Connecticut Statewide Transportation Study

Prepared by:
Karthik C Konduri, Ph.D.
Chloe Ritter
Joann Lynch
Nicholas E. Lownes, Ph.D.
Elizabeth Greene

Name of the performing organization
Connecticut Transportation Institute, University of Connecticut
Resource Systems Group, Inc.
ETC Institute.

Prepared for:
Connecticut Department of Transportation
Bureau of Policy and Planning
Coordination, Modeling and Crash Data
Travel Demand / Air Quality Unit

Maribeth Wojenski
Assistant Director of Policy and Planning

Submitted to:
Connecticut Department of Transportation
Bureau of Policy and Planning
Roadway Information Systems Unit
Research Section

Michael J. Connors
Assistant Director of Policy and Planning

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. CT-2296-F-17-8	2. Government Accession No. N/A	3. Recipients Catalog No.	
4. Title and Subtitle Connecticut Statewide Transportation Study		5. Report Date December 8, 2017	
		6. Performing Organization Code SPR-2296	
7. Author(s) Karthik C Konduri, Chloe Ritter, Joann Lynch, Nicholas E. Lownes, Elizabeth Greene		8. Performing Organization Report No. CT-2296-F-17-8	
9. Performing Organization Name and Address Connecticut Transportation Institute, University of Connecticut 270 Middle Turnpike, Unit 5202, Storrs, CT 06269-5202 Resource Systems Group, Inc. 55 Railroad Row, White River Junction, VT 05001 ETC Institute. 725 W. Frontier Circle Olathe, KS 66061		10. Work Unit No. (TRIS) N/A	
		11. Contract or Grant No. SPR-2296	
		13. Type of Report and Period Covered Final Report May 2015 – August 2017	
12. Sponsoring Agency Name and Address Connecticut Department of Transportation Bureau of Policy and Planning, Research Section 2800 Berlin Turnpike, Newington, CT 06131-7546			
		14. Sponsoring Agency Code SPR-2296	
15. Supplementary Notes A study conducted in cooperation with the Connecticut Department of Transportation.			

16. Abstract

Connecticut Department of Transportation (CTDOT) has embarked on the development of a new statewide travel demand model which will enhance evaluation of a variety of new transportation policy and planning applications, including transit oriented developments, multimodal public transportation systems and commuter rail options. A key input to successful implementation of a travel demand model is household travel survey data that provides detailed information regarding the different measurements of travel patterns of a representative subset of households/individuals in the study area. Household travel surveys also provide rich information regarding a variety of socio-economic, demographic and land use attributes that can be used to comprehend and analyze the travel choices people make. A household travel survey called the Connecticut Statewide Transportation Study was carried out between July 2015 and August 2017. The primary objective of the study was to conduct a comprehensive multi-modal statewide household travel survey that will facilitate the development of travel demand models by CTDOT and other planning agencies in the state. The report provides details regarding the design, development, data collection, data processing, and data preparation aspects of the study. The report also provides an overview of the travel behavior trends of Connecticut residents.

17. Key Words Household Travel Survey, Design, Development, Data Collection, Data Processing, Data Preparation	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA. 22161		
19. Security Classif. (Of this report) Unclassified	20. Security Classif.(Of this page) Unclassified	21. No. of Pages 97	22. Price N/A

DISCLAIMER

The contents of this report reflect the views of the author(s), who are responsible for the facts and accuracy of the data presented herein. The contents of this report do not reflect the official views or policies of the U.S. Department of Transportation, Federal Highway Administration or the Connecticut Department of Transportation.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Connecticut Statewide Transportation Study (CSTS), a comprehensive multi-modal statewide household travel survey, was conducted on behalf of the Connecticut Department of Transportation (CTDOT) between May 2015 and August 2017. The survey was conducted by the Connecticut Transportation Institute at the University of Connecticut and a consultant team comprising of Resource Systems Group Inc. and ETC Institute. A summary of the study is provided below.

Purpose

The primary objective of the CSTS was to collect complete travel information for a 24-hour weekday period from a representative sample of at least 7,500 households across the state of Connecticut. In addition to the travel characteristics, socio-economic characteristics, demographic factors, and vehicle ownership attributes were collected from each sampled household. The last household travel survey, in the state of Connecticut, was conducted in the late 1970's. Since that time, state demographics, employment, land use, and travel patterns have changed significantly. CSTS serves as a comprehensive and up-to-date database of travel behaviors of CT residents. Information from the survey will be used by CTDOT for developing a new statewide travel demand model (STDM) system, and for other transportation analyses.

Data

Each sampled household was asked to report detailed travel information for a pre-assigned travel day include travel mode, destination location, and destination purpose among other characteristics. In addition, socio-economic characteristics, demographic attributes, and vehicle ownership and usage information was requested from all individuals within the sampled household. The final dataset contained demographics and travel data for 17,481 persons residing in 8,403 households throughout the entire state of Connecticut. While residents of neighboring states (New York, Massachusetts, and Rhode Island) may also travel into and through Connecticut, residents of these states were not included in this study.

Respondents

A sample of households were invited from those residing in the state of Connecticut using a combination of simple geographic proportional random sampling, targeted oversampling (higher sampling rates) in selected areas, and upsampling (increased invitation rates) in other areas where higher proportions of hard-to-reach are identified. In order to obtain sufficient responses, invitation letters were mailed to 153,649 households across the state, and 8,403 households completed the study corresponding to a sampling rate of 0.6%.

Data Collection

Data was primarily collected using an online survey instrument. Sampled households could either self-administer the survey by logging on to the survey online or could call a toll-free number and provide their responses over phone. Call center operator used the same online survey instrument to enter the information provided by the callers. The survey contained two parts: recruit survey and travel diary survey. The recruit survey collected household-, person-, and vehicle-level information. Only one adult household member was required to complete the

recruit survey. Once this section was completed, respondents were shown (or read) a survey dashboard with further instructions about logging their trips on the assigned travel day and completing the travel diary survey. In order for a household to be counted as complete, travel diaries of all household members were required to be completed after their assigned travel date had passed. Individuals could self-report or a household member could report trips for others. Adults were asked to fill out travel diaries for children in the household.

Timeline

The main data collection effort began in February 2016 and ended in June 2016. The households invited to the survey were assigned one of 30 weekday "travel dates", spread over ten weeks in March, April and May 2016. One week in the middle was "skipped" to allow time for review and adjustment after the first few weeks of data collection. All travel dates were on a Tuesday, Wednesday, or Thursday of each week, due to the focus on typical weekday travel of residents. Travel dates were pre-assigned to households and invitations were evenly distributed over all the dates so that the recruitment and survey retrieval process could be easily managed. Distributing the sample evenly over all the travel dates also improved the final dataset as it contains a sufficient sample of each of the three weekdays and includes data for different months.

Travel Characteristics of CT residents

Below a summary of the travel characteristics of Connecticut residents are provided. The statistics were generated using travel survey data from the 2016 CSTS. All statistics reported are based on weighting analysis to match characteristics of CT residents from the 2009-2013 American Community Survey. Where appropriate, imputed, or logically estimated, values are utilized to supplement the collected data from the survey. As with any survey, the statistics are subject to error and bias.

Overall Trends

The approximately 3.4 million Connecticut residents, residing in 1.4 million households, made a total of 11.8 million trips. The average trip rate is 3.51 per person and 8.68 per household. Trip rates increase with household income, presence of children and household size. Women have lower work trip rates, but higher total trip rates than men. Persons 35-54 years of age have higher trip rates than the other age groups.

Trip Purpose

The largest proportion of trips (34.2%) are made to return home from an activity location. This is followed by work which accounted for 14.8% of trips and personal business which accounted for another 13.7% of trips. Shopping accounted for 10.4% of trips. The three purposes with the lowest shares are social recreation, meal, and other.

Trip Purpose Distribution

Pie chart showing trip purpose distribution:
- Home: 34.2%
- School: 7.4%
- Work: 14.8%
- Meal: 5.1%
- Shopping: 10.4%
- Personal Business: 13.7%
- Social Recreation: 5.4%
- Escorting: 7.4%
- Other: 1.4%
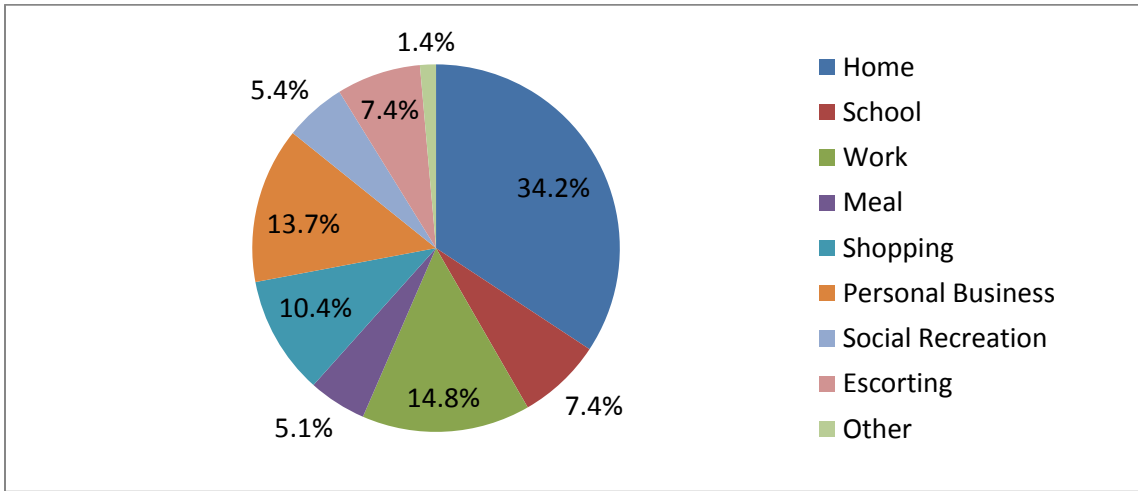
Trip Mode

Automobile is the dominant mode used for traveling accounting for nearly 82.8% of the trips. Out of the 82.8%, 48.2% of trips are made by Single Occupant Vehicle (SOV) mode, and the remaining 34.6% of trips involve some form of carpooling. Walk/bike is the next most popular mode accounting for 9.2% of all trips. Public transit serves 3.2% of all weekday trips in Connecticut.

Distribution of Primary Mode for Trips
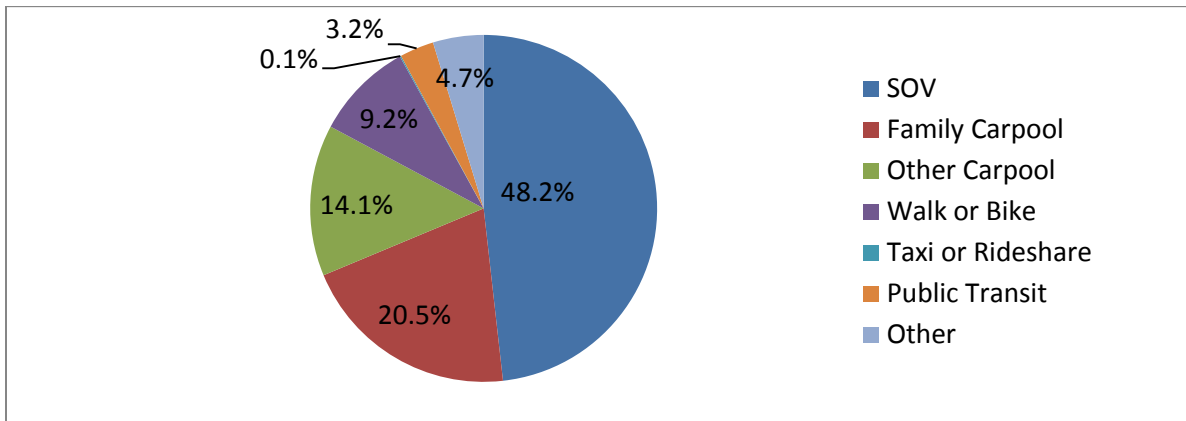
Pie chart showing distribution of primary mode for trips:
- SOV: 48.2%
- Family Carpool: 20.5%
- Other Carpool: 14.1%
- Walk or Bike: 9.2%
- Taxi or Rideshare: 0.1%
- Public Transit: 3.2%
- Other: 4.7%

SOV is the most popular mode for all types of trips except for school trips and escorting trips. 76.5% of work trips are made by SOV mode. About 50% of shopping, personal business and home trips are made by SOV mode as well. 28.4% of school trips are made by family carpool and 27.6% are made using school bus.

Distribution of Primary Trip Mode by Purpose

Legend: SOV, Family Carpool, Other Carpool, 'Walk or Bike, Taxi or Rideshare, Public Transit, Other

Categories (top to bottom): Other, Escorting, Social Recreation, Personal Business, Shopping, Meal, Work, School, Home

X-axis: 0%, 20%, 40%, 60%, 80%, 100%
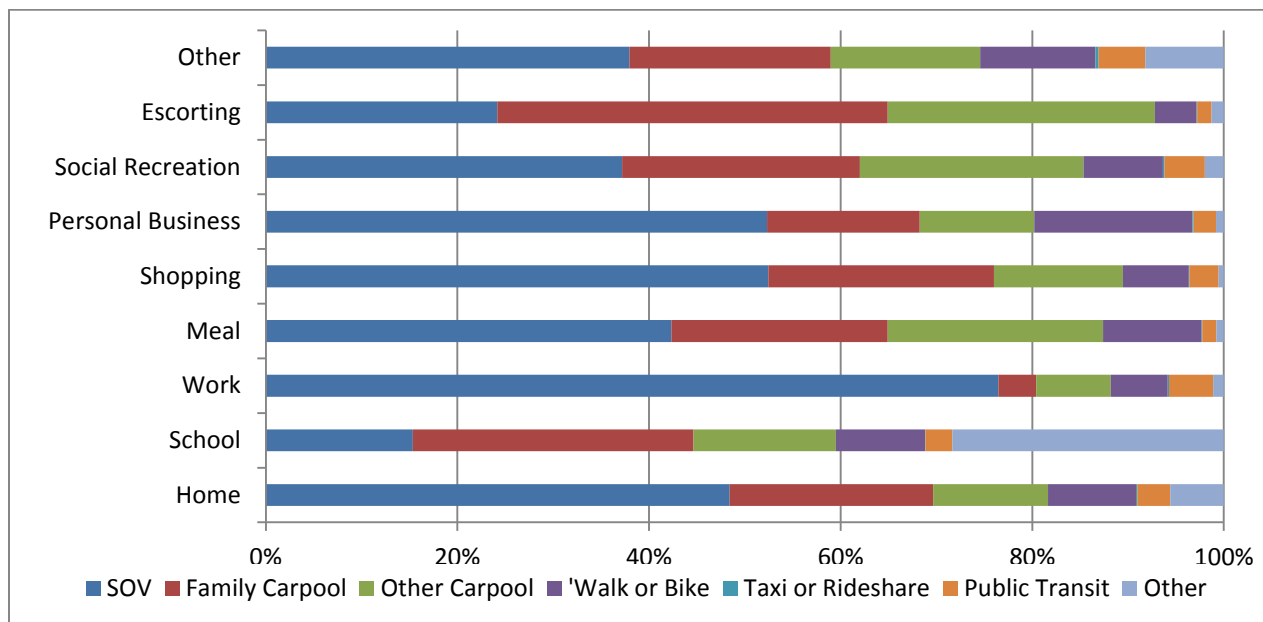
Spatial Distribution of Trip Ends

The average reported travel duration for all types of trips is 22.90 minutes. Work trips on average are longer than other trips. The average reported work travel duration is 29.66 minutes and reported non-work travel duration is 21.71 minutes. The average reported travel duration of trips by employed persons is 24.47 minutes whereas average travel duration of trips by unemployed persons is 20.09 minutes. It is interesting to note that the average travel duration of kids (age less than 16 years old) is slightly higher than the average duration for unemployed individuals.

A majority of trips made are within Connecticut. 84.7% of trips are intra-county and 11.4% of trips are inter-county of Connecticut. Only about 1% of trips are made from/to a location outside Connecticut. Intra-county trips account for the greatest proportion of both work trips and non-work trips.

Distribution of Trip Ends

Origin-Destination Pattern for Work Trips

- Intra-county
- Inter-county
- From Other State to CT
- From CT to Other State
- Out of CT

5.3% 1.3% 0.5% 18.9% 74.1%

Origin-Destination Pattern for Non-Work Trips

- Intra-county
- Inter-county
- From Other State to CT
- From CT to Other State
- Out of CT

0.7% 1.5% 1.1% 10.1% 86.6%
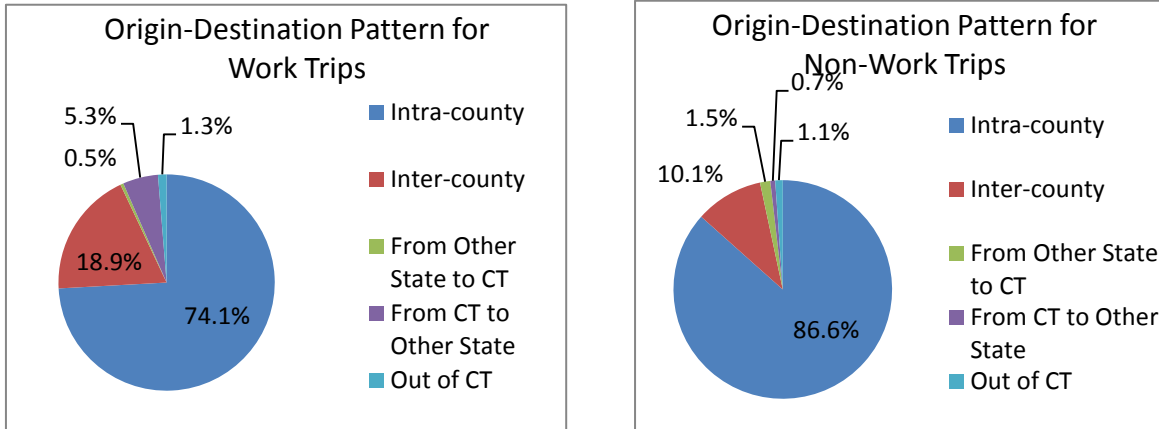
Travel Accompaniment

57.6% of all trips were made by individuals alone, while 43.4% of trips were made with others. Majority of work trips (85.1%) and personal business (65.7%) trips are made alone. Escorting trips, school trips and social recreation trips are more likely to be made along with others.

Other Key Travel Trends

❑ Profile of transit travelers:
  o 89% of the transit trips were made by residents of Hartford, Fairfield, and New Haven counties
  o Majority of the trips were made by females (53%) and by those who were in the 45 to 54 age group (20%)
  o 48% of the trips were made by those who were employed
  o 13% of the trips were made by students who were either enrolled fulltime or part time
  o 49% of the trips were made by those with a driver's license
  o Most common trip purposes were "Go home" (37%) and "Go to primary workplace" (19%)
❑ Profile of bike/walk travelers:
  o 81% of the bike/walk trips were made by residents of Hartford, Fairfield, and New Haven counties
  o Majority of the trips were made by females (51%) and by those who were in the 25-34 age group (18%)
  o 40% of the trips were made by those who were employed
  o 14% of the trips were made by students who were either enrolled fulltime or part time
  o 67% of the trips were made by those with a driver's license
  o Most common trip purpose was "Go home" (34%) and "Exercise" (19%)

CHAPTER 1 INTRODUCTION

With growing needs to evaluate a variety of new transportation policy and planning applications including transit oriented developments, multimodal public transportation systems, and commuter rail options among others, the Connecticut Department of Transportation (CTDOT) has embarked on the development of a new statewide travel demand model (STDM) system. A key input to successful implementation of a STDM is household travel survey data that provides detailed information regarding the different dimensions of travel engagement choices. Additionally, such information is sought from a representative subset of the households/individuals in the study area so as to make reliable inferences about all residents in the state. Household travel surveys also provide rich information regarding a variety of socio-economic, demographic, and land use attributes that can be used to explain the travel choices people make.

The last comprehensive household travel survey data collection in the state of Connecticut was conducted in the late 1970's. Since that time, state demographics, employment, land use, and travel patterns have changed significantly. While some of these changes (i.e. demographic changes, and limited travel measures) can be observed in Census records, a more comprehensive and up-to-date database of travel behaviors is needed in order for the travel model to be calibrated with this latest information. Subsequently, this will ensure a more accurate estimation of the impacts of various policy and planning applications.

The remainder of the chapter is structured as follows:

- ❑ Section 1.1 provides an overview of the 2016 Connecticut Statewide Transportation Study.
- ❑ Section 1.2 describes the pilot study portion of the project.
- ❑ Section 1.3 presents details regarding the main study phase of the project.
- ❑ Finally, Section 1.4 describes the report organization.

1.1 Connecticut Statewide Travel Survey Overview

The 2016 Connecticut Statewide Transportation Study (CSTS) provides data about travel patterns of residents within local communities, in regional planning areas, and across the state of Connecticut.

1.1.1 Focus of the Travel Data

The primary objective for the CSTS was to collect complete travel information for a 24-hour weekday period from a representative sample of at least 7,500[1] "complete"[2] households across the state of Connecticut. Each sampled household was asked to report travel information for a

---

[1] This number was determined based on a review of sample sizes from recent statewide survey efforts in other states. For instance, 2010-2011 Massachusetts Travel Survey collected travel data from 15,033 households out of the 2.54 million households in the state representing a sampling rate of 0.59%. In CT there are approximately 1.36 million households as per the Census 2009-2013 ACS 5-Year Estimates. The 7,500 target represents a sampling rate of 0.55%.

[2] In this study, a complete household is defined as one in which each member of the household has provided valid information for all data items collected in the survey.

pre-assigned travel day. In addition to the travel characteristics, a variety of other information was requested from all individuals within the sampled household including socio-economic characteristics, demographic factors, and vehicle ownership attributes.

Particular focus was placed on collecting daily "personal travel"[3] pursued by individuals and households on weekdays (i.e. Tuesday through Thursday). The study also sought to collect a sufficient sample of the types of households that are traditionally more difficult to reach, yet are important for transportation policies and planning applications. This included low-income households, low- or no-vehicle households, and households that frequently use transit. The study collected a broad range of possible household travel behaviors, including detailed trip purposes, all types of trip modes (e.g., driving, walking, bicycling, riding transit, etc.), and trips made by every household adult and child (both individually and jointly with other household members).

The household travel survey was designed, developed and deployed to conform to the state of practice in survey research methodologies to efficiently and effectively complete the CSTS, including:

- ❑ A stratified sampling plan to balance the objectives of representativeness and sufficient sample from sub-populations of interest
- ❑ An address-based recruitment strategy with multiple first-class mailings to invited households
- ❑ A multimodal data collection strategy including telephone retrieval and web survey technology
- ❑ An informative and aesthetically appealing public website with consistent branding and messaging throughout all official outreach materials and activities
- ❑ Applying state of research and practice methodologies for survey data cleaning and processing to prepare final datasets

1.1.2 Study Area
The CSTS collected data from households throughout the entire state of Connecticut, and included all travel reported by these households on their assigned travel dates (whether the travel took place within the state or elsewhere). While residents of neighboring states (New York, Massachusetts, and Rhode Island) may also travel into and through Connecticut, residents of these states were not included in this study.

1.1.3 Study Timeline
The CSTS began in June 2015. During fall of 2015, the study team evaluated the study needs, developed the survey instrument, developed administration materials and protocols, and designed a pilot sampling plan. In the winter of 2015, the pilot study was carried out. Following

---

[3] In addition to personal travel, there are a number of contributors to traffic on roadways in the state including inter-city travel from/to cities in surrounding states, and freight travel among others. The survey effort to collect data about these other contributors is very different from those aimed at collecting for daily personal travel. Therefore, information regarding the other contributors must be undertaken as separate surveys (e.g. long distance surveys, and freight surveys) in the future if so desired.

the pilot study, the instruments, materials, protocols, and sampling plan were refined and revised for the main study. The main study data collection began in spring 2016 and was completed by middle of summer. Finally, after data collection was complete, the study team processed and analyzed the data. A final version of the dataset was transmitted to CTDOT towards the end of 2016, to support the validation of the STDM system that was being developed. Additional data cleaning and data processing tasks were carried out to further enhance the quality and validity of the dataset beginning in the winter of 2016 through summer of 2017. The study team also developed long-term plans during the spring of 2017 in support of possible future data collection activities. Table 1 below illustrates the timeline of these activities.

Table 1: An Overview of the Study Timeline

	2015							2016												2017							
	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8
Project Kickoff	■	■																									
Preliminary Review and Needs Identification		■	■	■																							
Pilot Study: Development and Sampling				■	■	■																					
Pilot Study: Deployment and Data Collection						■	■	■																			
Pilot Study: Data Processing and Evaluation							■	■																			
Main Study: Development and Sampling								■	■	■																	
Main Study: Deploymetn and Data Collection										■	■	■															
Main Study: Data Processing, Weighting, and Analysis												■	■	■	■	■	■	X									
Main Study: Additional Data Processing																			■	■	■	■	■	■			
Long-term Data Collection Plan																			■	■	■						
Study Documentation																								■	■	■	X
X - Key Deliverables																											

1.2 Pilot Study Overview

The overall study can be divided into two major phases: the pilot study phase and the main study phase. The primary objectives of the pilot study phase were to develop, deploy, test, and revise the survey materials, and prepare for the main study. Specific emphasis was placed on testing the questionnaire and online survey instrument, evaluating the print and online materials used for survey recruitment and data collection, analyzing the survey administration processes and protocols, and studying the response rates. Based on the findings from the pilot study, necessary changes were made to various aspects of the survey and preparations were made for the main study. The pilot study spanned from July 2015 through January of 2016 and is discussed in detail in Chapter 2.

1.3 Main Study Overview

The primary objective of the main study was to collect responses from 7,500 "complete" households across the state of Connecticut. The main data collection effort began in February 2016 and "first version"[4] of the data was transmitted in November 2016 to support the model validation. Invitation letters were mailed to 153,649 households across the state, with 8,403[5] households completing the study. This resulted in a sampling rate of approximately 0.6%,

---

[4] In the report, two versions of the data are referenced corresponding to the two waves of deliverables transmitted in November 2016 and August 2017. The data delivered in November 2016 is referred to as first version and the data delivered in August 2017 is referred to as the final version.

[5] This number represents the final count of "complete" households with usable information.

exceeding the initial target of 7,500 responses and 0.55% sampling rate. Figure 1 illustrates the distribution of respondent households in towns across the state. As expected, areas shaded in darker red (representing higher number of responses) are generally concentrated in the densely populated regions in the southwest and central parts of the state. These areas follow the Metro North commuter rail routes toward New York, the I-95 corridor, and the I-84 and I-91 corridors between New Haven and Hartford. The main study approach and findings are discussed in greater detail in Chapter 3.

Figure 1: Complete Responses by Town
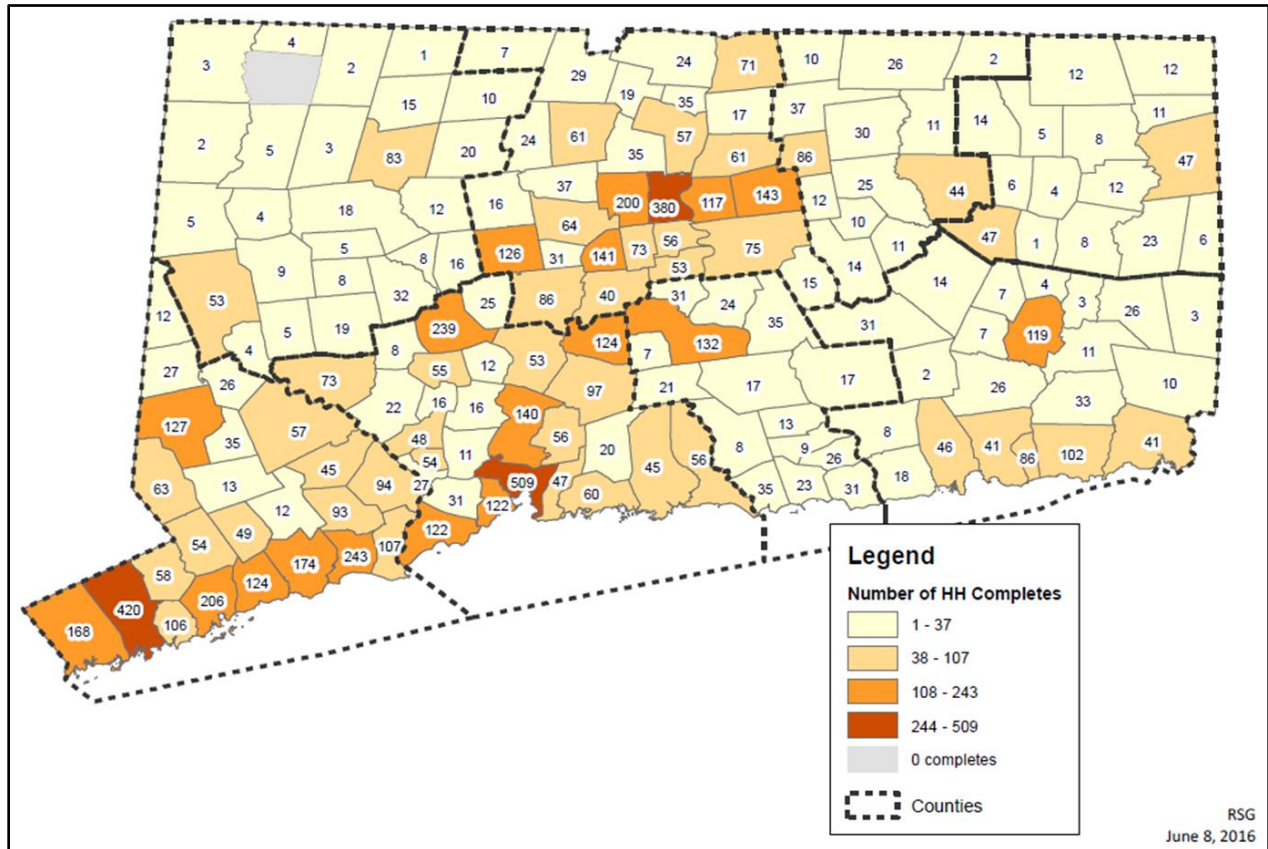
1.4 Report Organization
The pilot study and the main study are discussed in greater length in Chapter 2 and Chapter 3 respectively. The Appendices at the end of the report includes survey materials, memos, and documentation in support of the deliverables.

CHAPTER 2 PILOT STUDY

A pilot study was first carried out in order to evaluate survey response rates, study materials and survey design prior to the main study. The pilot study also explored the response rate implications of different types of incentive schemes (guaranteed incentive versus raffle) and mode of paper survey (online/phone versus paper diary). The goal of the pilot study was to collect data from 250 households.

The remainder of this chapter is organized as follows:

- ❑ Sampling plan is described in Section 2.1
- ❑ Incentive plan is presented in Section 2.2
- ❑ Questionnaire design and development is presented in Section 2.3
- ❑ Survey implementation is discussed in Section 2.4
- ❑ Development of survey design and branding materials is presented in Section 2.5
- ❑ Survey administration protocol and procedures are discussed in Section 2.6
- ❑ The approach to data processing and preparation is presented in Section 2.7
- ❑ Pilot study results are presented in Section 2.8
- ❑ Main study recommendations are discussed in Section 2.9

2.1 Sampling Plan

A stratified split sample design was used for the pilot study. To achieve the sample target of 250 complete households, those households residing in select regions from the state were divided into six segments based on geography (namely block groups). Subsequently, invitations were mailed to a random sample of approximately 8,500 residential households from within the geographies based on expected response rates, and survey protocol alternatives. It must be noted that the primary goal of the pilot was to assess the overall process (survey and material designs, protocols, and response rates). Subsequently, the information was used to refine the main study to be more effective and efficient. Therefore, the pilot results were not expected to be representative of the population.

2.1.1 Sampling Frame

Due to the smaller sample size, the pilot sample was invited from a smaller sub-area within the state instead of drawing a sample from the entire state. This provided for a higher pilot sampling rate[6], which in turn helped enable a better assessment of the pilot response. Three areas from within the state were identified for the pilot sample as shown in Figure 2. They were chosen to be representative of the diverse set of regions in the state. These areas included: 1) the town of Hartford –the state capital and a higher concentration of traditionally "hard-to-reach" households, 2) the town of Norwalk –located in the southwest part of the state and part of the New York City commute-shed, and 3) the county of New London –located in the eastern part of the state and less densely populated. The sampling frame for the pilot included all residential mailing addresses within the chosen areas in the state. The sample list was obtained

---

[6] Defined as the ratio of number of complete households to the total population for the chosen study area.

by randomly drawing from the sampling frame based on pre-specified spatial stratification and sampling rate. The sample list was obtained from Marketing Systems Group (MSG), a data vendor that maintains the United States Post Office's Computerized Delivery Sequence File – a list of all addresses.

Figure 2: Pilot Sample Geographies
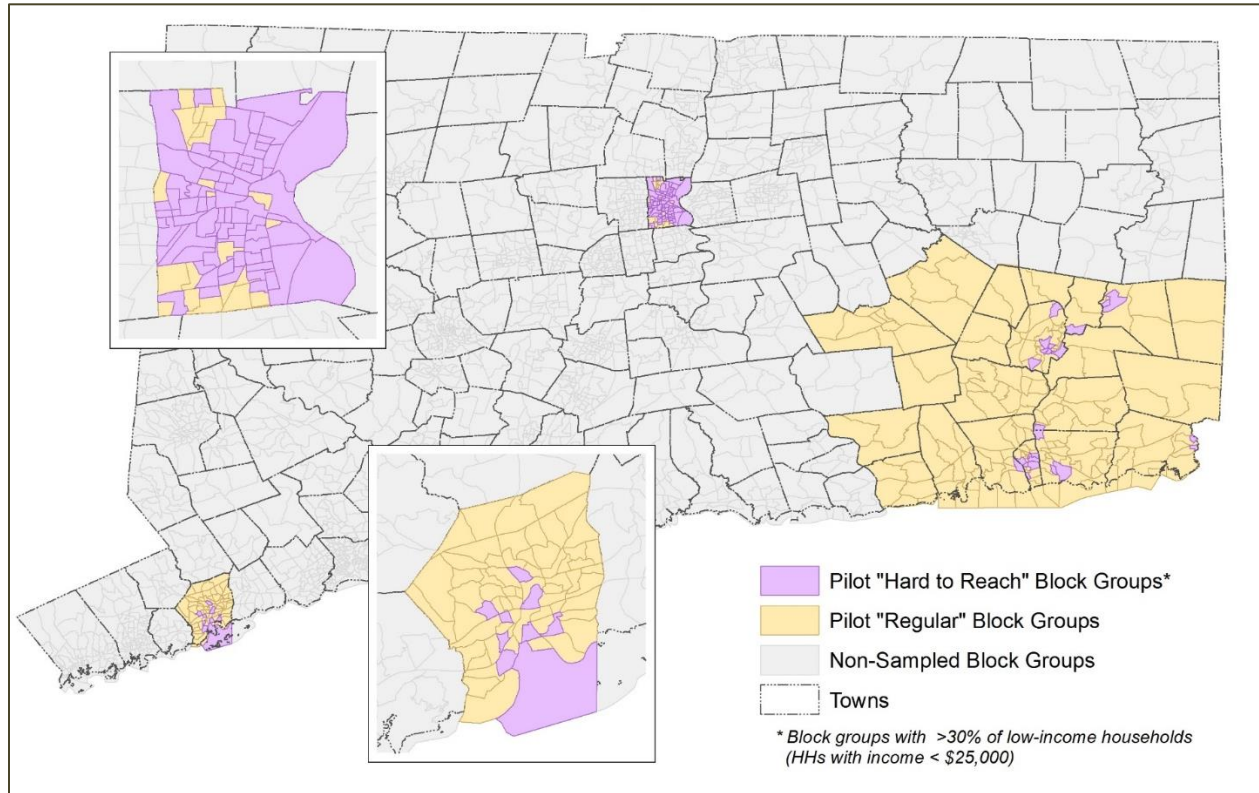
It was not expected that the sample distribution from the pilot sample areas would closely match statewide population demographic distributions (see Table 2 for a comparison of key demographic characteristics for the pilot sample areas and state as a whole). By controlling for an important determinant of response rates and collecting enough responses in the pilot study, however, the study team was able to evaluate response patterns of different types of households of interest and design a more comprehensive sampling plan for the main study.

Table 2: Demographic Summary of the Sample Areas and state of Connecticut as a whole

	Hartford Town	Norwalk Town	New London County	Connecticut
Total HHs	45,808	36,236	107,066	1,355,849
% Low-income HHs (<$25k)	44.3%	17.1%	16.0%	18.0%
% 0-Vehicle HHs	35.5%	9.4%	7.1%	9.0%
% Large HHs (4+ persons)	22.3%	19.1%	20.9%	23.1%
% Young HHs (householder <35yr)	27.5%	16.9%	17.2%	15.7%
% Low-income HHs (<$25k)	27.5%	11.7%	7.1%	8.7%

2.1.2 Sample Segments

As noted earlier, the sampling plan was designed at the block group level. The block groups from chosen pilot study areas in the state were categorized into six segments based on three primary considerations of the pilot study and were applied in the order presented below.

- ❑ *Expected response rates by demographic factors.* Within the three areas identified above, two types of households were defined as 1) hard-to-reach and 2) regular households. Characteristics of the two types of households are described below:
  - o Hard-to-Reach households: Typically, households with lower incomes respond to surveys at lower rates (Stopher 1996[7], Zimowski 1997[8]). Therefore, it is important to account for the differential response rates to avoid sample biases. This is customarily achieved by drawing additional samples in proportion to anticipated lower responses rates – a process also referred to as up-sampling. Within the pilot study areas, the designated up-sample geographies are block groups with more than 30% households in the low income category (i.e. income less than $25,000). It must be noted that there are other types of households who also respond at lower rates (e.g. large households and households with younger members among others). However, given the size of the sample target in the pilot, it was not feasible to account for the differential response rates due to other characteristics. The focus in the pilot study was limited to one of the important and widely acknowledged determinants of lower response rates, income. Potential differential response rates due to other factors were analyzed using data collected in the pilot and utilized to design the sampling plan for the main study[9].
  - o Regular households: All block groups in the pilot sample areas that were not identified as hard-to-reach were designated as the regular sample geographies. These block groups have 30% or fewer households in the low-income category. Households from these areas, therefore, were expected to respond at somewhat higher rates than those in the hard-to-reach geographies.
- ❑ *Expected response rates by incentive offered.* Within each of the hard-to-reach and regular block groups, the invited households were divided into two sub-groups. Half of the households were offered a guaranteed incentive, in the form of a gift card, upon completion of the study. The other half of the households were entered into a raffle for a predetermined prize. Each of these sub-groups were sent different invitation materials reflecting the differences in the incentives offered. However, both sub-groups received the same kinds of study information and reminders and

---

[7] Stopher, P. (1996). Household Travel Surveys: New Concepts and Research Needs. Conference on Household Travel Surveys. (Website: http://onlinepubs.trb.org/Onlinepubs/conf/1995/cp10/cp10-001.pdf, Accessed: October 15, 2017)

[8] Zimowski, M., Tourangeau, R., Ghadialy, R., and Pedlow, S. (1997). Nonresponse in Household Travel Surveys. (Website: https://www.fhwa.dot.gov/ohim/nonrespond.pdf, Accessed: October 15, 2017)

[9] More details regarding the adjustments for differential response rates in the main study are discussed in Chapter 3, Section 3.1 .

completed the exact same survey (online or over the phone only). It was anticipated that the raffle sub-group would participate at lower rates than the gift card sub-sample (Zimowski 1997). As a result, in the sampling plan, differential response rates were accounted for and invitations were adjusted accordingly.

❑ *Expected response rates by survey response method offered.* In order to test impact of survey mode on response rates, separate, non-overlapping invitations were drawn from the same hard-to-reach and regular block groups. These sub-groups were provided different invitation materials that included the option to fill out the travel diary portion of the survey using a paper diary in addition to the web and phone options. Households that choose to respond online or over the phone were allowed to use the same survey instrument as those who were not offered the paper option. On the other hand, households that prefer the paper diary were provided a paper diary along with instructions to complete the survey and mail it back. It was anticipated that those who were offered a paper diary option would complete the survey at lower rates than those required to complete online or over the phone (Zimowski 1997, Morency 2014[10]). As a result, the number of invitations was adjusted accordingly.

The above process resulted in a total of six sample segments as shown in Table 3 below. The number of invitations was selected based on anticipated response rates. Invitations for each of the six segments were randomly selected from households within the corresponding block groups. The study team tried to ensure that there was no overlap across the invitations (i.e. no duplicate addresses). Also, during the pilot study, each segment was monitored separately. It must also be noted that for the segments offered the paper option, they were entered into the same raffle as the other raffle invitation segments.

2.1.3 Invitations

Once the sample strata were determined, the next step was to determine the number of desired responses and resulting number of invitations needed for each segment. The desired overall sample size for the pilot study was predetermined as 250 households, or approximately a 0.13% sampling rate of the total households in the pilot sample areas. This is typically sufficient proportion to offer the type of insights motivating a pilot study. Next, response rates were assumed for the different sample segments based on a combination of demographics, incentives offered, and recent similar studies. The invitee count was then obtained by dividing the anticipated number of responses for each segment by the assumed response rate for the segment. In order to obtain the 250 responses, a total of 8,510 invitations were mailed out. Table 3 below shows the breakdown of the invitations by individual segments.

---

[10] Morency, C. (2014). Recent Development and Analysis on Household Travel Surveys. (Website: http://uttri.utoronto.ca/files/2014/10/6-Morency-Recent-development-and-analysis-on-Household-Travel-Surveys-in-Quebec.pdf, Accessed: October 15, 2017)

Table 3: Characteristics, Assumed Response Rates, and Target Responses for the Pilot Sample Segments[11]

ID	Sample Segment Description[12]	% of Sample	Sample Size[13]	Diary Mode Options	Sampling Rate[14]	Sample Area Block Groups Type	# Block Groups[15]	# HHs[16]	Response Rate[17]	Invites, # HHs[18]	Invites, % of HHs[19]
1	Hard-to-reach / Raffle	20%	50	Online/Phone	0.09%	Hard-to-Reach	106	58,245	2.0%	2,500	4.3%
2	Hard-to-reach / Gift card	20%	50	Online/Phone	0.09%	Hard-to-Reach	106	58,245	4.0%	1,250	2.1%
3	Regular / Raffle	20%	50	Online/Phone	0.04%	Regular	247	130,865	3.0%	1,670	1.3%
4	Regular / Gift card	20%	50	Online/Phone	0.04%	Regular	247	130,865	5.0%	1,000	0.8%
5	Hard-to-reach / Raffle/ Paper	10%	25	Online/Phone /Paper	0.04%	Hard-to-Reach	106	58,245	2.0%	1,250	2.1%
6	Regular / Raffle/ Paper	10%	25	Online/Phone /Paper	0.02%	Regular	247	130,865	3.0%	840	1.4%
Total		**100%**	**250**		**0.13%**		**353**	**189,110**	**2.9%**	**8,510**	**4.5%**

[11] In the table, the short form HHs is used to denote households.

[12] Defined by a combination of block group category, incentive type, and mode options offered to complete the diary portion of the survey.

[13] Sample size is the target number of households who have completed the entire survey (including the recruit survey and travel diary survey).

[14] Sampling rate is the target percent of total households who have completed the survey (i.e. number of responses / total number of households).

[15] This column notes how many block groups are within the particular category of block group type in the pilot sample areas.

[16] This is the estimated total number of households within the corresponding block groups based on 2009-2013 ACS 5-year estimates.

[17] In the CSTS, a complete household is defined as one where every person in the household answers every question (i.e. no missing responses). The predicted survey completion rates are based on demographics, incentives offered, and recent similar studies. Actual completion rates for the pilot were used to design the main study sampling plan.

[18] Number of households invited is determined by the desired sample size and predicted completion rate (sample size / completion rate).

[19] Percent of households invited is the percent of all households expected to receive study information.

2.1.4 <u>Travel Date Assignment</u>

The households invited to the survey were assigned to one of 6 weekday travel dates as shown in Table 4. All travel dates were on a Tuesday, Wednesday, or Thursday owing to the focus of the data collection on collecting information about travel behaviors on typical weekdays. This is a common approach for household travel surveys because travel on these days is expected to represent "typical" household travel patterns. CTDOT also confirmed their preference to focus on the typical weekday travel during this study, so that the study's primary objective of establishing a statewide database of travel behaviors can be achieved more efficiently.

Table 4: Travel Date Assignment by Segment ID

Assigned Travel Date	Segment ID[20]
December 8, 2015	1 through 4
December 9, 2015	1 through 4
December 10, 2015	1 through 4
January 12, 2016	5 and 6
January 13, 2016	5 and 6
January 14, 2016	5 and 6

Travel dates were pre-assigned to households and invitations were evenly distributed over all the dates. This also allowed the study team to manage resources for the recruitment and survey retrieval processes. Another reason to pre-assign travel dates is to minimize response bias and increase the ability of the dataset to describe a typical day for the region (even if it is not a typical day for a given individual). If participants were given a choice of travel dates or simply asked to report on a recent or a typical day, many might unintentionally (or intentionally) bias their response by choosing to report on a busy day or a day with unusual travel patterns (for example, the day they took the bus to work), or choosing a day with little or no travel to reduce their survey burden.

2.2 Incentive Plan

As noted earlier, a split sample design was proposed to test the efficacy of a raffle versus a guaranteed incentive. Nearly half of the invitees were considered to be part of the gift card group (i.e. segment IDs 2 and 4) and half of invitees were considered to be part of the raffle group (i.e. segment IDs 1, 3, 5, and 6). All respondents in the gift card group received a guaranteed gift card and all respondents in the raffle group were entered into a raffle for a chance to win a non-monetary award. Further, the plan for the gift card group in the pilot was implemented as follows:

❑ Contingent upon completion of the study, all gift card group households were offered a guaranteed incentive of $10

❑ Households could choose from an Amazon e-card, a Walmart e-card, or a Walmart mailed (physical) card

---

[20] Please see Table 3 for descriptions of the sample segments.

❏ Respondents were also offered the option to not receive an incentive (if they so choose)

❏ Hard-to-reach households were offered a differential incentive. Households were defined as hard-to-reach[21] (or not) based on two predetermined criteria: 1) household income less than $25,000, and 2) household size greater than or equal to 5. At the end of the recruitment survey, households determined as hard-to-reach were offered $20 instead of $10 to complete their travel diary portion of the survey. The hard-to-reach households were not informed of why they are being offered a higher incentive. Instead they were just told that approximately 1 in 10 households were selected to receive a higher incentive.

2.3 Questionnaire Design

The questionnaire was designed to collect travel information from all individuals in a household in a travel diary format (i.e. a record of all trips made by the household on a single randomly assigned day was gathered). The survey also collected demographic information and typical travel behavior from households and individuals to help explain variations in travel patterns. In addition to serving as explanatory variables in statistical models of travel characteristics, demographic information can be used to rectify biases through weighting analysis.

The survey was developed in three phases as described below:

❏ *Variable Identification*: A list of important variables for use in the travel demand model development and for performing other transportation analysis was developed. The enumeration of the variables was based on study team's experience and review of similar studies. The study team, along with Cambridge Systematics (CS) and CTDOT, reviewed and created the final list of variables based on importance to current and future transportation modeling needs, while also evaluating impact on overall survey burden and response rates. The final list primarily consisted of data elements required for development of the STDM that is currently underway and also to facilitate development of more advanced STDM implementations in the near future (e.g. activity-based model systems). Additionally, supplemental variables about commuting behaviors, travel preferences, and typical trip-making behavior over time were also compiled. These latter data items are not required for the STDM but can be helpful for future STDM development and for other transportation planning analyses.

❏ *Questionnaire Development*: Using the final list of variables, the questionnaire was developed including all question text, answer options, branching logic, question order, and survey instructions. The order of questions in the survey was carefully determined to provide a logical flow. The survey also included extensive question skip logic to ensure that respondents only saw questions that were applicable to them. During the questionnaire development, a balance was attempted between collecting detailed information versus survey burden and sensitivity, so as to

---

[21] This is different from the hard-to-reach block groups that were defined as part of the sampling plan segments.

minimize participants dropping out of the study and to promote response rates. After receiving feedback from CTDOT and CS, the questionnaire was finalized.

❑ *Survey Design and Implementation*: The questionnaire was then programmed as an online survey instrument. This online instrument was the primary data retrieval tool for collecting responses. For participants who preferred to complete the survey over phone, ETC (the call center operator) used the same online survey instrument to administer the survey verbally. CTDOT and CS had the opportunity to preview and comment on the survey instrument before the survey was fielded.

The survey was broken down into two distinct parts:

❑ *Recruit survey*[22] wherein information about the household, its members, and vehicles was collected
❑ *Travel Diary survey* wherein a one-day travel diary for each person in the household was gathered

The study overview shown in Figure 3 summarizes the two parts of the survey. Only one adult household member was required to complete the recruit survey. Once this section was completed, respondents were shown (or read) a survey dashboard with further instructions about logging their travel day trips and completing the travel diary survey. In order for a household to be counted as complete, travel diaries of all household members were required to be completed after their assigned travel date had passed. Individuals could self-report or a household member could report trips for others. Adults were asked to fill out travel diaries for children in the household. A household member was defined as anyone who lives in the same dwelling unit, including relatives, roommates, friends or household help.

Figure 3: A Graphic of the Study Overview
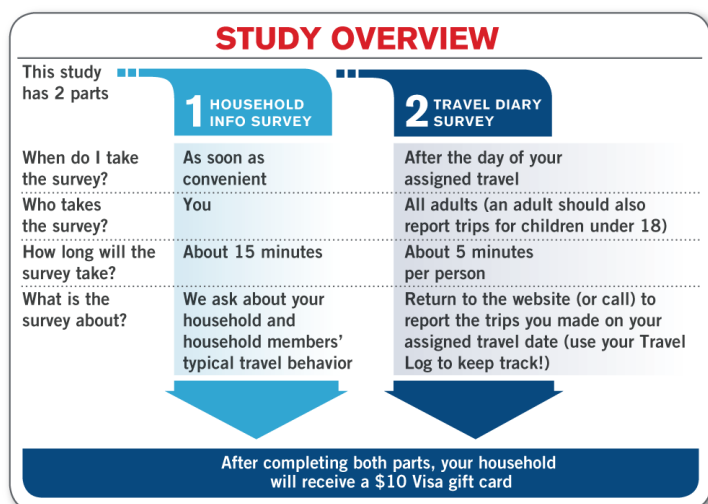
In the following two subsections, details regarding the recruit survey and travel diary are provided.

2.3.1 Recruit Survey

The recruit survey was designed as a stand-alone survey that could be completed by invited households. Households could complete this survey up to a week before their travel date, or could complete it after their travel date and then proceed directly to complete the travel diary survey. The recruit survey included questions about household-level and person-level demographic information as well as typical travel behaviors. Additionally, household vehicle information, and administrative details, such as contact information and incentive preferences, were collected to aid in the travel diary administration. Below is a summary of the household-, person-, and vehicle-level information collected along with screenshots of example questions from the online survey instrument. A copy of the questionnaire design in the form of a PowerPoint document is in Appendix A.

- ❑ *Household Information*: Household data collected in the recruit survey included the following variables:
  - o Household composition (number of members and relationships to householder)
  - o Household demographics (e.g., income)
  - o Current home location (Figure 4 shows screenshot of the online survey instrument for collecting location information), type, and tenure
  - o Number of household vehicles
  - o Administrative data (e.g., contact information and incentive preferences)

Figure 4: Example Home Location Question

- ❑ *Person Information*: The person-level details collected in the recruit survey included:
  - o Person-level demographics (e.g., age, gender, employment status, education level)

- o Work details (if employed) (e.g., work locations, occupation/industry, typical commute behavior); Figure 5 shows example questions of participants who commute to work
- o School details (if applicable) (e.g. school locations, typical school commute behavior)
- o Other travel behavior questions (e.g., driver's license possession, frequency of transit use)

Figure 5: Example Participant Commute Questions

- ❑ *Vehicle Information*: Respondents were asked how many motor vehicles were in their household (including leased or company-owned vehicles, but excluding recreational vehicles). For each vehicle, respondents were asked to provide the following details:
  - o Year
  - o Make
  - o Model
  - o Presence of a toll transponder (e.g., EZ Pass)

Information collected in the recruit survey was later used to prepopulate responses for questions in the travel diary survey. For example, the vehicle information was used to prepopulate responses when a household vehicle was used to complete a trip. Any vehicle listed in the recruit survey became a mode option for trips reported in the travel diary survey.

2.3.2 Travel Diary Survey

In the travel diary survey, details regarding all trips completed by members of the respondent household over a preassigned 24 hour period were collected. Additionally, this survey included a small number of follow up questions about other activities on the travel date (including reasons why no trips were made if such was the case and if it was a typical travel day or not). The travel diary survey was made available to respondents on the day after their assigned travel date had passed.

The travel diary survey was designed to collect detailed information about the trips. The first question was a proxy reporting question to determine whether the respondent was filling out his or her own survey, was present while another household member filled out the survey, or was not present while the survey was filled out by another household member. Respondents then reported where they started and ended their travel day (defined as 24 hours, beginning at 3 a.m. on the travel date). Respondents were then asked to provide a full list of all of the places they visited during the travel date – this is referred to as the trip roster (see Figure 6 for an example of trip roster on the online survey instrument). If respondents did not go anywhere, they skipped this roster, and instead were asked to select one or more reasons why they did not travel.

Figure 6: Example Online Trip Roster

After the trip roster page, a prompt question asked respondents to verify that they had reported all of their trip destinations. This question listed the types of trips that are commonly overlooked and provided respondents with the opportunity to return to their roster to add these or any other types of trip destination locations they may have forgotten to report. Commonly under-reported trips include short trips (e.g., stopping for gas or running a short errand on a lunch break) and loop trips (e.g., walking the dog or going for a run). Specific instructions for loop trips included a graphic that showed how those trips should be reported.

Once the roster was filled out, the travel diary survey then collected the location for each place that the person went on their travel day (located by searching for an address or placing a marker on a map). Respondents located each place that they listed in their roster. If they went home or to their usual school or work location (reported in the recruit survey), this location would be prepopulated and the respondent could confirm the address. If they went to the same place more than once (e.g., dropping a child off at an activity and then picking them up from the same place later), they only had to locate that location once.

Finally, the travel diary survey collected details about each individual trip (see an example in Figure 7). As previously noted, some questions were skipped when they were not relevant to a given trip.

- ❑ For all trips, respondents were asked:
  - o when they traveled (when they started traveling and when they arrived)
  - o the main purpose of their trip (e.g., go to work, personal errands, eat at a restaurant, drop off another person, shopping, etc.)
  - o the main way they traveled (e.g., in a household or other vehicle, riding transit, walking, etc.)
  - o who they traveled with if applicable (other household members, non-household members, or a combination of both)
- ❑ For private vehicle trips, respondents were asked about:
  - o the vehicle used
  - o the type of parking location and cost to park there (if applicable)
  - o whether they used a toll road on that trip (if applicable)
- ❑ For carpool or vanpool trips (i.e., vehicle trips with other people in the travel party), respondents were asked:
  - o whether they were the driver or passenger
  - o whether an High Occupancy Vehicle (HOV) lane was used (if applicable)
- ❑ For transit trips, respondents were asked:
  - o how they got to and from the transit stop
  - o the cost of the transit fare
  - o which specific transit system(s) and route(s) were used
- ❑ For taxi and rideshare trips, respondents were asked:
  - o the total cost of the fare for the trip

Figure 7: Example Set of Trip Detail Questions

An important feature of the travel diary was the ability for respondents to copy trips where they were reported as a member of the travel party with other household members. The first household member to complete the travel diary would provide full details for the joint trip with other household members. Subsequently, the other household members can copy the trips and thus avoid having to provide the same information again. They also have the option of altering the purpose of the copied trips as the purpose for one member may be different from the purpose of a different member (e.g. for a trip where a parent is dropping off their child to school, the purpose for the parent will be "drop-off" whereas the purpose for the child will be "school"). Once the trips have been copied, the other household members can add more locations to the trip roster to complete their travel diary.

2.4 Survey Implementation

2.4.1 Survey Instrument Development

Resource Systems Group, Inc. (RSG) programmed an online survey instrument as the primary data retrieval option for respondents. The online survey used RSG's proprietary survey software, rSurvey™, which includes rigorous protocols to protect data during and after data collection. The survey, conducted over the internet, is hosted on a secure https website; the rSurvey™ application uses Microsoft Azure to run the survey and store responses. Microsoft Azure employs numerous protocols and tools to ensure data security and data privacy.

As a further step to protect respondent privacy, all contact information was stored separately from response data. The only link between the mailed study invitations and the survey responses was a unique, randomly assigned password provided to respondents. This password was only used for administering the study and was not included in the final datasets, and only a small number of trained study administration staff had access to the passwords and contact information. Participating households entered their unique password and completed the survey through the online survey portal, which was accessible from the project website.

rSurvey™ also has several features that ensure data quality and minimizes respondent burden. One feature of rSurvey™ is that participants who stop midway through the survey arrive at the question they last answered when they return to the survey (with all previously provided data saved). Other functionalities to ensure data consistency and minimize respondent burden included real-time data validation, as described in Section 2.4.2 below.

ETC Institute (ETC) conducted all telephone activities for the survey using the same online survey tool to administer the survey over the phone if respondents preferred this option. Respondents who preferred to complete their survey over the phone were read the same web-based survey that web participants used. By administering the same survey to both phone and online participants, respondents who used the call-in option were fully integrated in real-time with all web respondents. ETC operators also had additional materials and information on hand such as the project Frequently Asked Questions (FAQs), the invitation letter, and a study overview document created by RSG, to interact with household members in an informed way and facilitate the process of data collection. Both English and Spanish speaking interviewers were available to assist callers.

Paper diaries that were offered to segment IDs 5 and 6 were designed and printed by UConn. The diaries were mailed to participants. Once the diaries were returned, the data was manually recorded.

2.4.2 Data Validation

An important step in collecting and preparing accurate, high-quality survey data is to ensure that respondents understand the questions and provide complete, accurate data as they take the survey. RSG's proprietary survey software, rSurvey™, includes built-in data checks that reduce respondent confusion, improve survey flow, reduce burden, verify response completeness and consistency, facilitate dataset preparation, and reduce the amount of data cleaning and up-coding required. A few examples of these built-in data checks include the following:

- ❑ Validation logic required that respondents answered all questions on a page (where a response was required) before continuing the survey, ensuring complete response records.
- ❑ Spatial validation, such as real-time geocoding of addresses, businesses, or location points on a Google map, ensured complete geographic data.
- ❑ Reported trip sequences were required to be spatially and temporally logical (i.e., one trip's end location was required to match the next trip's starting location, and a trip's starting time could not be before the previous trip ended).
- ❑ Filters to automatically show or hide certain questions based on previous responses helped reduce respondent burden (e.g., unemployed respondents were not asked commuting questions).
- ❑ Metadata collection (passive collection of data such as survey duration and browser type) was used to help troubleshoot survey errors and assist households that called or e-mailed for help. These metadata also helped inform improvements to the survey design between the pilot and main data collection periods.
- ❑ The "copy trips" feature in rSurvey™ allowed a household member to select and copy information for trips made with other household members who had already reported those joint trips. This feature, described in more detail in the previous section on survey design, ensures that jointly made household trips were reported with the same locations, modes, and trip times.

The above data validation features were not available for the data recorded from the paper diaries. These checks were manually done to qualify the completeness of the travel diaries returned by those who requested the paper survey mode.

2.5 Branding and Survey Materials

2.5.1 Survey Branding

A unique study branding was created with input from CTDOT. All study resources and materials (e.g., the study website and print materials) were branded in order to create a cohesive public profile for the study and to build legitimacy. The first step in this process was to develop the study name and logo.

Figure 8 shows the logo and the study name. This logo was then included on all official study materials. The study name, icons, fonts, and color schemes selected for the logo also influenced the design of all other aspects of the invitation and outreach materials.

Figure 8: Study Logo

2.5.2 Print Materials

Initial recruitment of households began with invitations delivered via First-Class Mail. Households received a pre-notice postcard informing them of the survey, an invitation packet inviting them to participate in the survey, and two reminder postcards. Below are more details about these printed invitation materials. A copy of the printed invitation materials are included in Appendix B.

- ❑ *Pre-notice Postcard*: A pre-notice postcard was sent and arrived approximately 7-8 days prior to the assigned travel date notifying each potential respondent household that a formal invitation would be arriving. The postcard included the study website and phone number, the household's password, and information about the study incentives. Households were invited to log onto the website or call the toll-free number in order to learn more about the study and to fill out the recruit survey.
- ❑ *Invitation Packet*: A formal invitation packet arrived shortly after the pre-notice postcard – approximately 4-5 days prior to the assigned travel date. The packet included a letter branded with the study logo and printed on CTDOT letterhead. The contents of the packet explained the purpose for collecting residents travel behavior, provided the household's assigned travel date, and described the steps necessary to complete the study. The packet also included reminders about the website address and phone number for providing responses. Other materials included in the invitation packet were travel logs and a sheet with study FAQs.
- ❑ *Reminder Postcard One*: A reminder postcard arrived approximately 1-2 days before the assigned travel date to encourage every household to complete the travel diary. It reminded households regarding the study phone number, website address, and participant travel date and login information.
- ❑ *Reminder Postcard Two*: A second reminder postcard arrived approximately 1-2 days after the assigned travel date as a final reminder to complete the travel diary. It was identical to the first reminder postcard.

UCONN's Connecticut Transportation Institute (CTI) address was established as the study address on all printed materials, and was used as the return address for any invitations that were undeliverable.

The study team established a print material QA/QC process to support smooth transitions between each stage of the invitation process (from address list development to print production to mail distribution). In the first stage, sampled address list purchased from MSG was reviewed to check for duplicates and out-of-state addresses. Duplicates and out-of-state addresses were subsequently removed. In order to keep the number of invitations at the level determined in the sampling plan, a small percentage of additional addresses were requested from MSG. The address list was then processed and prepared for bulk mailing (including identifying forwarding addresses for any households that had moved, eliminating duplicates, and mail sorting). Finally, the list was uploaded to the online survey instrument's database to be used for administering the survey online and over the phone. For each weekly mailing, electronic proofs were reviewed for quality, by conducting spot checks, and consistency, by comparing against the master database to ensure consistency between mailed materials and online materials for each password and address pair. Survey invitation materials were then mailed out to potential respondents.

2.5.3 Survey Website

The study team developed a project website specifically for the study. The website served two primary purposes. First, it hosted the online survey instrument where invited households could complete the recruit and travel diary surveys after entering their assigned household password. Second, it provided general information about the study including answers to FAQs, news mentions, and contact information. The informational portion of the project website was available to the public as well as invited households. The home page of the website is shown in Figure 9. The website was published in early fall before the pilot survey was fielded and was updated prior to the main data collection effort.

Figure 9: Study Website Homepage
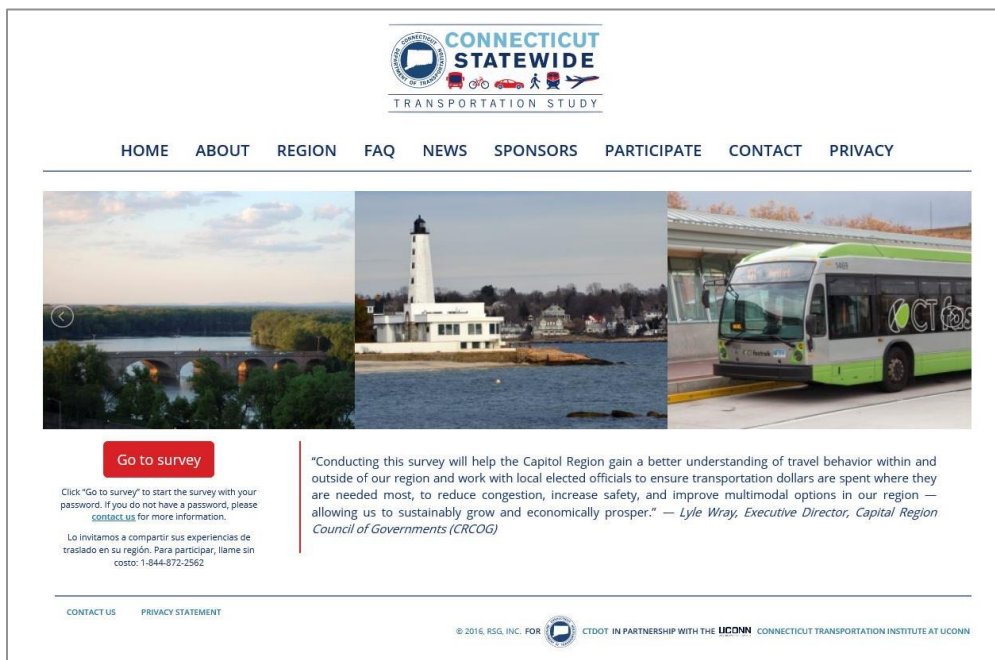
2.6 Survey Administration

2.6.1 Participation Methods

All respondents were recruited by USPS First-Class Mail. Potential respondents can self-recruit themselves by visiting the website or by calling the phone line. In the address list purchased from MSG, a subset of the households had an associated landline number. These households were actively recruited if they did not already self-recruit themselves. As noted in the previous section, several print materials were mailed to each invited household. Once households completed the recruit survey, they received email or telephone reminders before and after their assigned travel date, until they completed the diary (up to one week after the travel date). Consistent visual elements were used across all print and online materials (including the online survey, website, and reminder emails), with the intent of connecting all invitations, reminders, and other notices concerning the project in order to maximize response rates. All online and print materials included a toll-free telephone number, the website URL, the study email address, and the unique password assigned to each household. All aspects of this process were reviewed and refined with feedback from CTDOT.

Participants could complete the study through a combination of the three response methods, as described below.

❑ *Online Survey*: Households invited to take the survey were able to enter their unique password and complete the survey through the self-administered online survey instrument, which was accessible from the study website. rSurvey™ has several features that ensure data quality and minimize respondent burden as described in previous sections.

❑ *Telephone Survey*: ETC was responsible for all telephone communications. Each telephone interviewer underwent training for the 2016 CSTS, which included review of the online survey, print materials, and study website. Training documents including the questionnaire (screen by screen) for reference, guidelines for what operators should say, and a study overview sheet for quick reference while on the phone, were prepared. ETC also fielded incoming calls and made outbound calls to households with a known phone number. Both English and Spanish speaking interviewers were available to assist callers. Households could call to participate over the phone or call with specific questions. Respondents who preferred to complete their survey over the phone were read the same web-based survey that web participants used as described in previous sections.

❑ *Paper Survey*: For those households who belonged to segment IDs 5 and 6, and who preferred to complete the travel diary by paper mode, personalized mailings and instructions were distributed for completing the paper travel diaries. They were also provided a postage paid return envelope and instructions for returning the diaries.

2.6.2 Communication

The project team conducted several activities to engage invited households and the general public. The primary goal of communication activities was to encourage invited households to participate in and complete the study. Part of public engagement included responding to inbound questions from both invited households and the public. Residents (invited or not) were

welcome to provide comments about transportation issues or the survey process. Respondents continually received information about the study, and they were able to communicate with the survey team to ask questions or provide feedback. Multiple communication methods were utilized: the mailed invitations, the project website, and email and telephone contacts. Details regarding the different telephone communication activities are described below:

❑ *Outbound Telephone Recruitment*: After printed invitation materials were sent, households with known telephone numbers received calls encouraging them to participate in the survey. Households who agreed could complete the recruit survey over the telephone or could participate online on their own. ETC conducted recruitment calls up until the day before a household's travel date.

❑ *Outbound Telephone Reminders*: For households that preferred receiving reminders via telephone, reminder calls were made using the following process:
  o A telephone call was placed to the household on the day before their travel date to remind the household to track their travel the following day.
  o Additional calls were placed (for up to seven days after the travel day) to the households to remind them to complete the survey online or over the telephone. The timing and frequency of telephone calls varied based on households' previously expressed preferences for a call back and the ease of reaching the household.
  o After seven days from the travel date passed, no additional phone calls were placed to that household.
  o It was ensured that all reminder phone calls were placed on time and that scripted voice messages were left if they reached a voice mailbox instead of a live person.

❑ *Inbound Telephone Calls from Respondents*: ETC also responded to inbound telephone calls, including questions and requests to complete the survey over the phone. If an operator did not know the answer to a question or needed more information, then they reached out to RSG for guidance. In cases where a participant was having trouble completing the survey, the operator would help them complete the survey over the phone. Calls from households who wanted to report their travel on their travel date (rather than on the day after) were scheduled for callback. Calls to ask about the household's gift card status were forwarded to RSG for resolution. ETC would then call the household back with a status update.

Overall, call prioritization was based on the following criteria: 1) inbound calls, 2) reminder calls, 3) recruitment calls.

Details regarding different types of email communication activities are described below:

❑ *Outbound Email Reminders*: Email reminders were sent to households that had completed the recruit survey and preferred email contact. The emails were aimed at encouraging them to log and report trips for their assigned travel date. These emails also described the reporting process. Reminders included a link to the survey website, the household password, and the toll-free telephone number. If a

household completed the recruit and travel diary survey at the same time (after their travel date), they did not receive any email reminders. If a household had not reported travel after seven days past the assigned travel date, the survey was closed and no further email reminders were sent to the household. Approximately 53% of all recruited households received one or more email reminders. Depending on when a household was recruited and when they completed the last part of the survey, they received up to four reminder emails:

- o Reminder 1: sent the day before the assigned travel date (reminder to log travel the following day).
- o Reminder 2: sent the morning after the assigned travel date (reminder to report travel from the previous day).
- o Reminder 3: sent three days after the travel date (only if travel had not yet been reported).
- o Reminder 4: sent five days after the travel date (only if travel had not yet been reported).

❑ *Inbound Emails from Respondents*: RSG monitored and maintained project email accounts for the pilot and main surveys. As standard protocol, they responded to all email inquiries within one business day. Inquiries sent via email typically were for clarifications (e.g. how to fill out certain questions) or requests for information (e.g. study purpose).

Additionally, residents who had not been invited to the survey as part of the random sample were also welcome to complete the survey as volunteers. However, this volunteer data was distinguished from the statistically representative sample of invited households. Only those volunteer households that completed the entire main study (recruit survey and travel diary survey) were archived. Data for these volunteers are not included in the final dataset.

2.6.3 Incentives

For the CSTS pilot study, two different incentive plans were provided based on the segment to which the respondent belonged. Households belonging to segment IDs 2 and 4 were offered a guaranteed incentive, while segments with ID 1, 3, 5 and 6 were offered entry into a raffle for an iPad[23]. Both incentives were distributed only after completion of the recruit and travel diary portions of the survey by all members of the household. For households offered a guaranteed incentive, the payment was in the form of a $10 gift card from their choice of Amazon or Walmart. After completing the recruit survey, some pre-defined hard-to-reach households were offered an additional $10 incentive (for a total of $20) to help promote participation (see Section 2.2 for more information).

2.6.4 Response Monitoring

Once households began completing the study, the responses were monitored to ensure:

❑ Progress toward the total target of 250 complete households
❑ Progress toward the targets within each sample segment (discussed in Section 2.1)

---

[23] A 32GB Apple iPad Air Model MD7866LL with 9.7 inch display was offered as a raffle prize.

❑	Quality of the data and assessing errors with data collection

In order to monitor the responses, an online dashboard was created that was available for viewing by different partners of the study team and CTDOT. The dashboard included aggregate distributions along key factors of interest including response rates overall and by segment, person-level variables, household-level attributes, and travel characteristics. The tables in the dashboard were monitored to evaluate objectives described above.

2.6.5 Public Outreach
Since the pilot study was aimed at collecting a small number of responses, the study team did not engage in any public outreach efforts. The study team also did not want to cause potential study fatigue that could possibly impact response rates for the main study.

2.7 Data Processing and Preparation
In addition to the real-time checks used during data collection, a number of additional quality checks and data cleaning procedures were conducted after data collection was completed. The purpose of these checks was to confirm that the real-time controls worked correctly and consistently, and to evaluate any data issues that could not easily be resolved by real-time rules. Additionally, certain records and variables were excluded from the dataset to create a dataset that only includes complete valid responses and does not include personally identifiable information. Post-data collection checks included the following:

❑	Review of all data tables to check for missing data where responses were expected
❑	Inspection of potential outliers (i.e. responses that are allowed but are typically outside the normal/ expected distribution, such as trips with long reported travel times and short distances)
❑	Review of geocoded address fields for potential Google write-out errors
❑	Review of geocoded home locations (compared to the study area)
❑	Review of copied-trip details for potential inconsistencies

Household record exclusions checked for and implemented in this dataset included the following:

❑	Incomplete households (where some or all members did not complete the diary)
❑	Reviewer households (any records from UConn, RSG, CTDOT, or other stakeholders)
❑	Households outside of the study area
❑	Duplicate households

All person-, vehicle-, and trip-records associated with excluded households were also removed.

Variable exclusions from this dataset included the following:

❑	Sample mailing address information (purchased from the sample provider)
❑	Passwords (associated with the sample mailing address information, replaced with ID numbers)
❑	Email addresses and phone numbers (collected for survey administration only)

2.8 Results

2.8.1 <u>Response pattern analysis</u>

The pilot study resulted in complete responses from 250 households exclusively offered online and telephone options, and an additional 37 complete or partially complete responses from households offered a paper diary option. Table 5 summarizes the pilot response rates.

As expected, response rates were significantly higher from households offered a guaranteed gift card compared with raffle incentive. Additionally, between the hard-to-reach and regular households, the response rates were higher for the regular segments. When comparing the observed and target response rates, for the non-paper segments, response rates from the hard-to-reach areas were somewhat lower than the target, while response rates from the regular sample area were higher than target. The observed response rates for both paper segments were lower than expected for both regular and hard-to-reach groups.

Conversion rates for the non-paper segments were much higher (in the range of 73.8% to 83.9%) when compared to those segments where a paper survey was offered (41.1% to 46.7%). Within the non-paper segments, it was interesting to note that the conversion rate is higher for the hard-to-reach segment when offered a raffle versus when they were offered a gift card. On the other hand, for the regular segment the conversion rate for the gift card was higher than the raffle segment. It appears that for the hard-to-reach segments, the raffle with larger payout, even though it has lower odds, is more appealing whereas for the regular segments, the gift card with the more guaranteed payout is more appealing.

Upon analyzing the non-responses, the patterns were consistent with other studies. When only reviewing the data from the online survey instrument, it was found that 61 households dropped out of the recruit survey for a retention rate of 84 percent. Approximately 41 percent of the households actually dropped out without answering any questions – these households were likely curious about the survey but did not intend to participate. Out of all those who began the travel diary, 21 dropped out of the diary survey for a retention rate of 96 percent. The primary challenge was still encouraging households to begin the study. As a result, in the main study, the study embarked on numerous outreach efforts as discussed in Chapter 3.

Table 5: Summary of Responses

ID	Sample Segment Description	Diary Mode Options	Target Responses			Actual Responses			
			# Completes	Response Rate	Invites	# Recruited	# Completed	Response Rate	Conversion Rate[24]
1	Hard-to-reach / Raffle	Online/ Phone	50	2.0%	2,500	49	38	1.5%	77.6%
2	Hard-to-reach / Gift card	Online/ Phone	50	4.0%	1,250	61	45	3.6%	73.8%
3	Regular / Raffle	Online/ Phone	50	3.0%	1,670	91	68	4.1%	74.7%
4	Regular / Gift card	Online/ Phone	50	5.0%	1,000	118	99	9.9%	83.9%
5	Hard-to-reach / Raffle/ Paper	Online/ Phone/ Paper	25	2.0%	1,250	30	14[25]	1.1%	46.7%
6	Regular / Raffle/ Paper	Online/ Phone/ Paper	25	3.0%	840	56	23[26]	2.7%	41.1%
Total			250	2.9%	8,510	405	287	3.4%	70.9%

[24] Ratio of households who completed both recruit and travel diary to those who completed the recruit portion.

[25] 9 out of the 14 responded using paper mode for the travel diary. The data for these 9 households are either complete or partially complete owing to the mode used for data collection.

[26] 5 out of the 23 responded using paper mode for the travel diary. The data for these 5 households is either complete or partially complete owing to the mode of data collection.

Figure 10 shows the concentration of pilot responses in the different sample areas, aggregated to the town level. There are more responses per town in Hartford and Norwalk. The response rates in these towns (approximately 1.4% and 3.6%, respectively), however, are lower than New London County as a whole (approximately 4.6%). This can be attributed to the higher concentration of hard-to-reach households in Hartford and Norwalk (as shown in Figure 2), who report at lower rates.

Figure 10: Counts of Respondent Households by Towns
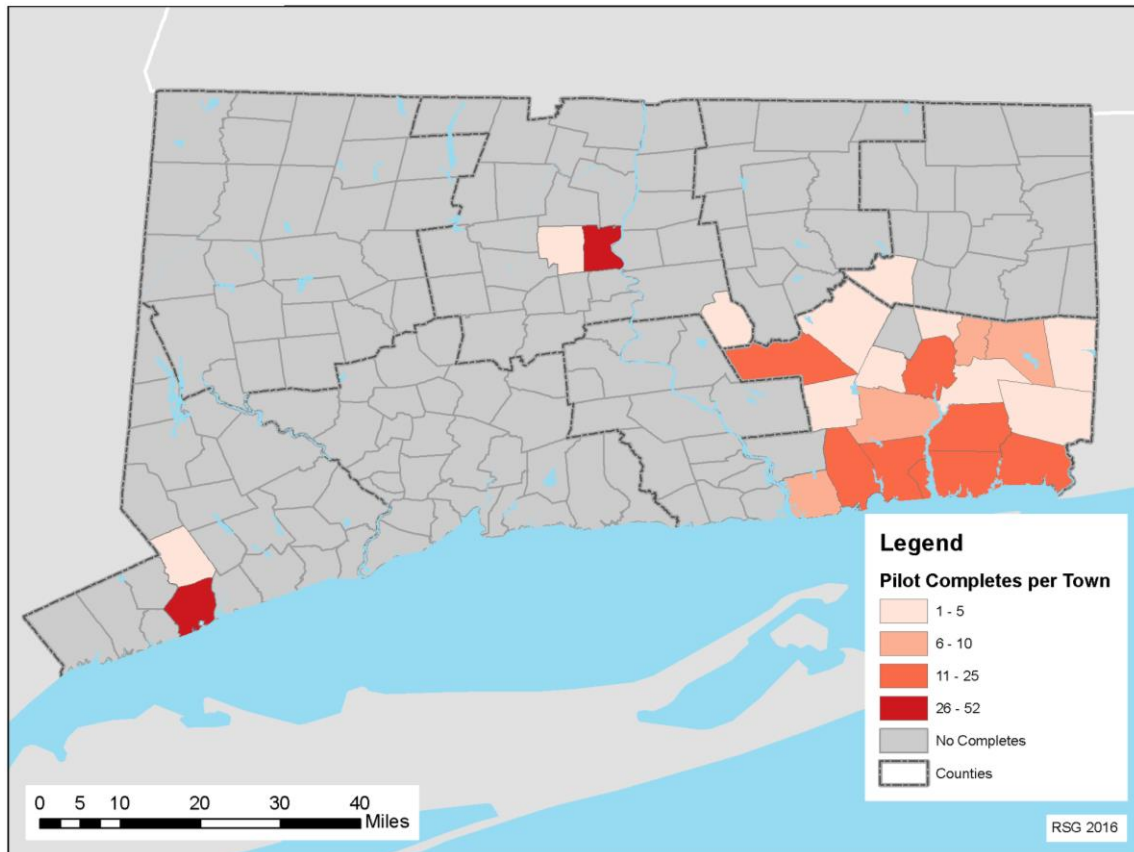
Based on findings from the pilot survey, the sampling plan was enhanced for the main study (as discussed in Section 0

2.8.2 Respondent Incoming Communication

In this subsection, inbound communications in the form of emails, phone calls, and open-ended feedback from the online survey instrument are summarized. A total of 20 emails and 37 phone calls were received. The majority of the inbound emails and phone calls sought clarifications to questions (e.g. how to answer certain questions), and requests for more information (e.g. study purpose). In addition, a total of 378 comments were received in the text boxes at the end of the online survey instrument. Nearly two thirds of the comments were non-specific (e.g. "no comment", "all was clear") or positive (e.g. "thanks for inviting me", "easy to complete", "glad I could copy my kid's trips"). Of the specific comments, nearly 30 were related to personal travel experiences and suggestions for transportation improvements. Another 100 were related to the

survey, which were analyzed further in an effort to improve the main study. Common themes of the survey comments include, concerns regarding length of the survey and privacy concerns, confusion about instructions, and feedback to specific sections of the survey. The comments aligned closely with expectations and helped to guide changes that were recommended and implemented in the main study (see Section 2.9 .

2.8.3 Survey Results

Household, person, travel and vehicle characteristics from the pilot survey were analyzed for reasonableness and to identify improvements to the main study. Since, the primary focus of the pilot survey was to evaluate the overall process, specific survey data results are not discussed in this chapter. Also, since no weighting analysis was performed, tabulations of sample characteristics will be subject to sampling biases and thus the results would not be interpretable. As a result, a presentation of the survey results is reserved for Chapter 3 utilizing the full data where sampling bias corrections were applied.

2.9 Main Study Recommendations

The pilot response rates and other observations from the pilot responses helped refine assumptions and survey design choices. A number of changes were made to the materials and methods for the main study. Key observations and the resulting changes included the following:

- ❑ Response rates in the pilot study were lower than initially predicted for areas with high concentrations of hard-to-reach households. Therefore, more conservative rates were used to estimate main survey responses and to predict the number of invitations needed to meet sampling targets for these segments.
- ❑ The pilot results showed that respondents who were offered a guaranteed incentive responded at higher rates than those who were offered a raffle. Multiplying this effect over the larger sample required for the main study and estimating the effect on invitation costs showed that guaranteed incentives would provide overall project savings compared to the raffle. Therefore, guaranteed gift card incentives were provided for all respondents who completed the main study.
- ❑ The portion of pilot respondents who were offered the option to fill out a paper diary responded at lower rates than those who were only offered the online and telephone options. Paper diary responses also allowed for skipped questions and more incomplete records. Therefore, the survey retrieval options were simplified for the main study to only include online and telephone response options.
- ❑ Pilot respondents provided feedback that questions about certain trip detail questions were burdensome to answer repeatedly for each trip, particularly when those details were about facilities that a person never used on their travel day (e.g., use of paid parking, toll roads or HOV lanes). Earlier questions had asked respondents about their typical weekly use of these facilities in order to filter these questions for people who never used these facilities, but did not reduce burden for many respondents who used these facilities occasionally. Therefore, in the main survey these questions were revised to ask respondents about their use of these facilities on their specific travel day, which then allowed more accurate filtering and a more effective reduction of survey burden.

In addition to these specific changes, a number of small changes were made to the study information and survey instruction text throughout the printed materials as well as the survey instrument in order to increase the clarity of instructions and improve response quality. These changes were made based on specific feedback from participants who identified parts of the survey that were ambiguous or confusing. For example, there were questions about how to use the geocoder to report their trip locations or whether non-auto trips should be included. Details on these types of administrative changes as well as other aspects of the study design are provided in Chapter 3.

CHAPTER 3 MAIN STUDY

The primary goal of the main study was to collect complete data from a representative set of 7,500 households to support the statewide model development and application. A secondary goal was to collect travel information from a sufficient number of transit-users and zero-vehicle households (from within the total sample) in order to better model travel behaviors of these groups in the statewide travel demand model. Furthermore, in comparison to the pilot study, additional data processing and data preparation tasks were pursued in the main study to prepare a final usable dataset.

The remainder of this chapter is organized as follows:

❑   Description of the sampling plan is provided in Section 3.1
❑   Incentive plan is descried in Section 3.2
❑   The questionnaire design and development process is presented in Section 3.3
❑   Survey implementation is presented in Section 3.4
❑   Survey design and branding materials are discussed in Section 3.5
❑   Survey administration protocol and procedures are presented in Section 3.6
❑   Data processing approach and preparation of the final data are presented in Section 3.7
❑   Results from the main study are presented in Section 3.8

3.1 Sampling Plan

An address-based sampling approach was adopted similar to the pilot study. A sample of households were invited from those residing in the state of Connecticut using a combination of simple geographic proportional random sampling, targeted oversampling (higher sampling rates) in selected areas, and upsampling (increased invitation rates) in certain areas where higher proportions of hard-to-reach households (typically report at lower rates) are identified. Details regarding the sampling frame, sample stratification scheme, sampling and invitation rates, invitations, and recommended travel date assignment are presented in the following subsections.

3.1.1 Sampling Frame

Similar to the pilot study, the geographic unit used for analysis and development of the main study sampling plan was block groups. These block group boundaries are as defined in the 2010 Census. Block groups are the smallest spatial resolution at which the detailed marginal distributions of various household and person variables are available from the US Census Bureau. In preparing the sampling plan, marginal distributions obtained from the 2009-2013 American Community Survey (ACS) were used. It must be noted that the marginal distributions used for the sampling plan are a little dated and do not reflect population distributions observed presently. While marginal distributions for more recent years are available from Census, they were only available at the higher spatial resolution of Census tracts. Census tracts, a coarser spatial resolution than block groups, offer less control of the sampling plan and are, as a result, less desirable. The choice of block groups as the spatial resolution (block groups) and the 2009-2013 ACS for the marginal distributions was made to develop a sampling plan that

utilized the maximum amount of information at the most disaggregate spatial resolution available.

The sampling frame for the main study included all residential addresses in the state of Connecticut. Unlike the pilot study, all regions in the state were considered in the sampling plan. The sample list of addresses for potential respondents was obtained by randomly drawing from all residential addresses based on pre-specified spatial stratification and sampling rate. The sample list was obtained from Marketing Systems Group (MSG), a data vendor who maintains the full list of residential addresses utilizing the United States Post Office's Computerized Delivery Sequence File. Additionally, matching landline information for the sample list of addresses was obtained where applicable.

3.1.2 Sample Segments

A stratified sampling approach was adopted and the sampling plan was designed at the block group level. More specifically, targeted oversampling[27] was used to obtain a larger proportion of responses from transit-using and zero-vehicle households. In addition, upsampling[28] was utilized for areas with more hard-to-reach households and lower predicted response rates in order to mitigate non-response biases.

To achieve the primary and secondary goals identified above, the block groups were categorized into six segments. Marginal distributions for five different variables from the 2009-2013 ACS were analyzed at the block group level to incorporate oversampling and upsampling considerations in the sampling plan. Moreover, using data from the ACS to stratify the random sample assists the process for weighting the final dataset after data collection is complete, since the same data source can be used to determine weighting targets.

Variables Used for Stratification
Two of the five variables used to address oversampling in the final sampling plan include:

❑ Percent of resident workers who use transit to travel to work (representing transit commuters)
❑ Percent of resident households that do not own a vehicle (representing 0-vehicle households)

---

[27] "Oversampling" is used when larger proportions of certain behaviors or demographics of interest may be desired in the final sample. This is mainly done because the behaviors or demographics of interest may be rare in the population (e.g., transit riders) and proportional sampling will lead to sample counts that may not be adequate for subsequent analysis (e.g. mode choice model estimation). In oversampling, higher targets for complete responses from certain subpopulation groups are established. Accordingly, additional invitations are sent out to meet the targets.

[28] "Upsampling" is used when a representative sample (proportional to population) is desired, but certain subpopulation groups are known to respond at lower rates compared to others (e.g. large households, younger individuals). Not accounting for the differences in response rates will lead to biases when drawing inferences using the data collected (also referred to as non-response bias). To account for the differences in response rates across different population subgroups, the number of invitations is based on the rates at which different groups respond.

Transit commuters and households without vehicles make up a small percentage of Connecticut's population (4.7% and 9%, respectively, though this varies by region within the state). However, the travel behavior of these types of travelers and households was deemed important for policy analysis and modeling needs from discussions with CTDOT. Therefore, the study team decided to oversample these types of residents in the final sample.

The remaining three variables used to address upsampling in the final sampling plan include:

❑ Percent of resident households with incomes under $25,000 (representing low-income households)
❑ Percent of resident households with the head of household under age 35 (representing young households)
❑ Percent of resident households with 4 or more members (representing large households)

These variables represent the types of households that typically respond at lower rates in travel surveys. The pilot study results[29] also confirm this tendency (as shown in Appendix C – the proportions of pilot study respondents with lower incomes, younger householders, and larger household sizes was lower than population proportions for the respective household types from the 2009-2013 ACS data. The variables identified above were used to stratify the block groups and accommodate oversampling and upsampling considerations.

Stratification Criterion
The first step to stratifying the block groups to address oversampling and upsampling was to evaluate the distribution of the five variables of interest across the block groups in the state. To assist exploration of the distribution of the data, the values of the variables for each block group were categorized into percentile ranges as described below:

❑ Range 1: Contains block groups in the 0 to 25th percentile
❑ Range 2: Contains block groups in the 25th to 50th percentile
❑ Range 3: Contains block groups in the 50th to 75th percentile
❑ Range 4: Contains block groups in the 75th to 90th percentile
❑ Range 5: Contains block groups in the 90th to 100th percentile

As can be seen, the first three ranges (1-3) of percentiles represent the bottom three quartiles. The highest quartile was split to focus only on the top 10 percent of block groups for each demographic measure, resulting in the last two ranges (4-5). Table 6 shows the values for each variable corresponding to the different percentile cutoffs. For example, a value of 7.0% under the "% low-income households" indicates that 25 percent of the block groups in the state have less than 7 percent of their households in the low income category. Similarly, 90 percent of the block groups in the state have less than 39.8 percent of their households in the low income category.

---

[29] It must be noted that only pilot study results from the Segments IDs 2 and 4 were used as the modes were limited to online and phone and the gift card incentive type was adopted in the main study.

Table 6: Percentile Ranges for Key Variables of Interest in the Main Study Sampling Plan

Percentile Value	% low-income households	% young households	% large households	% 0-vehicle households	% transit commuters
25th percentile	7.0%	5.9%	15.5%	0.5%	0.0%
50th percentile	13.7%	12.3%	22.9%	4.2%	2.0%
75th percentile	24.6%	22.3%	31.2%	11.9%	7.2%
90th percentile	39.8%	33.5%	39.0%	26.2%	16.7%
100th percentile (max)	100.0%	90.2%	100.0%	84.2%	63.9%

After exploring the distribution of the variables of interest, the first stratification of the block groups was performed to address oversampling. The approximate 90th percentile value of transit commuters and 0-vehicle households variables were used to identify block groups for oversampling. The block groups that exceed the 90th percentile values for these variables represent those block groups where the transit or 0-vehicle households are most likely to live. Subsequently, this suggests the areas where oversampling is most likely to be effective, since the chance is greater that a 0-vehicle or transit household will be randomly selected into the list of invitees. Therefore, the first stratification of the sample plan created two types of block groups:

❑ *Oversample block groups* where more than 15% of commuters take transit OR where more than 25% of households do not own vehicles
❑ *Regular sample block groups* where less than 15% of commuters take transit and less than 25% of households do not own vehicles

After the first stratification based on shares of transit commuters and 0-vehicle households to address oversampling, the block groups were further separated into groups based on expected response rates. The second level of stratification primarily addresses upsampling considerations to account for differential response patterns across subpopulation groups. The oversample and regular sample block groups were each separated into three groups of block groups as described below:

❑ *Low response block groups* where the most invitations should be sent were identified using the approximate 90th percentile values for low-income and young households (since low-income and young households had the lowest response rates in the pilot). In particular, block groups where more than 40% of households have low incomes (less than $25,000) or more than 35% of householders are young (under age 35)
❑ *Medium response block groups* were identified using the approximate 50th percentile value for low-income households and the approximate 90th percentile value for large households (since medium-income and large households had a moderate response rate in the pilot). More specifically, block groups where more than 15% but less than 40% of households have low incomes or more than 40% of households are large (have 4 or more members)

❑ *High response block groups* included all of the other block groups. As with the oversample criteria, this stratification suggests where upsampling is most likely to be effective. In particular, all other block groups (where less than 15% of households have low incomes, less than 35% of householders are young, and less than 40% of households are large).

When the oversample and upsample criterion were combined, the stratification created six types of block groups as follows:

❑ Oversample, low response
❑ Oversample, medium response
❑ Oversample, high response
❑ Regular sample, low response
❑ Regular sample, medium response
❑ Regular sample, high response

By stratifying the block groups this way, study invitations and other recruitment efforts were appropriately targeted to the areas in an effort to achieve the main study objectives. Figure 11 shows a map of the block groups in the state of Connecticut based on the above stratification scheme.

Figure 11: Map Showing Block Groups in the Study Area Defined by the Stratification Segment
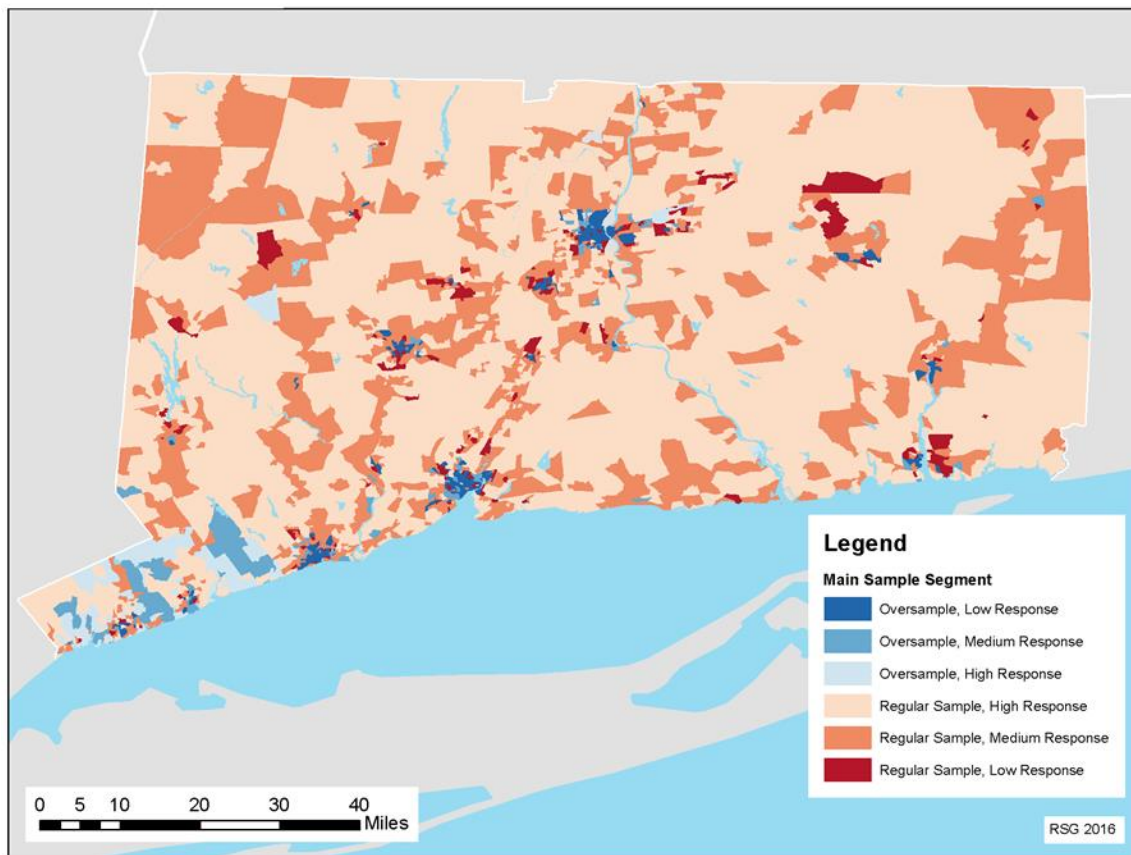
3.1.3 Invitations

Similar to the pilot study, once the sample strata were determined, the next step was to determine the number of desired responses and resulting number of invitations needed for each segment. The desired overall sample size was predetermined as 7,500 households, or approximately a 0.56% sampling rate of the total households in the state (typically a sufficient proportion to provide a statistically reliable sample). From this statewide average, the relative sampling rates for the oversample and regular sample segments were adjusted so that a higher proportion of responses would be obtained from transit users and zero-vehicle households (as previously discussed). The sampling rate for oversample segments was set to 1.1%, with the sampling rate of the regular sample naturally adjusted down to 0.45%. This aggressive oversampling rate (more than twice the regular sample rate) was set based on three factors: 1) previous experience, 2) information about the proportions of transit commuters and zero-vehicle households in these block groups (as not all respondents from high transit block groups would necessarily be transit users themselves), and 3) the cost implications.

For determining expected response rates, and the necessary number of invitations, the study team reviewed the pilot results as well as experiences from past studies. Also considered was the impact of the changes to the full-study incentive plan; since differential incentives will not be used in the main study. Table 7 shows the pilot response rate broken down by the full study sample segments.

Table 7: Observed response rate from pilot study for the full study sample segments[30]

Full Study Sample Segment	# Pilot Block groups in Segment	# Pilot HHs Invited	# Pilot HHs Completed	Pilot Response Rate	Assumed Response Rate
1-Oversample, Low response	73	884	28	3.2%	3.0%
2-Oversample, Medium response	25	202	7	3.5%	4.0%
3-Oversample, High response	9	27	3	11.1%	7.0%
4-Regular sample, Low response	25	180	6	3.3%	3.5%
5-Regular sample, Medium response	80	441	28	6.3%	6.0%
6-Regular sample, High response	137	585	72	12.3%	8.0%
Total	**350**	**2320**	**144**	**6.2%**	**5.1%**

It must be noted that the response rates presented in the table are based on invitations and completions from the gift card segments of the pilot study only because they are the most comparable to the full study plan. In the pilot study, the completion rates for the hard-to-reach segments are lower than what was observed from some previous studies, while the completion rates for other segments are higher than those observed in other efforts. Based on all of these factors, the expected response rates for each segment were adjusted from the pilot study as shown in Table 7. Also, assumed response rates closely follow observations from the pilot study

---

[30] In the table, the short form HHs is used to denote households.

while also accommodating contingencies in case pilot survey response observations were unusual.

Similar to the pilot study, the invitee count was obtained by dividing the anticipated number of responses for each segment by the assumed response rate for the segment. To obtain the 7,500 responses, a total of 150,006 invitations were mailed. The breakdown of the invitations by individual segments is shown in Table 8.

Table 8: Characteristics, Assumed Response Rates, and Target Responses for the Main Study Segments[31]

Sample Segment Description	Target Sample Size[32]	Sampling Rate[33]	# HHs[34]	Initial Response Rate[35]	Initial Invites, # HHs[36]	Initial Invites, % of HHs[37]
1-Oversample, Low response	1,367	1.10%	124,269	3.0%	46,647	37.5%
2-Oversample, Medium response	685	1.10%	62,256	4.0%	17,530	28.2%
3-Oversample, High response	361	1.10%	32,816	7.0%	5,273	16.1%
4-Regular sample, Low response	411	0.45%	91,414	3.5%	11,996	13.1%
5-Regular sample, Medium response	1,967	0.45%	437,164	6.0%	33,557	7.7%
6-Regular sample, High response	2,736	0.45%	607,930	8.0%	35,003	5.8%
Total	7,500	0.56%	1,355,849	5.0%	150,006	11.1%

The sampling targets and response rates were not further stratified by county, town or other sub-state geographies. Response by county, however, was monitored throughout the study. After the sampling targets and invitations were distributed among the block groups within each sample segment, the number of desired responses and invitations were tallied for each county. Expected response rates and sampling rates were calculated based on these distributions, as shown in Table 9.

This table reflects the number of invites at the start of the study (after processing/excluding duplicates and households that moved out of state), not including "extra" samples added later in the study (discussed below).

---

[31] In the table, the short form HHs is used to denote households.

[32] Target sample size is the number of households who have completed the entire survey (including the recruit survey and travel diary survey).

[33] The sampling rate is the percent of total households who have completed the survey (i.e. number of responses / total number of households).

[34] This is the estimated total number of households within the corresponding block groups based on ACS 5-year estimates (2009-2013).

[35] In the CSTS, a complete household is defined as one where every person in the household answers every question (no missing responses). The predicted survey completion rates are based on demographics, incentives offered, and recent similar studies. Actual completion rates for the pilot were used to design the main study sampling plan.

[36] Number of households invited is determined by the desired sample size and predicted completion rate (sample size / completion rate).

[37] Percent of households invited is the percent of all households expected to receive study information.

Table 9: Main Sample Plan, Targets and Initial Invitation Counts by County[38]

County	# HHs	Sampling Rate	Target Sample Size	Initial Response Rate	Initial Invites, # HHs
Fairfield County	332,655	0.64%	2,136	5.0%	42,801
Hartford County	347,874	0.55%	1,919	4.8%	39,710
Litchfield County	75,755	0.44%	332	6.3%	5,244
Middlesex County	66,141	0.48%	315	6.0%	5,209
New Haven County	328,013	0.56%	1,837	4.7%	39,012
New London County	107,066	0.50%	531	5.3%	10,090
Tolland County	54,327	0.46%	248	6.3%	3,950
Windham County	44,018	0.48%	210	5.3%	3,990
Total	**1,355,849**	**0.56%**	**7,527**	**5.0%**	**150,006**

The resulting expected sample rates per county from this distribution ranged between 0.44% - 0.64%. No single county had an aggressively high or low sample rate (compared to the statewide average). Fairfield, Hartford and New Haven counties had slightly higher rates than the rest of the counties as these counties contained higher concentrations of transit commuters and zero-vehicle households found in the "targeted oversample" segments.

Mid-Point Sample Adjustment
As responses came in, the study team closely monitored progress toward the final targets for the state as a whole and for each sample segment and county. After three weeks of data collection was complete, a detailed evaluation of progress toward these targets was conducted as well as comparisons to statewide demographic profiles. A number of adjustments were made to the study administration based on this evaluation. This evaluation and the implemented adjustments are discussed in more detail in Section 3.6 .

In addition to the administrative modifications, an adjustment was made to the invitation list. One of the findings in the evaluation was that the response rates from the "Oversample, High Response" sample segment and from Fairfield County were notably lower than initially predicted (1.5 and 0.3 percentage points, respectively). Projected through the end of the project, it was estimated that these rates would result in slightly fewer responses (approximately 80) than the 360 targeted in the "Oversample, High Response" segment (78% of target), and also fewer responses (approximately 150) than the 2,136 responses targeted in Fairfield County (93% of target). Note that there was significant overlap between Fairfield County and the "Oversample, High Response" segment (as can be seen in the distribution of light blue block groups in Figure 11). This is largely due to the fact that the "Oversample, High Response" segment could generally be characterized as higher income transit riders, such as residents who ride commuter rail into New York City.

---

[38] This table reflects the number of invites at the start of the study (after processing/excluding duplicates and households that moved out of state), not including "extra" sample added later in the study (discussed below). Also, in the table, short form HHs is used to denote households.

Given the importance of this segment of travelers in helping planners understand and model travel behavior, and given the high density of travelers, transportation facilities, and potential future investments in Fairfield County, the study team decided to add more invitations in Fairfield County and in the "Oversample, High Response" segments in order to ensure that the targets in these areas would be met. The final number of invitations by segment and by county (including the added invitations) is shown in Table 10 and Table 11 respectively.

Table 11 reflects the number of invites at the end of the study (after processing/excluding duplicates and households that moved out of state), including "extra" sample added later in the study (discussed below). It must be noted that final response rate estimates were adjusted (sample size target divided by final invites sent), however, sample size and sample rate targets were not modified.

Table 10: Characteristics, Assumed Response Rates, and Target Responses for the Main Study Segments[39]

Sample Segment Description	Target Sample Size[40]	Sampling Rate[41]	# HHs[42]	Final Response Rate[43]	Final Invites, # HHs[44]	Final Invites, % of HHs[45]
1-Oversample, Low response	1,367	1.10%	124,269	2.9%	46,771	37.6%
2-Oversample, Medium response	685	1.10%	62,256	3.9%	17,704	28.4%
3-Oversample, High response	361	1.10%	32,816	5.1%	7,101	21.6%
4-Regular sample, Low response	411	0.45%	91,414	3.4%	12,100	13.2%
5-Regular sample, Medium response	1,967	0.45%	437,164	5.8%	34,135	7.8%
6-Regular sample, High response	2,736	0.45%	607,930	7.6%	35,858	5.9%
Total	7,500	0.56%	1,355,849	4.9%	153,669	11.3%

---

[39] In the table, the short form HHs is used to denote households.

[40] Target sample size is the number of households who have completed the entire survey (including the recruit survey and travel diary survey).

[41] The sampling rate is the percent of total households who have completed the survey (i.e. number of responses / total number of households).

[42] This is the estimated total number of households within the corresponding block groups based on ACS 5-year estimates (2009-2013).

[43] In the CSTS, a complete household is defined as one where every person in the household answers every question (no missing responses). The predicted survey completion rates are based on demographics, incentives offered, and recent similar studies. Actual completion rates for the pilot were used to design the main study sampling plan.

[44] Number of households invited is determined by the desired sample size and predicted completion rate (sample size / completion rate).

[45] Percent of households invited is the percent of all households expected to receive study information.

Table 11: Main Sample Plan, Targets and Initial Invitation Counts by County[46]

County	# HHs	Sampling Rate	Target Sample Size	Initial Response Rate	Initial Invites, # HHs
Fairfield County	332,655	0.64%	2,136	4.6%	46,265
Hartford County	347,874	0.55%	1,919	4.8%	39,812
Litchfield County	75,755	0.44%	332	6.3%	5,257
Middlesex County	66,141	0.48%	315	6.0%	5,230
New Haven County	328,013	0.56%	1,837	4.7%	39,075
New London County	107,066	0.50%	531	5.3%	10,090
Tolland County	54,327	0.46%	248	6.3%	3,950
Windham County	44,018	0.48%	210	5.3%	3,990
Total	**1,355,849**	**0.56%**	**7,527**	**4.9%**	**153,669**

3.1.4 Travel Date Assignments

The households invited to the survey were assigned to one of 30 weekday "travel dates", spread over ten weeks in March, April and May of 2016. One week in the middle was skipped to allow time for review and adjustment after the first few weeks of data collection. Similar to the pilot study, all travel dates were on a Tuesday, Wednesday, or Thursday due to the focus on typical weekday travel of residents.

Travel dates were pre-assigned to households and invitations were evenly distributed over all the dates so that the recruitment and survey retrieval process could be easily managed. As the survey progressed, this made it more feasible to allow for adjustments due to possible impacts such as travel patterns being disrupted by unusual weather and other events or for variations in response over time. Distributing the sample evenly over all the travel dates also improves the final dataset as it contains a sufficient sample of each of the three weekdays and includes data for different seasons.

Similar to the pilot study, one of the reasons for pre-assigning travel dates was to minimize response bias and increase the ability of the dataset to describe a typical day for the region even if it is not a typical day for a given individual.

3.2 Incentive Plan

For the CSTS main data collection effort, the incentive plan was re-designed after the pilot based on feedback from CTDOT and other stakeholders. The guidelines below were followed in the main study for incentives:

❑   All households were offered $10, contingent upon completion of the entire study. Differential incentives (e.g. higher incentive values for hard-to-reach households) were not offered.

❑   All incentives offered were in the form of Visa or MasterCard gift cards (incentives from a neutral vendor) as opposed to gift cards from specific retail establishments.

---

[46] In the table, the short form HHs is used to denote households.

A raffle was not used, as the pilot study results demonstrated that the response rate for participants offered a guaranteed gift card was approximately double the response rate of participants offered the raffle. For more information on the incentive plan re-design process (i.e. effects of variable incentives on response rate, and vendor selection) please see Appendix D.

3.3 Questionnaire Design

Only minor changes were made between the pilot study and the main study. As a result, please reference Section 2.3￼ for information about the questionnaire design and also for an overview of the various components of the questionnaire. Major changes to the recruit and travel diary portions of the survey are referenced in the remaining section below.

- ❑ In the recruit survey, driving-age respondents are asked to report the following travel behavior details about their assigned travel day:
  - o whether they traveled on a toll road
  - o whether they traveled in an HOV lane
  - o whether they paid for parking at any location

    These details were collected prior to the set of questions asked about each individual trip. Subsequently, this information was used to reduce survey burden. More specifically, respondents were not asked related questions if they confirmed that they did not engage in these behaviors any time on their travel day.
- ❑ Question text, and responses were modified for a number of questions in an effort to enhance the presentation and avoid any misunderstandings.
- ❑ Responses to some of the travel diary items are auto populated based on day-level and previous trip travel behaviors.
  - o At the day level, if HOV is not used then the question about use of HOV for a particular trip is not posed.
  - o At the trip level, if the previous trip used a household vehicle and the specific household vehicle was identified, the same household vehicle was auto populated.

Also, some administrative changes were made which are described in Section 3.6￼. See Appendix E for a final copy of the questionnaire for the main study.

3.4 Survey Implementation
No changes were made in the approach to the survey implementation between the pilot study and the main study. The survey was still programmed as an online survey instrument using RSG's proprietary survey software – rSurvey™. While no new data validation checks were implemented, minor enhancements were made to the data validation implementation between the pilot and main study. As a result, please see Section 2.4￼ additional details about the survey instrument development and data validation. Screenshots of the various portions from the online survey instrument are included at the end of the report in Appendix F.

3.5 Branding and Survey Material

Only small changes were made to the branding and survey material between the pilot and main study. Consistent visual elements were used across all print and online materials, with the intent of connecting all invitations, reminders, and other notices concerning the project in order to maximize response rates. All materials (both online and print) included a toll-free telephone number, the website URL, the study email address, and the unique password assigned to each household. Please see Section 2.5 for more details regarding the approach to design and implementation. A summary of the changes are described below.

Minor edits were made to the survey material and the website in response to final choice of modes, incentives, and feedback from respondents and stakeholders. Additionally, the biggest difference between the pilot and main study was with regard to the updates to the survey website. Throughout the data collection period, the study team updated the website on a bi-weekly basis. Mainly these updates consisted of additions to the "News" page of the website that listed all references to the study by any news outlet and provided links to each article. A study video was added to the website for the main study which provides a brief description of its purpose and importance in regards to transportation planning and modeling throughout the state. Links to the study information web pages on the CTDOT and UCONN websites were also provided with information about the study sponsors and partners.

A copy of the print materials from the main study is included in Appendix G.

3.6 Survey Administration

Given the similarity in the survey administration, only a summary of the procedures and protocols implemented in the main study are highlighted. Similar to the gift card segments of the pilot study, survey administration comprised of the following key steps:

- ❑ All respondents were recruited by using four main types of mailings: pre-notice postcard, invitation packet, and two reminder postcards. All mailings were sent through USPS First-Class Mail.
- ❑ Selected households with a matching landline phone number that did not self-recruit within the first few days were also called and invited to participate.
- ❑ Once households completed the recruit survey, they received email or telephone reminders before and after their assigned travel date until they completed the diary (up to one week after the travel date).

3.6.1 Participation Methods

Participants could complete the study online, over the phone, or through a combination of both modes.

3.6.2 Communication

Depending on the preference, inbound and outbound phone calls were carried out as noted below:

- ❑ Telephone calls were placed to recruit households after the materials were mailed. Recruitment calls were made until a day before the travel date. Partway through the

main data collection efforts lower response rates for certain segments and counties prompted the study team to adjust the recruitment process. Recruitment calls were prioritized so that those segments and counties where response was below target were called first. The criteria for the prioritized call list were the following:

- o Over sample, high response (segment three) households
- o Fairfield County households
- o Over sample, medium response (segment two) households
- o Litchfield County households
- o All remaining households

Prioritized recruitment call lists were implemented for those households with travel dates of April 19th (travel week five) and later.

❑ A telephone reminder call was placed a day before their travel date to remind the household. Additional calls were placed (for up to seven days after the travel day) to the households to remind them to complete the survey online or over the telephone. Partway through the main data collection efforts, lower response rates for certain segments and counties prompted the study team to adjust the reminder process also. Beginning travel week five (travel dates beginning April 19th), reminder calls 3 and 4 were removed for households where the householder age was 65 or older and the household size was less than or equal to two. These criteria were implemented because older and smaller households responded at higher rates than other households without prompting, and ETC's efforts were prioritized for contacting hard-to-reach households.

❑ Additionally, ETC fielded inbound telephone calls from invitees and others interested in participating in the study or in learning more about the study.

❑ Overall, call prioritization was based on the following criteria: 1) inbound calls, 2) reminder calls, 3) recruitment calls.

In addition to phone calls, inbound and outbound communications were carried out over email as noted below:

❑ Up to five email reminder messages were sent up to a week after the travel day had passed. The first four followed the same schedule as the pilot study. However, a fifth reminder was added in the main study. The intention was to improve response rates by encouraging those who already finished the recruit survey to complete the travel diary survey. The fifth reminder email was sent six days after the travel date.

❑ Emails from invitees and others were monitored and responded to within one business day. Similar to the pilot study, inquiries sent via email typically involved questions about gift cards, requests for help with a specific part of the survey (such as how to use the map), requests to volunteer, or general comments or questions about the survey. Occasionally, households e-mailed comments about regional transportation issues. Emails from households with substantive comments regarding regional transportation or the study have been included in the Appendix L.

3.6.3 Incentives

As noted in Section 3.2   for the main study, a $10 gift card was offered to all households (except volunteers) as incentive for completing the survey. The $10 amount was printed on the invitation materials and included in the email text of outbound reminder emails. Households could choose between a virtual or physical Visa gift card—(sent via email or USPS mail). Alternatively, households could opt not to receive any gift card for participation.

3.6.4 Response Monitoring

Similar to the pilot study, an online dashboard was used to monitor progress and also to assess project responses for the remaining travel dates on a weekly basis. The dashboard provided real-time counts of recruited and completed households and persons. The page also provided overall counts as well as counts by county, sample segment, and several household, person, and trip-level variables. The page allowed the study team to monitor responses throughout data collection to ensure that a representative sample of the state was collected.

Bi-weekly response evaluations helped in guiding what adjustments were made to survey protocols. The biweekly reports also included a selection of maps and cross-classification tables to provide details beyond the online dashboard page. The bi-weekly report served as a formal method of response monitoring and enabled the study team to look at the response rates for certain target groups (e.g., young households, zero-vehicle households, and large households). In addition the bi-weekly evaluations, a somewhat more in-depth evaluation was conducted as part of "mid-point review". While minor changes were accommodated after each bi-weekly evaluation, higher-cost adjustments were only considered once during the "mid-point review".

Review of Database

The study team performed internal monitoring of the survey database throughout the data collection period. Key data checks included:

❑ Reviewing the dropout page frequency to confirm that respondents were not dropping out on a particular survey page due to an error

❑ Conducting spot checks to confirm that all variables were recording correctly (e.g. text strings were not truncated and values were cleared or overwritten if a survey response was changed)

❑ Comparing county demographics to ACS data to ensure that a representative sample of the region was collected

Mid-Point Review

At the mid-point review meeting, some segments were projected slightly over or under target. A larger gap was identified in the "Oversample, High Response" sample segment and in Fairfield County. The "Oversample, High Response" segment was largely made up of higher income households in transit-rich areas, such as the New York City commute-shed in southwestern Fairfield County, so there was significant overlap between these two targets (shown in the distribution of light blue block groups in Figure 11). Projected through the end of the project, it was estimated that these response rates would result in approximately 80 fewer responses than the 360 targeted in the "Oversample, High Response" segment (78% of target), and

approximately 150 fewer responses than the 2,136 responses targeted in Fairfield County (93% of target).

In order to increase completes from these segments, the study team first implemented small adjustments that could be made in the administration process (prioritizing telephone reminders and adding a fifth email reminder) with the goal of reducing deviation from the targets. Given that significant transportation investments are planned over time in the densely populated areas of Fairfield County, and that one of the sampling goals was to oversample transit riders, it was important to achieve or exceed the sampling targets for these segments. Therefore, the study team developed and implemented a plan to increase the number of invitations in Fairfield County and in the "Oversample, High Response" segment. The requirements for this additional sample were calculated in late April, and an additional 3,646 addresses were obtained and incorporated into the mailing list for the last study week (Week 10 travel dates of May 24-26). Half of the added sample addresses ordered were from block groups in the "Oversample, High Response" segment and the other half from other segments in Fairfield County. After processing, the resulting sample included:

- ❑ 1,820 additional addresses in the "Oversample, High Response" segment (the majority in Fairfield County, but a small number in other counties)
- ❑ 1,826 additional addresses in the rest of Fairfield County

Table 12 and Table 13 below summarize the additional invitations by sample segment and by county.

Table 12: County level Distribution of Additional Sample Invites for "Oversample, High Response" Segment

Sample Segment	County[47]	Added Sample
3-Oversample, High response	Fairfield County	1,621
	Hartford County	102
	Litchfield County	13
	Middlesex County	21
	New Haven County	63
Total Added Sample, Segment 3 (Oversample, High Response)		**1,820**

Table 13: Segment level Distribution of Additional Sample Invites for Fairfield County

County	Sample Segment	Added Sample
Fairfield County	1-Oversample, Low response	122
	2-Oversample, Medium response	171
	4-Regular sample, Low response	104
	5-Regular sample, Medium response	575
	6-Regular sample, High response	854
Total Added Sample, Fairfield County (excluding Segment 3)		**1,826**

---

[47] No "Oversample, High Response" block groups exist in New London, Tolland or Windham Counties

3.6.5 Public Outreach

The study team pursued an elaborate public outreach strategy within the available resources to encourage participation of invited households and supplement the pre-notice postcard, invitation packet, and reminder postcards that were sent directly by mail. Overall, public outreach aimed to:

- ❑ Ensure that residents recognized and opened invitations to participate in the survey by informing potential participants when the study was occurring and that it was a legitimate study sponsored by CTDOT in partnership with UConn.
- ❑ Encourage invited households to complete the survey by instilling confidence that the study benefitted their household as well as all residents of the state.
- ❑ Minimize voluntary (uninvited) participation in the survey by explaining the importance of collecting information from a random sample of households and by targeting most outreach to invited households rather than the general public.
- ❑ Prepare local officials and information-providers to answer questions by providing transparent information and inviting dialogue with local municipalities and regional entities.

The study team and CTDOT conducted a number of outreach efforts to local and statewide media before and during the main study effort, such as distributing press releases and conducting interviews.

3.7 Data Processing and Preparation

During and after data collection, responses must be reviewed and cleaned to assure the quality of the final data. This includes reviewing response frequencies for consistency; mapping location data; flagging, correcting or excluding records with survey errors; and deriving key variables for downstream data uses. This section discusses the data processing and preparation carried out in the CSTS and summarizes steps taken to prepare the final datasets.

3.7.1 Real-time Quality Control and Data Checks

An important first step in collecting high-quality survey data is to ensure that respondents provide accurate data as they take the survey. RSG's proprietary software (rSurvey™) that was used to build the online survey instrument includes capabilities for built-in data checks that verify response completeness and consistency, facilitate dataset preparation, and reduce the amount of data cleaning and up-coding required. A few examples of these real-time built-in data checks include the following:

- ❑ Validation logic required that respondents answered all questions on a page before continuing the survey (preventing skipped questions).
- ❑ Logic checking, such as hiding/skipping questions or answer choices that are not relevant (e.g. not asking employment questions for children); this also helps reduce respondent burden.
- ❑ Filters to automatically show or hide certain questions based on previous responses helped reduce respondent burden (e.g., if someone did not use HOV on travel day then the trip level HOV usage question was skipped).

- ❑ Metadata collection (passive collection of data such as survey duration and browser type) was used to help troubleshoot survey errors and assist households that call or email for help. These metadata also helped determine improvements to the survey design between the pilot and main data collection periods.
- ❑ The copy trips feature allowed a household member to select and copy information about joint trips already reported by another household member. This feature ensures that jointly made household trips were reported with the same locations, modes, and trip times.
- ❑ Reported trip sequences were required to be spatially and temporally logical (i.e., one trip's start and end location cannot be the same, and a trip's starting time could not be before the previous trip ended).

In addition to the real-time quality controls used during data collection, a number of additional quality checks and cleaning procedures were conducted after data collection was completed. This was done 1) to confirm that the real-time controls worked correctly and consistently, and 2) to evaluate any data quality issues that were not implemented using real-time checks in the online survey instrument.

3.7.2 Manual Quality Control and Data Cleaning

Additional checks were carried out to assess quality and validity. Based on the findings from these manual checks, data cleaning (described in Section 3.7.3) was carried out. The quality control and data checks can be categorized into three groups described below: Uniformity Checks, Consistency Checks, and Logical Checks.

- ❑ *Uniformity Checks*: The purpose of these checks was to ensure that the variables, and values contained in the different data files followed assumed notations for variable names, variable description, variable values, and value descriptions among others. One example is ensuring the values match the responses enumerated in the questionnaire document. Furthermore, the steps also ensure that all data files and the information contained in them are uniformly organized. For example, a value of -99 is used throughout to represent a "not applicable" and -999 is used to represent a "missing" value.
- ❑ *Consistency Checks*: The purpose of these checks was to ensure that the values reported for a given variable are consistent with information contained in another related variable in the same or different data files. For example, school type (schooltype) is only asked of adults (age >= 18) who are currently students (student=1 or 2). As can be seen, school type and student status provide different items of information but they are related, thus can be compared to evaluate the information contained in the variables. No consistency checks were performed for the vehicle information as there was no information in the dataset to compare against.
- ❑ *Logical Checks*: Unlike the consistency checks where the responses are assumed valid, the purpose of the logical checks was to evaluate responses for validity by using information contained in variables in the same data file or across other data files. For example, starting location type on assigned travel date (diary_st_loc) and

ending location type on assigned travel date (diary_end_loc) are reported in the person file. This information can be matched against the origin and destination purpose of the first and last trip reported in the trip file respectively. The focus of the logical checks is mainly on the travel responses.

Details regarding the specific uniformity, consistency and logical checks performed are enumerated in Appendix H.

3.7.3 Data Cleaning

Discarded Records

Household record exclusions implemented in this dataset included the following:

- ❑ Reviewer households (any records from UConn, RSG, DOT, or COG reviewers)
- ❑ 18 Households with reported home locations outside of the study area
- ❑ 34 Duplicate household responses (identified by phone and email)
- ❑ One household with multiple identified reporting errors

Trip record exclusions from this dataset included the following:

- ❑ 150 trips where the origin and destination place descriptions were identical and Google distance or duration were zero
- ❑ 63 trips where Google distance was zero and origin and destination place descriptions were both home
- ❑ 6 trips where the open ended response to the variable "d_purpose_other" clearly indicated that these were dummy trips[48]

Some trips with zero distance were left in the dataset, as participants could have intended to enter valid trips that started and ended in the same place (such as walking the dog or going for a bike ride). These trips may require additional inspection.

Deleting records requires making changes to other variables in the same file as well as to other files to ensure consistency across files. In response to household record exclusions, all person-, vehicle- and trip-records associated with excluded households were also removed. The total of 8,403 households is the sample size after removal of the problematic households. Similarly, deleting trips necessitates capturing the cascading impacts so that the information within a data file and across data files are consistent. In response to trip removals, all household-, person- and trip-variables were revised. Where appropriate, new variables with appropriately labeled variable names and descriptions were created to allow comparisons with original data.

---

[48] These trips may have resulted because of two potential causes. First, the individual incorrectly reported that they made trips at the beginning of the travel diary and this caused the survey instrument to start collecting trip responses. The individual instead of going back and changing their earlier response about the trips just continued to report dummy trips with a note in the "Other" purpose description in an effort to just complete the survey in the way they thought was most appropriate. Second, the way copy trip feature was implemented sometimes resulted in incorrect trips when the household had complex intra-household trip interactions. Individuals in this case again made a note in the "Other" purpose description to indicate that these are not valid trips and proceeded to complete the travel diary.

Variable Exclusions

Variable exclusions from the dataset included the following:

- ❑ Sample mailing address information (purchased from the sample provider)
- ❑ Email addresses and phone numbers (collected for survey administration only. Personally identifiable information data are not provided.)
- ❑ Passwords (associated with the sample mailing address information, replaced with ID numbers)

Purpose Cleaning

The list of locations reported by each individual on the assigned travel date comprises of a starting location (i.e. origin of the first trip) and a series of activity locations (i.e. destinations of the trips). For each location, individuals were required to provide a "name" that describes the location. The name information in turn was used by the online survey instrument to construct/present follow up questions in a way that was more meaningful and relatable to the respondents. The location name often contained information that could be used to confirm data that was collected about the location (e.g. purpose, or mode).

The name information was used to carry out two data cleaning tasks. First, the name information as used to confirm and clean the purpose of the starting location at the beginning of the day and the purpose of the primary activity at all trip destinations. The purpose cleaning was limited to only fixed activity locations (including home, work, and school) and took the form of series of programmable heuristics. The purpose corrections regarding fixed activity locations could be ascertained with a high degree of confidence using heuristics. Furthermore, different criterion were applied based on whether the location was the starting location or the destination for the trips on the travel day. Appendix I provides a description of the steps for purpose cleaning for different activity locations.

Second, the name information was used to check and confirm the purpose for all trips including the fixed activity locations. The cleaning of all purposes requires manual checks and corrections on a trip-by-trip basis. In an effort to overcome interpretation biases, a number of data quality procedures were applied. Two analysts first independently recoded potential trip purposes utilizing the name information. The recoded information was then reconciled by two senior members of the study team. The consistency of the recodes was also evaluated further by looking at all trips together using a visualization tool developed for this project. The tool allowed the analysts to confirm the initial recodes. Also, the tool allowed for consistency assessments in the recoded trip purposes across household members on joint trips. Heuristics applied for the initial recodes, and the visualization tool used to perform the joint trip checks are described in Appendix I.

Recoding the purpose of the trips necessitates performing changes to other variables due to the cascading impacts. For example, if the destination purpose of a trip is modified, it also means that the origin purpose of the next trip should also be modified. Cascading changes were applied in response to recoding the activity purpose at the location. Where applicable,

cascading changes made are recorded under new variables with appropriately labeled variable names and descriptions to allow comparisons with original data.

3.7.4 Data Preparation: Imputation

In surveys, valid values for some data items are missing because a respondent may not have provided a response – this is also referred to as item nonresponse or missing data. Data is often missing for sensitive items (e.g. income, race). Even in survey instruments that force respondents to provide responses, a valid skip is offered for such sensitive data items so as to avoid respondent drop-out and potentially risk not collecting responses for all other data items for which the respondent may be willing to provide an answer. Missing data are problematic and affect analyses with the survey data including the sample bias correction procedures discussed later.

In the CSTS, respondents were generally forced to provide a valid response for all but three data items, namely, exact age, income, and ownership status. For these three data items, respondents were offered "prefer not to answer" as a response (not a valid value that can be used for subsequent analyses). For the age question, respondents were required to answer a follow up question with coarser categories, therefore valid age information is available (not missing) for all respondents. For income, respondents were offered a follow up question where they were asked to respond to the same question on a coarser scale in the hope that those hesitant to provide a response on a detailed level may respond when the options were consolidated. The presence of these follow ups did improve valid responses. However, they were still offered the "prefer not to answer", as a result, there were still missing values for the income variable. For the ownership status no follow up option was available so there are observations with missing home ownership status information. The focus of the imputation analysis presented below is on the income variable which is critical for correcting sample bias (described in Section 3.7.5).

In order to impute the income, two different techniques were explored: stochastic regression technique and multiple imputation technique. In the stochastic regression approach, a multinomial logistic (MNL) model form was assumed to impute income. The MNL model formulation is appropriate for imputing income because responses were collected on a discrete scale in the CSTS (as opposed to continuous scale). For the multiple imputation technique, the Markov Chain Monte Carlo (MCMC) approach was applied. Unlike the MNL model, income was treated as a continuous variable. Comprehensive validation analysis was conducted to evaluate the two methods. Missing income observations were created artificially from valid records in the CSTS and the two methods were then used to make predictions for the artificially missing observations. Disaggregate and aggregate comparisons between the predicted values and actual values were performed to evaluate the two methods. It was found that the MNL model based stochastic regression imputation provided comparable results to the MCMC based multiple imputation. Additionally, the MNL model did not require making strong assumptions or manipulating the income variable to conform to the underlying model formulation. As a result, the MNL based approach was preferred owing to the comparable performance and limited number of assumptions involved.

A MNL based stochastic regression technique was used to impute detailed income information for the households with missing income information. To impute the income value a Monte Carlo procedure was applied. There are a number of questions in the literature on the best approach to prediction. In order to address this issue, an iterative stochastic regression approach is applied (Gold et al. 2000[49]). In this approach, a multinomial logit (MNL) model is estimated and applied iteratively to impute the income information until convergence is achieved. The process begins by estimating the MNL model first using the observations with valid income information. These first set of MNL model coefficient estimates are then used to stochastically impute the missing income values. In the second step, the observations with imputed income information are combined with observations with valid income information. The MNL model is re-estimated using this combined dataset (including observed and imputed income values). The income values are imputed again using the new coefficient estimates. This second step is repeated until convergence is achieved. In this project, stability in the parameter estimates (i.e. no change in parameter estimates across iterations) was used as the convergence criterion.

Results from Imputation Analysis

Table 14 provides a summary of the income imputation. The missing values in the original dataset are denoted by "Prefer not to answer" category. As can be seen, all the valid income values in the original variable were retained in the imputed income variable whereas the missing income values (i.e. "Prefer not to answer") are assigned a valid income value in the imputed variable. The highest counts of imputed income fall in the income category "$100,000 - $149,999". This is also the category with the highest number of observations in the original income variable. The least counts of imputed income are for the lowest income category i.e. "less than $10,000". Additional details regarding the model estimation and imputation are presented in Appendix J.

---

[49] Gold, M.S., and Bentler, P.M. (2000). Treatments of Missing Data: A Monte Carlo Comparison of RBHDI, Iterative Stochastic Regression Imputation, and Expectation-Maximization. Structural Equation Modeling, 7(3), pp. 319 – 355.

Table 14: Comparison between Values for the Original and imputed Income Variables

		Imputed Household Income												Total
		Less than $10,000	$10,000-$14,999	$15,000-$24,999	$25,000-$34,999	$35,000-$49,999	$50,000-$59,999	$60,000-$74,999	$75,000-$99,999	$100,000-$149,999	$150,000-$199,999	$200,000-$249,999	$250,000 or more	
Original Household Income	Less than $10,000	310	0	0	0	0	0	0	0	0	0	0	0	310
	$10,000-$14,999	0	291	0	0	0	0	0	0	0	0	0	0	291
	$15,000-$24,999	0	0	478	0	0	0	0	0	0	0	0	0	478
	$25,000-$34,999	0	0	0	487	0	0	0	0	0	0	0	0	487
	$35,000-$49,999	0	0	0	0	656	0	0	0	0	0	0	0	656
	$50,000-$59,999	0	0	0	0	0	560	0	0	0	0	0	0	560
	$60,000-$74,999	0	0	0	0	0	0	683	0	0	0	0	0	683
	$75,000-$99,999	0	0	0	0	0	0	0	986	0	0	0	0	986
	$100,000-$149,999	0	0	0	0	0	0	0	0	1294	0	0	0	1294
	$150,000-$199,999	0	0	0	0	0	0	0	0	0	556	0	0	556
	$200,000-$249,999	0	0	0	0	0	0	0	0	0	0	264	0	264
	$250,000 or more	0	0	0	0	0	0	0	0	0	0	0	395	395
	Prefer not to answer	61	64	83	86	130	112	135	161	288	127	68	128	1443
Total		371	355	561	573	786	672	818	1147	1582	683	332	523	8403

3.7.5 Data Preparation: Weighting Analysis

In most survey studies, the sample data collected is often not perfectly representative of the underlying population owing to two primary reasons.

- First, different population groups respond at different rates. Some differences in response rates are accounted for in the sampling plan by categorizing geographies into segments based on demographic profiles of residents in each geography. Accordingly, assumptions are made about response rates consistent with the demographic profiles, and households are drawn randomly from the geography in quantities that account for the assumed response rate. However, even with a carefully designed sampling plan, the data collected may not be perfectly representative because the response rates are based on experiences (from a pilot, other regions, or a different time point) and may not closely match the response patterns at the time. Also, while efforts are made to define segments such that they consist of homogenous groups of respondents, small variations still exist across household types in a segment.
- Second, in travel analyses, some infrequent types of households are of interest (e.g. transit users, zero vehicle households). If the infrequent household types in the population are sampled at the same rate as the frequent household types, this will lead to a survey dataset that doesn't have enough sample observations to make any reasonable inferences about the infrequent household types. As a result, the infrequent household types are oversampled[50] to provide a count of sample observations that can be used to make reasonable inferences. As can be seen, this process also results in a sample data that is not representative.

These reasons result in differences between the sample data distribution and the underlying population distribution for attributes of interest. This in turn leads to biases when inferring travel behaviors from the sample data collected, also referred to as sample bias. Not correcting for sample bias can lead to travel behavior inferences that will be erroneous.

Approach to Expansion

In an effort to correct for sample bias, two steps are typically applied: expansion and weighting. In the expansion step, a weight is assigned to each sample observation based on the sampling plan parameters and data collection results. The weight assigned to sample households during the expansion step is also referred to as the expansion weight. Typically segments (representing population subgroups) are defined in the sampling plan. An expansion weight for each sample observation k is given by the equation below:

$$w_k = \frac{T_h}{t_h}$$

Where, T_h is the total count of households in the population belonging to segment h to which the sample household k belongs, and t_h is the total count of sample households belonging to

---

[50] Higher number of responses for these groups are targeted in the sampling plan compared to their shares in the population

segment h. All sample observations belonging to a particular segment are assigned the same expansion weight.

Approach to Weighting

After the expansion step is complete, the weighting step is applied. In the weighting step, the expansion weights assigned to each sample household are readjusted such that weighted distributions of characteristics of interest match distributions of the characteristics in the population available from an external source (e.g. US Census). The most common approach to reweighting in the household travel surveys is the Iterative Proportional Fitting (IPF) procedure proposed by Deming and Stephan (1940)[51]. IPF is a heuristic procedure wherein the expansion weights are adjusted to match each of the marginal distribution of interest (from an external source) and the process is repeated ("iterated") until there is no further improvement in the match between the weighted distributions and known marginal distributions.

Typically in travel surveys, IPF procedure is applied separately at the household and person level to estimate household and person level weights respectively. Trip level weights are not generated separately by applying the IPF procedure because data about the trips made by the entire population in the region are often not readily available from an external source. Trip level weights are obtained by assigning the corresponding person weight to each trip. The independent approach to estimating household and person level weights has been adopted in most household travel surveys. However, this approach is problematic as this may introduce inconsistencies between travel characteristics implied by the households and that by persons. For example, for a given household, the person weights of members within the same household may not only be different but they may also not align with the household weight. Therefore, it is necessary to generate a single set of household level weights such that both the household and person level marginal distributions of interest are matched. This ensures that the travel characteristics derived at the household level are consistent with those derived at the person level.

Ye et al. (2009)[52] and Bar-Gera et al. (2009)[53] have developed algorithms (as part of synthetic population generation research efforts) that can be used to generate a single set of weights which match marginal distributions of both household and person level variables of interest. More recently Konduri et al. (2016)[54] proposed a new algorithm (again as part of synthetic population generation research) that can be used to control for marginal distributions of household and person variables at multiple spatial resolutions. In this study, the algorithm

---

[51] Deming, W. and Stephan, F. (1940) On least square adjustment of sampled frequency tables when the expected marginal totals are known. Annals of Mathematical Statistics, 6, pp. 427–444.

[52] Ye, X., Konduri, K.C., Sana, B., and Pendyala, R.M. (2009). A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations. Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, DC.

[53] Bar-Gera, H., Konduri, K.C., Sana, B., Ye, X., and Pendyala, R.M. (2009). Estimating Survey Weights with Multiple Constraints Using Entropy Optimization Methods. Proceedings of the 88th Annual Meeting of the Transportation Research Board, Washington, DC.

[54] Konduri, K.C., You, D., Garikapati, V.M., and Pendyala, R.M. (2016). Enhanced Synthetic Population Generation that Accommodates Control Variables at Multiple Geographic Resolutions. Transportation Research Record: Journal of the Transportation Research Board, 2563, pp. 40-50.

proposed by Konduri et al. (2016) was used to generate weights that match marginal distributions of household and person variables at two spatial resolutions.

In the CSTS, a single set of weights was generated so that both household and person level marginal distributions are simultaneously satisfied. In the next subsection, results from the application of the expansion step of the sample bias correction are presented. In the following subsection, the weighting step of the sample bias correction is discussed.

Results from Expansion Analysis

In the CSTS, an expansion factor was calculated for each of the six segments defined in the sample plan i.e., 1) Oversample Low Response, 2) Oversample Medium Response, 3) Oversample High Response, 4) Regular Sample Low Response, 5) Regular Sample Medium Response, and 6) Regular Sample High Response. The expansion factors were calculated for the state as a whole without further separating them based on regions or demographic characteristics. Where appropriate, data from the ACS 2009-2013 Census were used to obtain the counts of total population and counts of population subgroups. Expansion factors for the six segments are presented in Table 15.

The expansion weights were calculated using the equation presented previously. As expected, the expansion weight for the three oversample groups is smaller compared to the expansion weight for the three regular sample groups. This was anticipated because in the oversample groups, higher shares of these households were desired when compared to their shares in the population. As a result, the expansion weights are going to be smaller. For the three response groups within the oversample segments (low response, medium response, and high response), it is encouraging to see that the expansion weights are nearly equal, indicating that the assumed response rates capture the difference in the response patterns across these groups. There was a similar finding across the response groups for the regular sample segments, except for the regular sample, low response segment where the expansion weight is lower. This can be attributed to the large difference between the assumed response rate (3.4 percent) and the observed response rate for this group (4.7 percent). While there are differences between assumed response rate and observed response rate for other groups, the differences are smaller and more consistent across the groups, thus the similar trends in the weights.

Table 15: Expansion Factors for the Sampling Segments

Sample Segment	Total	Completed	Invited	Expansion Weight
1 – Oversample, Low Response Rate	124,269	1,547	46,774	80.33
2 – Oversample, Medium Response Rate	62,256	772	17,700	80.64
3 – Oversample, High Response Rate	32,816	423	7,093	77.58
4 – Regular Sample, Low Response Rate	91,414	565	12,099	161.79
5 – Regular Sample, Medium Response Rate	437,164	2,175	34,130	200.99
6 – Regular Sample, High Response Rate	607,930	2,921	35,853	208.12
Total	**1,355,849**	**8,403**	**153,649**	**161.35**

Results from Weighting Analysis

The expansion weight estimation step is followed by the weighting step wherein the expansion weights are modified to match known distributions of variables of interest obtained from an external source (e.g. US Census). As noted above, in the CSTS, a single set of household weights are generated such that the household and person level marginal distributions of interest are matched. Subsequently, the person weights and trip weights are simply the weights assigned to the household to which the person/trip belongs.

IPU methodology proposed by Konduri et al. (2016) was used in the weighting step. Unlike Konduri et al. (2016) that starts with unit weights for households, the expansion weights estimated in the previous step are used as initial weights. The list of variables used in the weighting step were identified based on four considerations, namely, account for variables that define household structures (important determinants of travel behaviors), review of four recent household travel surveys (including PSRC 2015[55], MassDOT 2012[56], CMAP 2010[57], DVRPC 2013[58]), availability of data from the 2009-2013 ACS Summary Files, and performance considerations of the weighting procedure.

In the weighting step, weights were readjusted at the region level. In other words, the weights of all sample households belonging to a region are adjusted such that region-level and state-level marginal distributions are matched[59]. Regions are defined by utilizing Metropolitan Planning Organization (MPO) and Council of Governments (COG) boundaries used by CTDOT. A total of 11 regions were identified as shown in Figure 12. Marginal distributions were obtained at the Census Tract level from the 2009-2013 ACS for variables of interest. The tract-level marginal distributions are then aggregated to obtain region-level and state-level control totals.

---

[55] PSRC (2015). 2015 Household Travel Survey Report. Puget Sound Regional Council. (Link: http://www.psrc.org/assets/14106/2015-Household-Survey-Tech-Report.pdf, Accessed: June 1, 2015)

[56] MassDOT (2012) 2012 Massachusetts Travel Survey: Report. Massachusetts Department of Transportation. (Link: http://www.massdot.state.ma.us/Portals/17/docs/TravelSurvey/MTSFinalReport.pdf, Accessed: June 1, 2015)

[57] CMAP (2010). 2010 Chicago Regional Household Travel Inventory: Final Report. Chicago Metropolitan Agency for Planning. (Link: https://datahub.cmap.illinois.gov/dataset/1f424666-3885-493d-8774-81ba7ac625f2/resource/22eb7767-b143-416f-bb0f-37e6755231fd/download/methodsfinal3111.pdf, Accessed: June 1, 2015)

[58] DVRPC (2013). 2012-2013 Household Travel Survey: Report. Delaware Valley Region Planning Commission. (Link: http://www.dvrpc.org/reports/14033.pdf, Accessed: June 1, 2015)

[59] The matching of state level marginal distributions is also a new feature of the weights generated in the CSTS compared to other household travel surveys.

Figure 12: Map Showing Definitions of Regions Used in the Weighting Step
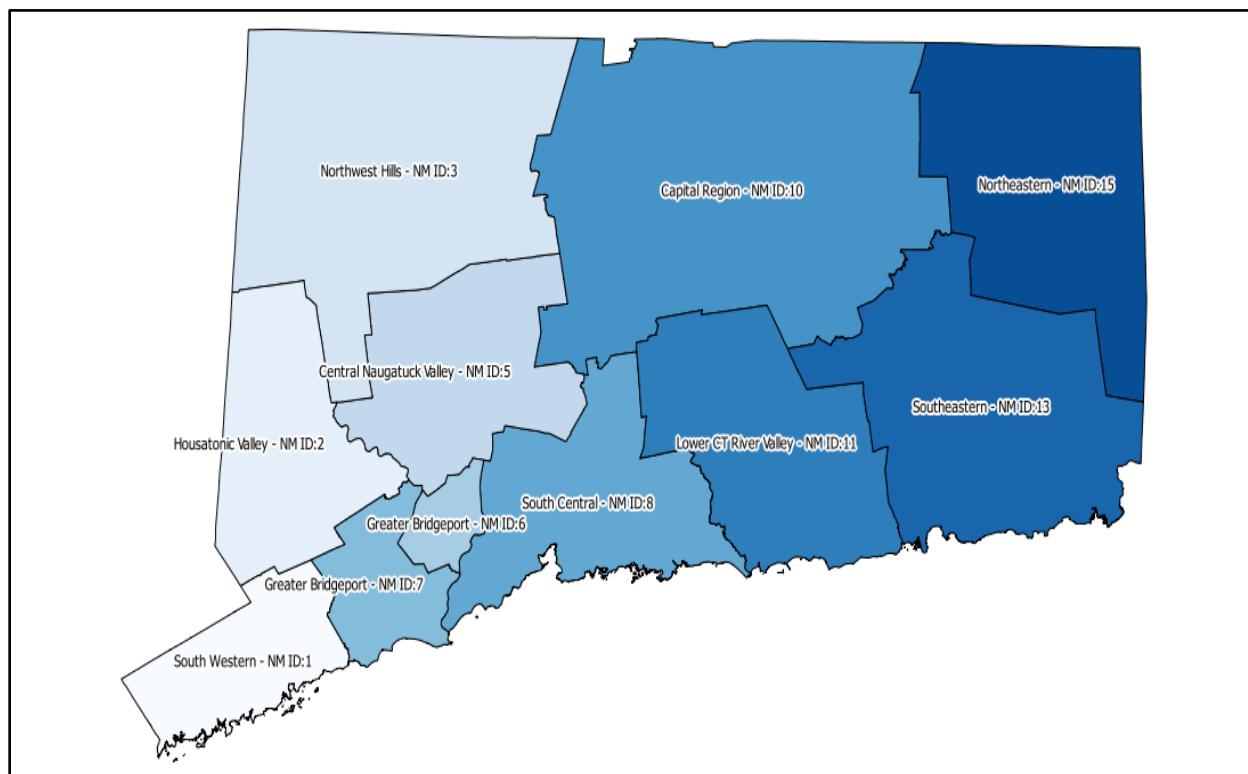
The marginal distributions for variables shown in the Table 16 were matched when adjusting the expansion weights. The table also shows the definitions of the values for each variable[60]. Only household income, household size, number of workers are controlled at the individual region level (i.e. marginal distributions for these variables will be perfectly matched at the region level) whereas the presence of children, household type, and number of vehicles was controlled only at the state level (i.e. marginal distributions for these variables will be matched but only closely). All person level variables are controlled at the state level[61].

---

[60] While data may have been available at a more disaggregate level in CSTS/2009-2013 ACS, it was consolidated to conform to the corresponding inputs or to address considerations of sparsity when applying the IPU procedure.

[61] A common problem with the weights generated using raking procedure (or any variant thereof) is that it can potentially generate extreme weights. The problem is even more magnified in the CSTS, as we not only attempt to match the given household marginal distributions but also person marginal distributions. Additionally, the choice of generating weights at the region level in the CSTS (as opposed to higher levels of spatial aggregation) potentially exacerbates the problem. In the CSTS, when all variables were controlled at the region level, it was observed that in an effort to match the marginal distributions of the household and person variables of interest, the Iterative Proportional Updating (IPU) algorithm assigned extreme weight values to many sample observations. Though the weights match the given marginal distributions (and thus rectify the sample bias), the extreme weights could lead to large variability when drawing population inferences. Two strategies were explored to rectify the extreme weight issue in the CSTS. First, trimming procedures were explored wherein adjusted weights during the raking procedure are limited to a valid range. Second, various configurations of the control variables were explored for the raking procedure (including altering the spatial resolution at which a control variable(s) are controlled, consolidating categories for control variables, and using subsets of control variables) in an effort to reduce the

Table 16: List of Household and Person Level Control Variables Used in the Weighting Step

Type of Variable	Variable Name	Category Value – Category Definition	
Household	Presence of children	1 – Children under age 18 present	
		0 – Children under age 18 not present	
	Household type	1 – Family households: Married-couple family	
		2 – Family households: Other family	
		3 – Nonfamily households	
	Household income	1 – Less than $25,000	
		2 –$25,000-$49,999	
		3 – $50,000-$74,999	
		4 – $75,000-$99,999	
		5 – $100,000- or more	
	Number of vehicles	0 – 0 vehicles	
		1 – 1 vehicle	
		2 – 2 vehicles	
		3 – 3 vehicles	
		4 – 4 or more vehicles	
	Number of workers	0 – 0 workers	
		1 – 1 worker	
		2 – 2 workers	
		3 – 3 or more workers	
	Household size	1 – 1 person	4 – 4 persons
		2 – 2 persons	5 – 5 persons
		3 – 3 persons	6 – 6 or more persons
Person	Age	1 – Under 5 years old	5 – 35-44 years
		2 – 5-17 years	6 – 45-54 years
		3 – 18-24 years	7 – 55-64 years
		4 – 25-34 years	8 – 65 years or older
	Gender	1 – Male	
		2 – Female	
	Employment	1 – Employed (including armed forces)	
		2 – Not employed (unemployed, not in labor force)	
		3 – Age less than 16	

Table 17 and Table 18 provide the match between the given marginal distributions and the weighted totals from the CSTS for the household and person variables that were controlled. As noted in Table 16 presence of children, household type, and number of vehicles were controlled at the state level whereas the household income, number of workers and household size were controlled at the region level. While attempts were made to include all variables at the region level, this was not pursued further as it affected the extreme weight issue negatively. It can be seen from Table 17 that the match with the marginal distributions for household control variables at the region level is perfect. On the other hand, the match for household

---

sparsity – an important consideration that affects the performance of the raking procedure. The trimming procedure was found to be ineffective. While it limited the maximum weight values to a valid range, it was causing other problems (including causing extreme weights at the lower end, affecting the performance in matching the marginal distributions, and skewing the weights towards the higher end of the valid range). On the other hand, trying various combinations of control variables was found to be more effective in addressing the extreme weight issue. Therefore, the latter approach was adopted in CSTS.

control variables that were specified at the state level is close with small deviations ranging from -3.0% to 3.3%.

The match for the marginal distributions for the person control variables is also close with deviations ranging from -8.6% to -3.6% (see Table 18). It must be noted that the deviations at the person level are owing to two reasons. First, the person totals include those residing in group quarters, however, we do not consider group quarter residents in the weighting. Second, the marginal distributions for person variables may not be consistent with marginal distributions for household variables. The IPU procedure tries to match marginal distributions of household variables (perfectly at the region level and closely at the state level) at the expense of person marginal distribution mismatch. Both these reasons contribute to the differences for person level marginal distributions. Nonetheless, the match with the marginal distributions is reasonable.

In addition to assessing the performance of the weights by comparing the match with marginal distributions for controlled variables (and controlled categories), it is desirable for the weights to also be evaluated against marginal distributions for uncontrolled variables (and uncontrolled categories). This ensures that all critical variables are included in the weighting and that no biases exist for key uncontrolled variables of interest. If there is observable differences against uncontrolled variables (and uncontrolled categories), then this calls for including these variables (and/or categories) in the weighting. Table 19 provides a comparison of the weights for one uncontrolled variable (namely commute mode to work) and for uncontrolled categories (namely unconsolidated age, and unconsolidated income). As expected there are deviations against both the uncontrolled variables and uncontrolled categories. The magnitude of the deviations are larger than that for the controlled variables. Attempts were made to include these variables (and categories) in the weighting analysis, but the extreme value issue was negatively impacted. As a result, the commute mode to work was completely excluded from the weighting and the age and income were specified in a consolidated form in the weighting analysis. The final choice of controlled variables attempted to strike a balance between minimizing the extreme weight issue while also matching the marginal distributions for critical variables of interest.

Table 17: Comparison of Weighted Distributions and Marginal Distributions from 2009-2013 ACS for Controlled Household Variables

Variable Name	Variable Category[62]	Given Totals from ACS 2009-2013	Weighted Total from CSTS	Difference	Percent Difference
Presence of children	0	913,487	926,552	13,065	1.4%
	1	442,362	429,297	-13,065	-3.0%
	Total	1,355,849	1,355,849	0	0.0%
Household Type	1	669,883	656,527	-13,356	-2.0%
	2	233,694	231,994	-1,700	-0.7%
	3	452,272	467,328	15,056	3.3%
	Total	1,355,849	1,355,849	0	0.0%
Household Income	1	243,901	240,168	-3,733	-1.5%
	2	254,982	252,550	-2,432	-1.0%
	3	223,674	226,459	2,785	1.2%
	4	178,175	179,841	1,666	0.9%
	5	455,117	456,831	1,714	0.4%
	Total	1,355,849	1,355,849	0	0.0%
Number of Vehicles	0	121,597	121,597	0	0.0%
	1	442,638	442,638	0	0.0%
	2	518,022	518,022	0	0.0%
	3	192,985	192,985	0	0.0%
	4	80,607	80,607	0	0.0%
	Total	1,355,849	1,355,849	0	0.0%
Number of Workers	0	339,644	339,644	0	0.0%
	1	515,182	515,182	0	0.0%
	2	401,798	401,798	0	0.0%
	3	99,225	99,225	0	0.0%
	Total	1,355,849	1,355,849	0	0.0%
Household Size	1	374,214	374,214	0	0.0%
	2	446,464	446,464	0	0.0%
	3	221,469	221,469	0	0.0%
	4	196,253	196,253	0	0.0%
	5	81,221	81,221	0	0.0%
	6	36,228	36,228	0	0.0%
	Total	1,355,849	1,355,849	0	0.0%

---

[62] Please see Table 14 for category definitions.

Table 18: Comparison of Weighted Distributions and Marginal Distributions from 2009-2013 ACS for Controlled Person Variables

Variable Name	Variable Category	Given Totals from ACS 2009-2013	Weighted Total from CSTS	Difference	Percent Difference
Age	1	197,395	180,590	-16,805	-8.5%
	2	606,650	556,234	-50,416	-8.3%
	3	335,755	306,887	-28,868	-8.6%
	4	428,258	399,942	-28,316	-6.6%
	5	469,746	440,649	-29,097	-6.2%
	6	568,510	534,997	-33,513	-5.9%
	7	456,963	433,136	-23,827	-5.2%
	8	520,284	501,771	-18,513	-3.6%
	Total	3,583,561	3,354,206	-229,355	-6.4%
Gender	1	1,745,364	1,597,341	-148,023	-8.5%
	2	1,838,197	1,756,865	-81,332	-4.4%
	Total	3,583,561	3,354,206	-229,355	-6.4%
Employment	1	1,768,014	1,643,217	-124,798	-7.1%
	2	1,111,136	1,064,989	-46,147	-4.2%
	3	704,411	646,000	-58,411	-8.3%
	Total	3,583,561	3,354,206	-229,355	-6.4%

Table 20 provides a variety of measures that were used to evaluate the degree of the extreme weights. The "Weights" column contains values for the various measures based on the final weighting analysis. It was also posited that the extreme weights may be resulting from specification of additional household and person level control variables. Results from a reference run wherein only three key household level control variables were used in the weighting analysis are also included in the table (see column "Reference Weights"). It can be seen that the degree of extreme weights is higher for the final weighting analysis compared to the reference weighting analysis. However, the differences are not that pronounced considering the fact that the revised weighting analysis addresses biases for additional household and person level variables of interest. It is interesting to note that the contribution of the top 5 percentile of weights towards the population is nearly 22.8% in the final analysis. While the percentage is smaller in the reference analysis, the value is still relatively large at 14.4%. This suggests that the sample is underrepresenting some household types that are present in the population but missing from the survey. As a result, even when a small number of variables are used to avoid extreme value issues, there are still considerable amount of large weights.

Table 19: Comparison of Weighted Distributions and Marginal Distributions for Household and Person Uncontrolled Variables and Uncontrolled Categories

Variable Name	Variable Category	Given Totals from ACS 2009-2013	Weighted Total from CSTS	Difference	Percent Difference
Age (Finer Categories)	Under 5 years old	197,395	180,590	-16,805	-8.5%
	5-17 years	606,650	556,234	-50,416	-8.3%
	18-24 years	335,755	306,887	-28,868	-8.6%
	25-34 years	428,258	399,942	-28,316	-6.6%
	35-44 years	469,746	440,649	-29,097	-6.2%
	45-54 years	568,510	534,997	-33,513	-5.9%
	55-64 years	456,963	433,136	-23,827	-5.2%
	65-74 years	269,422	327,537	58,115	21.6%
	75-84 years	164,260	121,463	-42,797	-26.1%
	85 years or older	86,602	52,770	-33,832	-39.1%
	Total	3,583,561	3,354,206	-229,355	-6.4%
Commute Mode to Work	Drove alone/Carpool or Vanpool	1,503,547	1,353,820	-149,727	-10.0%
	Bus or trolley bus/ Subway or elevated, Railroad	79,958	108,721	28,763	36.0%
	Taxicab/ Motorcycle/ Other means	15,778	13,417	-2,361	-15.0%
	Bicycle/ Walked	56,948	50,535	-6,413	-11.3%
	Worked at home	71,055	116,724	45,669	64.3%
	Unemployed	1,111,136	1,064,989	-46,147	-4.2%
	Age less than 16	704,411	646,000	-58,411	-8.3%
	Total	3,542,833	3,354,206	-188,627	-5.3%
Household Income (Finer Categories)	1 – Less than $10,000	76,257	72,550	-3,707	-4.9%
	2 – $10,000-$14,999	54,030	58,623	4,593	8.5%
	3 – $15,000-$24,999	113,614	108,995	-4,619	-4.1%
	4 – $25,000-$34,999	106,176	103,197	-2,979	-2.8%
	5 – $35,000-$49,999	148,806	149,353	547	0.4%
	6 – $50,000-$59,999	97,266	105,089	7,823	8.0%
	7 – $60,000-$74,999	126,408	121,370	-5,038	-4.0%
	8 – $75,000-$99,999	178,175	179,841	1,666	0.9%
	9 – $100,000-$149,999	227,122	236,243	9,121	4.0%
	10 – $150,000-$199,999	105,201	110,512	5,311	5.0%
	11 – $200,000 or more	122,794	110,076	-12,718	-10.4%
	Total	1,355,849	1,355,849	0	0.0%

Table 20: Measures for evaluating degree of extreme values in the revised weighting analysis

Measure	Weights	Reference Weights
Minimum weight value	0.1	23.0
Maximum weight value	2,216.4	1,188.6
Number of weights with value less than 1	2	0
Standard deviation of the weights	174.5	101.4
Weight value corresponding to 5th percentile	25.1	48.5
Weight value corresponding to 25th percentile	66.4	93.7
Median weight value	114.6	140.2
Weight value corresponding to 75th percentile	191.0	202.6
Weight value corresponding to 95th percentile	447.4	338.9
Sum of weights less than the 5th percentile	6637.7	17,405.6
Sum of weights less than the 25th percentile	85,880.8	136,500.8
Sum of weights less than the median value	279,546.5	386,111.2
Sum of weights greater than the 75th percentile	768,212.1	618,473.2
Sum of weights greater than the 95th percentile	308,770.5	195,248.3
Proportion of weights less than the 5th percentile	0.5%	1.3%
Proportion of weights less than the 25th percentile	6.3%	10.1%
Proportion of weights less than the median value	20.6%	28.5%
Proportion of weights greater than the 75th percentile	56.7%	45.6%
Proportion of weights greater than the 95th percentile	22.8%	14.4%

3.7.6 Data Preparation: Trip Linking

In the CSTS, individual legs of the travel occurrence were sometimes broken and reported as separate trips for multimodal journeys. For example, an individual going back home from office may have reported walking to the parking lot to get in his/her car and driving home as two separate trips when in fact they should have been reported as a single trip taken using the auto mode. The process of identifying these trip legs and consolidating them into a single trip is referred to as trip linking. Without linking trips, the trip rates will be inflated and subsequent travel analyses will be erroneous. For the above example, the trip count would be two when in fact it should be one.

In the final dataset, two types of trip linking efforts were carried out. First, trip legs were linked when purpose of intermediate trip legs (i.e. all trip legs except the last one) was reported as "Change Mode" (i.e. d_purpose_2 = 18). While it was desired that trip legs should be consolidated and reported as a single travel episode, it was also acknowledged that some individuals may want to report trip legs separately. In an effort to accommodate such instances, a destination purpose of "change mode" was included in the purpose options for the CSTS. This is the correct way of reporting trip legs, which allows trip legs to be identified and consolidated

into a single travel journey in a more straightforward manner. The following steps were used to identify a linked trip:

1) The trip records for each individual were sorted chronologically and considered for trip linking.
2) The first leg of a potential linked trip was identified by scanning for a record with reported destination purpose (d_purpose_recode_2) value of 18.
3) All intermediate trip legs of the multimodal travel episode were then identified by scanning for subsequent records with values of destination purpose value of 18. Subsequently, all these trip records were identified to be part of the same linked trip group.
4) The last trip leg of the linked trip group was then identified as the record following the last change mode trip record.
5) Efforts were also made to identify and group trips together where the destination purpose was reported erroneously. For example, a person traveling between home and work by "walking to the car – driving the car to the parking lot – walking to work from the parking lot" may have reported the destination purpose as "change mode – work – work" instead of reporting it as "change mode – change mode – work". In such instances, additional heuristics were applied to include all three trips and not terminate the trip at the second record. The heuristics proceed by identifying all consecutive trip records that have the same purpose as the trip identified in step (4) above. These trip records are then included as part of the linked trip group. The scanning for additional trip records was only limited to a subset of activity types where there is a small likelihood of consecutive trips for the same activity purpose occurring. For example, there is a small likelihood of back-to-back trips with the purpose of "Go Home". On the other hand, there is a higher likelihood that there are back-to-back shopping trips. Consequently the following set of rules were applied to scan for additional trip records that are part of the same linked trip group:
    o Identify additional trips that occur consecutively after and have the same purpose as the trip record identified in step (4).
    o Further, consider the additional trip records of the same purpose only if the purpose of trip identified in step (4) is home (i.e., d_purpose_recode_2 value 1), work (i.e., d_purpose_recode_2 value 2), exercise (i.e., d_purpose_recode_2 value 11), and vacation (i.e., d_purpose_recode_2 value 17).
    o Lastly, if the purpose of the trip record in step (4) is school (i.e., d_purpose_recode_2 value 3) or school related (i.e., d_purpose_recode_2 value 4) only consider additional trip records of the same purpose that have a dwell time that is within 10 minutes.
6) Once all trip legs of a linked trip were identified, the remaining trip records are scanned for the next potential linked trip by employing steps (2) through (5) on the remaining trips.

Second, trip legs were linked when the purpose of intermediate trip legs was not reported as "Change Mode". The identification of trip legs here requires heuristics because no direct

information is available. Potential trip legs can be inferred by analyzing purpose, mode, and dwell time information. Following steps were used to identify a linked trip:

1) The trip records for each individual are sorted chronologically and considered for trip linking.
2) Only those trip records that were not already linked in the previous trip linking task are included.
3) A set of consecutive trip records are compared to see if they share the same purpose and satisfy criteria for modes and dwell time thresholds across trips. In particular, for each pair of consequent trips:
   o The mode value changes and the dwell time at the destination of the preceding trip is not more than 10 minutes or
   o The mode of the subsequent trip is transit (i.e., mode_category value 6) and the dwell time at the destination of the preceding trip is not more than 40 minutes or
   o The mode of the preceding trip is transit (i.e., mode_category value 6) and the dwell time at the preceding activity location is not more than 25 minutes.
   o The purposes that were considered were limited to a small set based on manual confirmation of the linked trips.
      ▪ For trips where at least one transit trip was involved, the trip purpose was limited to home (i.e., d_purpose_recode_3 value 1), work (i.e., d_purpose_recode_3 value 2), work related (i.e., d_purpose_recode_3 value 3), school (i.e., d_purpose_recode_3 value 4), school related (i.e., d_purpose_recode_3 value 5), Go to restaurant/bar/take-out (i.e., d_purpose_recode_3 value 9), exercise (i.e., d_purpose_recode_3 value 11), medical visit (i.e., d_purpose_recode_3 value 13), leisure/entertainment (i.e., d_purpose_recode_3 value 15) and vacation (i.e., d_purpose_recode_3 value 17).
      ▪ For trips where a transit trip was not involved, the trip purpose was limited to home (i.e., d_purpose_recode_3 value 1), work (i.e., d_purpose_recode_3 value 2), work related (i.e., d_purpose_recode_3 value 3), school (i.e., d_purpose_recode_3 value 4), school related (i.e., d_purpose_recode_3 value 5), vacation (i.e., d_purpose_recode_3 value 17) and other (i.e., d_purpose_recode_3 value 97).
4) If they satisfy all conditions in step (3) then the subsequent trip is included as part of an active linked trip group. If a linked trip group is not currently active then a new linked trip group is created and the trip pair is added. When a pair of trips does not satisfy the conditions listed in step (3), the active linked trip group is terminated.
5) Steps (3) and (4) are repeated iteratively until all trip records for the individual have been scanned and potential linked trips identified.

Following the identification of linked trips, all linked trips were manually checked for validity. Subsequently, a new linked trip file was created wherein the trip legs were consolidated so that each record in the file represents a single travel episode between two activity locations,

including for multimodal trips. Section 3.7.8 and Appendix K provide information about the variables that are included in the linked trip file and how they were created.

3.7.7 Data Preparation: Tour Generation

As part of the data preparation, the study team also carried out the task of extracting tours from the trip file. Tours were generated to support the development of a disaggregate travel demand model (e.g. tour-based model) in the future. The tour extraction and tour variable generation were carried out in close consultation with CS and CTDOT. A tour file reporting the different tour episodes pursued by respondent households during the assigned travel date was generated.

During the tour extraction stage, two types of tours were extracted. First, home based tours were identified as a series of trips that begin and end at home. The specific steps applied to identify home based tours are as follows:

1) Identify the first trip for which the origin purpose is home (i.e., o_purpose_recode_3_linked value 1)
2) Identify the next trip for which the destination is home (i.e., d_purpose_recode_3_linked value 1)
3) Mark all the trips between origin as home identified in step (1) and destination as home identified in step (2), including the end trips, as belonging to the same home based tour with a unique tour id
4) Beginning with the trip following the last trip of the tour identified in step (3), repeat steps (1) through (3) until all trips reported by an individual on the assigned travel date have been scanned

Next, work based sub-tours were extracted. Work based sub-tours are identified only within a valid home based tour. A work based sub-tour identification is not carried out for the trips that are not included in any home based tour. The specific steps applied to identify work based sub-tours are described below:

1) Identify the first trip for which the origin purpose is primary workplace (i.e., o_purpose_recode_3_linked value 2)
2) Identify the next trip for which the destination is primary workplace (i.e., d_purpose_recode_3_linked value 2)
3) Mark all the trips between origin as primary work identified in step (1) and destination as primary work as identified in step (2), including the end trips, as belonging to the same work based sub-tour
4) Beginning with the trip following the last trip of the tour identified in step (3), repeat steps (1) through (3) until all trips within the home-based tour have been scanned

Following the identification of tours, they were randomly spot checked for validity. Subsequently, a new tour file was created wherein each record in the file represents a single tour or sub-tour episode. Additionally, tour level attributes were derived or calculated. Section 3.7.8 and Appendix K provide information about the variables that were included in the final tour file.

3.7.8 Data Preparation: Derived and Calculated Variables

The household travel survey dataset includes a combination of variables that were collected (asked directly of respondents), implicitly assigned (administrative variables such as id numbers), derived (calculated from directly collected variables) or recoded (modified from original values based on other information). Variables collected directly in the survey can be reviewed in the questionnaire. Implicitly assigned, derived, and recoded variables typically seek to calculate, combine, or reconfigure collected variables in order to provide a richer dataset with additional variables for analytical convenience. A summary of the implicitly assigned, derived, and recoded variables is provided below. A detailed listing of the variables in each file along with some metadata is provided in Appendix K.

Household Variables
- ❑ Household income category, aggregated from 2 questions (follow-up prompt if "prefer not to answer")
- ❑ Number of household adults, children, licensed drivers, students, workers, and travel day trips
- ❑ Household flag if any members of the household made transit trips on the travel day (public bus, train, paratransit, commuter rail, subway, intercity bus)
- ❑ Householder age (based on the person who responded to the recruit survey)
- ❑ Home location block group, town, county
- ❑ Implicit or passively collected data about the recruit survey (e.g. response mode, foreign language indicators, incentives offered, survey duration, etc.)

Person Variables
- ❑ Person age category, aggregated from 2 questions (follow-up prompt if "prefer not to answer")
- ❑ Number of person trips on travel day
- ❑ Travel day start and end locations (derived for members who began or ended their day at a copied trip location)
- ❑ Implicit or passively collected data about the diary survey (e.g. response mode, foreign language indicators, survey duration, etc.)

Unlinked Trip Variables
- ❑ Trip departure and arrival times, recoded into one-hour bins (for convenience)
- ❑ Reported trip duration
- ❑ Implied trip speed (reported duration / estimated distance)
- ❑ Household and total travel party size (from detailed list of household members and count of non-household members on trip)
- ❑ Number of transit transfers
- ❑ Trip origin purpose and overall trip purpose
- ❑ Passengers on trips copied from drivers
- ❑ Aggregate mode variable (e.g. group detailed modes into broader categories, including carpool and drive alone based on travel party)

Linked Trip Variables
- ❑   All variables in this file are derived from the unlinked trip file

Tour Variables
- ❑   All variables in this file are derived from the linked trip file

Vehicle Variables

- ❑   All variables in this file are collected

3.8 Results

After completing the data processing and preparation described in Section 3.7   valid and complete responses were collected from a total of 8,403 invited households. An invited household qualified as a valid complete when they filled out the recruit survey and every travel diary for all household members. Volunteer households and responses from student team members or other stakeholders are not included in the final count as they cannot be accounted for in weighting process. Also, records for incomplete households, those that completed recruit survey but not travel diary survey, are not included in these results.

Table 21 shows the final count of survey responses at each of the four different levels of data collected, including household, person, vehicle, and trip-level data. All of the data can be linked to individual households (the highest level), and trip data can be linked to individual people, providing the context for reported travel behaviors.

Table 21: Summary of Survey Responses

Data Type	Count
Households	8,403
Persons	17,481
Vehicles	14,540
Unlinked Trips	66,175
Linked Trips	65,103

3.8.1 Response pattern analysis

It can be seen from Table 22 that the sample target of 7,500 households was exceeded. This may be partly attributable to the higher observed response rates than the assumed response rates. It may also in part be driven by the addition of invitations at the mid-point review stage as discussed in Section 0. Targets for complete households were exceeded in all segments and counties as shown in Table 22 and Table 23. As can be seen from Table 22, observed response rates were higher than estimated response rates in all segments except "Oversample, High response" segment. The mid-point adjustment to add targeted invitations in the last week of data collection resulted in 228 completions overall. 117 of the 228 were in "Oversample, High response" bringing the total completes in this segment to 228. Of the 228 extra completes received, 219 were in Fairfield county. Thus the mid-point adjustments appear to have achieved the sampling objectives of adding more completes for two of the deficient areas identified at the time, namely "Oversample, High response" segment and Fairfield county.

Table 22: Distribution of Complete Responses by Segment

Sample Segment Description	# HHs[63]	Final Invites, # HHs[64]	Target Sample Size	Target Response Rate	Observed Completes, # HHs	Observed Complete Rate[65]	Observed Sample Rate
1-Oversample, Low response	124,269	46,771	1,367	2.9%	1,547	3.3%	1.24%
2-Oversample, Medium response	62,256	17,704	685	3.9%	772	4.4%	1.24%
3-Oversample, High response	32,816	7,101	361	5.1%	423	6.0%	1.29%
4-Regular sample, Low response	91,414	12,100	411	3.4%	565	4.7%	0.62%
5-Regular sample, Medium response	437,164	34,135	1,967	5.8%	2,175	6.4%	0.50%
6-Regular sample, High response	607,930	35,858	2,736	7.6%	2,921	8.1%	0.48%
Total	**1,355,849**	**153,649**	**7,527**	**4.9%**	**8,403**	**5.5%**	**0.62%**

Table 23: Distribution of Complete Responses by County

County	# HHs[66]	Final Invites, # HHs[67]	Target Sample Size	Target Response Rate	Observed Completes, # HHs	Observed Complete Rate[68]	Observed Sample Rate
Fairfield County	332,655	46,245	2,136	4.6%	2,289	4.9%	0.69%
Hartford County	347,874	39,812	1,919	4.8%	2,082	5.2%	0.60%
Litchfield County	75,755	5,256	332	6.3%	352	6.7%	0.46%
Middlesex County	66,141	5,229	315	6.0%	431	8.2%	0.65%
New Haven County	328,013	39,080	1,837	4.7%	2,086	5.3%	0.64%
New London County	107,066	10,089	531	5.3%	627	6.2%	0.59%
Tolland County	54,327	3,949	248	6.3%	325	8.2%	0.60%
Windham County	44,018	3,989	210	5.3%	211	5.3%	0.48%
Total	**1,355,849**	**153,649**	**7,527**	**4.9%**	**8,403**	**5.5%**	**0.62%**

The final dataset only includes responses from households that lived in Connecticut at the time of the study. As shown in Figure 13, responses are distributed across the entire state, with higher concentrations (in darker red) generally collected in the more densely populated areas of the state. This provides a relatively representative sample coverage across the state when evaluating the sample as a proportion of the population in an area (as noted in Table 23). Some rural areas only have a few responses per town and one town in the northwest corner of the

---

[63] This is the estimated total number of households within the corresponding block groups based on ACS 5-year estimates (2009-2013).

[64] This represents the final count of invites after the mid-point review adjustment

[65] This is defined as the ratio of number of complete responses observed to the total number of invites.

[66] This is the estimated total number of households within the corresponding block groups based on ACS 5-year estimates (2009-2013).

[67] This represents the final count of invites after the mid-point review adjustment

[68] This is defined as the ratio of number of complete responses observed to the total number of invites.

state, Canaan, had no responses. This is expected given the population densities in these towns. It must be noted that, while there are no observations for Canaan town, the weighting analysis does account for households and persons living in the town.

Figure 13: Distribution of Respondents' Home Location at the Town Level
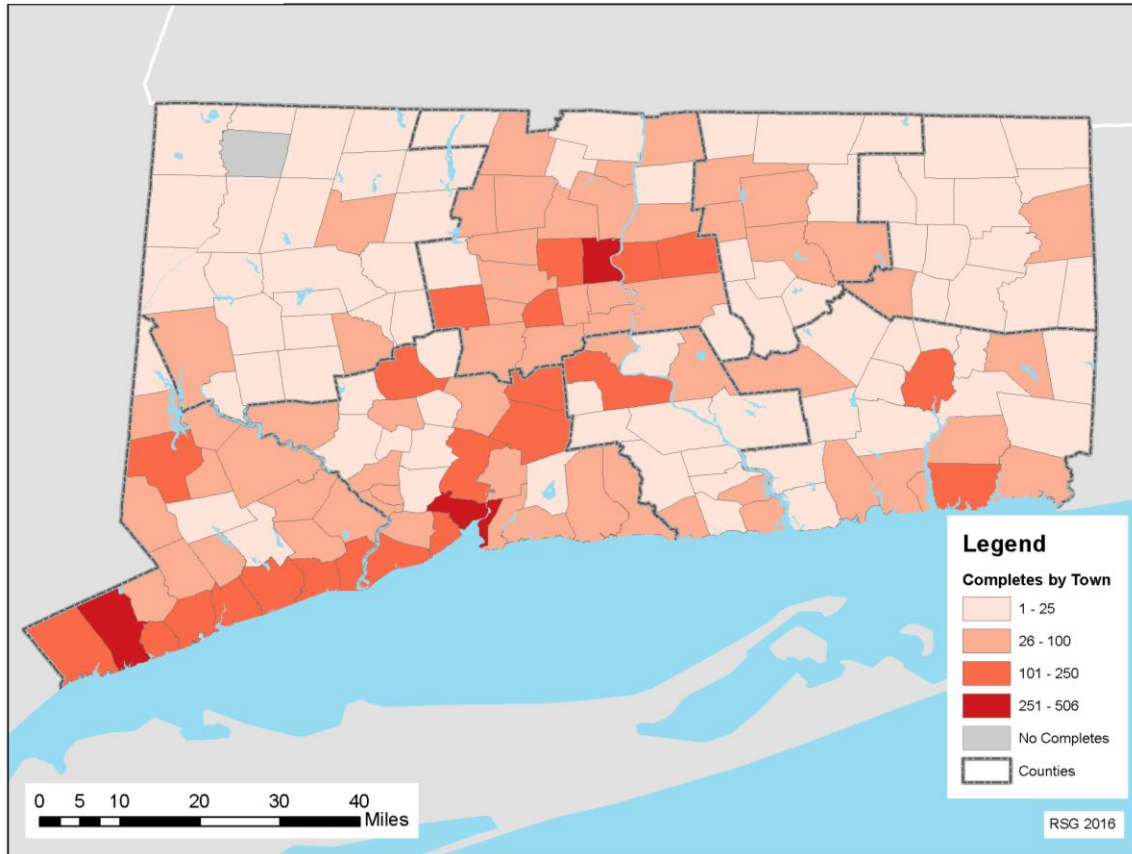
While the survey only includes households that live in the state of Connecticut, it did not restrict trip destinations to be limited to the state. Information about trip destination of CT residents outside the state was also collected. While the majority of survey respondents reported work locations and school locations in the state (Figure 14 and Figure 15), some respondents work or attend school out of state. Given the density and complexity of the greater New England region, it is important that planners in Connecticut and the surrounding states understand travel between and through the states in the region as well as travel within the state.

As shown in Figure 14, the majority of workplaces are concentrated in the southwestern and central part of the state (corresponding with the residential densities in these areas). However, 554 survey respondents (8% of employed persons) reported work locations out of state. The majority of these out-of-state job locations are in the neighboring states of New York, Massachusetts and Rhode Island, but a few people reported job locations as far as Maryland and Ontario, Canada. The vast majority of the out-of-state workplaces, however, are in nearby New York County (Manhattan) and Westchester County.

Similar to workplaces, school locations are primarily concentrated in the state (see Figure 15). Out-of-state school locations make up a smaller portion of responses (2.7% of surveyed adults and children who travel to school). Again, the majority of out-of-state school locations are in neighboring states, but reported school locations are somewhat more dispersed including locations as far away as California and Texas. The majority of out-of-state students are likely college students, who were either home on break or were reported as household members by a parent even though they may have been away at school.

Figure 14: Distribution of Respondents' Work Location at the Town Level
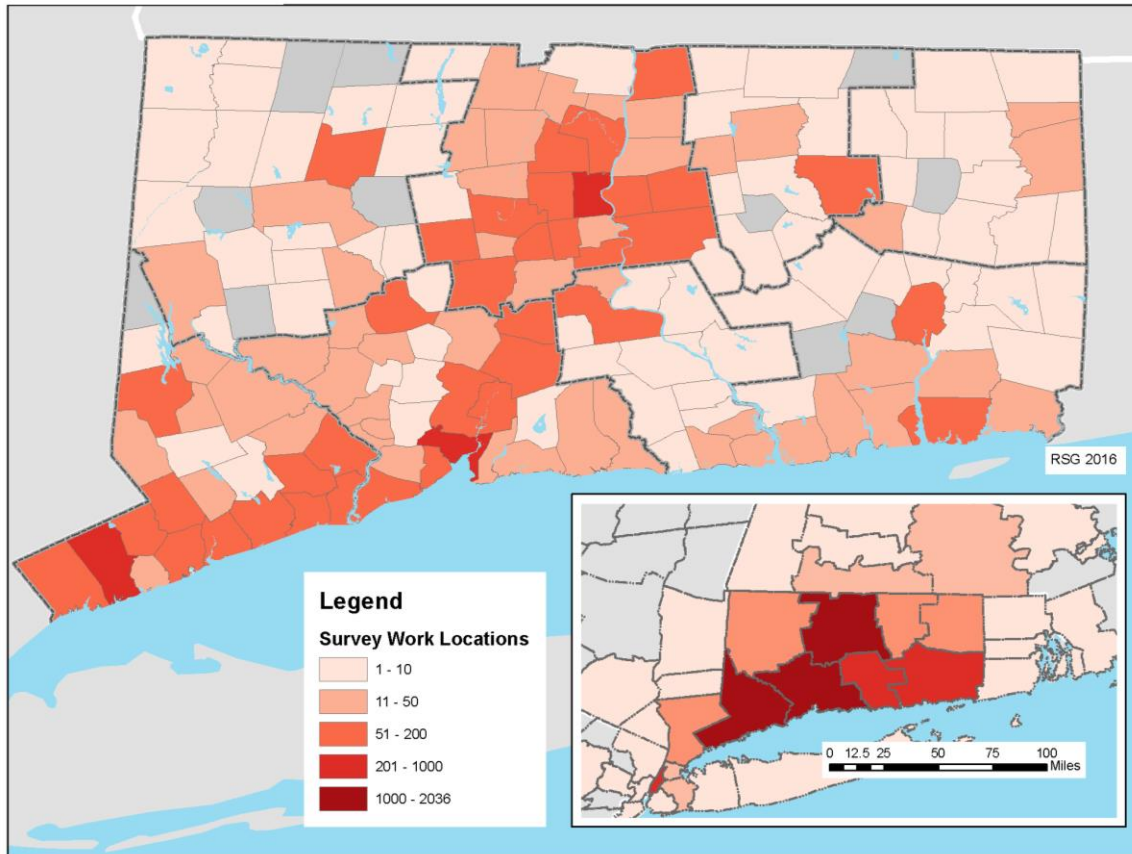
Figure 15: Distribution of Respondents' School Location at the Town Level
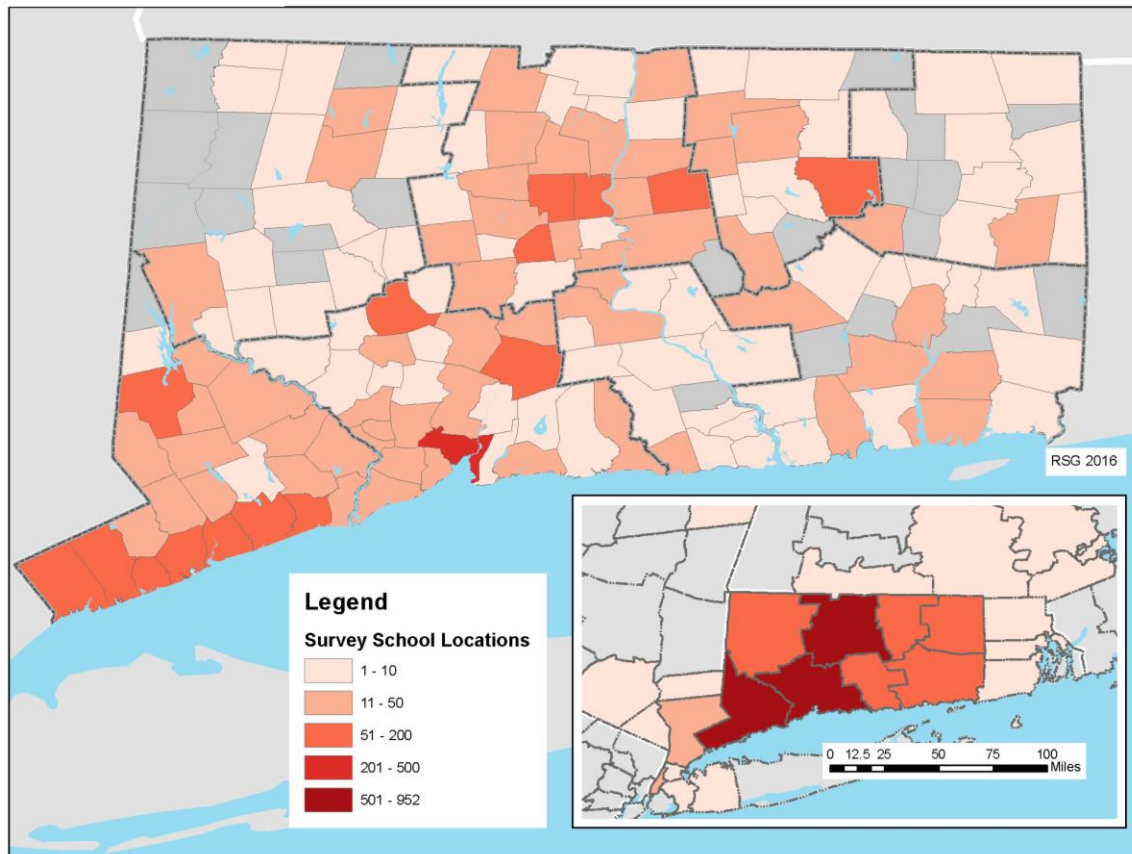
3.8.2 Survey Incentives

For the main study, a $10 gift card was offered to all households (except volunteers) as incentive for completing the survey. The $10 amount notice was printed on the invitation materials and included in the email text of outbound reminder emails. Households could choose between virtual or physical Visa gift cards (sent via email or USPS mail). Alternatively, households could opt not to receive any gift card for participation. Table 24 shows the distribution of incentive types chosen by completed households. Most household preferred the physical Visa gift card and received their incentive via mail (approximately 67%).

Table 24: Distribution of Complete Households by Incentive Type

Households by Incentive Type	Complete	Percent
Virtual Visa (by email)	2,432	29%
Physical Visa (by mail)	5,601	67%
Elected no incentive	370	4%
Total	**8,403**	**100%**

3.8.3 Respondent Communication

A summary of the respondent communication using phone and email during the main study are provided below:

- Approximately 36% of all invited households had a landline telephone number associated with the address. After printed invitation materials were sent, households with known telephone numbers received calls encouraging them to participate in the survey. For households that preferred receiving reminders via telephone, ETC conducted reminder calls. Throughout the main study, a total of 24,000 outbound calls were placed by ETC, including recruitment calls and reminder calls.
- Depending on when a household was recruited and when they completed the last part of the survey, they received up to five reminder emails. Approximately 53% of all recruited households received one or more email reminders during the main survey.
- The study team received over 600 emails during the main survey. Inquiries sent via email typically involved questions about gift cards, requests for help with a specific part of the survey (such as how to use the map), requests to volunteer, or general comments or questions about the survey. Occasionally households e-mailed comments about regional transportation issues. Table 25 provides a distribution of topics of the inbound email and redacted content of emails from households with substantive comments regarding regional transportation or the study has been included in Appendix L.
- Sixty households asked to volunteer and were provided with passwords, travel dates and survey instructions. 21 of these volunteer households completed the entire study. Volunteer data has been excluded from the final dataset.

Table 25: Distribution of Inbound Emails by Topic[69]

Inbound Emails by Topic	Count	Percent
General	165	24.6%
Incentive	157	23.4%
Volunteers	137	20.4%
Errors (self-reported, user)	111	16.5%
Survey administration	53	7.9%
Unsubscribe requests	37	5.5%
Privacy concerns	8	1.2%
Study concerns	3	0.4%
Total	**671**	**100.0%**

3.8.4 Survey Results

This section of the report summarizes the survey responses at household-, person-, vehicle-, and trip-levels. The results presented and discussed in this section summarize the dataset after going through the data processing and preparation steps outlined in Section 3.7 . Both weighted and unweighted results, only from invited households who have provided complete and valid information, are included. The survey measurements may be compared to demographic profiles of the population to assess representativeness and to identify biases. The ACS 2009–2013 five-year estimates were used for comparison because the sampling plan drew

---

[69] These only concern emails received by the study team on the study email address received as of August 15, 2016. A small number of emails were also fielded by UConn and also by CTDOT. The topics of these emails are similar to those received by the study email. These are however not included in this table.

upon population data from this source (which at the time of executing the study was the most current five-year estimates available). For the household and person variables, the results in the tables below include both unweighted and weighted results to show how the raw survey data compared to the region's population. This also is done to show that the after applying the weights (described in Section 3.7.5), the biases for important household and person variables have been addressed. Also, below only a small set of household-, person-, and trip-level variables are analyzed. Unweighted and weighted distributions of all variables are provided in Appendix P and Appendix Q respectively.

Comparison of Household Variables
In this subsection, distributions of key household level variables from the survey are compared against the population estimates from the 2009-2013 ACS. Income, household size, and vehicle ownership are three elements that typically play a significant role in household travel behavior analysis. The survey results and ACS estimates for these data items are shown below in Table 26, Table 27, and Table 28. As can be seen from the unweighted distributions, the household characteristics of the survey sample are relatively close to the regional characteristics with some observable differences. The observable differences can be attributed to typical non-response patterns seen in household travel surveys. There are fewer low-income and larger households in the sample as compared to the population. While only a small difference is observed, there are fewer zero vehicle households. These trends are reasonable and consistent with expectations and experiences of the study team. Low-income households are frequently underrepresented in many surveys, and larger households can be difficult to recruit and retain due to the additional burden per respondent for the household overall. The trend in zero vehicle households can partly be attributed to the underrepresentation of the low income demographic segment. Due in part to the sample design and recruitment and retention strategies (described previously), the differences between the unweighted responses and the state population are smaller than they might have been in the pilot study. Also, over the weeks of data collection, the differences improved owing to study administration adjustments.

There were higher proportions of one or two person households in the survey compared to distributions of the same types of households in the population, as shown in Table 26. To the extent possible, the sampling plan targeted more invitations for areas with higher proportions of large households, but after recruitment it was more difficult to retain larger households due to the additional burden per household member. Also, additional incentives that are typically offered to compensate for their higher burden of larger households could not be pursued in the study. However, efforts were made to decrease respondent burden by offering copy trip features to help ease reporting when joint trips are pursued. Targeted recruitment and retention strategies were also pursued as described in Section 3.6 to improve response rates of larger households.

Table 27 provides the distribution of household income. It must be noted that distribution shown is for the imputed income variable. Out of the 8,403 households 1,020 households (12.1%) selected "Prefer not to answer", thus opting not to report their household income category at any level. As noted, low-income households typically participate in surveys at lower than average rates, therefore the sampling plan targeted more invitations in low-income areas.

Due to this plan, a larger sample of low-income households was achieved, though it was still slightly lower than the statewide proportion of low-income households. Overall 15.3% of survey belonged to the incomes lower than $25,000 category, and 37.1% belonged to the incomes above $100,000 category.

The travel behavior of zero-vehicle households is important for policy analysis and modeling. However, due to the relatively small percentage of zero-vehicle households in the state, it was important to target these households to obtain a sufficient sample to analyze the behaviors of car-free households. Additionally, zero-vehicle households are often correlated with low incomes, as a result they are also more difficult to recruit. Table 28 shows the differences between household vehicle counts and the ACS household vehicle estimates. Compared to the low income households, the underrepresentation in zero vehicle households is relatively small.

As can be seen, there are sample biases due to both design choices (e.g. oversampling) and survey variability (e.g. anticipated response rates being different from observed response rates). Therefore, it is important to correct for these sample biases so that the inferences drawn are accurate. Weighting described in Section 3.7.5 rectifies the sample biases. Additionally, applying weights allows the survey results to be expanded to the entire population. As can be seen from the "Weighted Survey" portion of Table 26, Table 27, and Table 28, the distributions perfectly match the population distributions from the ACS and the sample biases have been rectified.

Table 26: Comparison of Household Size Distribution

Household Size Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
1 person	374,214	27.6%	2,929	34.9%	374,214	27.6%
2 people	446,464	32.9%	3,361	40.0%	446,464	32.9%
3 people	221,469	16.3%	1,046	12.4%	221,469	16.3%
4 people	196,253	14.5%	742	8.8%	196,253	14.5%
5 or more people	117,449	8.7%	325	3.9%	117,449	8.7%
Total	1,355,849	100.0%	8,403	100.0%	1,355,849	100.0%

Table 27: Comparison of Household Income Distribution

Income Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
Under $25,000	243,901	18.0%	1,287	15.3%	243,901	18.0%
$25,000–$49,999	254,982	18.8%	1,359	16.2%	254,982	18.8%
$50,000–$74,999	223,674	16.5%	1,490	17.7%	223,674	16.5%
$75,000–$99,999	178,175	13.1%	1,147	13.6%	178,175	13.1%
$100,000 or more	455,117	33.6%	3,120	37.1%	455,117	33.6%
Total	1,355,849	100.0%	8,403	100.0%	1,355,849	100.0%

Table 28: Comparison of Household Vehicle Ownership

Vehicle Count Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
0 vehicle	121,597	9.0%	639	7.6%	121,597	9.0%
1 vehicle	442,638	32.6%	3,100	36.9%	442,638	32.6%
2 vehicles	518,022	38.2%	3,147	37.5%	518,022	38.2%
3 or more vehicles	273,592	20.2%	1,517	18.1%	273,592	20.2%
Total	1,355,849	100.0%	8,403	100.0%	1,355,849	100.0%

Comparison of Peron Variables

Unweighted person-level results from the survey were also compared to ACS demographic estimates. As seen in Table 29, individuals under age 18 and between ages 18-34 participated at a fairly low rate compared to the ACS percentage of people in these age groups. For individuals under age 18, this is potentially due to increased survey burden for larger households (i.e. households with children). For individuals between ages 18-34, lower survey response rates may be a result of adults in this age group being more transient, less likely to have a permanent address, and possibly away at school, which leads to difficulty in achieving a representative sample of 18-34-year-olds. Residents aged 55-74 actively participated in the survey (35.9% of the total). Higher response rates within the 55-74 age group may be related to the high percentage of retired individuals within this age group who are easier to reach at home and who often have more time and interest in completing surveys.

From the unweighted results for gender as shown in Table 30, it can be seen that there is a slight overrepresentation of females in the survey and a small amount of underrepresentation of males. As shown in unweighted results in Table 31, 51% of respondents over the age of 16 reported being employed, meaning they were full/part-time or self-employed. Nearly 35.4% are unemployed (also includes not in the labor force).

Employed respondents who reported having a regular workplace (or who reported regularly commuting to different jobsites) also provided their "typical" commute mode as shown in the unweighted results of Table 32. The table excludes approximately 10% of employed respondents who work from home or drive for a living and do not typically commute to a fixed workplace. Not surprisingly, the most common way for commuters to get to work is by driving alone. Transit commuters were targeted for "oversampling" as noted in Section 0. Collecting data from transit users was deemed important for policy analysis and modeling by the study team, and the sampling plan set a high final target for transit commuters, which is reflected in the results, where 7.8% of survey participants reported transit as their typical commute mode in comparison to the 4.9% ACS estimates of the population.

Similar to the household variables, the sample contains biases as can be observed from the differences in the distributions between unweighted survey results and the ACS estimates. In the weighting procedure that was applied in the study, both household and person variables were controlled. As a result, the weighted distributions rectify the bias and match the population distributions from ACS very closely. It must be noted that unlike the household

variables, the distribution of person variable counts from the weighted survey results still show small differences when compared with ACS estimates. However, when the percentage distributions are compared they mirror ACS estimates very closely. This can be attributed to two reasons. First, the ACS estimates of person variables include those living in group quarters, whereas in the study only household units are considered (i.e. group quarters are excluded). Second, the weighting approach applied attempts to match distributions of household and person control variables of interest. However, in most real world data, there are always discrepancies between household and person control variables. As a result, it is not possible to generate weights that satisfy the household and person control variable distributions perfectly. In the weighting approach used in this study, the weights are generated so that distributions of household control variables are perfectly matched whereas person control variables are matched closely.

Table 29: Comparison of Person Age

Age Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
Under 18 years	804,045	22.4%	2,791	16.0%	736,824	22.0%
18–34 years	764,013	21.3%	2,659	15.2%	706,829	21.1%
35–54 years	1,038,256	29.0%	4,301	24.6%	975,646	29.1%
55–74 years	726,385	20.3%	6,274	35.9%	760,673	22.7%
75 or more years	250,862	7.0%	1,456	8.3%	174,233	5.2%
Total	3,583,561	100.0%	17,481	100.0%	3,354,205	100.0%

Table 30: Comparison of Gender

Gender Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
Female	1,838,197	51.3%	9,259	53.0%	1,756,865	52.4%
Male	1,745,364	48.7%	8,222	47.0%	1,597,341	47.6%
Total	3,583,561	100.0%	17,481	100.00%	3,354,206	100.0%

Table 31: Comparison of Employment Status of Individuals

Employment Status Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
Employed	1,768,014	49.3%	8,883	50.8%	1,643,217	49.0%
Not employed	1,111,136	31.0%	6,194	35.4%	1,064,989	31.8%
Age less than 16	704,411	19.7%	2,404	13.8%	646,000	19.3%
Total	3,583,561	100.0%	17,481	100.0%	3,354,206	100.0%

Table 32: Comparison of Typical Commute Mode

Typical Commute Mode Category	ACS Population Estimates		Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent	Count	Percent
Drive Alone/Motorcycle	1,360,635	82.2%	6,404	79.3%	1,178,894	78.0%
Carpool	144,497	8.7%	688	8.5%	154,821	10.2%
Transit	80,574	4.9%	630	7.8%	114,703	7.6%
Walk/Bike	56,948	3.4%	295	3.7%	51,367	3.4%
Taxi/Rideshare	1,381	0.1%	10	0.1%	2,856	0.2%
Other	12,196	0.7%	53	0.7%	9,358	0.6%
Total Commuters[1]	**1,656,231**	**100.0%**	**8,080**	**100.0%**	**1,511,999**	**100.0%**

Analysis of Trip Variables

In this subsection, an overview of the travel characteristics of Connecticut residents is provided based on the linked trip file.

Overall Trends

There were 65,103 valid linked trips were identified in the travel diary portion of the study. After weighting, the total number of valid trips is nearly 11.8 million. The 11.8 million trips are made by approximately 3.4 million Connecticut residents, residing in 1.4 million households. The average trip rate is 3.51 per person and 8.68 per household. Trip rates increase with household income, presence of children and household size. Women have lower work trip rates, but higher total trip rates than men. Persons 35-54 years of age have higher trip rates than other age groups.

Trip Purpose

Table 33 shows weighted and unweighted distributions of trip purpose. In the weighted and unweighted analysis, the two most common trip purposes is the same: go home, followed by trips to work. However, the order of the remaining purposes is different between the weighted and unweighted analysis. The next three common categories were run errands, pick-up/drop-off/accompany, and other shopping (e.g. mall, hardware store) in the unweighted analysis. Whereas in the weighted analysis, the next three common categories were pick-up/drop-off/accompany, run errands, and attending school/class. This appears plausible because from the comparison of the person demographics, there was underrepresentation of younger individuals who are also presumably school-going. After accounting for the underrepresentation of younger individuals and subsequently related trips, school trips jumped to one of the common purposes.

Trip Mode

Table 34 shows summary of travel modes. There are, again, differences between weighted and unweighted results. Corresponding with the trip purpose differences, there is a slightly higher share of school bus trips in the weighted results. In the weighted analysis, driving trips comprise approximately 82.8% of the total trips in the dataset. These trips can be split further in the following manner: 48.2% of trips are made by Single Occupant Vehicle (SOV) mode, with the remaining 34.6% of trips involve some form of carpooling. Walk/bike is the next most popular

mode accounting for 9.2% of all trips. Public transit serves 3.2% of all weekday trips in Connecticut.

Figure 16 shows distribution of the trip mode categories for different types of trips. The analysis in the figure is only limited to weighted data. SOV is the most popular mode for all types of trips except for school trips and escorting trips. 76.5% of work trips are made by SOV mode. About 50% of shopping, personal business and home trips are made by SOV mode as well. 28.4% of school trips are made by family carpool and 27.6% are made using school bus.

Table 33: Weighted and Unweighted Frequency Distribution of Trip Purpose

Purpose Category	Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent
Go home	21,943	33.7%	4,031,531	34.2%
Go to primary workplace	7,843	12.0%	1,385,468	11.8%
Go to other work-related place	2,081	3.2%	362,410	3.1%
Attend school/class	2,503	3.8%	664,923	5.6%
Other school-related activity	996	1.5%	210,097	1.8%
Grocery shopping	3,574	5.5%	601,176	5.1%
Do other shopping (e.g. mall, hardware store)	3,702	5.7%	619,111	5.3%
Run errands (e.g. bank, haircut)	4,647	7.1%	719,348	6.1%
Go to restaurant/bar/get take-out	3,649	5.6%	605,205	5.1%
Drop off/pick up/accompany other person	4,018	6.2%	871,347	7.4%
Exercise (e.g. gym, walk, jog/run)	2,947	4.5%	475,319	4.0%
Family activity (e.g. child's game)	700	1.1%	144,870	1.2%
Medical visit (e.g. doctor, dentist)	1,550	2.4%	251,184	2.1%
Social (e.g. visit friends/relatives)	1,763	2.7%	309,697	2.6%
Leisure/entertainment (e.g. movies)	927	1.4%	154,870	1.3%
Religious/civic/volunteer	1,143	1.8%	170,250	1.4%
Vacation/holiday/traveling (e.g. hotel)	166	0.3%	23,577	0.2%
Change travel mode (e.g. wait for bus, change planes)	66	0.1%	13,794	0.1%
Home other	460	0.7%	89,675	0.8%
Go to airport/switch airport	63	0.1%	11,550	0.1%
Other	362	0.6%	55,583	0.5%
Total	65,103	100.0%	11,770,987	100.0%

Table 34: Weighted and Unweighted Frequency Distribution of Mode Choice

Mode Category	Unweighted Survey		Weighted Survey	
	Count	Percent	Count	Percent
Walk/jog/wheelchair	5,721	8.8%	999,486	8.5%
Bicycle	422	0.6%	80,108	0.7%
Vehicle in household	52,314	80.4%	9,252,911	78.6%
Other vehicle (e.g. rental, friend's car)	2,322	3.6%	494,431	4.2%
Regular taxi or hired car service	69	0.1%	9,498	0.1%
Ride-share taxi/car service (e.g. Uber, Lyft)	35	0.1%	4,244	0.0%
Vanpool	55	0.1%	6,398	0.1%
School bus	1,789	2.7%	471,483	4.0%
Public (city) bus	1,226	1.9%	279,509	2.4%
Train (e.g. Amtrak)	86	0.1%	9,498	0.1%
Shuttle (e.g. a hotel's, an airport's)	100	0.2%	18,404	0.2%
Paratransit	59	0.1%	10,175	0.1%
Commuter Rail (e.g. Metro North)	536	0.8%	69,341	0.6%
Subway	40	0.1%	3,285	0.0%
Intercity bus (e.g. Greyhound)	19	0.0%	3,312	0.0%
Ferry/water taxi/boat	4	0.0%	376	0.0%
Airplane/helicopter	61	0.1%	11,194	0.1%
Other	245	0.4%	47,333	0.4%
Total	**65,103**	**100.0%**	**11,770,987**	**100.0%**

Figure 16: Distribution Trip Mode Category by Purpose
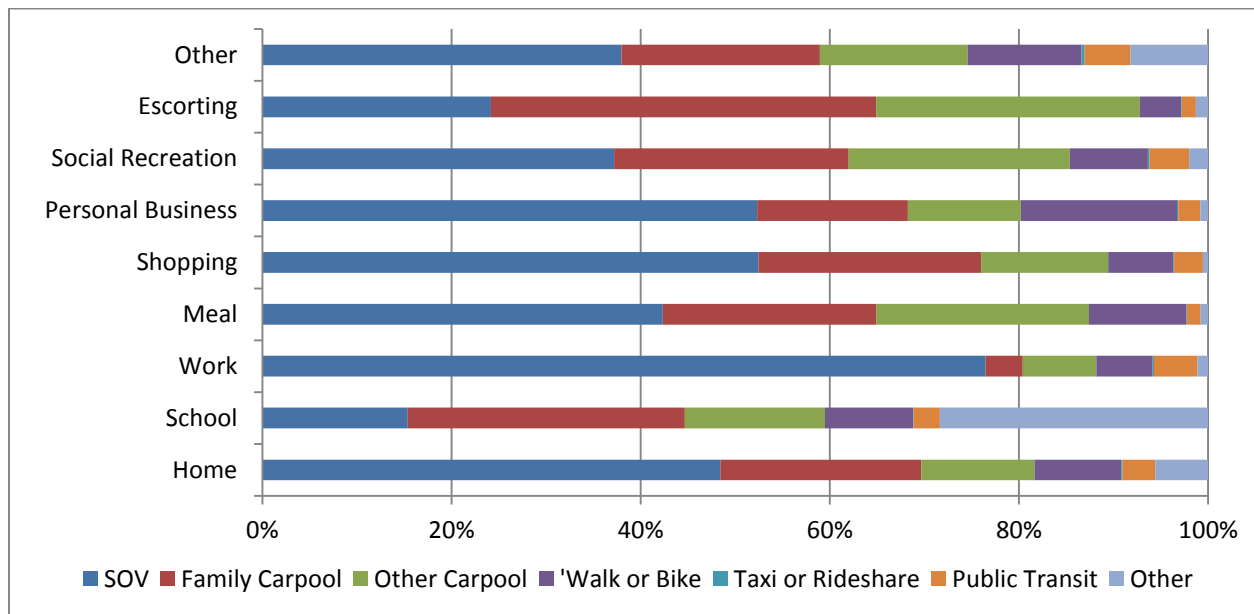
Spatial Distribution of Trips

Table 35 shows the average trip distance and duration obtained by trip purpose. The discussion below is limited to weighted data. Trip duration is based on the start time of the first leg and the end time of the last leg of the trip. The distance is obtained by aggregating the distance across all legs of a linked trip obtained using the Google Maps API. There are small differences between the weighted analysis and unweighted analysis. This can be attributed to the fact that average values are being reported. The average reported travel duration for all types of trips is 22.90 minutes. Work trips on average are longer than other trips with an average reported work travel duration of 29.66 minutes versus 21.71 minutes for non-work travel. When the travel durations are explored by employment status, it was found that the average reported travel duration of trips by employed persons is 24.47 minutes whereas average travel duration of trips by unemployed persons is 20.09 minutes. It was also interesting to note that the average travel duration of children (age less than 16 years old) is slightly higher than the average duration for unemployed individuals.

Table 35: Weighted and Unweighted Trip Distance Summary by Purpose

Purpose Category	Unweighted Survey		Weighted Survey	
	Average Duration	Average Distance	Average Duration	Average Distance
Go home	22.83	8.53	22.98	7.80
Go to primary workplace	29.27	12.57	29.66	12.50
Go to other work-related place	30.72	18.44	29.88	18.55
Attend school/class	23.18	5.18	24.99	5.13
Other school-related activity	23.44	6.80	24.25	7.41
Grocery shopping	17.02	5.52	17.98	5.12
Do other shopping (e.g. mall)	16.99	5.94	17.40	5.87
Run errands (e.g. bank, haircut)	16.42	6.73	16.61	6.02
Go to restaurant/bar/get take-out	18.27	7.81	18.10	8.09
Drop off/pick up/accompany other person	19.56	8.51	19.83	6.33
Exercise (e.g. gym, walk, jog/run)	20.79	3.88	21.21	4.05
Family activity (e.g. child's game)	24.04	12.76	22.22	8.97
Medical visit (e.g. doctor, dentist)	25.01	13.79	25.09	11.64
Social (e.g. visit friends/relatives)	25.92	11.67	25.45	9.35
Leisure/entertainment (e.g. movies)	23.58	9.85	21.44	8.44
Religious/civic/volunteer	19.25	5.84	19.5	6.29
Vacation/holiday/traveling (e.g. hotel)	88.71	147.11	85.16	156.24
Change travel mode (e.g. wait for bus, change planes)	33.71	11.48	22.74	5.91
Home other	28.53	13.90	24.70	9.57
Go to airport/switch airport	137.28	649.70	134.17	699.79
Other	32.33	15.22	27.42	11.65
Total	**22.68**	**9.76**	**22.9**	**9.05**

Figure 17 shows distribution of work and non-work trips by starting and ending locations. The analysis here is limited to weighted data and locations are characterized using definitions of counties in the state. A majority of trips made are within Connecticut. 84.7% of trips are intra-county and 11.4% of trips are inter-county of Connecticut. Only about 1% of trips are made from/to a location outside Connecticut. Intra-county trips account for the greatest proportion of both work trips and non-work trips.

Figure 17: Distribution of Trip Ends
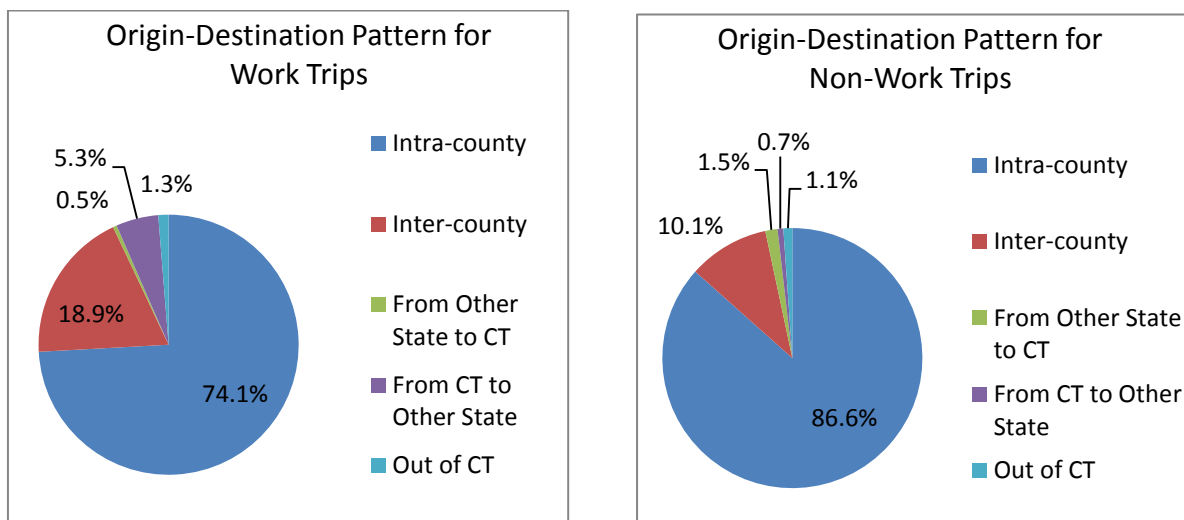
Travel Accompaniment

Table 36 shows the 57.6% of all trips were made by individuals alone, while 43.4% of trips were made with others. Majority of work trips (85.1%) and personal business (65.7%) trips are made alone. Escorting trips, school trips and social recreation trips are more likely to be made along with others.

Profile of Transit Travelers

Below socioeconomic and demographic characteristics of transit travelers are provided:

- ❑ 89% of the transit trips were made by residents of Hartford, Fairfield, and New Haven counties
- ❑ Majority of the trips were made by females (53%) and by those who were in the 45 to 54 age group (20%)
- ❑ 48% of the trips were made by those who were employed
- ❑ 13% of the trips were made by students who were either enrolled fulltime or part time
- ❑ 49% of the trips were made by those with a driver's license
- ❑ Most common trip purposes were "Go home" (37%) and "Go to primary workplace" (19%)

Table 36: Weighted and Unweighted Travel Accompaniment Summary by Purpose

Purpose Category	Unweighted Survey			Weighted Survey		
	Trip Count	% Traveling Alone	% Traveling with Others	Trip Count	% Traveling Alone	% Traveling with Others
Go home	21,943	63.8%	36.2%	4,031,531	58.4%	41.6%
Go to primary workplace	7,843	88.2%	11.8%	1,385,468	86.7%	13.3%
Go to other work-related place	2,081	82.8%	17.2%	362,410	79.1%	20.9%
Attend school/class	2,503	24.4%	75.6%	664,923	27.0%	73.0%
Other school-related activity	996	36.3%	63.7%	210,098	34.9%	65.1%
Grocery shopping	3,574	69.3%	30.7%	601,177	62.7%	37.3%
Do other shopping (e.g. mall, hardware store)	3,702	65.3%	34.7%	619,111	56.5%	43.5%
Run errands (e.g. bank, haircut)	4,647	74.8%	25.2%	719,348	71.1%	28.9%
Go to restaurant/bar/get take-out	3,649	53.8%	46.2%	605,205	49.0%	51.0%
Drop off/pick up/accompany other person	4,018	29.0%	71.0%	871,347	26.0%	74.0%
Exercise (e.g. gym, walk, jog/run)	2,947	69.4%	30.6%	475,319	65.9%	34.1%
Family activity (e.g. child's game)	700	33.6%	66.4%	144,870	27.8%	72.2%
Medical visit (e.g. doctor, dentist)	1,550	58.1%	41.9%	251,184	53.6%	46.4%
Social (e.g. visit friends/relatives)	1,763	61.5%	38.5%	309,697	54.0%	46.0%
Leisure/entertainment (e.g. movies)	927	50.3%	49.7%	154,870	43.7%	56.3%
Religious/civic/volunteer	1,143	67.5%	32.5%	170,250	60.7%	39.3%
Vacation/holiday/traveling (e.g. hotel)	166	16.3%	83.7%	23,577	16.1%	83.9%
Change travel mode (e.g. wait for bus, change planes)	66	66.7%	33.3%	13,794	49.4%	50.6%
Home other	460	60.0%	40.0%	89,675	62.8%	37.2%
Go to airport/switch airport	63	49.2%	50.8%	11,550	48.0%	52.0%
Other	362	45.9%	54.1%	55,583	39.2%	60.8%
Total	**65,103**	**63.2%**	**36.8%**	**11,770,987**	**57.6%**	**42.4%**

Profile of Bike and walk Travelers

Below socioeconomic and demographic characteristics of bike and walk travelers from the survey dataset are provided:

- 81% of the bike/walk trips were made by residents of Hartford, Fairfield, and New Haven counties
- Majority of the trips were made by females (51%) and by those who were in the 25-34 age group (18%)
- 40% of the trips were made by those who were employed
- 14% of the trips were made by students who were either enrolled fulltime or part time
- 67% of the trips were made by those with a driver's license
- Most common trip purpose was "Go home" (34%) and "Exercise" (19%)

3.8.5 Survey Validity Analysis

In order to assess validity of the results from the CSTS, results were compared against two other household travel surveys, namely, the 2009 National Household Travel Survey (NHTS) and the 2010/2011 Regional Household Travel Survey (RHTS). The focus of the comparison with the 2009 NHTS was to assess reasonableness of the survey results by comparing with a nationally representative household travel survey. On the other hand, focus of the comparison with the 2010/2011 RHTS was to conduct a more localized analysis to assess accuracy. While weighted and unweighted results were compiled, only the weighted results that rectify the sample biases are relevant and considered in the comparison analysis. A summary of the findings are included below, detailed tables and discussion of results along each dimension considered are presented in Appendix M.

*Summary of the Comparison between 2016 CSTS and 2009 NHTS*

Similar to the CSTS, the 2009 NHTS data is organized into four files including a household file, a person file, a trip file and a vehicle file. The study team compared results from the CSTS with NHTS along key household-, person- and trip-level attributes. Based on the comparison, it was found that the weighting process rectifies the bias in the survey sample. The weighted household attributes and person attributes from the two surveys are comparable, except for the household income, which does show some differences. The average person and household daily trip rates from the two surveys are also very close. When comparing trip purposes and travel mode distributions, differences were observed. These differences are reasonable because NHTS comprises of travel behaviors of a very diverse population across all of the US whereas CSTS provides information of only CT residents. Overall, the results were found to be plausible and consistent with expectations.

Summary of the Comparison between 2016 CSTS and 2010/2011 RHTS

In an effort to further evaluate the accuracy of the CSTS data, a more localized analysis was conducted by comparing results for only those households residing in the study areas that are common to both the 2016 CSTS and 2010/2011 RHTS. The 2010/2011 RHTS survey was sponsored by New York Metropolitan Transportation Council (NYMTC) and North Jersey Transportation Planning Authority (NJTPA). It was conducted between September 2010 and November 2011, and covered 28 counties across three states: New York, New Jersey, and Connecticut. The two counties in Connecticut were Fairfield County and New Haven County (as highlighted in brown color in Figure 18). In the RHTS, demographic and trip data were collected from 18,800 households, including a GPS sub–sample of 1,880 households. Respondents in the

GPS sub-sample provided additional travel data using wearable global positioning system (GPS) devices. The reason for including the GPS subsample was to improve accounting for short, non-motorized trips.

Figure 18: Overlapping Study Areas between 2016 CSTS and 2010/2011 RHTS
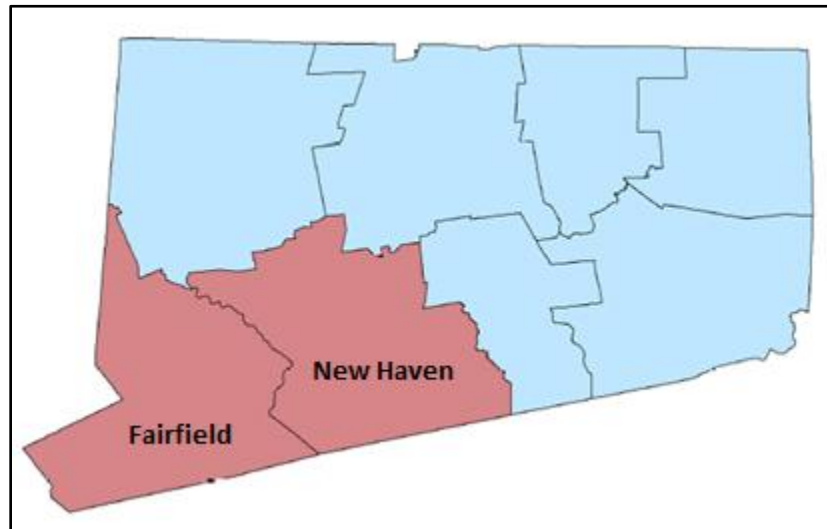
The localized comparison also involved a comparison of household-, person- and trip-level attributes of interest. The comparisons of household and person level attributes indicate that the weighting factor rectifies sample bias in the CSTS. The localized comparison of trip attributes indicates that the overall travel patterns from CSTS are similar to that from the 2010/2011 RHTS. This also provides evidence in support of the accuracy of the results. However, after comparing the GPS corrected results from the 2010/2011 RHTS, it was found that the 2016 CSTS may also be suffering from potential under-reporting of short trips. This is not unexpected, and is one of the primary reasons why more and more household travel surveys include a GPS sub-sample to account for potential underreporting of short trips. Therefore, in future editions of the CSTS, GPS sub-sample or smart-phone based survey approaches may need to be considered.