SEMCOG Travel Counts 15

Household Travel Survey

Final Methodology Report

Authors

Martha Wilaby, Associate Project Director
Jesse Casas, Project Director

September 29, 2016

Prepared for:
Southeast Michigan Council of Governments
1001 Woodward Avenue, Suite 1400
Detroit, MI 48226-1904

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

This page is intentionally left blank.

Table of Contents

Appendixes

Appendixes **Page**

Tables

Figures

Project Summary and Background 1

The Southeast Michigan Council of Governments (SEMCOG) is a government association that serves the Southeast Michigan region, made up of Livingston, Macomb, Monroe, Oakland, St. Clair, Washtenaw, and Wayne counties. SEMCOG is also designated as Metropolitan Planning Organization for the metro Detroit region to perform transportation planning and financing tasks.

SEMCOG, along with the Michigan Department of Transportation (MDOT), contracted Westat to survey randomly selected households. The purpose of the SEMCOG Travel Counts 2015 (STC 15) and the Michigan Travel Counts III (MTC III) was to provide much-needed travel behavior data for use in state and regional transportation models. Participants were recruited by mail and asked to fill out a web-based recruitment survey, were assigned a travel date, and were then asked to report the details of their travel in the web-based retrieval survey. They could also report by phone if they preferred.

At the conclusion of STC 15, a total of 12,394 households across in the SEMCOG region participated in the survey by providing household demographic information and then reporting their travel for 1 day. In addition, 1,106 households carried a GPS unit for a 3-day period.

The first statewide household travel survey (known as MI Travel Counts-MTC I) was conducted in 2004-2005, which collected activity and travel inventories from all members of 14,996 randomly selected households within the state of Michigan; of these, 2,249 households completed were within the Southeast Michigan region. To supplement this Southeast Michigan sample, SEMCOG commissioned a separate SEMCOG Travel Counts program, as an add-on component to MI Travel Counts. A decrease in travel, and the desire to determine if household travel had changed due to the downturn in the economy, prompted the second statewide survey (MTC II) and SEMCOG add-on in 2009.

SEMCOG will use data collected in the STC 15 to update, develop, and adjust regional and urban travel demand models. The primary use of the models is to estimate future travel pattern as well as travel demand for future project and investment priorities in the SEMCOG region.

The STC 15 was conducted between April and December of 2015. This report documents the methodology used to carry out the STC 15 survey. A separate report, the Travel Characteristics Report, provides information about the travel characteristics of Michigan residents.

Survey Design 2

This chapter describes the STC 15 household travel survey design. The survey design includes how the survey sample was selected randomly from addresses in the eight large geographic sample areas, the public awareness plan to promote the data collection, the recruitment process by mail, project staff available to assist respondents in completing the study, the project website, follow-up communication with the respondents, and the retrieval process where respondents reported their travel for both GPS and log-only households as well as the incentive structure.

2.1 Introduction

STC 15 used a random sample of households from all residential addresses in eight large geographical areas: East Wayne, West Wayne, Oakland, Macomb, Washtenaw, Monroe, St. Clair, and Livingston. Seven of these areas are counties, while East and West Wayne are one county split into two areas. Household were stratified by size and income because income could be used as an indicator for vehicle ownership and households with higher incomes may make more discretionary trips than less wealthy households with the same number of vehicles. A map with the number of the sampled households by area can be seen in Figure 2-1.

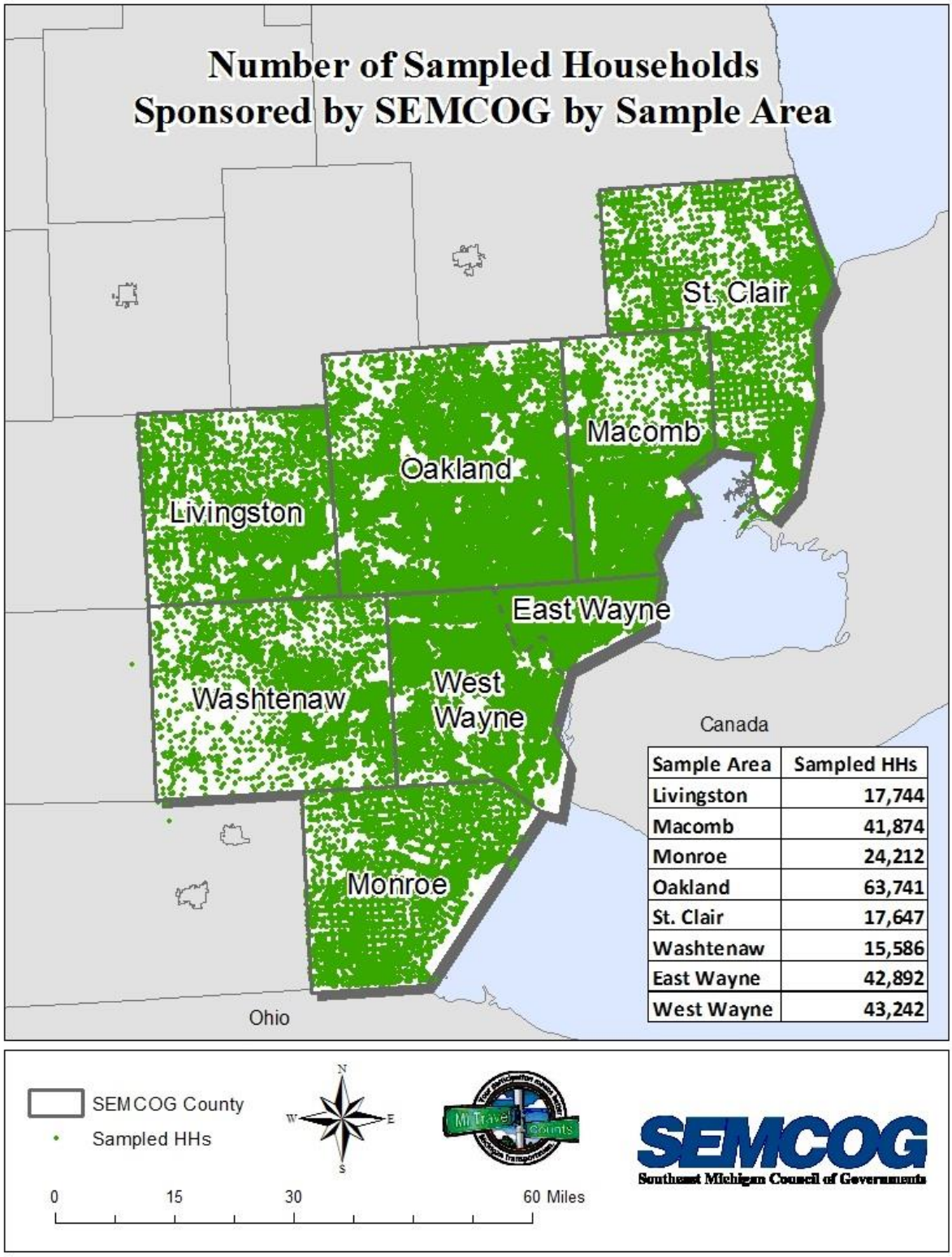
Participants were recruited by an invitation letter, were asked to fill out a web-based recruitment survey, were assigned a travel date, and were then asked to report the details of their travel in the web-based retrieval survey. Participants could also call a toll free number and complete recruit and/or report travel by phone. Some households were offered the opportunity to participate in a global positioning system (GPS) technology component of the study. In the GPS subsample, all household members ages 16-75 were asked to carry a wearable GPS device with them for 3 days.

It was important that all prospective participants be provided the ability to participate in the survey and to note that not all Michigan residents are proficient in the English language. The three most commonly spoken languages in Michigan are English, Spanish, and Arabic. The website text and survey materials were translated into Spanish and Arabic by a professional translation firm. A tagline was added to the bottom of the English recruitment letter in Spanish and Arabic, referring respondents to the webpages where the translated materials could be downloaded. The firm was also

available to translate if a respondent called or emailed needing assistance in completing either the recruitment or retrieval survey.

A pilot study was conducted from January to March of 2015. The main data collection was conducted in two phases. There was a spring data collection beginning in February 2015, followed after a break in the summer by a fall data collection beginning in September 2015. The Project Work Plan containing the project schedule can be found in Appendix A.

Figure 2-1. Map of SEMCOG region with number of sampled households



Sample Area	Sampled HHs
Livingston	17,744
Macomb	41,874
Monroe	24,212
Oakland	63,741
St. Clair	17,647
Washtenaw	15,586
East Wayne	42,892
West Wayne	43,242

2.2 Sampling

Westat used address-based sampling (ABS) to obtain a representative sample of households for each of the geographical areas in the SEMCOG region. The sampling frame was a database of addresses created by Marketing Systems Group (MSG) from the U.S. Postal Service's (USPS's) Computerized Delivery Sequence (CDS) file. Westat used MSG's geocoding for the initial frame selection, but re-geocoded all addresses in the frame to confirm location before final sample selection. In addition to containing non-vacant addresses, the sampling frame included: vacant residential addresses in each region except in areas where the vacancy rate was found to be higher than generally found in other areas across the state, and P.O. box addresses, which were excluded from the sampling frame. The sampling frame also included throwback addresses (i.e., street addresses for which mail is redirected by USPS to a specified P.O. box) but did not include seasonal addresses.

The first step in developing the sample design was to specify the target population. The target population for STC 15 was all households in the eight SEMCOG geographic areas: Livingston, Macomb, Monroe, Oakland, St. Clair, Washtenaw, East Wayne, and West Wayne. A household was defined as all the people who occupy a housing unit. A housing unit may be a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied as separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall.

In each region, Westat selected a stratified sample of addresses consisting of a high-density stratum and a low-density stratum. The high-density stratum contained a higher concentration of hard-to-reach households than the low-density stratum. Oversampling the high-density stratum helped in achieving specified sample size targets for post-stratification cells defined by respondent characteristics, such as household income, and by measures of travel, such as number of vehicles and household workers.

During the sampling process, addresses were randomly assigned a day of week and a release group for the initial mail-out. Assigning addresses with a day of week helped balance travel across the travel week, defined as Monday through Thursday. Release groups were used to manage the release of the sample and subsequently the flow of completed households throughout the field period. Westat released addresses on an arranged schedule based on expected response rates. Reserve sample was drawn at the same time the main sample was drawn as insurance in case response rates were lower than anticipated. Sample for the pilot and spring data collection were drawn at the same

time. Westat used the results of the spring data collection to inform the amount of sample and frequency of release groups required for the fall data collection.

Geocoding of the addresses in the ABS sampling frame permitted us not only to determine the set of addresses to be sampled for each sample area, but also permitted the oversampling of addresses in areas that, according to the ACS, have a high concentration of hard-to-reach households, as determined by highly related demographic characteristics such as the area's prevalence of households with no automobiles or the area's prevalence of four-person households.

Each address on the ABS sampling frame contained a number of administrative codes that have been assigned by the U.S. Postal Service to manage the delivery of mail. The following codes are also used by survey designers to decide which addresses should or should not be exposed to sampling:

- **P.O. Box and Only-Way-to-Get-Mail (OWGM) Codes.** These codes indicate if a residential address is a P.O. box, and, if it is, whether it is the only way for the residence to receive mail. In urban and suburban areas, P.O. boxes are usually not sampled because of the low prevalence of OWGM P.O. boxes. The middle column of Table 2-1 indicates the prevalence of OWGM P.O. boxes in the eight SEMCOG sample areas.

- **Vacant Address Code.** USPS indicates an address is vacant if it is unoccupied for 90 days or longer. Because new construction is considered by USPS to be vacant, but the USPS vacant address code may not be removed by the postal carrier immediately after occupancy, survey designers often sample USPS-designated vacant addresses. In national ABS surveys, approximately 50 percent of survey invitations sent to addresses that USPS considers to be vacant are returned as undeliverable. The other survey invitations do yield some completed surveys, though their survey return rate is much lower than the rate for sampled addresses that USPS does not designate as vacant. The disadvantage of sampling USPS-designated vacant addresses is that because of their lower yields their use increases the average cost per completed case. Table 2-1 also indicates the prevalence rate for USPS-designated vacant addresses for the eight SEMCOG sample areas.

- **Seasonal Address Code.** USPS designates an address as a seasonal address if mail is to be delivered to it only during a specified period of the calendar year. The rightmost column of Table 2-1 indicates that the seasonal-address rate for all of the eight SEMCOG areas was very low. All seasonal addresses were excluded from sampling.

- **Drop-Point Address Code.** A drop-point address is an address to which USPS delivers mail for multiple households and then a non-USPS person distributes mail to the individual households. For example, an apartment building in which the postal carrier leaves the mail at the front desk is a drop-point address, but an apartment building in which the postal carrier puts the mail for each unit in the unit's mailbox in the lobby is not a drop-point address. The individual households associated with a drop-point

address are referred to as *drop units*. Drop units were sampled at the same rate as non-drop-point addresses.

Table 2-1. Counts of residential addresses and prevalence of only-way-to-get-mail (OWGM) P.O. boxes, vacant addresses, and seasonal addresses

Sample area	# Residential addresses[1]	Prevalence of OWGM P.O. boxes[2]	Prevalence of Vacant addresses[3]	Prevalence of Seasonal addresses[3]
Eastern Wayne	375,740	0.00%	21.56%	0.02%
Western Wayne	432,438	0.00%	4.42%	0.00%
Oakland	527,293	0.17%	2.52%	0.02%
Macomb	357,337	0.00%	2.87%	0.02%
Washtenaw	146,537	0.25%	1.99%	0.01%
Monroe	61,578	0.30%	2.86%	0.00%
St. Clair	68,388	0.35%	3.40%	0.00%
Livingston	70,722	1.80%	0.54%	0.21%
Total	**2,040,033**	**0.15%**	**6.42%**	**0.02%**

[1] Includes vacant addresses and seasonal address, excludes P.O. boxes.

[2] Denominator is number of residential addresses, including OWGM P.O. boxes.

[3] Denominator is number of residential addresses, excluding P.O. boxes.

2.3 Public Awareness

The public awareness campaign for STC 15 was an effective tool to reach and engage audiences that may not otherwise have been involved in the survey data collection effort, and to bring awareness to the overall project. A relevant result of the public awareness campaign was that people who were asked to participate may have been more likely to do so because they had been exposed to the campaign through other public awareness tools and strategies.

Social media sites were leveraged to raise exposure, including the creation of accounts on both Facebook and Twitter. These sites provided links to the public project website (and vice versa). The Facebook page seems less effective. It had only a total of 72 "likes" by the end of data collection. Facebook averaged one post a day and included a mix of original posts and posts from other partners. Facebook was necessary to have a presence online but is increasingly hard to get grassroots engagement because of recent changes to algorithms favoring those who pay for Facebook advertising.

The Twitter account seemed more successful with 454 followers and over 1,484 tweets, an average of three tweets per day since the first tweet on January 23, 2015. Influential followers included:

Multiple MDOT accounts, SEMCOG, the *Arab American News*, Oakland County Road Commission, Detroit People Mover, Michigan Municipal League, County Road Association, ZipCar Detroit, Michigan VanPool, COMTO Michigan, Engineering Society of Detroit, MSU Black Alumni, Local Governments (Port Huron, Mt Clemens, Lathrup Village, Harrison Twp), Washtenaw County Area Transportation Study, Transportation 4 Michigan, Timothy Fischer, Ann Arbor Transit Authority, Michigan Public Transit Assoc., Safe Roads Yes, Dan Gilmartin (CEO of MM League), Michigan Trails and Greenways Alliance, Michigan Fitness Foundation, and Ann Arbor Get Downtown Program.

The public awareness also included press releases by SEMCOG and MDOT to coincide with the mailing of invitations. Concurrently, letters were mailed to local governments so that public officials would be knowledgeable about the survey. Officials were asked to encourage their invited constituents to take part in the survey.

Any media coverage was linked from the "News about the Survey" section of the public site. The goal with these efforts was to provide a sense of legitimacy for prospective participants. The public site also provided a means for participants or prospective participants to send questions or feedback to the survey team. A member of the team responded to each comment, and in cases where the message included a complaint or observation about the transportation system in Michigan, these were forwarded to project staff for review and response as appropriate.

The website also hosted a SEMCOG-created video featuring Kathleen Lomako, the executive director of SEMCOG, encouraging participation in the survey video as well as two MDOT videos explaining the study and encouraging participation. These videos were also available on YouTube. The memo deliverable describing the public awareness plan can be found in Appendix B.

2.4 Recruitment

Each sampled address was sent up to three pieces of mail requesting its participation in the study. The first contact was a letter, on letterhead using the study logo representing both sponsor agencies (SEMCOG and MDOT). The letter provided a brief description of the study, an overview of the requirements for participation, information about the incentive, the URL for the public website, and a personal identification number (PIN) that provides access to the survey. There were up to two reminder postcards. These provided a bit less information, but still provided the URL and PIN that

allowed access to the survey website. The first reminder postcard was sent 7 days after the initial invitation letter and was mailed to all sampled addresses. The second reminder postcard was sent 7 days after the first and was sent only to nonresponding addresses.

Sampled households were directed to the secure project website, where they completed the recruitment survey. This component of the survey process was where respondents indicated their willingness to participate and provide key demographic and contact data for the household. While the primary mode of participation at this stage was via web, a telephone recruitment option was provided for those participants who requested to complete their enrollment by phone. As part of the sample selection, Westat attempted to match each sampled address to a telephone number. This process typically resulted in approximately 40 to 50 percent of the sampled addresses having an associated telephone number. These telephone numbers were used to augment the contact information collected from each household in the recruitment survey (telephone number and email address), and to attempt to recruit households by telephone.

Each responding household was assigned a travel date at the end of the recruitment survey. Before the assignment of the travel date, survey participants were offered a choice of whether they want to have personalized travel logs (diaries) mailed to them or to print the logs themselves from the website. Households that chose to have the travel logs mailed were assigned a travel date at least 10 days in the future so the logs could be prepared and mailed to arrive a few days before the travel date. Those who opted to print their own materials were assigned the next available travel date.

The travel log packets for the log-only households were sent using first-class postage in 6 x 9 envelope and included a cover letter, individualized travel logs for all household members, and a $1 bill (as a "primer" incentive to further encourage completion of the travel day survey). The letter also reminded participants about the final $20 household incentive.

Appendix C contains recruitment survey instrument for the web, and Appendix D contains the recruitment survey instrument for CATI. The Data Memo (which includes Westat's recommendations for data elements to be included in the STC 15) is located in Appendix E. The respondent materials include the advance letter, 7-day postcard, 14-day postcard, travel log, travel log letter, example travel log, and long distance travel log. Recruitment respondent materials are in English, Spanish, and Arabic. These materials can be found in Appendixes F through Z.

2.5 Interviewer Staff and Training

Westat used two types of interviewing staff to conduct the STC 15 household travel survey. Westat's Rockville, Maryland-based Telephone Research Center (TRC) was responsible for retrieval contacts only. Another set of interviewers were located at an inbound call center operation. Inbound call staff received additional training targeted at assisting respondents with miscellaneous issues (e.g., troubleshooting Internet browser problems, tracking GPS shipping) and were responsible for responding to calls to the toll-free number and emails received through the public website. All call staff were trained to conduct the surveys, including contact procedures and refusal avoidance. The interviewer training manual is located in Appendix AA, and a template of email responses to frequently asked questions can be found in Appendix BB.

2.6 Project Website

Westat created a project-specific, public-facing website that was used for promoting the survey and communicating with the general population, and that served as an access point to the survey instruments. This website included highlights of the project, frequently asked questions, PDF versions of survey materials (e.g., travel logs and GPS instructions), links to "in the news" items and to the SEMCOG home page, and the toll-free project phone number and an email link to contact the study team. Westat coordinated with SEMCOG to ensure that the public awareness campaign and the project website were linked and in sync with each other.

The website was the access point for the self-administered web-based data collection instrument. The recruitment and retrieval surveys were supported through an integrated WebGeoSurvey (WGS)/TripBuilder Web™ (TBW) software system, a web-based survey system designed by Westat specifically for household travel surveys. TBW is a Web 2.0 application with an integrated online map (provided via the Google Maps API) that enables real-time geocoding and point-of-interest lookups to collect accurate travel details via a web browser.

TBW was used for computer-assisted self-interviews (CASI) and computer-assisted telephone interviews (CATI). Using this one system for the collection of all travel survey data ensured that all collected data went through the same underlying questions and corresponding response options, the same branching, and the same logic checks. The data were all stored in one database, with the reporting mode—web or telephone—captured as an additional data element.

Screenshots of the project website can be found in Appendix CC. The participant tutorials for the diary retrieval and GPS retrieval can also be found in Appendixes DD and EE, respectively.

2.7 Pre- and Post-travel Day Follow-up and Retrieval

Westat developed a sophisticated reminder protocol that was implemented in the STC 15 household travel survey. In the recruitment survey, each participating household was informed that they would be provided with reminders and asked how they would prefer to receive those reminders. The reminder options were telephone, text, and email. If by telephone, the first reminder occurred the day before the travel date. If by email or text, the reminder was sent 2 days before the travel day. If an electronic reminder failed for any reason, this approach provided time for the participating household to be contacted by other means (if they were available). Post-travel day email and text reminders were programmed to be delivered at 1, 5, and 10 days after the travel day.

In the recruitment survey, respondents were asked to indicate how they would prefer to complete the retrieval survey. Those preferring telephone participation were contacted by an interviewer the first day following the assigned travel date. Those choosing to respond online were given 2 days to do so before being contacted by telephone. All respondents were asked to provide a telephone number in the recruitment survey, even if a telephone match was obtained at the time the address was sampled. Westat made at least five retrieval calls per each household that had yet to complete their retrieval survey.

A completed household was one in which a household-level recruitment survey was completed, and a retrieval survey was completed that includes the trip data for each household member. In larger households (those with more than four total household members), the household may be considered complete if only one household member failed to complete the retrieval survey. The script for the retrieval instrument (both Web and CATI) can be found in Appendixes FF and GG, respectively. The data elements can be located in Appendix HH and the codebook can also be found in Appendix II while the text and email reminder scripts are located in Appendix JJ.

2.8 Incentive Plan for Log-only Households

Incentives have become an essential component of survey research, and determining a responsible level of incentive was vital to the success of the project. Based on Westat's experiences conducting

these surveys and observations from other types of survey experiments, a two-stage incentive approach was used for the travel log component of the STC 15.

The invitation letter described the survey process and the incentive offer of a $20 check per household upon completion of the retrieval interview. This incentive was consistent with rates used successfully in the past. To further facilitate a higher retrieval response rate, a $1 bill was included in the travel log package. This "primer" incentive was used to encourage households to keep track of and report their travel. More information on the log-only methodology can be found in the Diary Survey Methods Memo in Appendix KK.

2.9 GPS-based Prompted Recall

During the recruitment survey, 40 percent of the sample was flagged as GPS eligible, and those respondents were asked if they were interested in participating in a GPS technology component of the study. Depending on the assigned travel day, the number of GPS households was capped to achieve an overall 10 percent GPS subsample.

For households that agreed to participate in this component of the study, GPS devices were sent to all household members between the ages of 16 and 75. The GPS participants were asked to carry these wearable devices for 3 consecutive days and then to report their travel details for the first of the 3 days in a GPS-based Prompted Recall (GPS+PR) survey. The travel date assignment was consistent with the non-GPS households, with the first of the 3 GPS travel days starting Monday through Thursday, and continuing for 2 additional consecutive days.

GPS households were offered an incentive check for $25 per GPS-instrumented person upon completion of the GPS+PR interview and travel reporting for non-GPS eligible household members. This amount was consistent with amounts offered in Westat's other wearable GPS studies. This approach provided an incentive that was commensurate with the level of burden required to participate in the study, and that was enough to motivate households to return the GPS devices. GPS households received only the GPS person-based incentive and not the household-level completion incentive offered to non-GPS households. More information on the GPS methodology can be found in the GPS Survey Methodology Memo in Appendix LL.

2.10 The Retrieval Survey Process for GPS Households

Once the devices were returned to Westat, the GPS data were downloaded from the devices, and the data files were uploaded to a secure project website and processed into place sequences, the chronologically enumerated list of destinations identified from the GPS logger data on the travel date. After the data had been processed, the household was notified by text, email, and/or phone that their retrieval survey was ready to be completed. Respondents were encouraged to complete the survey online, but they could have also completed by speaking with an interviewer.

Instead of reporting all the places visited on the travel day, as was done by non-GPS households, GPS-eligible respondents were presented with a survey interface that showed the travel traces from their GPS device and were asked to confirm the list of places visited and to provide details about each place. They were also asked to use their completed travel log as a reference. Travel details for non-GPS-eligible household members were entered based on their travel log information.

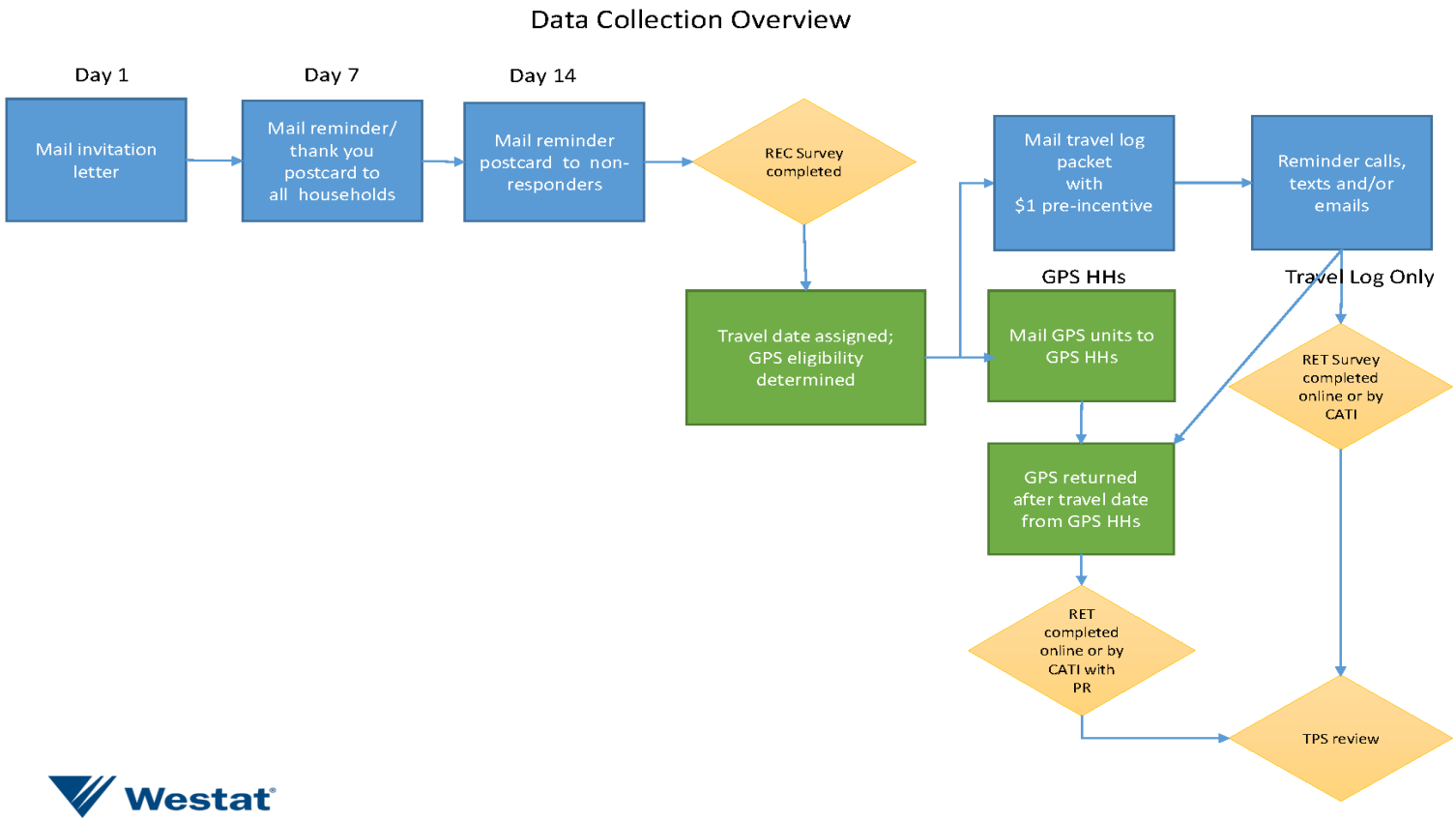
Westat processed the GPS data for days 2 and 3, identifying trip ends and trip attributes, and imputing travel modes and trip purposes. Westat used discrete choice modeling techniques for identifying trip purpose on days 2 and 3. This method included developing and estimating a multinomial logit (MNL) model for selecting trip purposes using the day one data and then applying it to the other 2 days. Because this model was estimated using day 1 data (which is a weekday), trip purpose was imputed only for GPS-based travel occurring on Tuesday through Thursday.

The purpose of the GPS+PR sample was twofold: (1) to calculate GPS correction factors for use with the data from the non-GPS sample, and (2) to generate calibration and validation datasets to improve the accuracy of the imputation algorithms and models used in generating trip details (such as trip purpose and travel mode) for the additional GPS days collected by the GPS+PR households. The GPS+PR design had the advantage of combining increased report accuracy and lower respondent burden compared to log-based self-report data collection methods, and it provided the ability to capture travel activities across multiple days. GPS respondent materials included GPS household letter, equipment usage log, device usage and return instructions, and the equipment retrieval letter. These materials were available in English, Spanish, and Arabic. These materials can be found in Appendixes MM through VV. The Trip Mode and Purpose Imputation Procedure Memo can be found in Appendix WW.

2.11 Data Collection Flow Chart

As part of the survey design, a data collection plan was developed. The overview can be found in Figure 2-2. The blue squares represent communication from Westat. The diamonds represent survey data. The green boxes are GPS related.

Figure 2-2. Data collection flow chart



Data Collection 3

This chapter presents the data collection schedule and process. A pilot study was conducted in early 2015. There was a spring data collection and after a summer break, data collection concluded in the fall of 2015. These are described here.

3.1 Schedule

The study was conducted in two phases. After a short pilot, there was a spring data collection, followed after a break in the summer by a fall data collection. The exact dates can be found in Table 3-1 below.

Table 3-1. Data collection schedule

Pilot	
First invitation letter	01/26/2015
First travel date	02/09/2015
Last travel date	03/04/2015
Retrieval closed	03/15/2015
Spring 2015	
First invitation letter	04/06/2015
First travel date	04/20/2015
Last travel date	06/04/2015
Retrieval closed	06/18/2015
Fall 2015	
First invitation letter	08/29/2015
First travel date	09/08/2015
Last travel date	11/24/2015
Retrieval closed	12/08/2015

3.2 Pilot Study

Work began on the design and development of the materials, instruments, software, websites, and public outreach efforts in October 2014, with the goal of delivering a simulation of the survey experience for a staff pretest in January 2015. Feedback from the staff pretest was then incorporated before a pilot survey was conducted in February 2015, in which 11,290 households were randomly

selected from the general public and invited to participate. The focus of the pilot survey was not to achieve a preset number of completed households; rather, the pilot study was undertaken as a means of testing the methods and systems developed in collaboration among staff from Westat, SEMCOG, and MDOT.

Specifically, these exercises were intended to identify weaknesses or shortcomings of the data collection tools, to seek out opportunities for enhancement of all processes, and to reveal any unanticipated procedural or data quality shortcomings. Therefore, the sample selection process did not use a proportionally representative approach but rather an even allocation of sampled households throughout the state. Households with targeted sociodemographic characteristics were also not included in the sampling for the pilot. Participation rates for the pilot were useful for assessing assumptions about recruitment and retrieval rates, and for adjusting the number of invited households for the spring data collection phase, but were not intended to evaluate the performance of the sample in specific two-way aggregate cells (e.g., household size by number of workers), or for comparison against American Community Survey (ACS) benchmarks.

Target and actual response rates of the overall pilot are summarized in Table 3-2, which shows that the actual rates for each category were better than anticipated. This resulted in a higher than expected number of completed pilot surveys, which in turn led to changes in the mailing strategy for the first phase of the main survey data collection.

Table 3-2. Sample performance rates for pilot; targeted versus actual

Rates	Targeted	Actual
Recruitment rate[1]	5.0%	7.5%
Participation rate[2]	62.0%	66.4%
Completion rate[3]	3.1%	5.0%
Completed surveys	350	560

[1] The percentage of households that joined the survey (based on the number of invitations mailed).

[2] The percentage of recruited households that reported travel.

[3] The product of the recruitment rate and the completion rate.

Improvements identified for the main data collection and survey processes such as re-wording questions, re-ordering the sequence for the long distance questions, and making changes to question skip logic within the instruments are detailed in the Pilot Report, which can be found in Appendix XX. The report contains a description of the changes, the survey element that was changed, and a note of explanation of each change.

3.3 Spring Data Collection

Following a successful pilot test, Westat fielded the STC 15 spring data collection. The first travel date of the spring data collection was April 20, 2015. The invitation letters were mailed by U.S. mail in four batches. The first letters were mailed on April 6, 2015, followed 7 days later by a reminder/thank you postcard to everyone. A second reminder postcard was sent 14 days after the initial invitation letter and 7 days after the first reminder postcard. This second postcard was sent to only those addresses that had not yet responded to the survey. The second batch of letters was sent on April 14, 2015. The third batch of letters was sent on April 21, 2015. The fourth batch of letters was sent on April 28, 2015 for a total of 87,375. Of those invited, 6,540 household completed the recruitment survey and were assigned a travel day. Of those who completed recruit, 4,113 households completed the retrieval survey.

The overall goal for the survey was to obtain a 10 percent subsample of complete GPS+PR households. There were no set targets at the sampling area or sponsor level. While the actual total recruit volume of non-GPS households exceeded the target for the spring, the GPS recruitment remained consistent throughout the data collection by setting the GPS household daily limit at 46. For informative purposes, the actual recruit values are included in Table 3-3.

Table 3-3. Spring recruitment results for GPS+PR households

Sponsor	Recruited
MDOT	923
SEMCOG	543
Overlap	168
Total	**1,298**

During the data quality control process, the GPS households that completed retrieval were reviewed. Of these, 20 percent were disqualified from delivery for various reasons, including cases where a participant changed the data in a fashion that violated basic speed checks by moving a GPS-based location away from the GPS-derived geocode, or cases where a participant ignored the times associated with a GPS-derived location and entered new arrival or departure times that violated basic speed checks. This resulted in a delivered rate of 58 percent, which is 2 percent lower than the anticipated rate of 60 percent. The shortfall from the spring was recovered in the fall. The Spring Data Collection Memo describes the results of the spring survey and the changes made for the fall based on these results. This memo can be found in Appendix YY.

3.4 Fall Data Collection

Westat used the response rates in the spring to predict the number of retrieval completes for the fall data collection. Westat increased the number of mailed letters in selected model areas and where it appeared to further decrease shortfalls and increase the oversampling of the area's high-density stratum. The size of the fall invitation letter mailing was 144,579 letters. This was implemented to reduce shortfalls for low-income households, three-person households, and households with four or more persons. Of those invited, 11,877 households completed the recruitment survey and were assigned a travel day. Of those who completed recruit, 8,281 households completed the retrieval survey. SEMCOG also implemented an in-person intercept approach of zero car and large households with an offer of an additional incentive. Strategies for improving GPS data quality from participants were added for the fall data collection and including prompts that warn participants when they are about to change the GPS data in a way that renders the results illogical and unusable.

Data Quality 4

This chapter presents the quality control plan, data processing and data cleaning throughout the study, and data checks, as well as logic checks that were built into the recruitment and retrieval instruments.

4.1 Quality Control Plan

Westat utilized the data coding and quality control manuals from the previous MI Travel Counts surveys and updated them to reflect changes in household travel survey methods and technologies that have occurred in recent years.

Quality control, by definition, includes a well-designed sampling plan, survey instruments, and respondent materials, as well as a carefully executed data collection operation. The designs for the STC 15sampling plan, survey instruments, and respondent materials were addressed in earlier chapters. The elements of the project management plan that were considered critical to quality assurance included the following:

- Effective involvement of SEMCOG staff to ensure survey outcomes meet the agency needs
- Monitoring the areas of design where biases and errors can occur
- A dedicated team of supervisors, trainers, and interviewers
- Establishing and maintaining a detailed project work schedule
- Developing and maintaining detailed data collection protocols
- Effective selection, training, and debriefing of interviewers, explaining the effort and their responsibilities
- Periodic monitoring of interviews and calls in the helpdesk, accompanied by continuous supervisor monitoring and feedback in the form of supplemental training
- Electronic tracking and monitoring of interviewers' performance—dialing statistics, completed interviews, refusals, non-contacts, and average interviewer lengths

- Implementing an appropriate public information effort
- Establishing measures to protect respondents' data privacy and ensure confidentiality of survey data
- Secure processing, storage, and eventual disposal of survey data, equipment, and materials

Quality control requires that each household sample element be individually tracked through its life cycle in the survey from sample release to completion or to final disposition. For this purpose, Westat applied its Survey Management System (SMS), which is specifically developed to handle multi-mode household travel surveys. This web-based sample management system provided the up-to-date status of each household sample element through all steps of the survey. The system's built-in call scheduler and disposition monitoring generated continuous information to ensure that each household received appropriate attention so that quick remedial action could be taken as needed. Furthermore, the SMS system was tightly integrated with the survey's secure public website, its online and CATI instruments, the automated reminder program that sent emails and text messages to households to help them along the data collection process, and the continuous automated checking system. This integration made it possible for Westat to develop reports that were continuously updated and always reflected the current state of the survey as a whole.

4.2 Data Checks

Westat's household travel survey data collection process included several layers of automated data checks. The first set of checks was included in the online instruments and was followed by automated checks performed nightly on all households that completed the retrieval interview. The final round of checks happened once data delivery files were extracted from the project's database.

Data processing and data cleaning were conducted on an ongoing basis throughout the study. Updates were made to variables that affected data collection during the administration of survey (e.g., the removal of a household member who was originally reported in the recruit interview, but no longer lives at the household location) and at the conclusion of data collection for data that did not affect the flow of the survey (e.g., recoding trip purposes based on "Other, specify" responses).

A series of automated edits, range checks, and consistency checks were performed within the survey instrument, and data preparation staff performed frequency reviews and problem resolution to

monitor, correct, and update the data and online programs as needed. Automated checks were run to evaluate the validity of reported trip data and household location data.

Frequency reviews were conducted at each data delivery to ensure that all data were being properly captured in the survey database. Additional reviews were performed following any changes to the survey instrument after the start of the survey.

4.2.1 Logic Checks Built into Online Instruments

Logic checks were programmed into the recruit and retrieval instruments to ensure that questions were answered as accurately as possible. These included requiring that key questions be answered, even if the answer is "don't know" or "prefer not to answer," and forcing the data type (e.g., requiring a number for numeric questions, U.S.-formatted telephone numbers for phone questions, etc.).

Data range checks were conducted to guarantee that the data fell within the expected range for a given question (e.g., 0–110 for AGE). Skip logic was programmed into the instruments such that questions that did not apply were not asked (e.g., participants were asked if they have a driver's license only if they were at least 16 years old). Specific information on these built-in checks and skip logic features can be found in the instrument design documents. In addition, the following checks were built into the online instruments:

- If a household member reports being employed, the number of jobs reported must be at least one
- The household respondent (person 1) cannot report that they are under 18 years of age and complete recruitment
  - Must confirm they are at least 18 years of age at the beginning of recruitment
  - Cannot enter an age or age category under 18 years of age
- If an age is reported as zero, a follow-up question confirms that the person is an infant
- Households are required to provide details on as many members as they report in their household size; after they are done rostering the household, they are asked if there are any other household members that they need to report

- The household must confirm their home address (or P.O. box) matches the sampled address
- Number of household members on a trip is not asked of one-person households
- Households that report zero vehicles are not asked if a household vehicle was used on a trip

Westat's online travel data retrieval tool, TBW, integrated several data consistency checks within its user interface (UI). This ensured that collected places passed basic data completeness and consistency requirements. One of these checks included the requirement that all visited places be associated with a geocoded location. That way, all records deemed complete at the end of the retrieval instrument will have all destinations geocoded. Other checks built into the TBW UI include:

- All places must have valid arrival and departure times (i.e., times must increment)
- All places after the first one must have a travel mode associated with them
- First place of the day must have an arrival time equal to 3:00 am
- Last place of the day must have a departure time equal to 2:59 am
- If the trip to a place took longer than expected (based on captured distance and minimum mode speed), a follow-up question asked for a reason
- If a participant reported a single place on their travel date, a follow-up question asked for the reason why no travel was reported

In addition, the following checks were performed on the captured long-distance trips:

- The start date must be before and within 3 months of the current date (i.e., when the survey is collected)
- Must have an end date that is the same or is after its start date
- A geocode for the destination is captured
- The 3-month and 12-month frequency questions trigger confirmation questions if reported at values higher than 30 and 100, respectively

4.2.2 Automated Edit Checks

Travel data were processed through Westat's trip processing system (TPS). TPS included a series of consistency checks on reported travel data. When a TPS edit check failed, an analyst reviewed the

data to determine whether adjustments to the data could be made based on information provided by another household member or if the household needed to be re-contacted to resolve the inconsistency in the data. Whether the data were updated by an analyst or an interviewer as a result of a re-contact with the household, the entire household record was reprocessed through the TPS checks. Each case was subjected to this process until it cleared TPS without any failures. Only households that successfully passed these edits were considered as complete in the final dataset

4.2.3 Upcoding and Cleaning

At the conclusion of the data collection period, open-ended and "other, specify" responses were reviewed and upcoded or collapsed as appropriate. The upcoding of responses was the activity of recoding an open-ended response into a categorical response option (e.g., recoding "slept" to "typical home activities"). The process included removing the "other, specify" (open-ended) text response.

Westat also combined or collapsed other responses that were similar to each other. These responses appeared in the original dataset as independent responses ("one-offs") due to a misspelling or different letter spacing in the response or capitalization issues. Combining these text responses allowed for more efficient analysis.

4.3 Item Nonresponse

The Westat data quality team evaluated the levels of item nonresponse for key variables (e.g., age, income, employment status, trip purpose, etc.) at the household and person level to ensure data completeness. Relevant variables were compiled and calculated for review by the project management team. These variables included the total number of items asked, the number of items answered and refused, the resultant item nonresponse rate, and an enumerated list of refused items by household and household member. Using these data, the project management team defined thresholds to be applied to the dataset that were used to disqualify households from delivery as complete households.

4.4 Geocoding Methods and Quality Checks

Home locations were initially geocoded as part of the address-based sample and were updated during the online interview, with the exception of P.O. box sample addresses.[1] These locations, along with their geocoded coordinates, were then transferred to the retrieval instrument once the household completed the recruit interview.

The online instruments collected place data for frequently visited locations, referred to as habitual locations, in the recruitment survey, and there were integrated advanced online ("on-the-fly") geocoding and points of interest (POI) resources in the WGS platform and the TBW travel data retrieval tool. These resources made it possible to geocode person-level habitual locations (e.g., work and school) as part of the recruit interview. The recruit instrument required that a geocode was collected for all habitual locations. The long-distance component of the retrieval instrument also required that a geocode of the destination be provided; wording on the question indicated to participants that only a city-level location was expected to be provided.

The web-based UI used in TBW to retrieve travel data allowed participants to select a habitual location from a list or to enter address information using free-form text. The input fields were capable of processing place name (e.g., Starbucks), street address, or intersection information to assist the respondent in providing the address of the trip destination.

Only after an initial search was done were the inputs sent to the Google Maps geocoding and Google Places Search application-programming interfaces (APIs). The address data were used to search for geocode matches, while the place name was used to search for POI locations around the provided address or map center if no address is provided. A list of potential matches was then displayed to the participant in a list box and added to the electronic map using "pushpins."

Once a result was selected, the geocoded coordinates were saved using longitude and latitude values in decimal degrees (WGS84 datum) along with the address components, the place name (in the case of POI results, the place name returned by the search was saved), and a geocode type. In addition to these search capabilities, the geocoding UI allowed participants to click on the map to set the geocode coordinates, or to drag and drop a previously geocoded result; in these cases, a reverse-geocode lookup using the updated coordinates was performed using the Google Maps API. This ensured that address information was derived for the manually adjusted coordinates.

[1] If a P.O. box was sampled, the physical home address was captured in the recruitment survey.

When returning coordinates for address information, Google APIs automatically offset coordinates from the street centerline; in regions where Google had access to parcel data, it geocoded coordinates placed on the center of the parcel. Westat automated a process that flagged geocodes that fell closer than 25 feet from the MI Geographic Framework road network for adjustment. Westat consulted with MDOT to ensure that the latest version of the road network was incorporated into this process.

Westat performed speed checks on the collected geocodes and travel data as part of the automated review process completed through TPS. These checks detected geocoded locations that were not consistent with the reported times and travel modes. The distance used to perform these checks was derived from shortest path routes that were computed by Google Maps and integrated into the TBW UI, and the selected travel mode was used to tell Google Maps whether to compute an auto, bike, or walk route (transit modes were treated as auto for the purpose of route generation). This capability removed the need to conduct further processing on a GIS tool such as TransCAD to derive network distances.

Within TPS, each travel mode had a maximum speed that was compared against the average trip to place speed, computed using the shortest network path distance and the travel time derived from the provided arrival and departure times (Table 4-1). Places that had computed average speeds above the maximum travel mode value, or below 5 mph for motorized modes or 1 mph for non-motorized modes, were flagged for analyst review.

When a check failed, analysts reviewed the geocoded results and place name and address information and used it to identify a possible alternate match using Internet search resources such as Google, Bing, and Yahoo! In cases where additional information was needed to address a data inconsistency, the household record was sent to "research." In the research status, cases were re-contacted by a specially trained member of the interviewing team to address the inconsistencies and seek a resolution from the household member. During the research call, interviewers had the ability to perform those same speed checks to verify the quality of the recollected geocode or adjusted time and travel mode information.

Table 4-1. Travel modes and maximum speed values

Mode ID	Name	Max speed (mph)
1	Walk	13
2	Bicycle	20

3	Motorcycle/moped	95
4	Auto/ van/truck (as the driver)	95
5	Auto/van/truck (as the passenger)	95
6	Carpool/vanpool	95
7	School bus	50
8	Public transit/local bus	50
9	Dial-a-ride/paratransit	50
10	Private bus or shuttle	50
11	Taxi/limo	95
12	Train/Amtrak	95
13	Detroit People Mover	50
14	Airplane	600
15	Boat/ferry	50

Finally, all collected geocodes were checked for completeness of address attributes (i.e., address or intersection, city, state, and zip code). Whenever necessary, reverse geocoding using Google Maps and Bing was used to fill in missing address components.

4.5 GPS Processing and Checks

The GPS data collected by the participants were imported into TBW, and the UTC (Universal Time Coordinate) date and time stamps in the GPS point data were translated to local date and time. The wearable GPS data loggers were programmed with speed filter settings that screened out all points with zero values for speed, and points with speeds greater than zero were recorded using a three-second interval. The initial process of creating a sequence of places based on the input stream of GPS points is summarized in the following steps:

- Filter out GPS data based on the assigned travel date range.
- Create first place of the day so that the display can be place-based; the arrival time of this place is pre-set to 3:00 am and its departure time is set to the start time of the first identified GPS trip; the coordinates for this place are also set based on the first identified trip's origin.
- Split GPS points into trips leading to places using a 120-second dwell time criteria.
- Look for mode transitions within each detected trip (non-motorized to motorized and vice-versa) and further split trips based on these detected mode transitions.
- Determine if each identified trip can be attributed to real movement or GPS noise using heuristics-based criteria.

- Check resulting trips to places against minimum travel time and distance constraints (distance covered ≥ 100 meters and travel time ≥ 1 minute); if a trip fails to meet these constraints, its points are aggregated onto the next place.
- Compare each place's destination point against the household's geocoded habitual locations and associate trip ends with location if the distance between them is less than 75 meters.

Once initial GPS processing in TBW was completed, the resulting place sequences were visually reviewed by analysts to screen out traffic delays and other falsely identified stops with dwell times of 120 seconds or more, and to add stops with short dwell times not identified in the initial processing but clear stop characteristics—for example, a brief stop (under 120 seconds) at a post office/mailbox or a fast food drive-thru. These data were pushed to the TBW and the participating household was contacted to complete the prompted recall survey online (or by telephone).

Following the prompted recall survey completion, the GPS-derived places for all other days were processed and reviewed according to the following criteria:

- Review places with durations greater than 60 minutes or shorter than 2 minutes.
- Review places with distances greater than 60 miles or shorter than 0.25 miles.
- Ensure that travel mode is present for each place greater than place 1.
- Trip passes all speed checks.
- Maximum point speed is not greater than 85 mph.
- Average speed is not greater than maximum point speed.
- Review places with a gap distance greater than 0.5 miles (gap distance is the distance between the previous destination and the first GPS point on the trip to the current destination).
- Review GPS places that are outside the travel date range.
- Review last GPS place of each day if it does not end at participant's home.

More information can be found in the Data Quality Control and Geocoding Procedures Memo in Appendix ZZ.

4.6 Weekly Reports

During the two data collection periods, Westat produced weekly reports used in managing the ongoing survey operations. See Appendix AAA for the final weekly report. The weekly reports indicated adherence to, or departures from, the sampling plan in the areas of; the final weekly report shows the actual and target number of completed households by sample area and household size and income; number of completed cases by sample area, substrata, and post-strata (subpopulations to which completed cases are assigned following data collection); recruitment rates by sample area and substrata; retrieval rates by sample area, substrata, and post-strata; and achievement rates for the sample areas and post-strata, where an achievement rate is the ratio of the number of completed cases in a cell to the associated data-collection goal for the cell.

4.7 Data Deliveries

Westat delivered a series of unweighted interim datasets after the pilot, spring, and fall data collections and a final weighted dataset after the fall data collection ended. The delivered files included household-level, person-level, location, place, and long-distance files and updated codebooks and GPS data. The interim deliveries were reviewed by SEMCOG to identify any problems. Problems identified during reviews were addressed in subsequent deliveries. Each delivery included the cumulative data retrieved and checked to that point. The data delivery observations in a table of changes can be found in Appendix BBB. A schedule of the data deliveries is listed in Table 4-2.

Table 4-2. Data Deliveries

Data Deliveries	
Spring interim (unweighted)	07/24/2015
Spring final (unweighted)	09/18/2015
Fall interim (unweighted)	11/20/2015
Fall draft final (unweighted)	01/25/2016
Final data (unweighted)	03/04/2016
Final data (weighted)	03/22/2016

5 Assessment of Survey Targets

This chapter summarizes how survey targets were met and provides an assessment of both recruit and retrieval goals versus actuals, and overall response rate.

5.1 Sampling

The sample design for STC 15 assumed a recruitment rate of 6 percent and a retrieval rate of 65 percent for an overall participation rate of 3.9 percent. These rates would require 239,231 invitations to recruit 14,354 households and retrieve 9,330 households. This section provides details of the actual response rates achieved in the data collection effort for each sample area and for the sample overall.

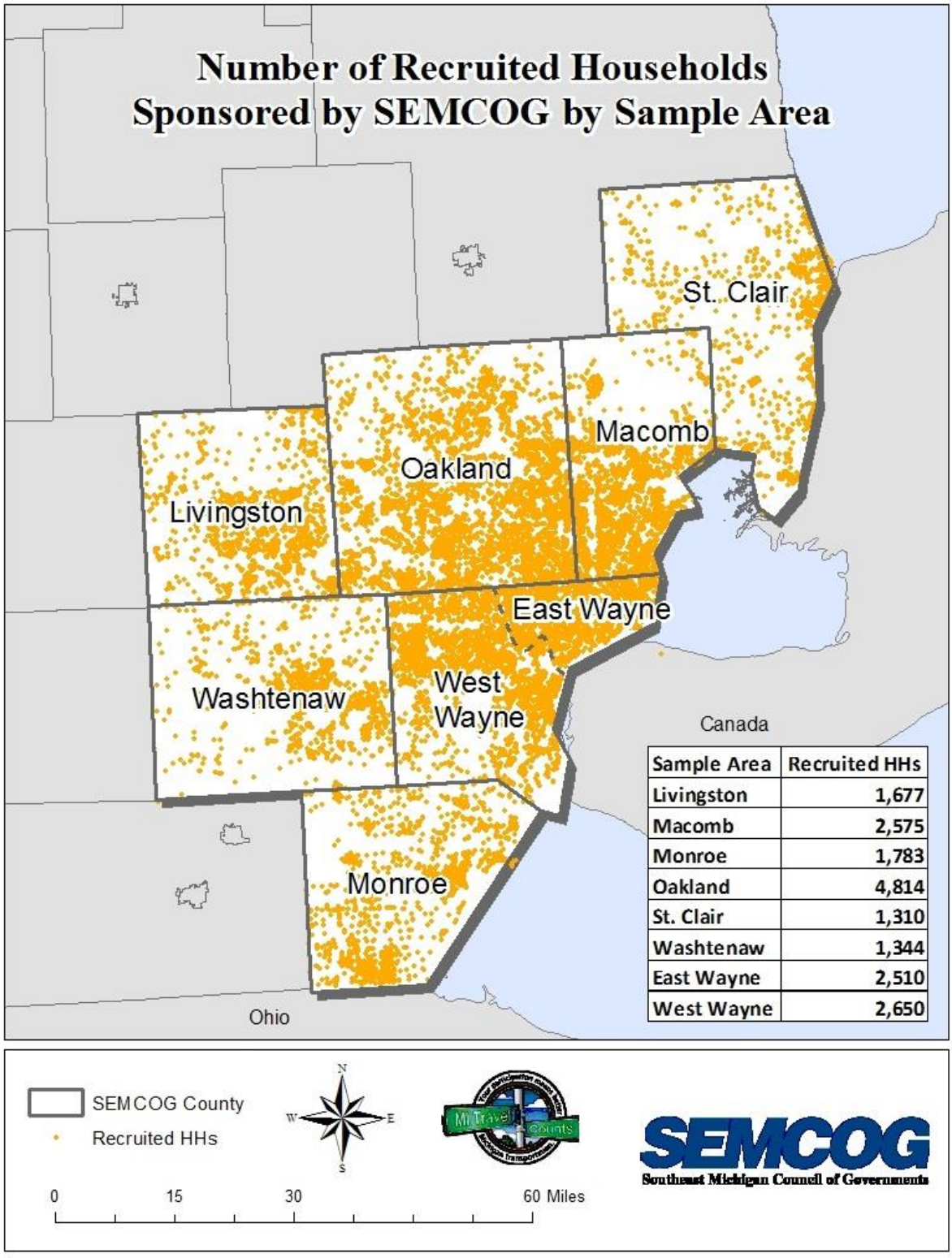
5.2 Recruit Goals Versus Actual by Sample Area

Table 5-1 shows the samples invited, actual recruits and targets, alongside the recruitment rates that were observed in data collection for each model area and for the overall sample. Excluding the intercept sample, the recruitment rate ranges from 6.7 percent in the East Wayne County to 10.7 percent in Livingston. On average, the model areas achieved a 7.9 percent recruitment rate. This necessitated a decrease in the volume of mailed invitations for the fall phase of data collection. Figure 5-1 shows a map of recruited households by sample area.

Table 5-1. Recruitment rates by model area

Sample area	Released	PND	Actual	Target	Percent	Rate
East Wayne	37,379	1,073	2,486	1,954	127.2%	6.7%
West Wayne	36,137	559	2,610	2,554	102.2%	7.2%
Oakland	56,286	953	4,785	3,354	142.7%	8.5%
Macomb	36,728	595	2,545	2,262	112.5%	6.9%
Washtenaw	12,590	334	1,310	1,231	106.4%	10.4%
Monroe	21,794	361	1,757	1,000	175.7%	8.1%
St. Clair	15,255	337	1,280	1,000	128.0%	8.4%
Livingston	15,285	169	1,637	1,000	163.7%	10.7%
Intercept	500	0	7	0	0.0%	1.4%
Total	**231,954**	**4,381**	**18,417**	**14,354**	**128.0%**	**7.9%**

Figure 5-1. Map of number of recruited households by sample area



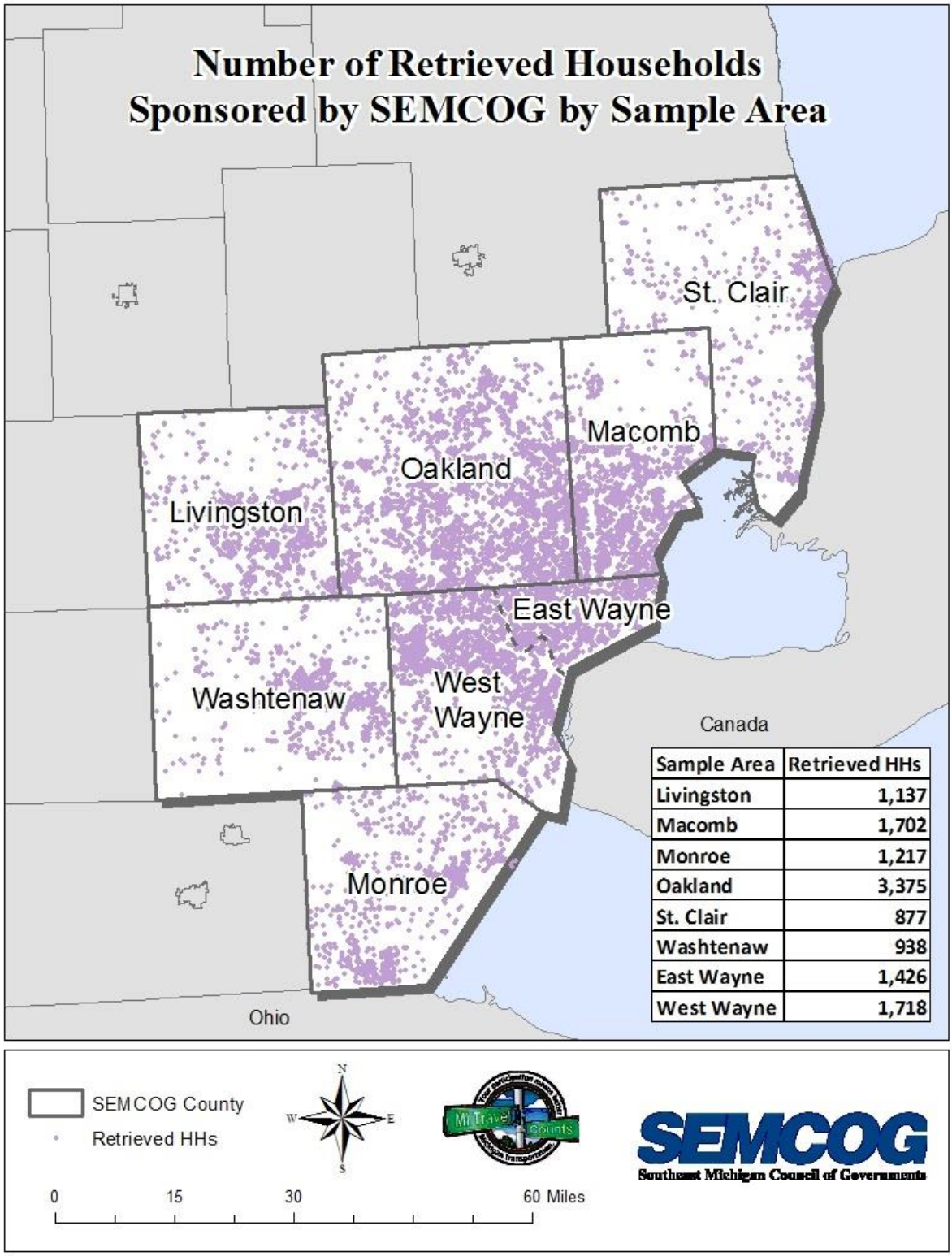
5.3 Retrieval Goals Versus Actual by Sample Area

Table 5-2 shows the retrieval results for the STC 15 sample areas. The average retrieval rate of 67.3 percent was slightly better than the anticipated rate of 65 percent. Figure 5-2 is a map of the number of retrieved households by sample area.

Table 5-2. Retrieval and participation rates by model area

Sample area	Actual	Target	Percent	Rate	Participation rate
East Wayne	1,426	1,270	112.3%	57.4%	3.8%
West Wayne	1,718	1,660	103.5%	65.8%	4.8%
Oakland	3,375	2,180	154.8%	70.5%	6.0%
Macomb	1,702	1,470	115.8%	66.9%	4.6%
Washtenaw	938	800	117.3%	71.6%	7.5%
Monroe	1,217	650	187.2%	69.3%	5.6%
St. Clair	877	650	134.9%	68.5%	5.7%
Livingston	1,137	650	174.9%	69.5%	7.4%
Intercept	4	0	0.0%	57.1%	0.8%
Total	**12,394**	**9,330**	**133.0%**	**67.3%**	**5.3%**

Figure 5-2. Map of number of retrieved households by sample area



5.4 Overall Response Rate

The combined increase in the recruitment and retrieval rates led to a final delivery of 12,394 households or 133% of the goal for the project. This translated into an overall participation rate of 5.34%. Independently, the AAPOR standard for calculating the minimum response rate was used (American Association for Public Opinion Research, 2016). In particular, Response Rate 3 (RR3) was selected as it "estimates what proportion of cases of unknown eligibility is actually eligible." Given the frame of the sample is the USPS distribution file (which is updated monthly), our assumption is, postal non-deliverable (PND) addresses are not considered to be known ineligible. Therefore, only a small fraction of cases with known ineligibility can be used to calculate the proportion of eligible households for e.

$$RR3 = \frac{I}{(I+P)+(R+NC+O)+e(UH+UO)}$$

Where:

Variable	Description	Count
RR3	Response rate	–
I	Complete interview	12,394
P	Partial interview	9,227
R	Refusal and break-off	172
NC	Non-contact	0[1]
O	Other	1,387
UH	Unknown if household/occupied HU	4,334
UO	Unknown, other	205,430
–	Known ineligible	8
e	Proportion of known eligibility	99.997%

[1] It is presumed that all households sampled and released in the survey received a mailing unless the materials were returned as a Postal Non-deliverable (PND).

Using these parameters, the STC 15 survey final response rate is 5.32 percent. Note again that this nearly the same as the calculated participation rate due to the fact that the known ineligible share of contacted households is only .003 percent, meaning e is effectively 1.

The overall goal for the survey is to obtain a 10 percent subsample of complete. There were no set targets at the sampling area or sponsor level. Table 5-3 presents the recruited, retrieved and delivered results for GPS+PR households as well as the retrieval rate and percentage of overall GPS target.

Table 5-3. Recruited, retrieved and delivered results for GPS+PR households

Sponsor	Recruited	Retrieved	Delivered	Retrieval rate	Percent of target
MDOT	2,130	1,368	1,325	62.2%	–
SEMCOG	1,705	1,137	1,106	64.9%	–
Overlap	393	266	258	65.6%	–
Total	**3,442**	**2,239**	**2,173**	**63.1%**	**103%**

Data Weighting 6

This chapter discusses types of weights, calculation of weights and adjustment for nonresponse on the STC 15 data collection.

6.1 Types of Weights

When different strata have different sampling rates or different response rates, it is necessary to use weights[2] when analyzing the collected data. If this is not done, then areas with higher sampling rates or higher response rates will be overrepresented relative to areas with lower rates.

Because survey data can be reported at different levels—for example, household level or person level—weights are often labeled with their associated reporting unit. For example, *household weights* are used to compute estimate from data reported at the household level, whereas *person weights* are used to compute estimates from data reported at the person level.

For the SEMCOG-funded samples, Westat delivered separate files of household weights and person weights for the retrieval completes in the SEMCOG area. The file of household weights contained a full-sample household weight and an associated set of 100 replicate household weights for each household in the SEMCOG area that completed a recruitment interview. The file of person weights contained a full-sample person weight and an associated set of 100 replicate person weights for each eligible person in a household in the SEMCOG area that completed a recruitment interview. Each file will contain weights for both MDOT- and SEMCOG-funded samples. Analysts will be able to use the provided full-sample weights to compute a weighted estimate, denoted $\hat{\theta}$, of a population characteristic, denoted θ, and will be able to use the 100 replicate weights to compute the sampling variance of $\hat{\theta}$, using the following formula:

$$v(\hat{\theta}) = \sum_{g=1}^{100} \left(\hat{\theta}_{(g)} - \hat{\theta} \right)^2 ,$$

[2] Also referred to as *household expansion factors*.

where $\hat{\theta}_{(g)}$ is the estimate of θ computed by using the replicate weights from the g^{th} replicate sample. Standard errors of estimates can be computed by taking the square root of the sampling variances, and confidence intervals can be computed for estimates from the associated standard errors.

The above formula is implemented in a number of statistical analysis packages capable of analyzing data obtained from surveys with complex survey designs; for example, SAS, SUDAAN, STATA, and WESVAR.

Weights (both full sample weights and replicate weights) were developed in a series of steps to compensate for unequal selection probabilities, nonresponse, non-coverage, and sampling fluctuations from known population totals. The steps in the weighting process for the samples for the SEMCOG area will be the following:

- Construction of household base weights (the reciprocal of the probability of selection of each sampled address) for the samples in the SEMCOG area;
- Adjustment for nonresponse at the household level, yielding nonresponse-adjusted household weights, for the MDOT- and SEMCOG-funded samples;
- Creation of composite nonresponse-adjusted household weights for the recruitment completes in the SEMCOG area;
- Further adjustment of the composite nonresponse-adjusted household weights to achieve consistency with characteristics for the universe of households (achieved by *poststratifying* or *raking* the composite nonresponse-adjusted household weights to independent household-level totals, referred to as *control totals*);
- Calculation of initial person weights, which will be the composite nonresponse-adjusted household weight applied to each eligible person in a responding household in the SEMCOG area; and
- Further adjustment of the initial person weights to achieve consistency with characteristics for the universe of eligible person (achieved by *poststratifying* or raking the initial person-level weights to independent person-level control totals).

The above six steps were used to develop both full-sample weights and replicate weights. For the full-sample weights, each step was performed using all respondents in the full sample for a particular area. For the 100 replicate samples, each step was performed 100 times, using each time the respondents in a different replicate sample for a particular area.

6.2 Calculation of Base Weights

The household base weight reflects the probability of selection for a sampled household and is calculated simply as the reciprocal of its probability of selection. That two samples were selected—a spring sample and a fall sample—complicates the calculation of the household base weights, however, because MSG updated their address fame several times between the selection of the spring and fall samples. If only one sample had been selected, the following formula would be used to compute the base weight for each substratum within each stratum:

$$\text{Base weight} = N_h / n_h,$$

where

N_h is the number of addresses in MSG's sampling frame for substratum h, and
n_h is the number of addresses sampled from substratum h.

Because two samples were selected, however, the denominator of the household base weight becomes the sum of the unduplicated number of addresses selected in either the spring or fall. For the numerator, some surveys would use the stratum's average of the spring and fall frame sizes. Westat used the fall sample's frame size for the numerator of the household base weight because the date of the address frame used to select the fall sample was closer to the middle of 2015 than that of the address frame used to select the spring sample and should thus better represent the universe of addresses in 2015. Household base weights were calculated for all sampled addresses, not just for respondents. Table 6-1 below contains the calculated base weights by strata and substrata for the SEMCOG samples.

Table 6-1. Base weights for SEMCOG samples

Sample area	Base weight by sub-stratum	
	Low density	High density
Eastern Wayne	24.458	4.054
Western Wayne	23.768	7.483
Oakland	29.899	4.758
Macomb	35.239	5.026
Monroe	6.710	1.570
St. Clair	14.412	2.122
Livingston	14.822	2.407

6.3 Adjustments for Household-level Nonresponse

At the end of data collection, each sampled address was assigned a disposition code, which for weighting purposes can be assigned to one of the following four categories:

- *Undeliverable address*;
- *Eligible respondent*, i.e., a retrieval complete;
- *Other-ineligible case* (i.e., other than being an undeliverable address), e.g., a business address instead of a residential address; and
- *Nonrespondent*, e.g., (1) a recruitment refusal, (2) a recruitment complete that fails to complete a retrieval interview, or (3) a nonrespondent to the recruitment interview, with no evidence that the address was undeliverable.

The household base weights of the *eligible respondents* were multiplied by a nonresponse adjustment factor to compensate for the non-completion of recruitment interviews by the *nonrespondents*. The base weights for the *other-ineligible cases* will also be adjusted to reflect that some *nonrespondents* may be ineligible. The base weights for the *undeliverable addresses* were not adjusted, and the adjusted base weights for the *nonrespondents* were set to zero.

The calculation of the nonresponse adjustment factors required that each sampled address that was not an undeliverable address be assigned to a set of sampled addresses, called a *nonresponse adjustment cell*. Sampled addresses assigned to the same nonresponse adjustment cell should have similar response rates. Both respondents and nonrespondents are assigned to nonresponse adjustment cells, so the assignment process must be based on information available for both respondents and nonrespondents. For the MDOT-funded samples in the WATS area and in the SEMCOG minus WATS area, we used the four sampling substrata (the high- or low-density substrata within WATS or non-WATS) to define the nonresponse adjustment cells. For the SEMCOG-funded samples, we used the 14 sampling substrata to define the nonresponse-adjustment samples. Additional nonresponse adjustment cells based on survey wave (i.e., spring or fall) and match status (i.e., whether or not an address was matched to a landline telephone) were created within the cells based on substrata. We evaluated the suggestion by Norman and Sigman[3] to use the type of delivery point (i.e., city delivery, rural delivery, P.O. boxes, and highway contact) to define nonresponse adjustment

[3] Norman, G. and Sigman, R. (2007). Using addresses as sampling units in the 2007 Health Information National Trends Survey, Proceedings of the Survey Research Section, American Statistical Association, Alexandria, VA.

cells, but because of resulting small differences in response rates among the resulting cells, we did not use the type of delivery point to define nonresponse adjustment cells.

After the nonresponse adjustment cells were created, the following formula was used to compute household nonresponse adjustment factors for each cell:

$$f_c^{(1)} = \frac{\sum_{i \in ER_c} w_i + \sum_{i \in OI_c} w_i + \sum_{i \in NR_c} w_i}{\sum_{i \in ER_c} w_i + \sum_{i \in OI_c} w_i},$$

where,

$\sum_{i \in ER_c} w_i$ is the sum of base weights for *eligible respondents* in weighting cell *c,*

$\sum_{i \in OI_c} w_i$ is the sum of base weights for *other-ineligible cases* in weighting cell *c,* and

$\sum_{i \in NR_c} w_i$ is the sum of base weights for *nonrespondents* in weighting class *c*

6.4 Calculation of Composite Household Adjusted Base Weights

In the six counties of the SEMCOG area other than Washtenaw County, there were MDOT-funded completes as well as SEMCOG-funded completes. Following nonresponse adjustment of the household base weights, the sum of these weights for the MDOT-funded retrieval completes in a county were approximately equal to the number of households in the county, and the sum of the adjusted household base weights for the SEMCOG-funded retrieval completes in a county were also approximately equal to the number of households in the county. Consequently, if the nonresponse adjusted base weights were used to compute weighted totals from all of MDOT- and SEMCOG-funded retrieval completes in the SEMCOG minus WATS area, the resulting estimated totals were approximately twice the value they should be. The reason for this is that in the SEMCOG minus WATS area, the MDOT-funded samples provide estimates of population totals and the SEMCOG-funded samples also provide estimates of the population totals. To prevent overestimation of

population totals when combining MDOT-funded samples and SEMCOG-funded samples, a household composite weight will be calculated using the following formula:

$$\text{composite household weight} = \alpha \text{ (MDOT-funded adjusted household base weight)} + (1 - \alpha) \text{ (SEMCOG-funded household base weight)}$$

where

α, the *compositing factor*, is computed as follows:

$$\alpha = r(\text{MDOT}) / [\, r(\text{MDOT}) + r(\text{SEMCOG}) \,],$$

$r(\text{MDOT})$ = number of MDOT-funded retrieval completes in the county, and

$r(\text{SEMCOG})$ = number of SEMCOG-funded retrieval completes in the count).

A given retrieval complete has only one adjusted household base weight. Within the SEMCOG sample areas other than Washtenaw County, however, some of the retrieval completes have a MDOT-funded adjusted household base weight and other retrieval completes have a SEMCOG-funded adjusted household base weight. Hence, the above formula for the composite household adjusted weight has only one positive term on the right-hand side—for MDOT-funded retrieval completes, it is the first term, whereas for SEMCOG-funded completes, it is the second term.

GPS Correction Factors 7

This chapter discusses the GPS correction factor plan as well as applying trip correction factors.

7.1 Correction Factor Plan

To account for possible under-reporting of travel in MTC III, 10 percent of the recruited households were instrumented with person-based GPS data loggers. This captured all of the actual travel movements of respondents on a given travel day. Consequently, these data were used to assess the magnitude and pattern of under-reporting of travel in the larger, non-GPS main survey sample. They were also used to estimate correction factors that can be applied to more fully account for full sample travel. In other words, the GPS correction factor analysis attempted to identify the factors that significantly affect trip under-reporting so that the resulting information can be used to derive a set of weights (i.e., correction factors) for more accurate estimation of household trip rates.

It is widely acknowledged that respondents, when self-reporting or reporting for others in their household, typically under-report trips in household travel surveys. One common method to measure and correct for trip under-reporting is to deploy GPS data loggers to a sample of households and compare the self-reported trips to those collected passively by the GPS loggers. Analysis is conducted to identify circumstances that lead to under-reporting, and correction factors are calculated. Because there is substantial variation that occurs in under-reporting, some households may not require a correction factor, whereas others will require a fairly large weight (see Zmud & Wolf, 2003); applying a single correction factor to all log-reported travel is not appropriate. There are several factors or conditions that can contribute to trip under-reporting. For example, trip length has been shown to be correlated with trip under-reporting. Trips of short duration are often missing from respondent logs more frequently than trips of long durations (Zmud & Wolf, 2003). Variables examined in the past have included trip duration, household size, reported vehicle ownership, household income, respondent age, employment status, student status, and presence of children under 18 in the household.

As noted above, the traditional travel correction factor generation process applies to study designs where GPS and survey data are collected concurrently and independently in what is often referred to

as the dual-method design. This approach does not apply to the STC 15 prompted recall design, in which the GPS data serve as a basis for travel reporting and confirmation, but we believe it is helpful to explain the traditional approach Westat used in previous studies so that differences in the two approaches are known and understood.

Under the dual-method design, the process for developing travel correction factors begins with a database of GPS trip records used to test a model of trip misreporting. A dummy variable is used to indicate if a trip record was "missing" when compared to the GPS data. A logistic regression model is then used to determine which of the variables associated with the trip record (e.g., household size, household income, employment status, age, etc.) has a statistically significant effect on travel under-reporting. Based on the logistic regression analysis, adjustment factors are developed for specific combinations of household, person, and trip types. These factors are then applied to all households, allowing for more accurate estimation of household trip rates for a wider population.

The process for developing correction factors for STC 15 is notably different from the traditional trip correction process because of the project's use of the GPS prompted recall (GPS-PR) design. Instead of comparing two "streams" of data from the same sub-sample of households to identify underreported travel, this process involves a comparison of GPS-captured travel made and confirmed by GPS households with the reported travel behavior equivalents from non-GPS households. The thought is that the participants using the GPS-PR retrieval method are more likely to capture all travel movements on the travel day. Under this assumption, data from the GPS households are used to identify and quantify travel under-reporting in non-GPS households, and to estimate correction factors that can be applied to more fully account for the travel of the full sample.

7.2 Development Process

Westat structured the GPS correction derivation process around tour and stop frequencies. However, Westat did not use tour generation models to determine statistical significance of travel under-reporting in the non-GPS portion of the sample. Instead of following this approach, a process tailored to the MTC III and STC 15 was developed to compute travel correction factors (TCFs). It included the following general steps:

- Prepare tour data files that translate place-based survey database to tours, as well as linked trip files; these allow the estimation of tour and tour stop frequencies.

- Identify tour segmentations (e.g., tour type, person type, and tour mode combinations) for which statistically significant tour frequency distribution differences exist between GPS and non-GPS groups.
- Use a similar exploratory approach to identify tour stop frequency segmentations (e.g., by tour mode and stop counts) that show under-reporting in the non-GPS group.
- Utilize person-level weights to expand the GPS and non-GPS tour frequencies for the segmentations found to have significant under-reporting and compute tour frequency TCFs (i.e., compute weighted tour frequencies using the person level weights for GPS and non-GPS and use them to calculate a ratio).
- Compute tour stop frequency TCFs by tabulating tours using person weights multiplied by the tour frequency travel correction factors derived in step 4 (i.e., compute weighted tour frequencies classified by number of stops using the product of person level weights and tour frequency TCFs for GPS and non-GPS groups and use them to calculate a ratio).

7.3 Applying Corrections

The computed TCFs should be used whenever summarizing travel attributes, such as traveled distance in conjunction with existing household or person weights. For example, to compute average trip distance by person, one should multiply individual place distances by the product of the person weights and the place TCFs, sum the resulting values, and then divide by the sum of the product of the person weights and TCFs, as noted in the following equation:

$$\bar{d} = \frac{\sum d \cdot w_p \cdot TCF_{freq} \cdot TCF_{stop}}{\sum w_p \cdot TCF_{freq} \cdot TCF_{stop}}$$

where:

d : trip distance,

w_p : person level weight,

TCF_{freq} : tour frequency travel correction factor, and

TCF_{stop} : stop frequency travel correction factor.

More details can be found in the Data Integration and GPS Correction Factors Plan located in Appendix EEE.

Lessons Learned and Future Recommendations 8

This chapter looks at a few lessons learned and recommendations for future data collections.

The data collected in 2015 during STC 15 constitute a rich resource for examining travel behavior across all of Michigan. The dataset gives insight about the travel and activities for 12,394 households while providing proportional coverage of this geographically and demographically diverse region. The experience of the sample design, data collection, and analysis of the resulting data revealed some lessons that can be applied to future survey projects.

The survey included a comprehensive effort to engage the public with press releases, informational videos, social media coverage, and a clean, modern survey website. The two sponsoring agencies were featured prominently in the materials and all communications, lending credibility to invitation mailings and other forms of contact made on behalf of the survey. The project logo was the same logo used in the last two surveys. All of these factors combined led to a higher than anticipated response rate across every sample area in the study. Maintaining and building upon this approach to public outreach in future surveys will be critical to their success. It is foreseeable that an even greater role for social media may exist in future surveys as the market penetration of the use of these services approaches 100 percent.

The SEMCOG-created videos available on both the public website and YouTube received relatively few views considering the number of people invited to participate in the study. The cost of creating the video versus the benefit of the limited number of views should be considered for future studies. While a nice feature to have if budget permits, these could also be reduced to one video or even cut completely.

STC 15 translated materials from English to Spanish and Arabic to make them accessible to citizens who speak the three most common languages in Michigan. Although we have no quantitative evidence that this affected the response rate in any particular language, anecdotally, negative feedback was received from over a dozen English speakers, particularly regarding using funds to translate materials to Arabic. The translation service received only two calls from Arabic speakers and around six calls from Spanish speakers. While it is important to be inclusive, there is a cost associated with the translation off all of the materials. The lack of response from Spanish and Arabic

speakers was disappointing. In the future, it may be more cost effective to translate only a few key materials, such as the lead letter and travel log, while including a tagline on all other materials in Spanish and Arabic with the phone number to the interpreter service for additional assistance.

As a part of the dataset validation, trips are typically aggregated by person type. This requires data on worker status (full- or part-time) and student, retiree, and other statuses that are taken from various person-level questions in the survey. For STC 15, the question of hours worked was not asked, and therefore, the full- or part-time status of workers could not be validated against surveys from other sample areas that had a means for separating the two types of people. Including this variable in future surveys is recommended. Other variables to consider (especially if there is any interest in migrating to an activity-based model) would include the collection of a household vehicle list and the collection of the user of the vehicle for each trip made.

Appendices 9