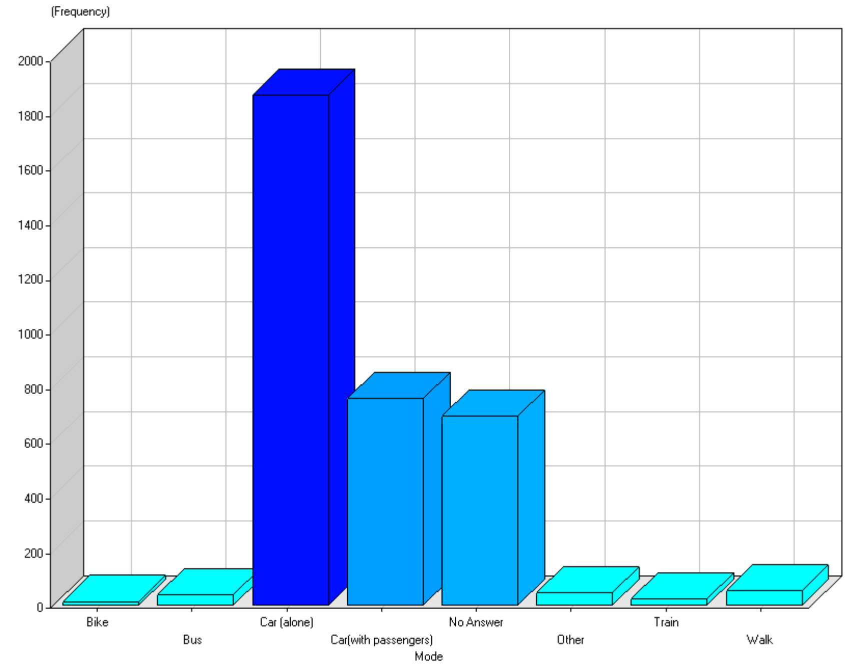


Figure C-7: Mode Used for Trip

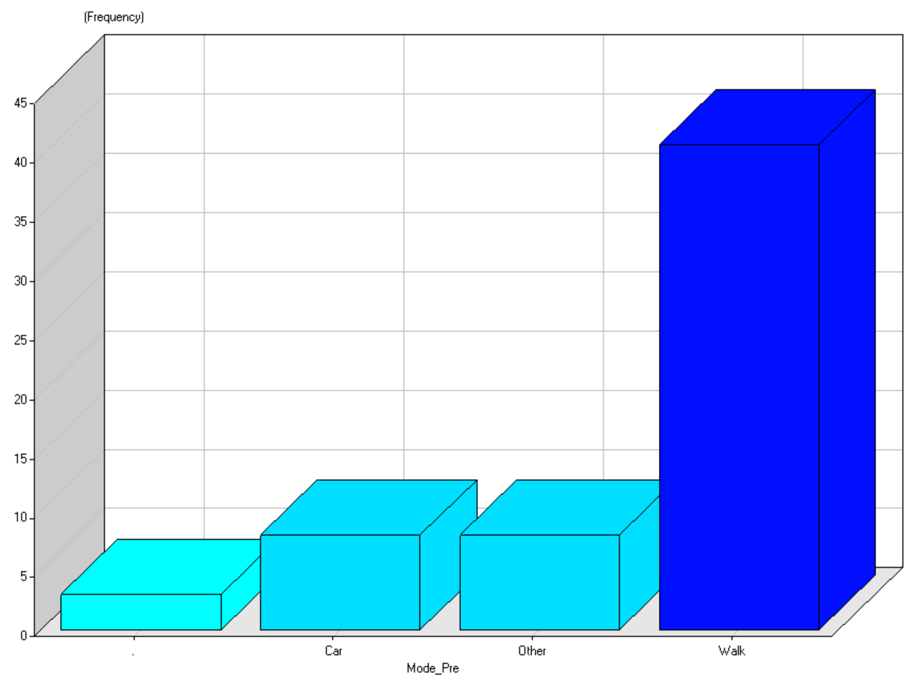


Figure C-8: Mode Used to Get to Bus or Train Stop

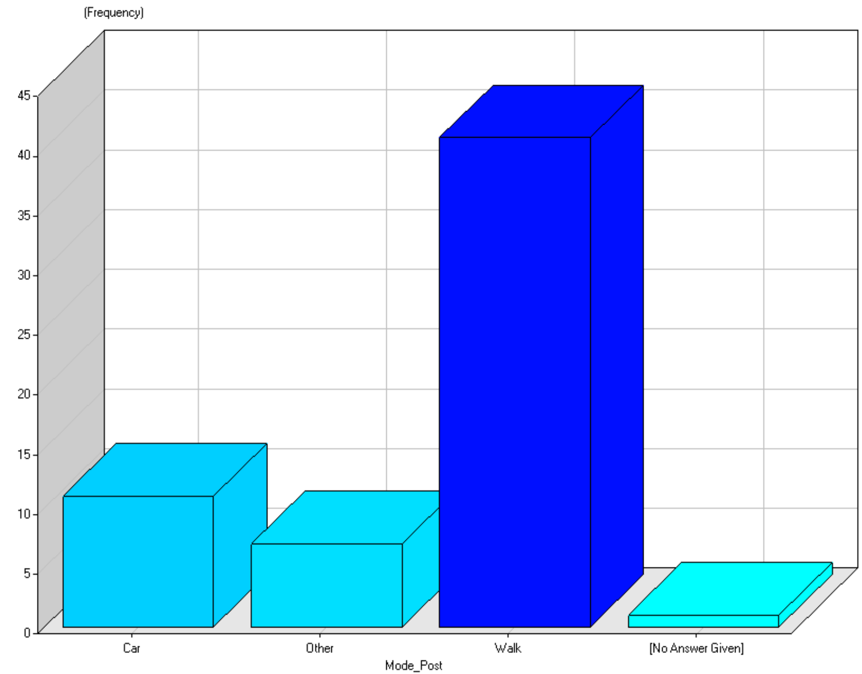


Figure C-9: Mode Used to Leave Bus/Train Stop

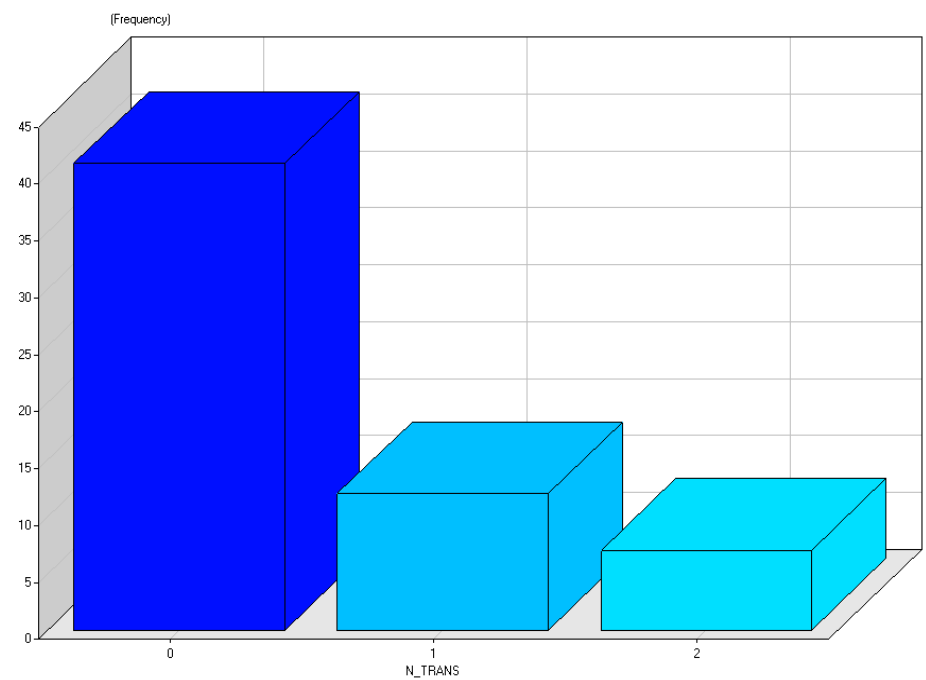


Figure C-10: Number of Transfers on Public Transit

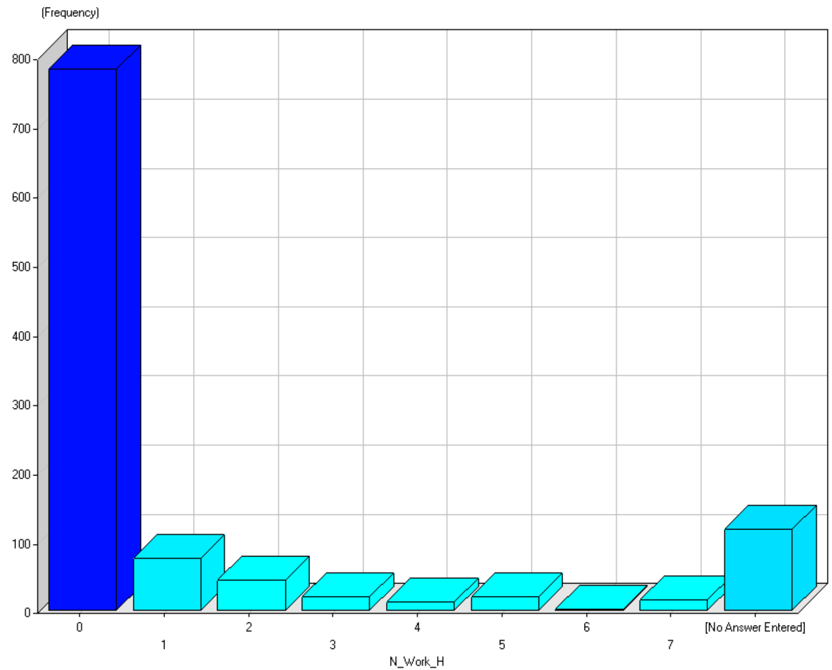


Figure C-11: Number of Days per Week, Worked from Home

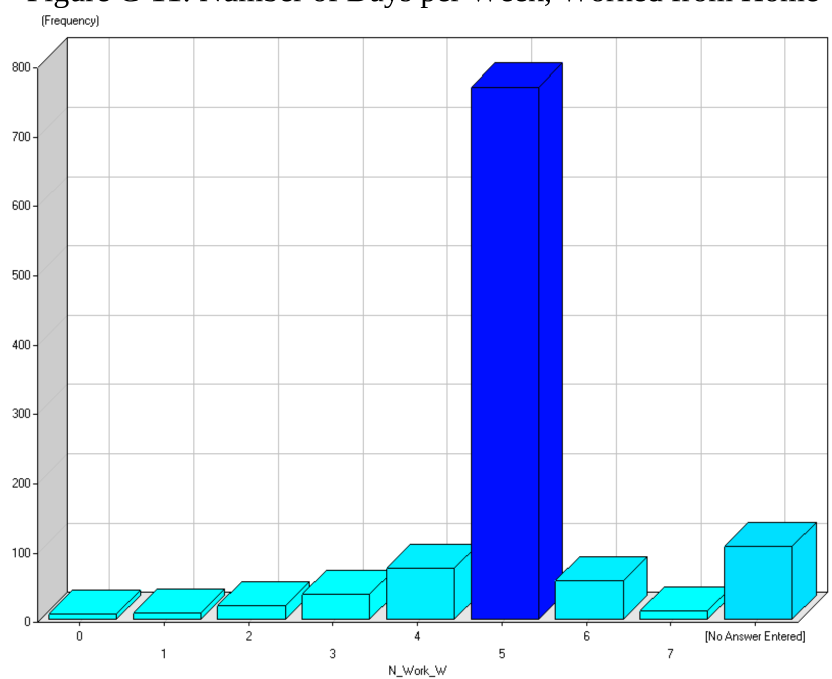


Figure C-12: Number of Days per Week, Worked at Work