

A HOUSEHOLD TRAVEL SURVEY DATA COLLECTION PLAN

1. INTRODUCTION

Many Metropolitan Planning Organizations (MPOs) around the country are interested in collecting detailed household travel survey data for their planning regions to obtain up-to-date information on the socio-economic, demographic, and travel characteristics of the population. This information is often used by MPOs to update the travel demand forecasting model system that is used for a variety of multimodal planning studies and long-range transportation planning purposes. Even if a household travel survey may have been conducted in the recent past (say, within the past 10 years), many metropolitan regions should consider conducting a new comprehensive household travel survey due to rapid growth and demographic, socio-economic, and transportation infrastructure changes that may have occurred or are projected to occur in the foreseeable future. The household travel survey continues to be the most reliable means of obtaining up-to-date information about travel characteristics in the region.

Collecting household travel survey data takes on added importance in light of increasing interest on the part of numerous metropolitan areas to compete for FTA funding under the New Starts program for implementing light rail or expanding existing rail systems. Up-to-date survey data will help agencies update model systems, calibrate models to local conditions and modal choices, and more accurately predict ridership and user-benefits that may be realized from a light rail investment. Also, household travel survey data collection efforts will provide the opportunity to survey mode usage prior to the start of the new/expanded light rail service. By combining the “before” data with data from a subsequent survey conducted after the new/expanded rail service has commenced, agencies will be able to determine the impact of light rail service on mode usage and multimodal travel demand. Such impact assessment is now required by the Federal Transit Administration (FTA) as part of federally funded New Starts Rail projects.

Agencies should also plan to conduct a comprehensive on-board transit survey in the region, particularly in the context of planning applications for FTA New Starts funding. If the household travel survey can be conducted at about the same time to allow coordination between the two surveys as appropriate, they will together provide a comprehensive picture of transit usage and overall household travel demand in the region. The data from these two surveys will prove valuable in updating travel demand model components and assessing transit markets.

The household travel survey should provide data sufficient to assist agencies in developing and deploying new state-of-the-practice travel demand modeling methods. Over the past few years, several metropolitan areas in the United States including Portland, San Francisco, Sacramento, Puget Sound, Denver, Atlanta, Tampa, Columbus, Dallas-Fort Worth, and others, have developed or initiated the development of tour/activity-based models of travel demand. In light of these developments and the increasing recognition of the merits of these approaches, it is likely that the transition to tour- and activity-based modeling systems will gain momentum in the near future. Therefore, the household travel survey data should include information necessary to facilitate the development of and transition to activity-based modeling methods.

This document is aimed at offering a rather general household activity-travel survey data collection plan that MPOs around the country can use as a starting point for planning their own surveys within the context of their planning needs.

2. WORK PLAN

The overall objective of this effort is to collect detailed household activity-travel survey data for a sample of about 2,500 – 3,000 households in a region. Methodologically advanced and sophisticated survey procedures must be deployed to ensure that the sampling plan is robust and survey design is devoid of biases. The sample must be representative of the population (with exceptions for the specialized over-sampling described later) and any biases potentially introduced by using a single mode of survey administration, survey non-response, and hard-to-reach populations must be minimized to the extent possible. The survey must be pre-tested to ensure that the data obtained from this overall effort will be useful for travel demand model development and transportation planning and policy analysis at large. Local agency staff will play a key role in ensuring quality control and will provide oversight on all project tasks.

2.1 Development of Detailed Survey Methodology

A detailed survey methodology will be developed and documented. The survey methodology will describe, in detail, all of the steps and tasks that will be undertaken as part of the effort. At a minimum, the document will include all of the tasks identified in this household travel survey data collection plan. Subsequent tasks can be undertaken only after the entire survey methodology is discussed and approved by agency staff. The document should answer the following key questions:

- What information will be collected as part of the survey?
- What are the different methods that will be employed for administering various components of the survey?
- What is the sample frame and how will the sample be drawn?
- When should the survey be conducted and for how long?
- What days of the week and months of the year will be covered by the survey?
- How will hard-to-reach populations and specially targeted groups be recruited to participate in the survey?
- In what languages will the survey be administered?
- How will technology be used to help design and administer the survey?
- What will be the sample size for the survey pretest or pilot?
- What incentives will be provided to maximize response rate?
- How will non-response bias be minimized and what type of non-response follow-up will be undertaken?
- What methods will be employed for geocoding location information?
- What methods will be used to develop weights and expansion factors for the survey sample?
- What is the final format of the databases that will be delivered to the agency?

As mentioned earlier, the survey is intended to aid agencies in gradually transitioning towards the development of tour- and activity-based travel demand models. In addition, the data will be used for a variety of transportation planning and policy analysis including transit analysis, HOT lane evaluation, and non-motorized transportation mode use. Information to be collected in the survey includes:

Household Information

- Household size
- Number of individuals by age group, gender, working status, school status, driver license holding (a detailed household member matrix)
- Number of vehicles by type
- Number of bicycles
- Home address
- Housing unit type and year built
- Size of home (sq ft of living space)
- Household income
- Home ownership status
- Year moved into current residence
- Place/address of last residence
- Phone number or e-mail for re-contact
- Number of computers in home
- Type of internet access
- Distance to nearest shopping center
- Distance to nearest recreational park
- Distance to nearest transit stop
- Number of visitors in the household
- Permanent vs temporary residence in Arizona

Vehicle Information

- Make of vehicle
- Model of vehicle
- Year of vehicle
- Year of acquiring vehicle
- Fuel type of vehicle
- Primary user/driver of vehicle
- Secondary user/driver of vehicle
- Mileage on vehicle
- Approx annual miles driven

Person Information

- All information from household member matrix (age, gender, driver license, working status, school status, relation to householder, occupation, education level)
- Place of work (address)
- Number of years worked at current location
- Distance to nearest transit stop at work

- Place of school (address)
- Distance to nearest transit stop at school
- Race/Ethnicity
- Number of paid jobs
- Hours worked last week at each job
- Days of week usually going to work/school
- Flexible work hours/alternative work schedule
- Telecommuting eligibility/availability
- Frequency of working from home
- Employer transportation benefits (specify details – parking, transit, mileage)
- Employer provided vehicle
- Number of days used personal vehicle while at work/school
- Disability status (mobility related)
- Number of days worked at workplace last week
- Number of days worked at home last week
- Number of days went to school/college last week
- Number of times commuted by each mode last week to work/school/college
- Parking availability and cost at work/school
- Availability of secured bicycle parking at work/school
- Typical number of hours of internet use per week for work, school, shopping, entertainment/recreation, browsing/searching for news and information
- Typical frequency of serve-passenger or serve-child trips during week
- Constraints:
 - When (what hours of the day) do you have to be at home?
 - When (what hours of the day) do you have to be at school/work?
 - When and where do you have to serve-passenger or serve-child?
 - Do you have a vehicle available to you during day at home? At work? At school?
- Typical travel time/distance to work/school
- Typical number of stops on journey from home to work/school (specify purposes)
- Typical number of stops on journey from work to home/school (specify purposes)
- Average time spent inside home per weekday and weekend day on the following:
 - Personal hygiene and care
 - Preparing and cooking meals
 - Eating meals/snacks
 - Watching TV, Video, DVD
 - Sleeping
 - Taking care of other household members (children, elderly, etc.)
 - Reading
 - Work/work related business
 - Personal business (paying bills, correspondence, etc.)
 - Pet care
 - Studying/school work
 - Playing/recreation
 - Internet use/browsing
 - Cleaning/arranging house, laundry, etc.
 - Home projects

- Auto care
- Hosting/entertaining friends/visitors/relatives

Activity-Trip Information (24 or 48 hour activity-diary)

- Did person undertake activities outside home?
- Activity location (address or landmark or nearest cross streets)
- Primary and secondary activity type/purpose (detailed categories – approx 20)
- Activity beginning time
- Activity ending time
- Means of travel to activity (detailed modal categories – approx 20)
- Transit fare/pass/discount
- Tolls
- Parking cost/pass/who pays
- Location of parking
- Vehicle used
- Vehicle occupants (number and household members)
- Percent of trip on freeway (identify)
- Distance traveled to activity location
- Total time taken to travel to activity location
- For transit
 - Wait time
 - Walk/drive access time/distance
 - Walk/other egress time/distance
 - Location of transit stop at both ends of trip
 - Route number/identification
 - Transfers (number, locations, routes, fares)
- Other/alternative modes available for getting to activity
- Other/alternative destinations/locations for conducting activity
- HOV lane usage

Attitudes and Values

- Attitudes and values about residential location
 - Access to activities/opportunities
 - Crime
 - Schools
 - Transportation system performance/options
- Attitudes and values towards transportation choices
 - Ideal commute time
 - Most and least preferred modes
 - Ranking of solutions to transportation problems
 - Cost and willingness to pay (e.g., higher taxes)
 - HOT lanes and tolls
 - Congestion pricing/value pricing/parking pricing
- Attitudes and values towards environment
 - Perceptions of modes of transport and their contribution to pollution
- Attitudes and values towards technology

- Alternative fuel and hybrid vehicles
- Intelligent transportation systems (traveler information systems)
- Substitution of travel with telecommunications (telecommuting, teleshopping, online gaming, online news and book reading, etc.)

Stated Preference

- Consider a trip from activity-travel diary and construct scenarios to help determine preference for and impact of the following:
 - Light rail service (with associated cost, time, etc. attributes)
 - Parking cost and availability changes
 - HOV lanes with guaranteed ride home
 - HOT lane implementation
 - Toll charges/congestion pricing
 - Bus rapid transit service (with associated cost, time, etc. attributes)
 - Bicycle paths/lanes/pedestrian walkways

2.2 Design Multimode Household Travel Survey

The survey effort should employ a sophisticated design that takes advantage of multiple modes of survey design and administration. The survey design process should be laid out in detail with all of the steps and components clearly identified. For each step of the survey process, the method or mode by which the participants will be sampled/recruited and the survey will be administered needs to be identified. For example, sample recruitment may be done using phone and post-card mailout methods. Survey instruments can be administered through the mail and/or the phone. Information retrieval may be undertaken over the phone or via the internet through a web-based survey data retrieval system. Here are a series of items that need to be considered:

- Identify the languages in which the survey should be available and design the survey for each language. At a minimum, it is anticipated that the survey will need to be designed and administered in English and Spanish.
- Determine whether the survey should be a 24- or 48-hour activity-diary survey. It would be ideal to collect information for a 48-hour period, but respondent burden may make it difficult to realize this. It would be nice if a 48- and 24-hour activity survey could be designed and pre-tested and a final decision made following an evaluative comparison of respondent burden/fatigue
- Determine the distribution of respondents across days of the week; in general, it would be preferable to obtain information about travel demand for both weekdays and weekend days. As such, it is probably best to distribute the respondent sample evenly across all days of the week.
- Determine the months of the year in which the survey should be administered. It would be ideal to have a pre-test/pilot in Fall and then conduct the real survey from February through August. In this way, various seasons are captured including winter, spring, and

summer (and the pretest will cover Fall). In general, the target should be to get about 300-400 households completed per month.

- Determine the steps and available modes of administration for each survey step/component. For example, one could envision the following:
 - *Step 1 – Sample recruitment: Phone (Random Digit Dialing) and Postcard Mailout*

In this step, households are recruited for participation in the survey. Standard random digit dialing (RDD) procedures can be employed to recruit households on a random basis. In addition, pre-paid post-cards with a pre-paid tear-out reply portion can be sent to randomly selected addresses from an address sampling frame to recruit households that do not have phones, have only cell phones, or screen their calls very carefully.
 - *Step 2 – Household, Vehicle, and Person Information Mailout:*

In this step, households are mailed a detailed questionnaire on household, person, and vehicle information.
 - *Step 3 – Household, Vehicle, and Person Information Retrieval:*

It is conceivable that this first set of information can be retrieved in three different ways. In this day and age of survey overload, it is desirable to offer people the maximum number of options and choices for providing the data. In this particular instance, respondents could mail back a completed questionnaire, fill the information online in a web-based retrieval system, or provide the information over the phone. The survey contractor should follow-up with the household by phone or post-card to make sure that the household has provided the information and maximize response rates.
 - *Step 4 – Activity Diary Mailout:*

Based on the information obtained from the socio-economic and demographic/vehicle questionnaire, the survey contractor will mail out the requisite number of activity diaries to the household. The diary will clearly identify the specific days (with a set of alternate dates) on which the household members should provide activity-travel information. The diaries should be mailed out about one week prior to the designated activity-travel dates for the household.
 - *Step 5 – Reminders:*

Following the mailout of the activity-travel diaries, the survey contractor should provide one or two reminders (say, a day or two before the designated travel dates). The reminders could be in the form of a phone call/message and/or a postcard.

- *Step 6 – Activity-Travel Data Retrieval:*
As with the socio-economic and demographic/vehicle information, there are presumably three different ways in which the data can be retrieved. The household can mail back completed diaries. Alternatively, they could fill the diaries and then input all of the information into a web-based data retrieval system. Third, the survey contractor could collect all of the information over the phone through a CATI (computer-assisted telephone interview). It would be preferable to have all three methods available to households; if that is not possible, then it would be nice to have at least two methods of retrieval available. Again, it is important to have multiple modes of data collection simply to offer respondents flexibility and choices.

- *Step 7 – Attitudes/Perceptions/Stated Preference Mailout:*
The attitudinal and stated preferences questionnaires should be customized to the particular respondent and then mailed out to the household. It is likely that this questionnaire will be administered to one adult per household (as opposed to all household members in each household) to decrease respondent burden. Moreover, this would still yield a respondent sample of about 3000 questionnaires for this section of the survey. The entire questionnaire, tailored to the particular context of the individual, should be mailed out.

- *Step 8 – Attitudes/Perceptions/Stated Preference Reminder:*
The survey contractor should follow-up with a phone and postcard reminder to maximize response rates.

- *Step 9 – Attitudes/Perceptions/Stated Preference Data Retrieval:*
Once again, a three-pronged data retrieval system could be implemented. The respondents could mail back completed questionnaires, provide the information through a web-based data retrieval system, or furnish the information through a CATI.

- Determine the incentives that will be provided to respondents. It would be nice to offer all households a two-step post-completion incentive. They could get \$10 or \$20 after completing the activity-travel diary portion of the survey and then another \$10 or \$20 after completing the attitudinal and stated preference questionnaire. This will keep the respondents motivated and rewarded for providing the requested data. It is also possible to offer the incentive at the time of initial mailout, but it is probably preferable to offer post-completion incentives to reward effort. It is more easily justified to political constituents.

- Determine how non-respondents will be contacted for follow-up to measure and minimize the impact of non-response bias. Design a specific survey instrument and administration method for non-respondents. More details regarding non-response follow-up are provided under item 2.7.

- Determine how data will be collected for children under the age of 12. In general, it is assumed that anybody above the age of 12 will directly respond to the survey while a responsible adult will be asked to provide the information for all children 12 and under.

2.3 Develop Sample Frame and Detailed Sampling Plan

The overall respondent sample size of the household travel survey is envisioned to be about 2,500 to 3,000 completed households where a completed household is defined as one where at least 50% of the household members have provided complete activity-travel information (along with socio-economic, demographic, and vehicle information). Data will be collected for all household members as per the specifications in the survey design.

In general, the sample frame should constitute the population of the entire planning region together with any adjoining areas that fall within the boundaries of the travel demand model. It may be advisable to develop an address-based sampling frame to permit differential sampling rates by geographic jurisdiction and to mitigate sample coverage bias that results from the exclusion of households without land line telephones in traditional Random Digit Dialing (RDD)-based travel surveys.

Once the sampling frame is developed, a geographically stratified random sampling scheme should be employed to ensure adequate geographic coverage by the overall respondent sample. The geographic stratification may be done by city, planning district, or any other appropriate geographic jurisdiction that would allow the implementation of a stratified sampling scheme. If necessary, a disproportionate stratified sampling scheme may have to be adopted to ensure adequate sample units (respondents) from certain geographic jurisdictions that may be very small or have very low population (sampling frame). Weights can be developed to account for biases introduced by the disproportionate stratified sampling scheme.

In addition to serving as a general purpose travel survey for the region, this survey effort is designed to provide specific information about mode usage, particularly for certain population subgroups. In this context, it may be prudent to enrich the geographically stratified random sample with additional samples specifically drawn to facilitate special modal analyses. The following over-sampling of special population groups should be considered, as appropriate (choose those relevant to the specific context):

- *Households Within Five Mile Buffer of Proposed Initial Light Rail Corridor*
Households that reside within a five mile buffer of an initial proposed light rail line should be specially sampled to participate in the survey. After identifying all of the addresses in the buffer, a random sample of households should be drawn (say, through random digit dialing and postcard mailout) such that about 200-250 respondent households are obtained. These households may need additional mechanisms to maximize response rate and minimize non-response bias. The inclusion of these households in the sample will help assess the impact of the light rail system on household travel behavior specifically within the light rail buffer area. To best accomplish this objective, it would be ideal if this sample of 250 households could be surveyed

repeatedly on an annual or semi-annual basis. This “panel” sample component would serve as a valuable assessment tool for light rail impact studies.

- *Households with Workers Within Five Mile Buffer of Initial 20-mile Light Rail Corridor*
This is very similar to the previous over-sampling suggestion except that the focus shifts to the workplace. Within the five mile buffer of the initial proposed light rail line, identify large and small employers and ask them permission to distribute recruitment postcards and/or complete survey packets to their employees. Appropriate follow-up mechanisms will need to be put in place to ensure that response rates are maximized. These households will not necessarily reside along the light rail line. However, at least one worker in the household works at an employer located along the light rail line. It is hoped that about 100-150 households with at least one worker working at an employer located along the light rail line can be obtained into the respondent sample. This 100-150 households, together with the previous 200-250 households (that reside along the light rail line), would constitute a panel sample that is repeatedly surveyed to assess light rail impacts on travel behavior. Such a panel sample of 300-400 households is a great start and this can be expanded over time if the experiment works well and there is funding commitment to continue the panel.
- *Households Participating in On-Board Transit Survey*
As mentioned earlier, it would be ideal to conduct a comprehensive on-board transit survey in the region to obtain detailed information about transit travel behavior. The on-board transit survey will offer valuable information about transit riders and their trips. However, as with virtually all transit surveys, only information about selected transit trips undertaken by the rider will be obtained. In other words, a complete picture of the transit rider’s travel behavior, and that of his or her household, will not be obtained. An analysis of transit rider characteristics can best be accomplished by viewing their transit travel in the overall context of their travel behavior and mobility needs. Moreover, the geographically stratified random sample of households drawn for the household travel survey is not likely to yield a sizeable number of transit users. It is necessary to augment the transit sample in the household travel survey with additional transit households so that the final transit user sample in the survey sample is large enough to permit specialized analysis of transit travel behavior. The survey contractor should contact respondents from the on-board transit survey and recruit them to participate in the household travel survey. About 200-250 additional transit user households should be targeted for inclusion in the overall survey sample through this mechanism.
- *Zero-vehicle Households*
In the context of assessing mobility needs of the transportation disadvantaged and the transit market potential, one group of particular interest is “zero-vehicle households”. The household travel survey random sampling procedure is likely to yield some, but not nearly enough, zero-car households that would allow a specialized analysis of this market segment. It would be desirable to augment the survey sample with an over-sample of zero-car households. These households could potentially be identified and recruited from the on-board transit survey sample as well. There may be other mechanisms to identify

and recruit these households; the overall target is to have about 250-300 zero-car households in the overall final household travel survey sample.

- *Households in Transit-Oriented Developments (TODs) or Mixed Use or Neo-Traditional Neighborhoods*
Transit-oriented developments (TODs) are gaining much attention for their potential to encourage transit use, bicycling, and walking, and reduce automobile use. As metropolitan areas move forward aggressively in expanding transit options, and light rail in particular, it is likely that new transit-oriented developments will be proposed and built by developers. Although research has shown that people residing in TODs make fewer vehicle trips and use transit, bicycling, and walking more, the exact contribution of the TOD environment to this phenomenon is not clear due to residential self-selection issues that potentially mask the true impacts of the TOD environment. It is conceivable that people who choose to live in such neighborhoods have different attitudes, lifestyle preferences, and proclivity to using alternative modes of transportation compared to those who choose to live in suburban residential neighborhoods. Thus, there is a self-selection mechanism at play and it is important to account for this phenomenon to accurately assess and measure the impacts of TODs on travel behavior. It is desirable to identify about 2-3 TODs in the planning region and recruit about 200 households from these TODs to measure their travel behavior in the context of their attitudes and preferences towards mobility and residential location choice. Analysis of this survey sample can significantly help inform future policy directions in the development of TODs and assessing their impacts on travel demand.

The over-sampling procedures described here will result in the overall survey sample being biased and no longer representative of the population. This is not a problem. Appropriate methods are available to weight choice-based or endogenously enriched samples so that biases are eliminated during the statistical analysis and model estimation phase of the project. The survey contractor will need to develop these weights and ensure that the weighted sample is representative of the population (see item 2.10).

2.4 Conduct Pilot Test of Survey and Analyze Results of Survey Pretest

As with any survey, it is absolutely critical that the entire survey protocol, administration procedures, and overall design are thoroughly tested in the field prior to conducting the full-fledged survey. In addition, the pilot test will offer the ability to compare alternative designs (say, a 24-hour vs a 48-hour activity-travel survey) and make a decision regarding the final design specifications to be adopted for the full-fledged survey. In the Fall, the survey contractor will field test the survey. The target respondent sample size for the pretest should be about 300 households and should offer the survey contractor the ability to test the survey design and administration in different languages. Geographically stratified random sampling should be adopted for the pretest as well (although no over-sampling needs to be implemented at this stage).

The survey contractor should geocode all location information from the pretest and thoroughly analyze the survey pretest sample to assess various aspects of the survey design and

administration. At a minimum, the analysis should focus on assessing response rates, examining socio-economic and demographic profiles/characteristics, and describing activity and travel behavior characteristics in detail. The results of the pretest should be thoroughly documented to help identify any shortcomings that need to be addressed and enhancements that can be made.

The pretest should also include a follow-up survey of about 200 non-responding households who were sampled in the pre-test, but did not respond to the request to participate in the survey or, after initially agreeing to participate, did not fully complete the household/person/travel day interviews. The results of the non-response follow-up survey are to provide insight on the reasons for survey non-participation and identify ways and means by which response rates can be maximized (and non-response bias can be minimized) during the full-fledged administration of the household travel survey.

As part of the survey design and administration procedure, the following issues and questions will need to be addressed:

- Interviewer selection and training procedures
- How non-English speaking households will be handled
- How household requests to be contacted in specified ways and at specific times will be handled
- How household call backs and re-contacts will be handled
- Under what conditions will proxy interviews be considered acceptable
- The techniques that will be used to encourage reluctant households to participate
- Special procedures to reach hard-to-recruit populations (low income, transportation disadvantaged, zero-car, college/university students, elderly, residents of dorms and group quarters)
- How the data collected will be checked, coded, edited, and validated
- How missing data items will be handled
- How the non-respondent follow-up survey will be handled; and
- The survey management and quality control procedures that will be used.

2.5 Revise Survey Methodology and Design Based on Results of Pretest

Following the analysis of the data and results from the survey pretest, the survey methodology, design, and administration procedures will be suitably revised to address shortcomings and enhance the survey to the extent possible. All survey materials, administration procedures, and protocols will be finalized in this step.

2.6 Administer Full Household Travel Survey

The household travel survey is administered in its entirety. The survey contractor will implement the sampling scheme including the over-sampling of special market segments and target a final respondent household sample of about 2,500 – 3000 completed households. All agreed upon protocols and procedures must be strictly followed throughout the survey process. The main survey should be undertaken in the Spring and Summer with about 300 households

obtained per month. Data will be coded, entered, and analyzed on an ongoing basis to allow continuous quality control, adjustments and revisions to survey procedures as needed, and constant monitoring of the statistical profile of the respondent sample.

2.7 Conduct Non-Response Follow-up

A comprehensive follow-up of non-respondents will be undertaken on a continuous basis simultaneous with the actual administration of the survey. In each month, the survey contractor will administer the survey non-response follow-up procedures to contact non-responding households. The non-response follow-up is a very short procedure in which the household is asked to provide a few basic socio-economic and demographic variables (say, household size and vehicle ownership) and reasons for non-participation. The non-respondent survey should be short and sweet to obtain information from as many non-responding households as possible. This will provide valuable information for assessing the extent to which non-response bias affects the overall survey respondent sample and for developing post-processing methods that minimize the impact of non-response bias on sample statistics and model parameters.

2.8 Geocode Location Information

All residence, workplace, and activity location information should be geocoded using automated address matching capabilities now available in commercial software packages. Geocoded location information is very useful for mode choice modeling, transit access and egress measurement, and land use – travel behavior interactions analysis. It would be ideal to have the location information geocoded to a variety of geographical measures to facilitate the merging of a number of secondary data sets and variables in the future. Geocoding of location information should be done to provide the following:

- Regional Model TAZ and Super TAZ
- Zip code
- Census Block
- Census Blockgroup
- Census Tract
- Planning District
- Latitude and Longitude

All locations that are not matched automatically should be geocoded manually. Where there is missing address or location information, the survey contractor should make every effort to impute the information logically from other records of the person/household or, in the event that logical imputation is not possible, should contact the household/person and follow-up as needed to obtain clarification regarding the ambiguous location information. It is envisioned that more than 95 percent of locations will be successfully geocoded.

2.9 Develop Relational Databases and Detailed Documentation

A set of relational databases will be developed together with detailed documentation describing the variables and the coding schemes adopted. To the extent possible, variables will be treated as numeric fields with string fields reserved exclusively for textual entries such as address information. The relational databases may be developed in Microsoft Access and should include a set of tables including a person table, household table, vehicle table, activity-trip table, non-respondent follow-up table, and location (geocoded) information. Separate tables should be developed for the survey elements on attitudes/perceptions/values and stated preferences/choices. These are very important components of the travel survey and offer valuable information about attitudes and perceptions that influence people's travel choices/behavior and the likely behavioral patterns/choices that may emerge under hypothetical, but realistic, scenarios. The relational databases should be set up such that the tables can be related to one another and integrated/merged/matched with relative ease based on unique key identifiers. The tables should be in standard rectangular format. All missing value codes should be thoroughly documented. The relational databases may need to be developed by integrating data entry from numerous sources; extreme care will need to be taken to ensure that the database assembly is robust. For example, in a mixed-mode survey, data may be received through the mail, by phone, and over the web. All of this data will need to be integrated carefully to create the final data file.

It would be useful to also have an exclusive file that focuses on joint trip making among household members. It is often the case that individual household members report trips that they undertook with others in their respective travel diaries. However, when records are checked against one another, one usually finds discrepancies between records where two or more household members are presumably on the same trip (for example, the beginning time of the trip or the length of the trip will not exactly match up). It would be nice if the survey contractor could isolate all joint trips/activities and prepare a data file of these activities and trips where all information is completely reconciled so that the records of people on the same trip are perfectly consistent with one another (note that the activity purpose may differ between people on the same trip).

The documentation should follow standard formats and constitute an exhaustive metadata for the survey. The 2001 National Household Travel Survey data documentation format may be used as an example. Other well-known and established data documentation standards and metadata formats have been adopted recently and implemented in the context of the Metropolitan Travel Survey Data Archive (<http://www.surveyarchive.org>) and in European data archives. The survey contractor should research these carefully and implement the most user-friendly data documentation system for the survey.

2.10 Develop Weights and Expansion Factors

The household travel survey data set will be used to obtain up-to-date statistical profiles about the socio-economic, demographic, and activity-travel characteristics and preferences of the population in the region. Due to survey non-response, over-sampling of special population segments, and hard-to-reach populations, the survey sample is unlikely to be representative of

the population. In order to use the survey sample for statistical analysis and model development, suitable weights need to be developed and included in the data set so that the weighted sample is representative of the population. The survey contractor will need to identify a select number of demographic and socio-economic variables/dimensions for which marginal and joint distributions will be compared between the survey sample and population census tables. This comparison will provide the basis to develop and compute weights for the survey sample. Both household-based weights and person-based weights will need to be developed and incorporated into the dataset. In addition, a set of weights to account for choice-based over-sampling (e.g., transit households) will also need to be developed and included in the survey datasets. The data documentation should clearly describe how the weights were computed and the appropriate use of the weights for analysis purposes.

Likewise, a set of expansion factors will be developed and included in the databases to expand the survey sample to the overall population figures. These expansion factors will also be developed by comparing sample statistics against those obtained from population census tables for the region.

One possible idea which deserves consideration is the possibility to introduce a GPS-based travel survey component within the context of the overall household travel survey. It is often noted that trips are under-reported in standard activity-travel surveys due to respondent burden and fatigue or simply because people forget to report certain activities and trips. By introducing a GPS travel survey component, in which about 100 households are asked to place GPS trackers on their vehicles, one could potentially develop correction factors (similar to weights) to account for potential trip under-reporting. By comparing the GPS travel logs against the actual activity-travel diaries, one can assess and quantify the degree to which under-reporting is taking place and then develop a set of correction factors (as a function of household and person socio-economic and demographic characteristics perhaps) for inclusion in the final datasets. These correction factors, when applied during analysis and model development, will help compensate for any trip under-reporting that may be taking place and more accurately capture the full range of travel. The level of detail and sophistication employed in the GPS travel survey component can be determined based on available budget, technology, and logistics.

2.11 Submit Final Report and Databases

A comprehensive final report that documents the entire travel survey and database assembly will be prepared. The final report should also include basic tabulations of statistical profiles, detailed statistics on non-response, and numerous comparisons with census tabulations. The final report should also include a detailed analysis of the attitudinal and stated preference components of the survey. Based on the analysis, the final report should include guidance on the use of the data for a wide range of applications. The final report should provide an assessment of the quality of the data, particularly in the context of the stated preference data that is often eyed with suspicion. The final databases should be packaged and delivered along with comprehensive documentation. It would be useful to offer the travel survey dataset to the Metropolitan Travel Survey Archive (<http://www.surveyarchive.org>) so that the data is archived forever. In this way, the dataset is available with documentation to any entity interested in analyzing activity-travel characteristics

in the region and users can also run tabulations and analysis online (similar to that currently feasible with the 2001 NHTS data and several datasets at the Metropolitan Travel Survey Archive). This online archiving and analysis capability will serve the agency well in the long run; data will never be lost and people who request data can simply be referred to the archive. Of course, the data that is posted online at the archive will need to be stripped of all information in which individual households or persons can be identified. The master datafile that includes such information should reside exclusively with the agency, which can then provide such data to users on a case-by-case as-needed basis.

3. SCHEDULE AND BUDGET

The design of the household travel survey and sampling plan should commence in May/June. The pretest of the survey should be done in October/November at least two weeks prior to Thanksgiving and in close coordination with a transit on-board survey (if possible) to provide the ability to test the proposed over-sampling procedure. Survey materials revision and enhancement should take place through January and the full fledged survey can commence in February and continue through August. Final preparation and delivery of databases and documentation can take place in Fall. Thus, the entire survey effort will likely take a little more than one year. On the other hand, the survey data collection period can be shortened to as little as two months (say, just February through March/April) and all final databases and documentation can be prepared and delivered by May. In this scenario, the survey effort will take no more than one year to complete.

With a budget of \$500,000 for the survey, it should be possible to have a respondent sample of about 2,500 to 3,000 completed households, although it depends on the cost of over-sampling and implementing a GPS component, if desired. This is generally sufficient to meet most travel demand model development and statistical analysis needs. One potential idea is to have individual jurisdictions and municipalities “purchase” add-on samples for their respective areas. These additional add-on samples can help enhance the overall sample size and ensure that individual jurisdictions will have sufficient sample to meet local transportation planning needs. It would be nice if the regional planning agency can enter into a series of inter-agency agreements with cities and other jurisdictions to enhance the overall budget and sample size. An annual budget will need to be put into place to help continue and potentially expand the panel component of the survey where households that reside and/or have a worker in a five-mile buffer of the proposed initial light rail line will be tracked over time.

A final note is due here regarding the use of the NHTS Add-on approach for accomplishing the household travel survey data collection project. If the NHTS Add-on approach is used, then it is likely that there are established procedures, protocols, and sample design elements for conducting the household travel survey. The extent to which the survey can be customized or altered to meet the needs and desires of an individual agency is unknown. It is likely that the NHTS group will accommodate some requests for customization, but it is unlikely that they will be able to introduce new elements or accommodate major departures from standard procedures without a significant added cost. In that case, the NHTS services can be utilized to obtain the household travel survey data set according to standard NHTS procedures and a parallel

coordinated effort could be undertaken to obtain additional data elements and sample additional households as noted in this plan. There will be cost implications associated with this hybrid approach.

APPENDIX

A question of interest is the level of precision and confidence that one can obtain for various travel characteristics with a sample size of 2,500 – 3,000 households. This appendix offers some insights into this.

According to sampling theory, the sample size needed for a precision level of d and confidence level $(1-\alpha)100\%$ (in the context of determining the mean of a continuous variable) is as follows:

$$n = \left[\frac{Z_{\alpha/2} S}{d} \right]^2$$

where

n = desired sample size

$Z_{\alpha/2}$ = the value of the normal standard distribution corresponding to the desired confidence level

S = standard deviation of the characteristic of interest

d = desired level of precision

In the case of a population proportion (i.e., the characteristic of interest is a population proportion or share), the sample size needed for a precision level of d and confidence level $(1-\alpha)100\%$ is as follows:

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{d^2}$$

where

p = proportion of the population exhibiting a certain characteristic.

Let the characteristic of interest be the “average daily person trip rate”. The mean daily person trip rate is often around 4-5 trips per day and the standard deviation is often about 2.75. If we want to estimate daily person trip generation with a 95% confidence, then the level of precision that we are getting is:

$$d = \frac{Z_{\alpha/2} S}{\sqrt{n}}$$

$$d = \frac{1.96 \times 2.75}{\sqrt{6250}} = 0.068$$

Note that a sample size of 2,500 households will yield information for about 6,250 persons (assuming an average household size of 2.5 persons per household). Hence ‘n’ for purposes of this calculation is 6,250. In other words, daily person trip generation will be estimated with 95% confidence at a precision of ± 0.07 trips. That is, the true average population person trip generation rate will lie within ± 0.07 of the sample estimate obtained from the survey. This is pretty good. Note that the degree of precision will be less for trip rates by purpose, time of day, mode, etc. because variance increases as one deals with travel characteristics at a finer resolution.

Let the characteristic of interest be the “percent of households owning 2 or more cars”. This is typically about 55-60% of the households. Using a value of 55% for p, the level of precision in estimating this population characteristic is computed as follows:

$$d = \frac{z_{\alpha/2} \sqrt{p(1-p)}}{\sqrt{n}}$$

$$d = \frac{1.96 \sqrt{0.55(1-0.55)}}{\sqrt{2500}} = 0.0195 = 1.95\%$$

In other words, the percent of households with 2 or more cars will be estimated with 95% confidence at a precision of $\pm 2\%$. That is, the true population proportion will lie within $\pm 2\%$ of the sample estimate obtained from the survey. This is pretty good too. Note that the level of precision will be worst for characteristics where the proportions are nearly equal in the population (close to 50%) such as the one considered here. The level of precision will be greater (better) when dealing with characteristics where proportions are more uneven (example, transit mode split, etc.).

The main concerns with the 2,500 sample size are that there may not be enough samples to perform analysis on small population segments and in localized sub-areas or jurisdictions. The over-sampling procedures proposed in this plan should address the first concern. If local jurisdictions could purchase add-on samples for their respective areas, that would address the second concern.