

On-Board Transit Rider Surveys – Synthesis of Practice

Contents

- 1 Introduction 1
 - 1.1 Motivations 1
 - 1.2 FTA Requirements 1
 - 1.3 Synthesis of Practice Topics..... 2
- 2 Fielding of Survey – Logistics..... 2
 - 2.1 Self-Administered Methods..... 2
 - 2.2 Personal Interviews 4
 - 2.3 CATI Retrieval..... 7
 - 2.4 Expansion Factor Implications 9
 - 2.5 Web-Based Retrieval 10
 - 2.6 Incentives..... 10
 - 2.7 Implications for Small Transit Systems..... 11
- 3 Selected Data Items – Lessons Learned..... 11
 - 3.1 Origin and Destination Reporting 12
 - 3.2 Route Sequencing 12
 - 3.3 One-Way Trip..... 13
 - 3.4 Income 13
 - 3.5 Access/Egress Modes 14
 - 3.6 Value of Personal Interviews..... 14
- 4 Expansion Methodologies 14
 - 4.1 Survey Bias: Causes and Consequences 15
 - 4.2 Traditional Practice 16
 - 4.3 Control Counts..... 17
 - 4.4 Issue with Traditional Practice..... 18
 - 4.5 Advanced Practice..... 18
- 5 Emerging Use of APC Data..... 20
- 6 File Manipulation & Preparation..... 20
 - 6.1 Trip Purpose Designation 20
 - 6.2 Production-Attraction Format..... 21

6.3	Missing Responses	22
6.4	Illogical Responses.....	22
	Home to Home Trips and Trips with Identical Origins and Locations	22
	Transit Centers Reported as Home	22
	Illogical Route Sequences	23
	Access and Egress Reporting Errors.....	23
	Inconsistent Number of Transfers.....	23
6.5	Data Element Imputation.....	24
6.6	Modal Hierarchy and Trip Table Construction	24
7	Assignment of the Survey	26
	7.1 Metrics and Analysis	27

1 Introduction

1.1 Motivations

The purpose of the Transit Modeling Update project is to specify, within FSUTMS and associated support systems, the revisions necessary to improve the preparation and quality of transit forecasts to a point consistent with federal expectations, and to incorporate supporting state of the practice techniques and tools. The calibration and validation of a regional mode choice model must extend well beyond matching aggregate control totals. It must demonstrate that the model adequately grasps each of the key transit markets within the region and reflects the characteristics and behavior of riders within those markets. The availability of “current” transit on-board survey data is the key source of data that is required to inform the specification, development, calibration, and validation of the model.

The on-board rider survey data provides useful insight into the nature of those markets, and often reveals that a standard regional model is not sufficient to address all of the unique markets served by transit. For example, the transit system may serve significant numbers of university students, travelers to special events (sporting, cultural, general entertainment), air passengers, or visitors. Or as has been the case in many metropolitan regions, the urban rail system provides the mobility that facilitates a large Central Business District circulation movement.

Although a standard use of the on-board survey data is the development of target values for use in model calibration, it is the creation of district (or sector) level matrices, stratified in various dimensions (trip purpose, socio-economic strata, access/egress mode, etc.), that is used to evaluate the strengths and weaknesses of the model, and ultimately judge the adequacy of the model calibration.

Another valuable use of on-board survey data is the formation of observed trip matrices and the assignment of those matrices to the transit network. The process assists in analyzing and evaluating (1) the integrity of the survey, (2), the integrity of the network, and (3), the sufficiency of the transit path building parameters. The results of this analysis may result in revisions or corrections to survey geo-coding, network coding procedures and assumptions, or path building parameters.

1.2 FTA Requirements

In view of the critical role of on-board transit rider survey data in not only the calibration and validation of travel forecasting models, but in understanding the existing rider population, FTA requires that any model used in the New Start planning and design process, rely upon “current” on-board rider survey data. Use of the word “current” is a deliberate attempt to insure that the survey data is an up-to-date representation of system supply, ridership levels, travel and behavior. As such, “current” will vary from region to region depending upon how much, or how little, the system has changed over time.

FTA is currently developing guidelines that will provide important principles that should be considered in the design and implementation of an on-board survey. These guidelines are anticipated to include important revisions based upon current practice and more detailed elaboration of the general guidelines

developed in 2006¹. It is also anticipated that these guidelines may provide various options for implementing these principles.

1.3 Synthesis of Practice Topics

The purpose of this technical memorandum is, therefore, not to duplicate the guidelines under development at FTA, but rather to offer a brief synthesis of good practice in a number of topical areas related to on-board surveys. The memorandum begins (Section 2) with reflections on historical methods used in fielding these surveys and highlights emerging new and advanced methods that are becoming more widely used in practice. This is followed by a selected set of lessons learned (Section 3) with regard to important data items, along with possible solutions that may assist in avoiding common pitfalls. Sections 4 through 7 focus on use of the collected data to expand the survey to universe of travel (Section 4), a discussion of the positive implications of APC data in supporting the survey weighting process (Section 5), preparation of the file for use in model development (Section 6), and finally, assignment of the survey to the network (Section 7).

Use of on-board survey data in the calibration and validation of a regional mode choice model is contained in a separate technical memorandum².

2 Fielding of Survey – Logistics

2.1 Self-Administered Methods

For at least the last 40 years, a vast majority of on-board rider surveys have been self-administered. In other words, a rider is given a survey form to complete upon boarding the vehicle and is expected to complete the survey during their ride on the vehicle and return the form upon exiting. In most instances, there is also a mail-back option should the rider choose to use that form of return. The self-administered method is fraught with a wide variety of potential response biases – inadequate time to complete the survey, language and/ or educational barriers, length of trip, level of crowding, question phrasing and wording, and general confusion about what constitutes a trip. Of equal concern is the quality and accuracy of the individual question responses. Many are left blank, are illegible, or simply illogical. And the specificity of information given for each of the geo-codable data items makes it often impossible to accurately locate the location. These realities, in combination with the advent of more technologically advanced survey methods (described in the succeeding sections), clearly indicate that self-administered survey methods can no longer be considered best practice and are not recommended for use in any Florida region.

There may be some potential exceptions to this conclusion. Express routes, longer distance markets, and trips that generally do not require transfers could still be candidates for a self-administered survey.

¹ Discussion-piece #17, “Methods for Collection of Data on Transit Riders”, Federal Transit Administration, June 6, 2006

² “Principles of Transit Model Calibration and Validation”, Task 6, Transit Modeling Update Project, Florida Department of Transportation, prepared by Parsons Brinckerhoff, March 2012.

Although these types of markets are somewhat less inclined to exhibit many of the response biases cited earlier, they are still be subject to the same reporting problems as outlined above.

With any survey, it is the number of “usable” surveys that dictate the usability and confidence in the reliability of the survey for data exploration, model calibration and validation, and forecasting comparisons. The definition of what constitutes a useable sample has evolved over time and a survey record is generally considered to be “usable” when all of the following variables are available:

1. Production/Attraction Zone
2. Boarding/Alighting Location
3. Access/Egress mode
4. Purpose from/to
5. Key demographic attributes
6. Route sequence

Statistics are often provided which demonstrate that for any individual one of these variables, the response rate might be 90% or higher. However, when taken all together, the numbers of records that survive this test typically become drastically reduced. Sometimes as little as only one-third or less of the original number of records. In one Florida case, only 10 percent of the records survived this critical screen.

Say, for example, that an overall “usable” sample size of 7.5% is desired. If only one-third of the completed surveys will ultimately be usable, then a completed sampling return of over 20% would be required. Which in turn, implies that the number of trips to be sampled increases significantly just to get enough potential completed return to reach the usable sample rate?

In general, there are two strategies available to overcome the problem of missing items in sample surveys: (1) nonresponse weighting adjustment, and (2) imputation. In weighting adjustments for nonresponse, missing or incomplete units (or items) in the sample are ignored and the sampling weights of the responding units (or items) are inflated (or referred to as adjusted) to account for the nonresponding units. This implicitly assumes that the responding units are representative of the overall universe of riders, and that may not necessarily be the case. However, in standard self-administered surveys there is little choice but to remove the records that cannot be considered usable.

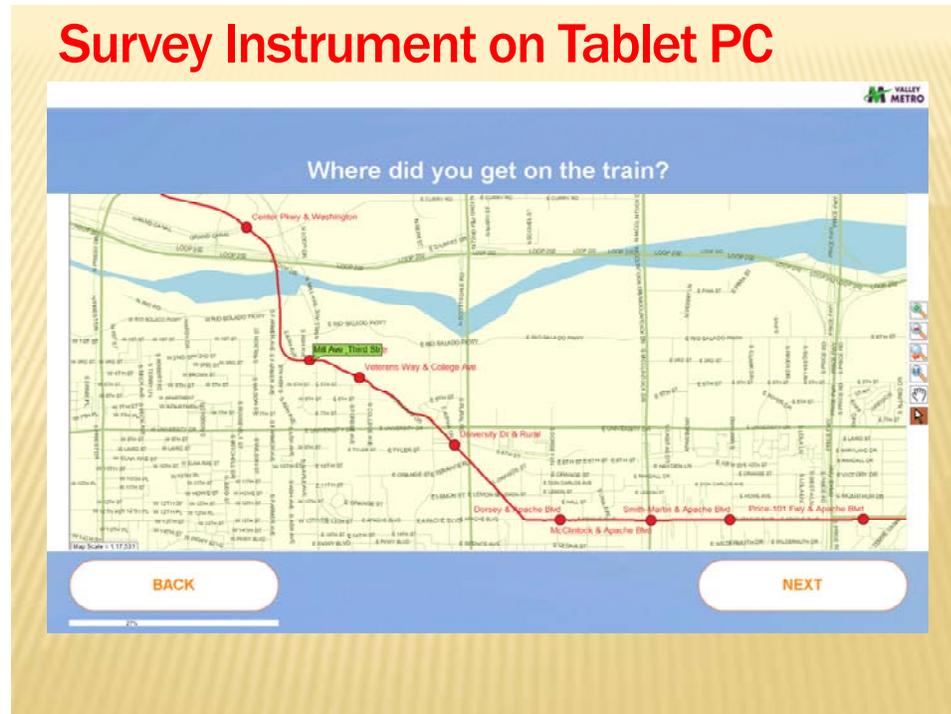
In the imputation approach, incomplete or missing data items are included in the sample with missing values replaced by imputed values and multiple records created to represent the distribution of the missing value based upon known attributes, the sum of weights adding up to the original weight for the record. This is an improvement over eliminating the entire record, but this method only is limited to key demographic variables.

Another major drawback of self-administered surveys is the inability to verify responses in real-time. This is particularly true of geo-coded locations. But also extends to questions such as “purpose from” and “purpose to” which very frequently are both coded as “home”.

2.2 Personal Interviews

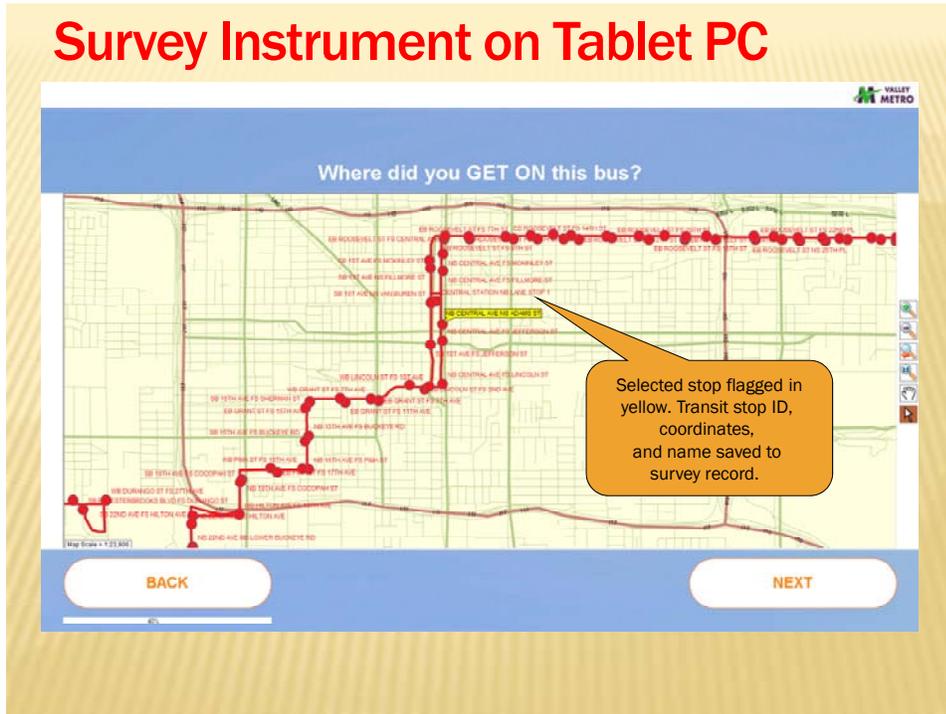
Over the last few years, personal interviews have become the accepted standard in survey fielding. Atlanta, Phoenix, Nashville, Orlando, Palm Tran, and Cincinnati are recent examples of personal interview surveys. The technology that most often supports personal interviews (Tablet PC's, PDA's, iPad's, etc) provide the ability, in real time, to validate all geo-coded locations and the actual route sequence.

From Phoenix, this is a screen shot of a rider identifying their urban rail boarding location:



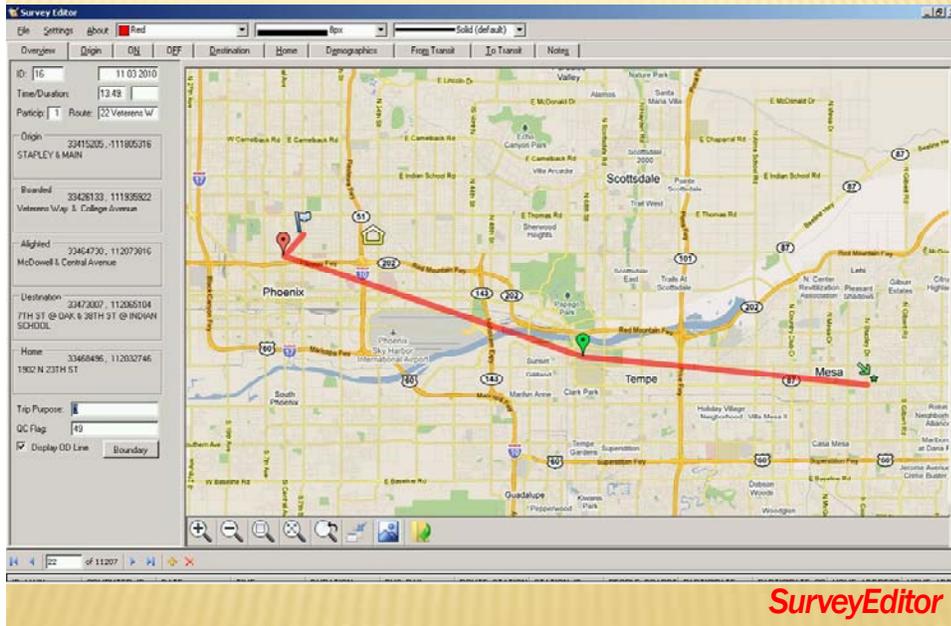
This data is then immediately entered into the record:

Survey Instrument on Tablet PC



And during edit checking, a google interface can be used to evaluate the logic of the reported trip pattern:

Sample Trip – overview screen



Use of a hand-held device to record the responses to each question substantially improves the quality of the data and lends itself to a series of internal logic checks, particularly for those passengers who will transfer one or more than once to reach their final destination. As such, there is consistent quality in the data responses and the data can also be used to evaluate and control interviewer performance.

To further illustrate the improved response rates and quality of data using personal interviews and supporting technology, the following are statistics from three locations where the previous survey was self-administered and the most recent survey was personal interview:

Tablet PC vs. Self Administered

Measure	Tablet PC	Self Administered
% All Questions Completed	95%	67%
% Geocoded to 5 Key Addresses	92%	35%
Time to Administer the Survey	4.3 minutes	7.2 minutes
% Provided Income Data	96%	76%

Combined results of pilot tests in Atlanta (2009), Phoenix (2010), and Nashville (2011)

2.3 CATI Retrieval

Atlanta and Los Angeles are both regions that have very effectively utilized special discretionary FTA Section 5339 funding to field on-board rider surveys exploring and using advanced methods. Atlanta was a 100% personal interview survey and much of what is described in Section 2.2 was based upon that experience³. Los Angeles took a slightly different approach that combined use a boarding and alighting card to retrieve basic information and also serve as a recruiting tool for the second stage of the survey, a computer-aided telephone interview (CATI).

³ "ARC Regional On-Board Transit Survey", Final Report, prepared by ETC Institute, June 2010.

The boarding and alighting card is shown below:

Why do you use the bus or train? Let us know to serve you better.

1. Where are you coming FROM now?

<input type="checkbox"/> Work or Work Related	<input type="checkbox"/> Hotel (guest only)
<input type="checkbox"/> School (K-12) (student only)	<input type="checkbox"/> College or University (student only)
<input type="checkbox"/> Social or Recreational	<input type="checkbox"/> Shopping
<input type="checkbox"/> My Home	<input type="checkbox"/> Airport (airline passenger only)
<input type="checkbox"/> Other: _____	

2. Where are you going TO now?

<input type="checkbox"/> Work or Work Related	<input type="checkbox"/> Hotel (guest only)
<input type="checkbox"/> School (K-12) (student only)	<input type="checkbox"/> College or University (student only)
<input type="checkbox"/> Social or Recreational	<input type="checkbox"/> Shopping
<input type="checkbox"/> My Home	<input type="checkbox"/> Airport (airline passenger only)
<input type="checkbox"/> Other: _____	

Please provide us with your name, phone number, and best time to call you. Once we complete the brief phone interview, you will be added to a drawing for **\$500!** Thank you.

Name: _____

Phone: (____) _____ - _____

Cell/Mobile Landline Home Landline Work

3. When is the best day/time to call you?

Weekday Weekend Daytime Evening



Per office use only:

24001-02-03 - 07/2010

Español →

Finished filling out this form? Please hold onto this card and return it ONLY as you get off.



1. Please take card



2. Fill form

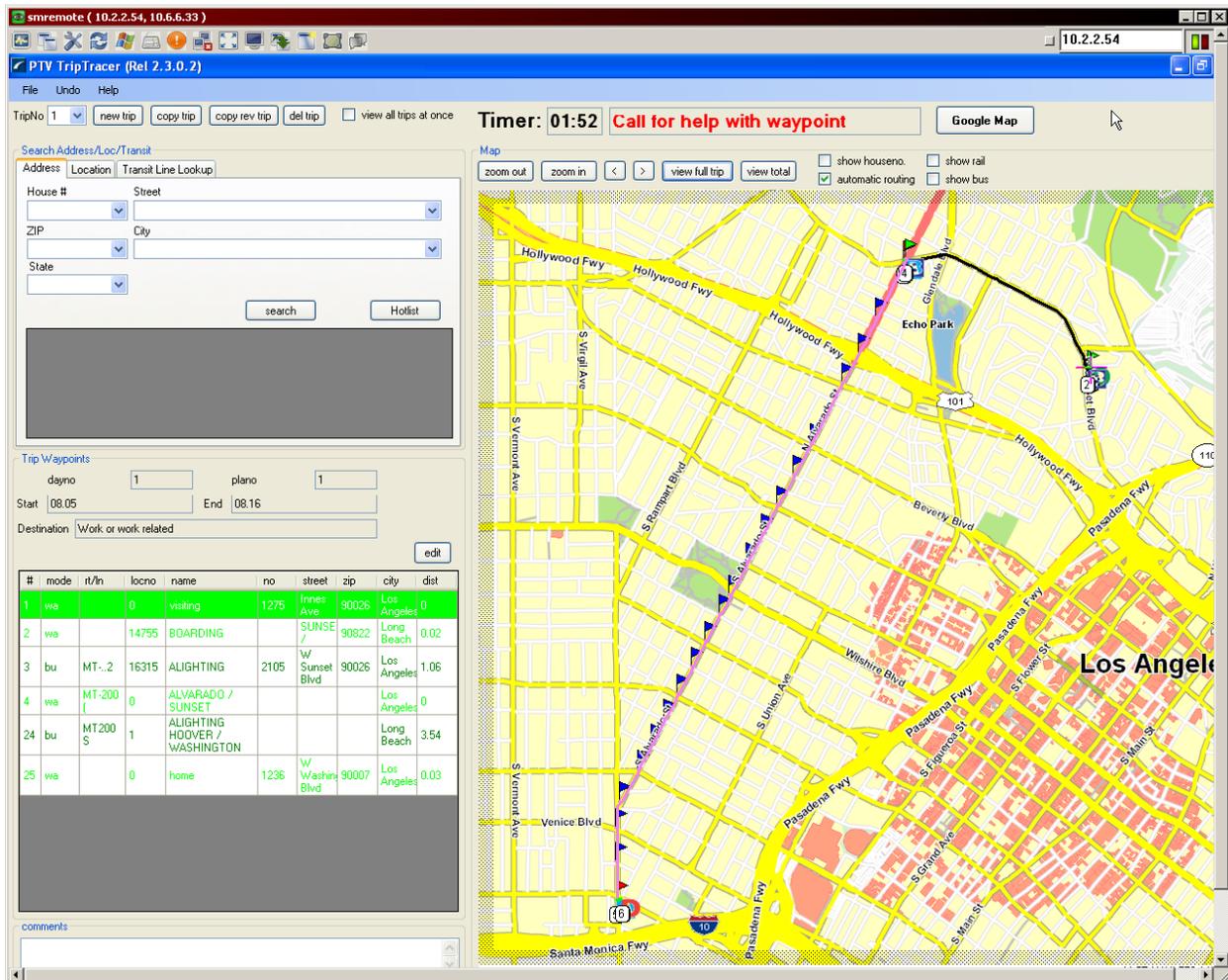


3. Give it back as you leave

Phase I of the Los Angeles project explored a wide range of survey and recruitment methods⁴, and it was this two-stage survey method that was adopted for use in Phase 2. Like Atlanta, this more personalized and technology supported survey has resulted in markedly improved response rates and data quality. The CATI retrieval software expands upon the technology used in Home-Interview Survey CATI software.

⁴ "System-Wide On-Board O-D Study, Draft Phase I Pilot Report", prepared for the Los Angeles County Metropolitan Transportation Authority, prepared by NuStats, October, 2009.

An example of a screenshot from the CATI retrieval software is shown below:

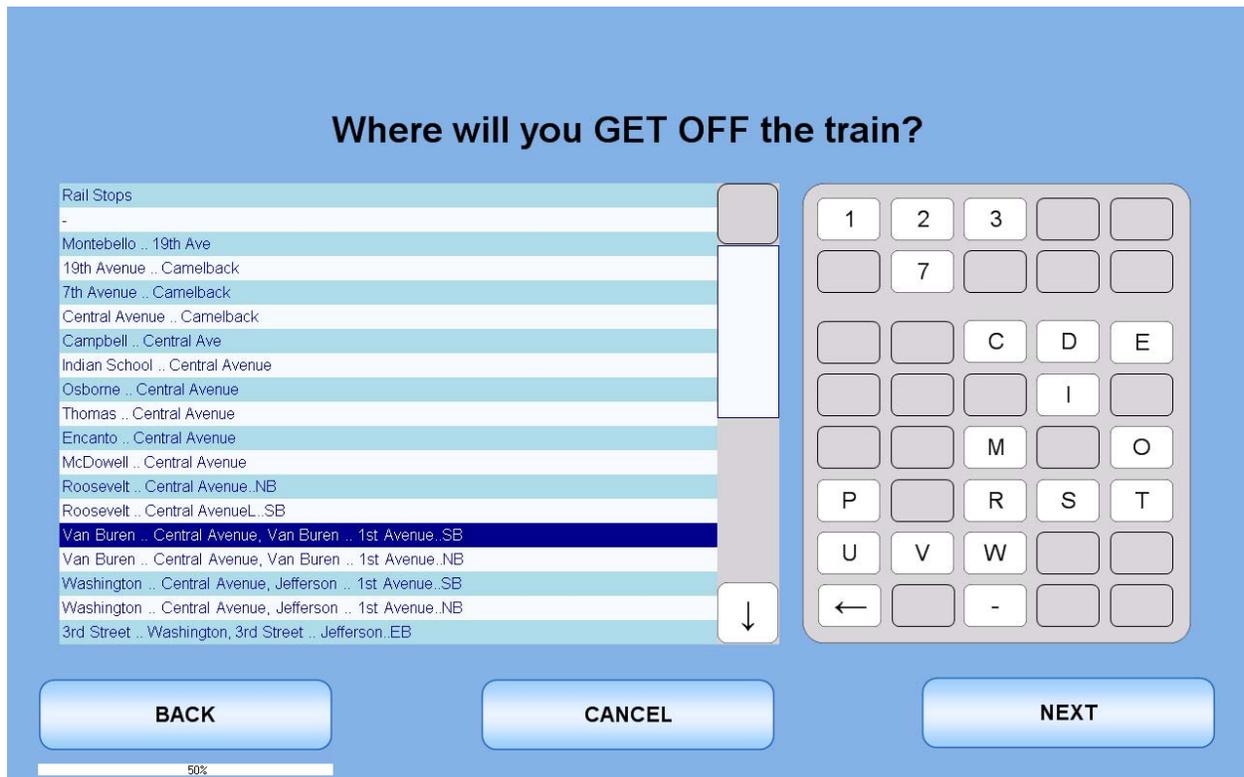


One major improvement in the two-stage survey method in Los Angeles compared with previous self-administered surveys was a tabulation that indicated that 40% of completed responses were obtained in Spanish.

2.4 Expansion Factor Implications

As described in more detail in Section 4, expansion of the survey by stop on and off (in addition to route, direction, and time-of-day) has emerged as the preferred method. The two-stage card and CATI retrieval method directly obtains stop-to-stop observed data for this purpose from the cards. The Atlanta survey had rail station-to-station data from their automated passenger counting system. In Phoenix, this data was collected with a simple stop on and off survey, that like the Los Angeles cards have a very high response rate.

The Phoenix on and off survey was performed using hand-held units as shown below:



As such, there are strong parallels between the personal interview method (on-to-off survey combined with computer-aided personal interview) and the two-stage card and CATI retrieval method. Both methods result in the ability to more accurately target and compute a desired sampling plan, provide more complete and accurate data, and provide the ability to properly expand the resulting data.

2.5 Web-Based Retrieval

Web-based retrieval has been used successfully in a wide range of contexts. In the realm of travel model development, web-based retrieval has been most widely used in stated-preference surveys. However, for on-board rider surveys, use of this method has resulted in generally very poor levels of participation. A good example was a survey that was performed for the University of Michigan, where the thought was that a web-based survey would yield large cost savings for a population that was very computer-literate. However, the survey was not directly usable on smart phones and only 1 question per screen was possible. In this limited context, it is difficult to ascertain all of the reasons for the very low response rates. In general, there is such limited experience with web-based retrieval that it is difficult to determine if this might be a viable option.

2.6 Incentives

There has always been considerable debate about the importance and value of incentives designed to increase survey response rates. Contractors who perform rider surveys unanimously endorse the use of incentives. In Los Angeles, for example, 20 random drawings were held with a prize of \$500 each. It does seem clear that cash prizes are far more effective than discounts or short-term transit passes.

2.7 Implications for Small Transit Systems

The size of the system, in terms of ridership or service provided, should have no bearing on the survey method utilized for performance of the survey. Although, the total cost may be somewhat higher for smaller systems (as there is no opportunity for economy of scale savings), the cost per usable sample with a personal interview or CATI retrieval is comparable to the cost per usable sample in self-administered surveys when all of the unusable survey records are removed.

Selected survey costs reported in May of 2011⁵ clearly indicates that these advanced survey methods are very cost-efficient:

Unit Costs Per Useable Survey Using the Tablet PC/Personal Interview

Location	Year	Method	Survey Administration Cost*	Complete Useable Surveys	Unit Cost Per Useable Survey
Atlanta	2010	Tablet PC/Interview	\$ 1,624,300	43200	\$ 37.60
Phoenix	2011	Tablet PC/Interview	\$ 618,500	15450	\$ 40.03
Nashville	2011	Tablet PC/Interview	\$ 165,000	3956	\$ 41.71

*Survey Administration Costs Do Not Include Survey Design, Sampling Plan, Pilot Test, Analysis, etc

The addition of a stop on and off survey adds somewhere in the neighborhood of \$8-\$9 per usable sample. Survey administration costs also add another \$7-\$8 per usable sample.

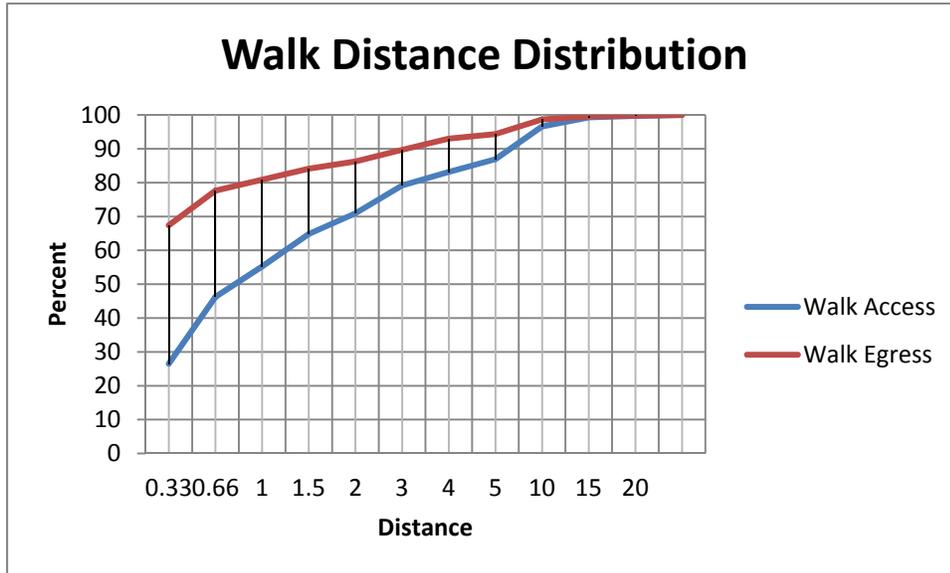
3 Selected Data Items – Lessons Learned

Historical use of self-administered surveys have encountered many problematic issues with question sequencing, phrasing, and respondent interpretation. Significant efforts have continually been made to resolve these issues with limited success. A few examples are discussed in this section. All of these examples pertain solely to the use of self-administered surveys.

⁵ “Best Practices in Transit Rider Survey Data Collection”, presented by Chris Tatum at the TRB Applications Conference, ETC Institute, May 11, 2011

3.1 Origin and Destination Reporting

Most often respondents were given the choice to enter an exact address (if known), the nearest intersection, or a place name. The accuracy of the resulting geo-coded information can vary significantly depending upon the data provided. In the 2007 Phoenix (self administered) survey, walk access and egress distances to transit were summarized based upon the expanded survey records:



The walk access cumulative frequency distribution indicates that over 45% of transit riders walk more than 1 mile to a bus stop and 20% walk longer than 3 miles. Obviously, this is not a logical outcome and reflects the inaccuracy of the respondent address data. The walk egress distribution is reasonable only because the alighting locations were imputed based upon the final destination.

3.2 Route Sequencing

Considerable effort has been invested in testing and evaluating a range of methods to obtain the sequencing of routes used by the rider for their entire trip. An example from the recent Cincinnati on-board rider survey:

3. List all the route numbers (or names) you will use for this one-way trip.

1st Route	2nd Route	3rd Route	4th Route
Example: 43 OR READING RD	5 OR HOLMAN		
<input checked="" type="radio"/> Metro	<input type="radio"/> Metro	<input type="radio"/> Metro	<input type="radio"/> Metro
<input type="radio"/> Tank	<input checked="" type="radio"/> Tank	<input type="radio"/> Tank	<input type="radio"/> Tank
<input type="radio"/> CTC	<input type="radio"/> CTC	<input type="radio"/> CTC	<input type="radio"/> CTC
<input type="radio"/> Metro	<input type="radio"/> Metro	<input type="radio"/> Metro	<input type="radio"/> Metro
<input type="radio"/> Tank	<input type="radio"/> Tank	<input type="radio"/> Tank	<input type="radio"/> Tank
<input type="radio"/> CTC	<input type="radio"/> CTC	<input type="radio"/> CTC	<input type="radio"/> CTC

A variation on this theme was used in the 2005 Honolulu survey:

4. Was this the **FIRST bus you used for **THIS ONE-WAY** bus trip?**

Yes, this is my first bus on this one-way trip No, I transferred from Route #:

For office use only:

5. Will you **TRANSFER from the bus you are on right now to **ANOTHER** bus?**

Yes, I will transfer to Route #: → **After that bus, will you transfer again?** Yes, to Route#:

No, I will not transfer to another bus No, that will be my last bus

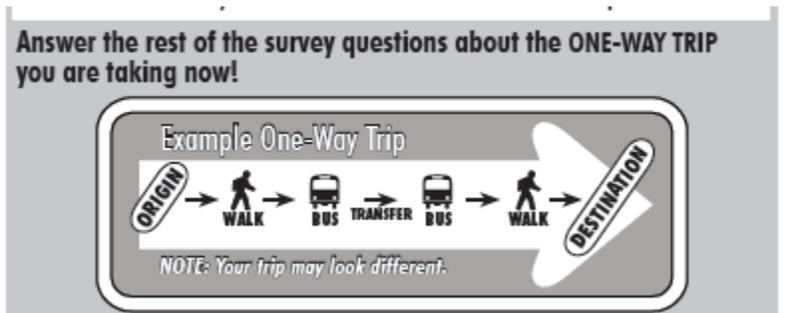
In this example, respondents often provided all of the routes they could have used. Or still became confused with fully understanding the sequence when the route on which they were surveyed was not the first or last route used to complete the trip. And of course, confusion between a one-way trip and a symmetric two-way trip further complicates the accuracy of the response to these questions.

3.3 One-Way Trip

Distinguishing between a one-way trip and a round trip has continued to confound riders in responding to a self-administered survey. This is particularly evident when encountering records in which both the origin and destination are provided as “home”. Various techniques have been used to define and emphasize the need for data for this one-way trip. A typical example is:

Please tell us about the **ONE-WAY BUS TRIP** you are making now. Return the completed survey to the surveyor, OR place it in the special box located near the exit door of this vehicle OR drop it in any mailbox (*no postage required*).

A graphical representation has often been used to assist the respondent in visualizing the one-way trip:



The later does seem to be more successful than the former.

3.4 Income

For a variety of reasons, respondents are hesitant to provide information on their household income. Most often, this question is placed near the end of the survey in the hope that if all other responses are completed prior to this question, the survey record would still provide useful information. One technique that has been used successfully to encourage lower income riders to respond is to make the first category artificially low. Somewhere in the range of 50 percent of the lowest income category threshold.

This is an example of using this technique:

What was your total household income in 2009 before taxes?

<input type="radio"/> Less than \$10,000	<input type="radio"/> \$25,000 - \$34,999	<input type="radio"/> \$75,000 - \$99,999
<input type="radio"/> \$10,000 - \$14,999	<input type="radio"/> \$35,000 - \$49,999	<input type="radio"/> \$100,000 or more
<input type="radio"/> \$15,000 - \$24,999	<input type="radio"/> \$50,000 - \$74,999	

3.5 Access/Egress Modes

Obtaining useful information on the first and last mode used to complete a one-way trip is challenging when transfers occur. What has become standard in self-administered surveys is the inclusion of an additional question(s) to obtain this data for the very first and last transit service used. A typical example for the first route used:

How did you get from your origin location to the first bus stop on your trip?

<input type="radio"/> Walked/Wheelchair 	<input type="text" value=""/>	<input type="text" value=""/>	<input type="text" value=""/>
# blocks (please answer 0 if less than 1 block)			
<input type="radio"/> Dropped off 	<input type="radio"/> Bicycled 	<input type="radio"/> Taxi 	
<input type="radio"/> Drove alone 	} Parking Lot Name/Cross Streets _____		
<input type="radio"/> Carpooled/ Private shuttle 			

3.6 Value of Personal Interviews

This set of five selected lessons learned all stem from the challenges of using the self-administered method of data retrieval. In personal interviews, there is an opportunity for evaluating the logic of responses in real-time, assisting with geo-coding and route sequence answers, and generally improving the quality of the data received.

4 Expansion Methodologies

The chapter summarizes the state of the practice in transit on-board survey expansion methodologies. Before doing so, a brief summary is required of the surveying process and how it relates to sample size. By definition, an on-board survey instrument is posited to a subset of all travelers. The resulting “raw” dataset is processed for reasonableness and any automated post-processing, such as geo-coding locations, is applied. The “processed” dataset then reflects the best, most detailed information about the surveyed travelers.

If the survey data items are posited to all travelers, and all travelers respond correctly, then the survey reflects the surveying population (i.e., in statistical terms, the “universe”) and no survey expansion is needed.

Much more common is that the data items are posited to a subset of all travelers. This is in part due to the costs associated with interviewing or distributing survey forms to high volumes of riders and the resulting processing. The final survey dataset then reflects a subset of travelers, and the dataset must be

expanded to reflect all travelers. The methodologies of survey expansion are summarized in this chapter.

Expanding the survey can lead to bias in the results simply by itself or in combination with the survey sampling plan. The following section describes the causes of survey bias.

4.1 Survey Bias: Causes and Consequences

There are two causes of bias in transit surveys:

- (1) The inadequacy of the assumption or assertion that all surveyed travel variables and characteristics are normally distributed, and
- (2) The practical consequence that surveys are completed and returned at a rate unequal to their distribution rate.

It should be noted that a combination of the first and second causes occurs in some instances.

Statistically significant sample sizes are correlated to the distribution of the variables to be analyzed. This distribution is either known or estimated through previous research or previously-conducted surveys. In cases where the distribution is impossible to know *a priori*, the distribution is typically asserted. In the social sciences, variables are typically assumed to be normally distributed. Normal distributions have the practical consequence of requiring very small sample sizes, thereby minimizing the resources needed to conduct a survey. For some variables, such as gender and income, this assumption has been found to be accurate. For other key information, such as travel flows, this assumption has been found to be inadequate in many cases. Unfortunately, the sample sizes are not increased to address the non-normality of these variables, usually for budgetary and schedule reasons but also if those non-normally distribution variables are not a primary focus of the survey.

A similar simplifying assumption is made to the completion and return of surveys. Again, for some variables this assumption is a good one while the assumption falls short for others. Short transit trips are typically identified as one trip category that likely suffers from this assumption. This is due to the fact that the traditional survey instrument cannot be reasonably completed in less than 7-10 minutes, so trips of that length or shorter are unable to complete it. Mail-back and similar procedures fail to generate significant response rates to compensate. Other trips that have been found to suffer from this assumption are:

- Walk-access riders, who typically respond less frequently than auto-access riders,
- Riders who are less than 16 years of age, as these are unlikely to be surveyed for policy reasons,
- Afternoon and evening riders, who may suffer from “survey fatigue” after completing an identical survey form for their trips earlier in the day, and
- Trips occurring in heavy load conditions, which make the process of distributing and completing surveys very challenging.

If survey bias is evident, three situations can occur:

- (1) The O/D survey has a meaningful number of records matching the rider or trip characteristic, but these are known to or suspected of being severely biased in some way,
- (2) The O/D survey has a modest or insufficient number of records matching the rider or trip characteristic, and
- (3) The O/D survey is devoid of any records matching the rider or trip characteristic.

It is important to note that a survey dataset can have one of these situations for each survey variable. Therefore, a practical conclusion is that all survey datasets have some level of bias. What is important is to address bias for the variables that are most important to the analysis at hand. The purpose of survey expansion then, beyond simply extending the survey dataset to larger values, is to guard against or, preferably, eliminate survey bias.

4.2 Traditional Practice

The traditional expansion method accounts for three dimensions of response bias: route, time-of-day and direction. This method is more commonly known as the RTD method, with the first letter in each dimension forming the acronym. In this method, an expansion factor is developed for each combination of route, time-of-day and direction categories. For instance, assume a system has 20 routes and the model system accounts for four time periods. Bus routes almost always operate in two directions (e.g., northbound and southbound, eastbound and westbound or clockwise and counter-clockwise). So for $20 \times 4 \times 2 = 160$ expansion factors are computed:

$$ExpansionFactor_{rtd} = \left(\frac{NumBoardings_{rtd}}{NumSurveys_{rtd}} \right)$$

Where:

R is the bus route or rail line,

T is the time period,

D is the direction of the bus route or rail line,

NumBoardings is the number of boardings for the RTD category,

NumSurvey is the number of surveys in the RTD category, and

ExpansionFactor is the resulting expansion factor for the RTD category.

Each survey is assigned an expansion factor for its appropriate category. In some cases, this method also includes a fourth dimension for single or groups of bus stops or rail stations. Survey expansion occurs when the expansion factor is applied to all survey records in each category.

The traditional expansion process requires an accurate accounting of transit boardings across each of the three dimensions. The number of boardings is determined either by using boarding and alighting counts from Automated Passenger Counters (APCs) or through a manual counting process. Boardings are collected across different routes, route directions, time periods and stops/stations. Modern APC systems routinely collect boarding and alighting data by route, time, direction and stop.

This method can be supplemented with additional control counts (beyond boarding counts) to better address response bias in certain areas. Control counts that are typically collected are described in the following section.

4.3 Control Counts

In order to expand the survey properly, there must be an accurate accounting of system-wide ridership across various dimensions. This information is collected independently of the on-board survey effort; i.e., no direct information from the survey is used to develop this information. These counts are known as control counts, as they “control” the expansion. Common control counts are described in this section.

Alightings – Alightings can be collected in the same manner as transit boardings.

Rider Types – Some riders respond to surveys at different rates than other types of riders. Consequently, it can be highly advantageous to collect information on different rider types. Different types of riders are counted usually through a manual process. Electronic fare media may offer opportunities through the different pass types offered by some transit agencies (e.g., university/college fares). For example, the 2008 Central Ohio Transit Authority (COTA) survey collected counts for 4 rider types: university/college students, riders 16 or fewer years of age, riders 16+ years of age but public school age, and other non-student adults. These four dimensions were used to improve the survey expansion, especially for university/college students, which responded to the survey at a much higher rate than other rider types, and riders 16 or fewer years of age, as the surveyors were forbidden to distribute surveys to those riders.

Vehicles – Park-riders tend to respond to surveys at a higher than riders who access the transit system by walking or getting dropped off. Consequently, survey expansions need to account for this response bias by identifying the number of riders who park-ride at each stop or station. This data is commonly collected through manual counts, but some major systems can gather this information through their electronic fare media since it is used as a debit card for parking. The 2008 Tri-Rail survey (in Southeast Florida) collected vehicle counts at many of the 18 park-ride locations. The counts were performed just after the AM peak period, as most of the park-riders in that system occur in that time period. Other areas may experience parking spaces used multiple times over the day. In these cases vehicle counts should be collected throughout the day.

Transfer Movements – Counts of riders as they transfer between rail or bus platforms are also helpful to properly account for important or popular rider movements. These counts are typically collected via manual observation unless they can be collected electronically via fare media.

Access/egress counts – Rider counts can also be conducted for various access and egress modes by station to provide more accurate estimates of these movements. These counts can be challenging if the stop and station location provides multiple entry and exit points. Also, it can be time-consuming to correctly identify the egress mode if riders mingle around the station waiting to be picked up. For the 2008 Tri-Rail survey, the Florida Department of Transportation conducted access and egress counts to provide an improved representation of non-traditional markets. An onboard survey a year earlier identified several of these markets, including sizeable auto-egress and drop-off markets, but was unable to accurately provide reasonable estimations of those markets by station.

4.4 Issue with Traditional Practice

The traditional RTD method has been found to be helpful in providing an accurate accounting of rider types and quantities. However, it has been generally recognized that the method is generally insufficient to properly reflect travel patterns (i.e., station-to-station, or stop-to-stop flows) and riders who respond to surveys at a low rate. While control counts can help the traditional method to mitigate some aspects of response bias, they provide limited value if the number of collected survey records is sparse or nonexistent.

In addition, a practical issue with the traditional method is that the survey results can change drastically with each additional dimension. The extent of the change depends on the amount of bias in the survey dataset, although with some markets (e.g., auto- vs. walk-access riders) the bias is usually substantial. Some surveys undergo a series of expansions as control count data is made available. Such practices, while well intended, can easily confuse decision-makers and reflect poorly on the survey collectors.

It should be noted that these issues occur frequently with traditional practice because the onboard surveys' sample size and sampling plan typically do not account for travel markets or low-response riders. Rather, the historical practice is to develop the sampling plan based on route-level boardings and bus service, which are very aggregate measures. Since travel markets and rider characteristics occur at a disaggregate level, the aggregate nature of the sampling measures inevitably lead to weaknesses in the final survey dataset.

4.5 Advanced Practice

Current practice requires testing and applying weighting strategies beyond the standard RTD process. For example, iterative proportional fitting techniques (IPFT) for expansion are being used to detect and, to the extent possible, correct for non-response bias related to trip distance. Examples include Robert Farley's initial analysis for the 2007 LA Metro Blue Line survey (rail) and AECOM's analysis of Routes #1 and #2 in COTA's 2008 survey (bus), the Palm Tran survey, along with many others.

While procedures such as these address the issue of response records, it is recognized that the optimal solution is to improve the sampling strategies so that an adequate number of surveys records is available for expansion. Modeling practitioners and survey collection firms are now working together in

this direction. Advanced practice requires a strong data collection effort *before* developing the survey sampling plan. With this *a priori* knowledge of rider characteristics and travel patterns, modeling practitioners and survey collection firms can tailor the sampling plan accordingly.

The initial data collection consists of gathering all available data sources; where possible, collecting control counts before the onboard survey; and analyzing the data for rider characteristics and travel patterns. Key data to be collected include: a previous onboard survey, station-to-station and/or stop-to-stop travel flows, boarding/alighting counts, and other control counts mentioned the earlier section.

Recent developments in fare media and surveying methods have made collecting station-to-station and/or stop-to-stop travel flow data much more practical. Certain types of fare media can record the entry and exit station used by the rider. In these situations, the transit agency has the ability to develop station-to-station flows for the entire rail system. Note that this process should be used only if a sufficient number of riders use the electronic fare media and the rider is required to use their card to enter and exit the station area.

It can be challenging to gather this information for bus systems since the rider is typically not required to use a card in order to exit the bus. To collect this information, a survey can be distributed that requests only boarding and alighting stops. Because the survey is very simple to answer, the response rates are much higher than those of traditional onboard surveys. If the survey is administered to a majority of bus runs, the resulting data set will provide a reasonable representation of bus travel patterns.

Two recent survey efforts represent this advanced practice. One is the 2009 Metropolitan Atlanta Regional Transit Authority (MARTA) survey. To address non-response bias from certain rail travel markets, the Atlanta Regional Commission (ARC), the sponsoring agency for the survey, ensured that the survey results would match the actual distribution of riders by reviewing rail station-to-station flows and bus boardings by stop and segment. MARTA had this data available from its Breeze card system. The sampling plan was tailored to MARTA's travel flows by placing more crews at the appropriate time and direction on rail lines that carry major travel markets, and fewer crews on the other lines (and direction and times). The ARC's survey consultant went one step further by continuously monitoring sampling rates by travel flow during the fieldwork. Crew assignments were adjusted if actual sampling rates did not correspond with expected rates. That is, the survey leaders directed crews *away* from over-reported flows and re-directed crews *towards* under-reported flows.

The benefit of such an approach became evident during post-survey analysis, which found that, because the survey was well-sampled and the sampling plan well-designed, that the survey results did not alter significantly when different expansion methodologies were used.

The other survey effort was in Los Angeles and discussed in section 2.3. The Los Angeles Metropolitan Transit Authority (LA Metro) conducted a system-wide onboard survey using an innovative two-part procedure. The first part is a systemwide survey to capture stop-to-stop (and station-to-station) movements on its bus and rail routes. The instrument captures only trip purpose, contact phone number, boarding location and alighting location.

The survey results are processed, tabulated and reviewed. A subset of records are selected, and within a few days after the survey is collected, a call center contacts the rider and gathers the detailed travel information. The call center is equipped with geo-spatial and transit information to validate the riders' responses during the call. There is no onboard survey in the traditional sense.

This survey approach avoids the under-reporting bias of short trips – through the shortened stop-to-stop survey. It also offers the possibility to significantly reduce survey bias since the call center can target under-reported movements or riders in a similar manner as in the 2009 MARTA survey. And the survey data quality is improved tremendously through the use of in-interview validation.

5 Emerging Use of APC Data

The emerging installation and use of Automated Passenger Counting (APC) systems eliminates the need for manual counts and can be directly used in survey weighting (see Chapter 4 above). The availability of APC data can provide significant cost savings compared with manually obtaining boarding and alighting counts, the cost of which can range between \$20 and \$25 per hour. If APC data is available, then the vehicle trip factor can be eliminated as the actual ridership on each run not sampled is accurately known. In addition, APC data can be used to measure and adjust for daily, seasonal, and annual variations in route level ridership. And finally, the survey could conceivably be re-weighted over time in response to fare change, gas price increases, or service reductions.

6 File Manipulation & Preparation

The raw trip records from an on-board trip survey always require refinement and enhancement before use in model development and testing. This chapter describes the data enhancements that are required. In general, these enhancements address both quality control and the addition of attributes that are necessary for model development.

6.1 Trip Purpose Designation

In raw form, the On-board survey trip records will contain the origin purpose and destination purpose. These data items are typically in response to multiple-choice questions such as “What were (or will) be doing there?” or “What kind of place is that?” Typical responses include “Home”, “Work”, “School”, “Shopping” or “Personal Business”. The overall or “general” trip purpose should be based on the combination of these answers. For example, a trip from home to work should be classified as a Home-Based Work trip (in production-attraction format). Note that the general purpose is not sensitive to direction, so a trip from home to school or from school to home are both Home-Based School trips. In application, a lookup table that returns a general purpose code given a valid origin or destination purpose code should be used to automate the process. Suggested general purpose codes include:

1. Home-Based Work (HBW)
2. Home-Based School (HBSCH)
3. Home-Based Shopping (HBSH)
4. Home-Based Other (HBO)

5. Non-Home Based Work (NHBW)
6. Non-Home Based Other (NHBO)

The analyst should summarize expanded trips by general purpose, and compare with typical values from past surveys, and from other similar systems to verify reasonableness. Home-Based work trips for transit riders are a larger portion than in the general trip-making population, typically comprises 35%-75% of all trips, with non-home trip making being very small, often less than 20%.

Remove any trip records that have invalid origin or destination purposes.

6.2 Production-Attraction Format

The raw data from the on-board survey are in origin-destination format. That is, the starting zone is the origin zone and the ending zone is the destination zone. For trip-based model use, the production and attraction ends of the trip need to be defined. This is most easily done by adding additional fields identifying the production and attraction-related attributes that correspond to the raw trip-end attributes. The test for converting origin/destination data into production/attraction information is simple:

- Check to see if the destination zone is home
- If the destination zone is home, set a "PtoA" field to "false"
- Otherwise, set the "PtoA" field to "true"
- For all records where "PtoA" is TRUE, set all origin-related field data to equal the corresponding Production-end fields and all destination-related field data to equal the corresponding Attraction-end fields
- For all records where "PtoA" is FALSE, set all destination-related field data to equal the corresponding Production-end fields and all origin-related field data to equal the corresponding Attraction-end fields

Trip-end attributes that should be created in P/A format include

1. Start and end zones
2. Start access modes and boarding locations (origin end)
3. Ending egress modes and alighting locations (destination end)
4. Trip end purposes and area types

As a quality control check, the analyst should summarize expanded trips by time of day, purpose and P to A/A to P directions (the logical value PtoA mentioned above). The expected diurnal patterns should show pronounced AM dominance of P to A direction work trips and PM dominance of A to P direction work trips. Non-Home Based Work trips should peak during midday, without a directional imbalance. In addition, plotting the locations of productions and attractions, with an overlay of the route system, should allow the analyst to inspect the reasonableness of these locations with respect to available transit service within specified buffers, given the assumed access mode. A similar analysis could be done with boarding and alighting locations, which should fall very near available transit lines. A final

check should examine egress modes, and look for auto mode egress – this is unusual but not impossible and should, in any case be examined.

6.3 Missing Responses

The on-board survey data should be cleared of invalid responses. A trip record may be invalid if it is missing valid responses for any of the following:

1. Origin or destination location
2. Boarding or alighting location
3. Route, time and direction of trip
4. Trip end purposes
5. Access/egress modes
6. Auto availability
7. Household size or Number of Workers
8. Household Income

Items 1-5 above are essential to the trip's usefulness for model development. Items 5-8 are also required, though in some cases can be imputed or synthesized.

6.4 Illogical Responses

The survey data should be systematically checked for logic errors. These include errors that are evident due to logical inconsistencies. These responses should be corrected, if possible by the analyst using standard assumptions and procedures, or the record should be eliminated. In many recent on-board surveys, the home phone number is requested so that some of these errors might be corrected by re-contacting the respondent. Types of illogical responses are discussed below:

Home to Home Trips and Trips with Identical Origins and Locations

A respondent may report a trip as having an origin and destination at the same place, often both at home. This response reveals a misunderstanding of what the survey is asking for in terms of the trip definition. Often travelers think of a trip as a round-trip, for example a trip from home to shopping and back to home might be recorded as starting at home and ending at home. Boarding and alighting locations or origin or destination activities may give some indication of what the one-way trip actually was for and where it ended, but short of this, the only way to salvage such a trip record is to contact the respondent, if that is possible. Survey instruments and survey personnel should be prepared to minimize this common misconception.

Transit Centers Reported as Home

Similar to the error discussed above, respondents will sometimes designate the transit center as "home", presumably in the sense that it is a sort of home-base for the transit trip itself, or possibly the location of a transit transfer point. A more general error is failure or refusal to list the actual trip origin or destination (especially if it is home) and instead simply list the boarding or alighting location – which

becomes redundant with the actual boarding or alighting location asked on the survey. Again, short of re-contacting the respondent, it is difficult to identify the actual trip origin or destination in cases such as this, especially if the access mode is drive or transfer from another bus. In some cases, if walk is listed as the access or egress mode, and a zone-level geocode is all that is required, it may be possible to estimate the actual origin or destination from the given boarding or alighting stop, if that is considered a reliable location. In surveys that use gps-enabled tablets to code boarding and/or alighting locations, this might be a possibility, though the analyst will need to use their own judgment and local knowledge in synthesizing these locations.

Illogical Route Sequences

When a respondent's trip involves two or more transit vehicles (i.e., transfers) the analyst should check that the sequence of routes is feasible. For example, that the stated routes actually intersect during the time reported. If this is the case, an evaluation of initial boarding and final alighting locations may suggest the actual route sequence. This error can be systematically caught by assigning the trip records to a current transit network and comparing the model-estimated routes used with those reported. These survey assignment results can also be a source for correcting the survey record, if the reported route sequence is clearly in error.

Access and Egress Reporting Errors

Logical errors in access and egress can be revealed by computing the distance of each from the reported coordinates of the origin and boarding locations; and destination and alighting locations. A trip length distribution of access and egress distances for walk, bike and auto should be prepared and examined for instances of extremely long reported distances. Straight-line distances of over 1 mile for work, over 5 miles for bike, and over 30 miles for auto should be flagged and the individual trip records examined for reporting errors. Also, drive access locations may be in error if the reported origin and boarding locations show that other logical park and ride locations are bypassed. In such cases, reporting or geocoding errors may be present, or the respondent failed to report an intervening transit leg. Again, the model results from an assignment of the on-board survey may provide clues as to missing transit legs. An examination of the distribution of observed drive-access distances can also inform the specifications of the drive-access builder routine in the model.

Inconsistent Number of Transfers

In some on-board surveys, the respondent is asked to report both the number of transit vehicles used in the trip, and the actual sequence of routes. These two responses should be compared with respect to the implied number of transfers. If they are inconsistent, the number of transfers based on the reported sequence of routes is usually preferred. Once again, the number of transfers is a statistic that can be checked against model results after assigning the on-board survey. Usually this is an aggregate statistic, used to compute the transfer ratio (number of linked trips/number of unlinked trips) as well as the total number of linked and unlinked trips.

6.5 Data Element Imputation

At times, missing or otherwise invalid data can, in certain cases, be imputed or synthesized thus making an otherwise unusable trip record useful for model development. Any imputation or synthesis of data should be clearly marked in the database, so that other users can choose to include or remove that record with full knowledge of its source. Candidates for data imputation and suggested approaches to imputation include:

1. Income – if income is given but only in terms of a wide range, it may still be possible to accurately classify the respondent within model-defined income ranges. Alternatively, missing or incomplete income information can be refined based on correlation with home location, auto ownership and/or family size, if those are reported. This correlation can be established from trip records that do contain a full set of valid information.
2. Auto Ownership/Auto Availability – In some surveys, questions are asked about both auto ownership as well as if a vehicle was available to the rider for the trip being surveyed. If the former question’s response is missing, it might be informed by the response from the latter, or vice-versa. If neither is provided, auto ownership might be imputed based on home location, income and family size, based on correlations established from other, fully-reported records.
3. Household Size – if household size is not reported, it is difficult to impute based on correlations with other variables. However, minimum household sizes can be implied if a non-adult is the respondent (i.e., at least 2) or from the auto ownership response.
4. Trip end and boarding/alighting location imputation – the zone of the origin or destination may sometimes be reasonably estimated if the corresponding boarding or alighting location is known with good certainty, as is the case for gps-coded surveys. This is true if the access or egress mode is walk, but not if the mode is auto, since the possible locations are much more widespread. Conversely, for walk access or egress, the boarding location might be reasonably estimated if the origin or destination coordinates are known. This imputation is much more reliable in lower-density locations served by few transit lines such as in suburban areas, as opposed to high-density locations served by multiple routes, such as in the CBD.

6.6 Modal Hierarchy and Trip Table Construction

A very useful product of the survey processing is the creation of observed transit trip tables, ready for assignment in the model. The following variables, shown in Table 1 are necessary for the construction of these trip tables:

Table 1: On-Board Survey Trip Records Necessary for Trip-Table Construction

Item	Description
1	Production and attraction zones
2	Linked and un-linked trip expansion factors
3	Time of day codes (usually peak and off-peak designations)
4	Mode (based on Mode hierarchy assumptions in path-builder and mode choice mode, see below)
5	Trip General Purpose (only if the model uses a classical work-peak/nonwork-off-peak stratification)

Items 1,-4 and 6 should be easily available from the completed and processed trip records. Note that it is very important that the zones represent the production and attraction locations, not the originally-reported origin and destination zones, since almost all transit assignments rely on P/A format for assignment.

The modal hierarchy is based on definitions implicit in the model’s mode choice formulation. This identifies what is meant by a commuter rail, urban rail, express bus, BRT, Light Rail or local bus trip in the model and the corresponding trip tables must faithfully follow these definitions.

In most mode choice models, the following hierarchy is used:

- | <u>Index</u> | <u>Mode</u> |
|--------------|---------------|
| 1. | Commuter Rail |
| 2. | Urban Rail |
| 3. | Express Bus |
| 4. | BRT |
| 5. | Local Bus |

For any transit trip, the mode used with the lowest index (and highest priority) above identifies the mode it is defined as in the model. For example, a trip which uses a local bus to transfer to an urban rail would be considered an urban rail mode trip. In on-board survey records, therefore, it is necessary to identify all the modes of a multi-mode trip, and apply priority to each. For each trip, the mode with the lowest index (i.e., highest priority) defines the kind of trip it will be classified as for the purposes of building the observed model trip table.

To verify a particular model’s mode hierarchy, inspect the allowed modes for each type of transit mode assignment and note which modes are allowed for each dominant mode assignment. For example, it is typical that a commuter-rail path-building script will allow all other modes to be included. This indicates that commuter rail has the top priority.

To build the observed trip tables, identify a table sequence based on time of day, access mode and hierarchical modes. For example, the table set below addresses a simple transit system and model with only local and express buses, walk and drive access and a peak and off-peak time stratification:

<u>Table</u>	<u>Time of Day</u>	<u>Access mode</u>	<u>Hierarchical mode</u>
1	Peak	walk	Express Bus
2	Peak	walk	Local Bus
3	Peak	drive	Express Bus
4	Peak	drive	Local Bus
5	Off-Peak	walk	Express Bus
6	Off-Peak	walk	Local Bus
7	Off-Peak	drive	Express Bus
8	Off-Peak	drive	Local Bus

Once defined, add an attribute to each observed trip record identifying the table number. Then generate a database that contains the following fields:

1. Production zone
2. Destination zone
3. Table
4. Linked trip expansion factor
5. Unlinked trip expansion factor

One observed trip per record is generated. This can then be used to create a CUBE/Voyager format set of trip tables using the CUBE “matrix” module using the “pattern” keyword set to “IJM:V” will generate the appropriate table matrices from trip records into one matrix file. Two sets (i.e. two files of multiple tables) of trip tables should be generated – one based on the linked trip expansion factor, which will be used in the model assignment, and one based on the unlinked trip expansion factor. The latter is not assigned, but can be compared with a “boardings” skim from the model assignment of the linked trips to compare specific transfer rates by zone-to-zone geography at a district level. The latter analysis might reveal transit coding or path-building parameter adjustments. Once assigned, the loadings by route should be summed and compared with independent estimates of transit boardings by route. Unassigned trips should also be noted, as these reveal either errors in geocoding previously undetected, or errors in transit network coding or path-building parameters.

Finally, with reference to the updated trip records, further summaries should be generated to reveal frequencies and distributions of transit trip-making across mode, time period, access mode and geographic orientations. These summaries should be reviewed for reasonableness against other known observations.

7 Assignment of the Survey

One of the most powerful uses of the survey data is to assign the expanded dataset to the transit network. This technique has many benefits. It can help assess the survey’s integrity, detect network coding errors and issues with network coding schemes, and validate transit path structure and weighting of the travel components. Each of these benefits is discussed in this section. It should be noted that a well-executed survey and geocoding process is required to maximize the benefits of this technique.

Executing the technique is straightforward:

1. Create a series of trip tables from the survey records by access mode, sub-mode and time period. Further stratification by market segment (e.g., auto ownership, income levels) may also be desired,
2. Assign the survey trip tables to the relevant transit paths and time period,
3. Report and analyze the resulting metrics, and
4. Repeat the second and third steps as issues are detected and addressed.

7.1 Metrics and Analysis

Unassigned Records. A key metric in assessing the survey's integrity is the number of observed trips (or survey records) unable to be assigned to the network. This metric can be stratified by agency, line-haul mode, access mode, and time period to help the analyst identify patterns that help to quickly identify issues.

Reviewing unassigned records may reveal issues with address reporting or geocoding procedures, or route sequence reporting. Survey records with incorrect address reporting can report zero or very high walk- (5+ miles) or auto- (20+ models) distances that may not be actually experienced in the reality.

Network coding issues may also be discovered. For example, a walk-access transit trip may be unassigned because the origin zone does not have a walk access connector. If a sufficient number of such trips exist in the dataset, this could either be a problem with the walk-access connector procedures and assumptions or network coding. The procedures could be unintentionally prohibiting the creation of a connector. Alternatively, the network coding could be incorrectly representing the transit system in the area around the origin zone.

Finally, path characteristics can be verified by reviewing the unassigned records. Trips may be unassigned due to path settings, For example, actual trips that exceed the model's maximum threshold for travel time or the number of transfers may not assign to the network.

Estimated Transit Boardings and Transfer Rate. These metrics can be stratified by agency, line-haul mode, access mode, and time period. When compared to the corresponding observed values, the analyst can learn insights about potential issues.

Strong differences between observed and estimated transfer rates and transit boardings may reveal survey reporting issues. For instance, through reviewing these metrics it has been discovered in some surveys that riders' reported all potential routes rather than the routes used on their specific trip.

Reviewing these metrics can also reveal network issues. In one city, the assignment of the survey trip table produced too many local bus boardings as compared to express bus boardings. This occurred because the speeds for local and express buses were identical, although express buses had about 20% faster than the corresponding local bus service. The local bus service had more frequent service than the express services, so the assignment process allocated more riders to local buses. The solution to this problem was to correctly reflect the faster speeds of the express buses.

Reconsideration of path parameters and weights could also arise if differences in boardings and transfer rates exist. For example, high transfer rates may imply that the transfer penalties are too low (vice-versa for low transfer rates). High estimated transfer rates could also suggest that transit fares are not correctly being reflected in pathbuilding.

Records Using Correct Line-Haul Mode. This metric can be stratified by time period and access mode. The line-haul mode used per the model is compared to the line-haul mode reported from the survey. This metric can be used to confirm route sequencing, geo-coding, and network coding issues in the survey as discussed above.