University Transportation Research Center - Region 2

Final Report

Innovative Travel Data Collection Recommendations

Performing Organization: State University of New York (SUNY)

December 2016

University Transportation Research Center - Region 2

The Region 2 University Transportation Research Center (UTRC) is one of ten original University Transportation Centers established in 1987 by the U.S. Congress. These Centers were established with the recognition that transportation plays a key role in the nation's economy and the quality of life of its citizens. University faculty members provide a critical link in resolving our national and regional transportation problems while training the professionals who address our transportation systems and their customers on a daily basis.

The UTRC was established in order to support research, education and the transfer of technology in the field of transportation. The theme of the Center is "Planning and Managing Regional Transportation Systems in a Changing World." Presently, under the direction of Dr. Camille Kamga, the UTRC represents USDOT Region II, including New York, New Jersey, Puerto Rico and the U.S. Virgin Islands. Functioning as a consortium of twelve major Universities throughout the region, UTRC is located at the CUNY Institute for Transportation Systems at The City College of New York, the lead institution of the consortium. The Center, through its consortium, an Agency-Industry Council and its Director and Staff, supports research, education, and technology transfer under its theme. UTRC's three main goals are:

Research

The research program objectives are (1) to develop a theme based transportation research program that is responsive to the needs of regional transportation organizations and stakeholders, and (2) to conduct that program in cooperation with the partners. The program includes both studies that are identified with research partners of projects targeted to the theme, and targeted, short-term projects. The program develops competitive proposals, which are evaluated to insure the mostresponsive UTRC team conducts the work. The research program is responsive to the UTRC theme: "Planning and Managing Regional Transportation Systems in a Changing World." The complex transportation system of transit and infrastructure, and the rapidly changing environment impacts the nation's largest city and metropolitan area. The New York/New Jersey Metropolitan has over 19 million people, 600,000 businesses and 9 million workers. The Region's intermodal and multimodal systems must serve all customers and stakeholders within the region and globally.Under the current grant, the new research projects and the ongoing research projects concentrate the program efforts on the categories of Transportation Systems Performance and Information Infrastructure to provide needed services to the New Jersey Department of Transportation, New York City Department of Transportation, New York Metropolitan Transportation Council , New York State Department of Transportation, and the New York State Energy and Research Development Authorityand others, all while enhancing the center's theme.

Education and Workforce Development

The modern professional must combine the technical skills of engineering and planning with knowledge of economics, environmental science, management, finance, and law as well as negotiation skills, psychology and sociology. And, she/he must be computer literate, wired to the web, and knowledgeable about advances in information technology. UTRC's education and training efforts provide a multidisciplinary program of course work and experiential learning to train students and provide advanced training or retraining of practitioners to plan and manage regional transportation systems. UTRC must meet the need to educate the undergraduate and graduate student with a foundation of transportation fundamentals that allows for solving complex problems in a world much more dynamic than even a decade ago. Simultaneously, the demand for continuing education is growing – either because of professional license requirements or because the workplace demands it – and provides the opportunity to combine State of Practice education with tailored ways of delivering content.

Technology Transfer

UTRC's Technology Transfer Program goes beyond what might be considered "traditional" technology transfer activities. Its main objectives are (1) to increase the awareness and level of information concerning transportation issues facing Region 2; (2) to improve the knowledge base and approach to problem solving of the region's transportation workforce, from those operating the systems to those at the most senior level of managing the system; and by doing so, to improve the overall professional capability of the transportation workforce; (3) to stimulate discussion and debate concerning the integration of new technologies into our culture, our work and our transportation systems; (4) to provide the more traditional but extremely important job of disseminating research and project reports, studies, analysis and use of tools to the education, research and practicing community both nationally and internationally; and (5) to provide unbiased information and testimony to decision-makers concerning regional transportation issues consistent with the UTRC theme.

Principal Investigator(s):
Catherine T. Lawson, Ph.D.
Chair, Geography and Planning Department
Director, Lewis Mumford Center
Albany Visualization And Informatics Lab (AVAIL)
University at Albany (SUNY)
Albany, NY 12222
Tel: (518) 442-4775
Email: lawsonc@albany.edu

To request a hard copy of our final reports, please send us an email at utrc@utrc2.org

Mailing Address:

University Transportation Reserch Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: 212-650-8051
Fax: 212-650-8374
Web: www.utrc2.org

AVAIL
Albany Visualization And Informatics Labs

Z-14-04
Innovative Travel Data Collection

Recommendations
December 6, 2016

Albany Visualization And Informatics Lab (AVAIL)
Lewis Mumford Center, University at Albany, SUNY
Catherine T. Lawson, Ph.D.
Chair, Geography and Planning Department
Director, Lewis Mumford Center
Director, Masters in Urban and Regional Planning (MRP) Program
AS 218 1400 Washington Avenue
Albany, New York 12222
(518) 442-4775
lawsonc@albany.edu

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The contents do not necessarily reflect the official views or policies of the UTRC or the Federal Highway Administration. This report does not constitute a standard, specification or regulation. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

1. Report No. Z-14-04	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Innovative Data Travel Collection		5. Report Date Dec. 6, 2016
		6. Performing Organization Code
7. Author(s) Catherine Lawson, Ph.D		8. Performing Organization Report No.
9. Performing Organization Name and Address Albany Visualization And Informatics Lab (AVAIL) Lewis Mumford Center, University at Albany SUNY Albany AS 218 1400 Washington Avenue Albany, New York 12222		10. Work Unit No.
		11. Contract or Grant No. 49198-42-27 55606-02-27
12. Sponsoring Agency Name and Address NYMTC 25 Beaver Street New York, NY		13. Type of Report and Period Covered Final, Jan. 2015-March 2017
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This study had the following objectives:

1. To identify and clarify these two emerging effects – real time data and changing culture,
2. To identify the shifts in data collection and transportation modeling that must take place to
assist in identifying and forecasting travel behavior, and
3. To discuss the impacts of such operational shifts, both in cost and outcomes to provide NYMTC with the cost and efficacy impacts of incorporating these emerging tools."

To address these objectives, the research team at Albany Visualization and Informatics Lab (AVAIL), led by Dr. Catherine Lawson, PhD., from the University at Albany, conducted a literature review; a cost benefit analysis of current and emerging transportation data surveying and modeling methodologies; and produced a set of recommendations for the near-term and the longer-term.

The literature review and cost benefit analysis revealed certain facts about the state of travel data collection in the United States. The paper travel diary remains the predominant instrument for collecting travel data despite its well-documented shortcomings and high cost. GPS devices have grown in popularity but are used primarily as a supplement for paper travel diaries. The value of travel surveying via smartphones is no longer a strictly academic question as numerous agencies have used the smartphone in a travel survey, either as the primary survey instrument or in a subsample (Indiana, Oregon, Singapore, Boulder, etc.). Origin-destination tables used in travel demand models can be constructed from social media posts and call-detail records though these datasets often lack valuable information (such as reliable trip purpose or mode), present incomplete pictures of travel (especially in the case of social media), are prohibitively aggregated, or are not representative. Finally, the research team found it is possible to develop and deploy a travel demand model that does not use travel survey data for inputs, opening the possibility of eliminating or reducing the size of travel surveys.

This study introduces two categories to classify data collection efforts:

1. Active Data Collection – the use of self-report and surveying to generate data.
2. Passive Data Collection – the acquisition of existing data.

The research suggests three orientations toward travel data collection, each with their own risks and advantages that could satisfy NYMTC's modeling needs, while enabling future cost savings and/or increases in data quality. Two of these three orientations (or pathways) emphasize Active Data Collection strategies while the third emphasizes Passive Data Collection. These pathways are fluid and dynamic. They are not intended as a step-by-step guide to the future. Instead, they are intended to illuminate the data collection trajectory, highlighting opportunities and delineating the consequences, both positive and negative, of various data collection decisions.

Briefly, these pathways are:

1. Paper and Online Diary Travel Survey (with GPS or Smartphone Supplement)
2. Smartphone Diary Travel Survey (with Online Supplement)
3. Passive Data Collection (with Smartphone Supplement)

17. Key Words Data Collection, forecasting travel behavior, travel surveys		18. Distribution Statement	
19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 81	22. Price

Form DOT F 1700.7 (8-69)

CONTENTS

List of Figures and Tables

1 EXECUTIVE SUMMARY

The purpose of this study, as specified by New York Metropolitan Transportation Council (NYMTC), is to:

> "…Identify and describe rapidly emerging new methods of personal travel data collection, a first step in the development of travel models relevant to the mid-21st Century.

> This task assignment has the following objectives:

> 1. To identify and clarify these two emerging effects – real time data and changing culture,
> 2. To identify the shifts in data collection and transportation modeling that must take place to assist in identifying and forecasting travel behavior, and
> 3. To discuss the impacts of such operational shifts, both in cost and outcomes to provide NYMTC with the cost and efficacy impacts of incorporating these emerging tools."

To address these objectives, the research team at Albany Visualization and Informatics Lab (AVAIL), led by Dr. Catherine Lawson, PhD., from the University at Albany, conducted a literature review; a cost benefit analysis of current and emerging transportation data surveying and modeling methodologies; and produced a set of recommendations for the near-term and the longer-term.

The literature review and cost benefit analysis revealed certain facts about the state of travel data collection in the United States. The paper travel diary remains the predominant instrument for collecting travel data despite its well-documented shortcomings and high cost. GPS devices have grown in popularity but are used primarily as a supplement for paper travel diaries. The value of travel surveying via smartphones is no longer a strictly academic question as numerous agencies have used the smartphone in a travel survey, either as the primary survey instrument or in a subsample (Indiana, Oregon, Singapore, Boulder, etc.). Origin-destination tables used in travel demand models can be constructed from social media posts and call-detail records though these datasets often lack valuable information (such as reliable trip purpose or mode), present incomplete pictures of travel (especially in the case of social media), are prohibitively aggregated, or are not representative. Finally, the research team found it is possible to develop and deploy a travel demand model that does not use travel survey data for inputs, opening the possibility of eliminating or reducing the size of travel surveys.

This study introduces two categories to classify data collection efforts:

1. Active Data Collection – the use of self-report and surveying to *generate* data.
2. Passive Data Collection – the *acquisition* of existing data.

The research suggests three orientations toward travel data collection, each with their own risks and advantages that could satisfy NYMTC's modeling needs, while enabling future cost savings and/or increases in data quality. Two of these three orientations (or pathways) emphasize Active Data Collection strategies while the third emphasizes Passive Data Collection. These pathways are fluid and dynamic. They are not intended as a step-by-step guide to the future. Instead, they are intended to illuminate the data collection trajectory, highlighting opportunities and delineating the consequences, both positive and negative, of various data collection decisions.

Briefly, these pathways are:

1. Paper and Online Diary Travel Survey (with GPS or Smartphone Supplement)
2. Smartphone Diary Travel Survey (with Online Supplement)
3. Passive Data Collection (with Smartphone Supplement)

2 INTRODUCTION

Travel surveying efforts are currently a critical component in transportation planning as they provide the necessary data for transportation planning models and travel behavior analysis. Recent technological advances have contributed to changes in the procedures used for collecting travel data, and subsequent improvements in data quality. Looking forward into the next two decades, these technological advances could bring further changes, and therefore need to be reviewed, compared to current practice, and evaluated for appropriateness for deployment in the near-term and beyond.

This study is organized as follows: Section 3 provides a brief background on the history of travel surveys in the US, including NYMTC's recent data collection efforts. Section 4 describes the approach taken to identify current and emerging data collection options. It describes a data collection dichotomy (Active and Passive Data Collection) and presents a review of literature on various strategies within those paradigms. Section 5 provides an analysis of strategies/approaches that could be adopted and deployed as part of NYMTC's data collection process today. It includes a comparison of prominent active data collection strategies, such as paper travel diary, online diary, and smartphone diary. Additionally, Section 5 lays out in detail, the major considerations in effective travel surveying. Section 6 outlines the pathways for three general orientations toward data collection. Section 7 contains information on key pilot studies designed to address specific concerns relating to the adoption of novel data collection strategies. Section 8 includes future considerations and additional recommendations, followed by a section with conclusions.

3 BACKGROUND

3.1 INTRODUCTION

Travel surveys have been conducted in the United States for more than 40 years. For most metropolitan areas, travel surveys are the largest routine expenditure made from planning budgets (Stopher et al., 2008). Early survey efforts relied exclusively on participant's recall of a day's trips and required surveyors to mail or call participants to gather data. The development of computer-assisted personal interviewing (CAPI), computer-assisted telephone interviewing (CATI), and computer-assisted self-administered interviewing (CASI) helped streamline the survey process. In the mid-1990s, the emergence of GPS devices offered new opportunities for practitioners and academics to gather increasingly accurate spatial and temporal travel data while reducing the burden on respondents. Most recently, the proliferation of smartphones equipped with GPS and other sensing technologies across the globe promises to bring travel survey efforts to a new height of accuracy and participation.

Data collected in travel surveys is valuable to the management of transportation systems and informs major policy decisions conducted at all levels of government. Metropolitan Planning Organizations (MPOs)

across the United States use the data collected in surveys in travel demand models. In turn, output from travel demand models helps MPOs make informed decisions on how to prioritize major transportation projects. Moreover, travel demand models help MPOs satisfy the requirements of federal transportation regulations.

In response to stricter federal regulations, increasing demands on and for a diverse transportation infrastructure, and opportunities created by advances in computing, practitioners have developed complex travel demand models to predict traveler's behavior. Such models require high quality data across a wide range of variables. Practitioners seek to know not just when and where people are traveling but why they travel. Further, they seek to know the details of traveler's sociodemographic profile as well as their travel decision-making process. Such information is valuable in forecasting the demands made on the transportation infrastructure in response to social, economic, and physical changes in a region.

Nevertheless, declining cooperation rates threaten survey efforts. Cooperation rates suffer for a host of reasons—prior unpleasant experiences with surveys, skepticism regarding the efficacy of survey efforts, association of surveys with solicitation. Furthermore, many potential respondents are concerned about government use of data and are unwilling to share information with government agencies. Finally, many potential participants feel that they simply do not have time to fill out a survey.

Advances in survey instruments, location sensing technologies, and large-scale data collection have enabled new methodologies for capturing travel data. These new methodologies hold great promise in capturing more accurate and diverse data at reduced cost and effort. Although many have not undergone the rigorous testing required to replace current data collection strategies, preliminary evidence hints at the value these strategies may provide to data collection and modeling efforts.

3.2   NYMTC TRAVEL DATA COLLECTION ENVIRONMENT

Data collection in the NYMTC region must be able to account for the unique sociodemographic and physical characteristics of the greater New York metropolitan region in order to be effective. Gong et al. (2012) writes, "With eight million people living in Manhattan, Brooklyn, Queens, Bronx, and Staten Island, NYC is the largest city in the United States. It has a population density over 10,000 per square kilometer (Demographia, 2010) and the most comprehensive transit system that carries an annual ridership of over 1.5 billion by subway and 0.7 billion by bus (MTA, 2009). Manhattan, in particular, is highly developed, with 220 high-rise buildings over 150 m in height (Emporis, 2010) and a population density of about 27,000 per square kilometer (Demographia, 2010)."

New York's high density and prevalence of tall buildings and a complex multimodal transportation system with an extensive subway network makes gathering location data from GPS units and call detail records (CDRs) challenging. For traditional paper travel diaries with GPS supplements, GPS traces are invaluable for obtaining an accurate and complete record of travel (Wolf, Bachman, Oliveira, Auld, Mohammadian, & Vovsha, 2014). Therefore, data collection strategies reliant on GPS and CDR may require supplemental survey instruments to achieve an acceptable data quality.

Data collection efforts are further frustrated by the interconnected, multimodal nature of the transportation network. Users of the transportation network routinely use multiples modes to travel to and from work, home, and other locations. Traditional self-reporting through paper travel diary can place great burden on survey participants when asked to recall multiple modalities. Novel data collection strategies attempt to address this challenge by shifting the burden of reporting trip characteristics to passive location-sensing technologies and algorithmic inference. Despite improvements, however, the inaccuracies of algorithmic inference—especially for distinguishing certain modalities (e.g. private car and taxi)—limit data quality (Cottrill et al., 2013).

Finally, travel data must be collected from New York City's large commuter population. In the past, NYMTC has coordinated survey efforts with New Jersey Transportation Planning Authority (NJTPA) to account for the large number of home-to-work trips that begin in the NJTPA jurisdiction and end in the NYMTC jurisdiction. Future data collection efforts will likely necessitate ongoing coordination with the NJTPA to accurately capture travel behavior in the region.

3.2.1   Review of NYMTC Travel Survey 2010-11

The New York Metropolitan Transportation Council (NYMTC) and the New Jersey Transportation Planning Authority (NJTPA) partnered in a 28-county Regional Household Travel Survey (RHTS) in 2010-11. The RHTS was conducted to obtain demographic and trip data from a sample of the ~7.9 million households in the NY-NJ-Connecticut metropolitan area. Data from the RHTS is used to drive the New York Best Practices Model (NYBPM), the travel demand model that helps NYMTC meet federal requirements and make important transportation decisions for the region.

Surveyors collected travel data for a 24-hr period from 18,965 households (HH), representing 0.24% of the HH in the NYMTC region. Demographic variables for individuals and households were also collected. A subsample of 1,930 HH agreed to wear GPS devices in addition to reporting their travel behavior. Respondents were recruited by computer-assisted telephone interview (CATI) or mail. Their data was retrieved by CATI, mail, or TripBuilderTM, proprietary software that facilitates travel data collection. Survey materials were offered in English, Spanish, Russian, and Chinese, the four main languages spoken in the region.

NYMTC HOUSEHOLD TRAVEL SURVEY 2010-2011

Households in NYMTC Region: ~7.9 Million

Households Invited to Participate: 711,551

Households Completed the Recruitment Interview: 31,156

Households Completed Survey: 18,965

Household Survey Yield: 2.67%

Households Completed of Recruited HH: 61%

Survey Sample as Percent of Regional Households: 0.24%

NYMTC Survey Region Population: ~12 Million

Total Participants in RHTS: 43,558

Survey Sample as Percent of Population: 0.36%

Type of Survey: Paper Travel Diary, Telephone Retrieval

GPS Subsample for Weighting/Correction: 1,930 Households

Number of Data Elements: ~150

Cost: ~$3.9m

Figure 1 -- NYMTC Household Travel Survey 2010-2011 Statistics

3.2.2 New York Best Practices Model (NYBPM)

NYMTC utilizes the New York Best Practices Model (NYBPM) for modeling trips based on travel survey results. The NYBPM is NYMTC's activity-based travel demand model. Inputs into the NYBPM dictate which data needs to be collected about travelers in the region. Presently, the NYBPM requires numerous trip characteristics and demographic data including Origin Location, Destination Location, Departure Time, Arrival Time, Trip Purpose, Mode, Parking Info, Toll Info, Travel Party, Household Size, Vehicle Ownership, Income, etc. In the past, these data were collected in massive regional household travel surveys (RHTS).

Survey data is used to "revise the behavioral relationships that comprise the core choice models of the NYBPM (auto availability, tour frequency, and destination, mode, and stop choice)" (NYMTC 2014). These models give the probability of a traveler selecting activities and making travel choices based on availability and utility. Therefore, the collection of behavioral data that drives these core choice models is necessary for executing model runs with the NYBPM and forecasting future travel behavior.

Until recently, these data could only be gathered through surveys. It may be possible, however, to revise these core choice models through the acquisition and fusion of large data sets. Researchers from Transport Foundry and Parsons Brinckerhoff, for instance, used "household-level data, firm-level data, origin-destination data, travel time data in traffic, and the National Household Travel Survey" to create a simulation of person-level tour-based travel (Kressner et al., 2016). Such an approach to modeling and data collection, if effective, could drastically reduce costs and labor. This "Big Data" approach is discussed further below.

3.2.3 NYMTC Data Needs

Data collected in the RHTS is used primarily to revise the statistical estimation of choice models and secondarily to calibrate and validate the model.

Data collected in the RHTS is stored in eight relational database files (household, person, vehicle, place, unlinked trips, linked trips, tour, and subtour). There are over 100 data items in the household, person, vehicle, and place files. Table 1 provides a list of these data items.

Table 1 -- List of Travel Survey Variables Collected and Computed in NYMTC RHTS 2010-11

Household		
Mode Recruitment	Household Income	Home Zip Code
Mode of Retrieval	Day of Week of Travel	County Group: Level 1
County of Residence	Date of Assigned Travel	Regional Boundaries Level 2
Sampling Bin (Geographic Flag)	# Household Vehicles	Political Boundaries: Level 3
GPS Sub-Sample Flag	# Household Students	Home TAZ (NYMTC)
Sample Type	# Household Workers	Home Census Tract 2010
Language of Interview	# Household Driver's Licenses	Flag for Partially Completed Households
Residence Type	# Household Children	Level 2 Weights
Household Size	# Household Trips (Computed)	Household Structure

Land-Based Telephone Service	Willingness to Participate in Future Surveys	# Household Trip by different modes
Household Language	Home State	# Household Trip by Work / Non-Work Purpose

Person		
Gender	Work Location TAZ (NYMTC)	Mode of Transport to School
Age	Work Location Census Tract (2010)	Frequency of Bike Travel to School
Driver's License Status	Work Location County	Typical Travel Time to School
Availability of Cellular Phone	Mode of Transport to Work	Travel Diary Completed
Relationship to Head of Household	Frequency of Bike Travel to Work	Have Completed Diary
Hispanic Origin	Typical Travel Time to Work	Proxy Reporting Flag
Race/Ethnicity	Commercial Driving	#Person Trips (Computed)
Disability Status	Household Vehicle Number Commercially Driven	Reason for No Travel
Type of Disability	Employer Transportation Benefits Offered	Flag for Partially Completed Persons
Employment Status	Work Start/End Times	Level 2 Weights
Volunteer Status	Flexible Work Schedule	County Group: Level 1
Work Status (computed)	Work Start/End Time Variation	Regional Boundaries: Level 2
Unemployment Status	Student Status	Political Boundaries: Level 3
Job Hours	Grade Level Attending	Life Cycle Status
Telecommute Hours	School Location (Home/Other)	#Person Trip by Different Modes
Compressed Work Week	School State	# Person Trip by Work/Non-Work Purpose
Industry	School Zip Code	
Occupation	School Location TAZ (NYMTC)	
Employer	School Location Census Tract (2010)	
Work Location (fixed/varies)	School Location County	
Work State		
Work Zip Code		

Vehicle		
Year of Vehicle	E-Z Pass Tag	Level 2 Weights
Body of Vehicle	Vehicle Used on Travel Day	
Type of Fuel	Reason Vehicle Not Used	

Place		
Place Name	# of Toll Facilities Used	State
Primary Trip Purpose	Name of Toll Facilities	Zip Code
Secondary Trip Purpose	Interchange Used to Enter Toll Facility	Speed Check Flag
Transportation Mode	Interchange Used to Exit Toll Facility	Flag for Long Trip Duration
Total People Traveling	Toll Payment Method	Location Imputed by NYMTC/NJTPA
Household Members on Trip	Route/Line # for Transit Trips	Missing Transit Information
Person Number on Trip	Transit Service Used	Note for NYMTC/NJTPA Data Review
Non-Household Members on Trip	Transit Fare Type	Flag Indicating Record was Changed During Data QC
Vehicle Number	Transit Fare Cash Amount	Level 2 Weights
Exit Vehicle at Place	Transit Pass Type Used	County Group: Level 1
Parking Location	Type of Unlimited Ride MetroCard	Regional Boundaries: Level 2
Parking Description	Transit Pass Cost/Unit	Political Boundaries: Level 3
Pay to Park	Arrival Time	County FIPS
Amount Paid to Park	Departure Time	TAZ (NYMTC)
Pay to Park Unit	Activity Duration (calculated)	Census Tract 2010
Paid Fee for Toll road/Bridge/Tunnel	Trip Duration (calculated)	
Toll Road or Bridge/Tunnel	Trip Distance (Straight-line)	

The important variables collected in the household travel survey are demographic variables for individuals and households (including income, vehicle ownership, gender, etc.) and trip characteristics (such as origin, destination, activity, and mode) have traditionally been obtained directly from individual households, but could also be gathered using recent technological advances.

4 EMERGING DATA COLLECTION OPTIONS

4.1 METHODOLOGY REVIEW

The research team conducted a literature search using Google Scholar to search for instances of planning organizations where novel or innovative travel data collection methods were employed. Search terms included: smartphone household travel survey; mobile phone trace household travel survey; GPS household travel survey; Facebook OR Twitter OR Instagram OR Foursquare AND household travel survey; and Google Location History household travel survey. Additionally, a search of the Department of Transportation website for Metropolitan Planning Organizations (MPOs) of comparable size and population to the New York Metropolitan Transportation Council was conducted. The method used for travel data collection in each of these organizations most recent HTS was identified.

4.2 ACTIVE DATA COLLECTION STRATEGIES

New technologies and shifting attitudes are likely to drive a paradigm shift in HTS in the near future. Smartphones, with their numerous sensing devices and ability to connect to the internet, facilitate the collection of high quality location data. Moreover, smartphone applications and data processing algorithms can simplify the collection of demographic variables and trip elements that are traditionally harder to capture through existing passive sensing technologies, such as standalone GPS devices.

In recent years, numerous agencies throughout the world have used smartphones to supplement HTS and gather mobility data. There is a growing list of smartphone travel survey applications including: Future Mobility Survey (FMS), SITSS, rMove, CycleTracks, GPS-ATD, Florida Trip Tracker, Mobile Market Monitor, Motion X GPS, and DVMobile. These smartphone travel survey applications can reduce respondent burden, reduce the time spent filling out a survey, and provide higher quality and more accurate data.

4.2.1 Using Distributed GPS Devices in Travel Survey

The advent of GPS technology fundamentally changed the collection of travel data. In 1996, the Federal Highway Administration (FHWA) assisted the Lexington Area Metropolitan Planning Organization (MPO) in the first GPS-supported household travel survey (HTS) in the United States (Wolf et al., 2014). In the ensuing years, comparative analysis of data from GPS devices and traditional pen-and-paper travel diaries revealed a number of human errors in self-reporting of travel behavior (Wolf et al., 2014). Nevertheless, the adoption of GPS devices as a replacement for pen-and-paper travel diaries was and remains limited by methodological challenges.

Use of early GPS devices was limited by their physical and technical properties. GPS devices were bulky with high rates of battery consumption. These properties largely restricted their use to in-vehicle. As the automobile is only one mode of transportation within the larger transportation infrastructure, the findings were of limited use to planners (Huang & Wang, 2014).

Another challenge facing MPOs seeking to use GPS devices for travel surveys is the cost associated with purchasing, distributing, collecting, and replacing devices used by participants. Though prices have dropped considerably for GPS devices in recent years they remain an expensive budget item for MPOs seeking the higher quality data they provide.

Additionally, GPS devices do not remove the possibility for human error in data collection. A wearable GPS still requires the participant to remember to bring the GPS with them on the day of their travel. This may become especially problematic for MPOs seeking travel data for more than one day.

Lastly, the efficacy of GPS devices can be limited by the lack of clear signal pathways from the device to the satellites that inform their data. In large metropolitan areas, this may be especially problematic for capturing travel data from users of underground transit systems as well as those who travel in areas with a high density areas of tall buildings (Gong et al., 2012). The urban canyon effect created by skyscrapers is well-documented across numerous disciplines.

GPS devices are now an important supplement to the large travel surveys conducted by most MPOs every 5-10 years. The dual method approach, so called because it requires a subsample of the survey population to complete a travel diary and use a GPS device (Wolf et al., 2014), is widely regarded as the best way to collect high quality travel data. In NCHRP Report 775 Applying GPS Data to Understand Travel Behavior the authors point out the numerous advantages afforded by using GPS technology:

> GPS-enhanced surveys provide a more accurate and detailed account of the spatial and temporal aspects of personal travel than what survey respondents are able to recall and report, and GPS data sets have been used to correct significant trip underreporting errors associated with pen-and-paper or phone-based activity survey … GPS-enhanced surveys should have less respondent burden for capturing travel details by leveraging passive GPS data collection while collecting more information and more accurate information. In addition, by further reducing respondent burden through the use of automated activity type, location, timing, and travel mode identification routines, GPS-based prompted-recall surveys allow for more complex questions to be asked. …the combination of more accurate spatial-temporal data along with reduced respondent burden allows for multiday data collection, which in turn enables more in-depth aspects of travel behavior to be studied, including variability in travel patterns, route choice, activity location selection, and mode selection. Furthermore, multi-day data collection can support reductions in required sample sizes, thereby offsetting some, if not all, of the additional costs inherent in GPS-enhanced and GPS-based travel surveys (Wolf et al., 2014).

4.2.2 Smartphone

Practitioners and academics throughout the transportation industry view the smartphone as the next evolution of passive sensing technologies that support reduction in respondent burden while improving the quality of travel data. Practitioners note that smartphones can reduce the common implementation problems associated with GPS-based travel surveys by, "(a) eliminating the need to ship out and retrieve GPS loggers, (b) shortening the time between travel date collection and data review, and (c) reducing costs associated with equipment loss" (Wolf et al., 2014). Furthermore, their increasing market penetration, now is greater than 58% nationally ("Smartphone ownership, over time," 2014) coupled with the fact that users rarely travel anywhere without them, make smartphones an ideal tool for collecting travel data. However, it may still be necessary distribute and collect smartphones from underrepresented populations.

Supplementary sensing technology present within most smartphone devices can attenuate the clear pathway limitations of GPS devices. Accelerometers, GSM, gyroscopes, and compasses, amongst other sensors, provide valuable data on the orientation and movement of the phone and, by proxy, the participant. By combining data from these different sensors and processing this data with various algorithms, researchers can make compelling inferences in cases where GPS data is limited.

Figure 2 -- Smartphone Ownership, Over Time.

Source: [Online image]. (2014) Retrieved December 3, 2015 from http://www.pewinternet.org/2014/02/27/part-1-how-the-internet-has-woven-itself-into-american-life/

In recent years, numerous agencies throughout the world have used smartphones to supplement household travel surveys (HTS) and gather mobility data. In Singapore, the Future Mobility Survey (FMS) app was developed and administered to a subset of nearly one thousand participants of the nation's Household Interview Travel Survey (HITS) (Cottrill et al., 2013). Atlas II, an app developed for the New Zealand Smartphone-based Individual Travel Survey System (SITSS), collected data from 73 users as part of an exploratory analysis of the feasibility of a nationwide smartphone-based travel survey (Safi et al., 2015). In Sydney, smartphone sensing technology was used to examine the impact of the new bicycle infrastructure on inner city travel patterns (Greaves et al., 2014). Finally, in Indianapolis, approximately 300 participants, in a 2014 HTS, also participated in the "In the Moment" Travel Study administered via smartphone and characterized by a "random moments" sampling paradigm (Greene et al., 2015.)

By comparison to traditional pen-and-paper HTS, travel surveys using smartphones reduce respondent burden, increase accuracy of trip start/stop times and trip duration, decrease underreporting of short and walking trips, and increase cost-effectiveness of collecting multiple days of travel data (Cottrill et al., 2013; Safi et al., 2015; Greaves et al., 2014; Abdulazim et al., 2013).

Collecting usable travel data through smartphones must address the following user and agency/modeler concerns:

- demands on battery and data plan usage;
- privacy and management of respondent data;
- representativeness of sample population;
- data gaps in user travel day(s); and
- travel and activity inference/detection errors (Cottrill et al., 2013; Nitsche et al., 2012; Zhao et al., 2015a; Zhao et al., 2015b).

In order to elucidate the typical architecture of smartphone travel data collection and processing, the research team review a representative example, the Future Mobility Survey (FMS) system. The FMS system collects and processes data using three primary components: smartphone app, servers, and a web browser (Cottrill et al., 2013; Zhao et al., 2015b). The app collects data from the phone's sensing technologies and periodically sends it to the server. The server uses algorithms and contextual information to process the data. The web interface is the portal through which the user accesses and validates the processed data.

The FMS app used phased sampling, a technique that creates sleep/awake cycles for the GPS, to address battery consumption concerns (Cottrill et al., 2013). During awake cycles the app collected high-frequency (1-Hz) GPS data. During sleep cycles the GPS sensor is off. Phased sampling was further enhanced by using the phone's
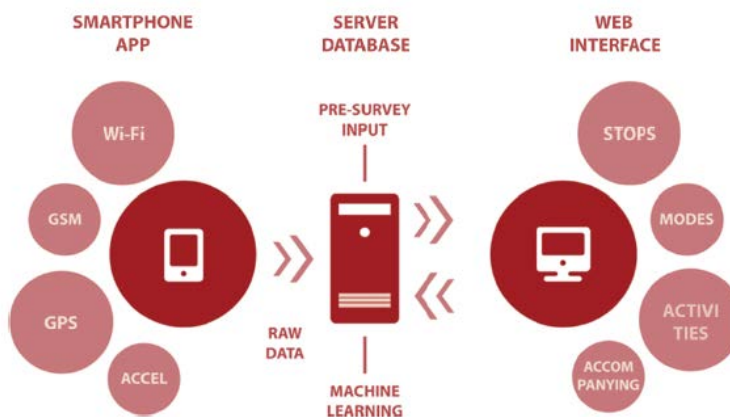
Figure 3 -- Survey Architecture

Source: Cotrill et al., 2013, Future Mobility Survey. Transportation Research Record: Journal of the Transportation Research Board, 2354, 61.

other sensors to trigger the GPS to collect data upon trip start. Finally, GPS was adjusted to longer sleep cycles at spatiotemporal points where the user was likely to remain for long periods of time (e.g. home or work). Despite nuanced strategies to manage the trade-off between battery consumption and data needs a reduction in data quality during sleep periods was observed (Cottrill et al., 2013). Participant data plans were managed by providing the user the option to upload sensor data using mobile, Wi-Fi, or both (Cottrill et al., 2013).

Development of a clear and accessible privacy agreement can alleviate concerns over data collection. Though some users expressed privacy concerns by participating in the minimal amount of days or not inviting other members of their household to participate in the survey, few users expressed concern over their data being used for other research purposes (Cottrill et al., 2013).

The use of smartphones for travel surveys introduces a sampling bias toward a younger population. Such a bias is largely consistent with analysis of GPS representativeness documented by Nitsche et al. (2012) who noted that willingness to use GPS is biased toward younger, male, and healthy participants as well as households with high income, car ownership, and tech equipment. However, such a bias could be mitigated by using smartphone travel surveys as a supplement to traditional travel surveys which typically capture elderly populations better than younger populations. Furthermore, offering online or telephone support to less tech-savvy participants may drive up response rates and decrease bias. Additionally, GPS loggers could be distributed to underrepresented portions of the population (Nitsche et al., 2012).

Smartphones and GPS reduce respondent burden by passively providing location data. However, the contributions location data can make to emergent activity-based transportation models are limited unless mode choice and trip purpose can also be inferred (Ghorpade

Figure 4 -- FMS Web-interface – Activity Diary

Source: Zhao, F., & Ghorpade, A., 2015, Stop detection in smartphone-based travel surveys. Conference Paper, 5.

et al., 2015; Kim et al., 2015). Thus far, attempts to address this challenge have employed algorithms and contextual information to infer mode selection and trip purpose from the acquired data. Despite elegant data manipulation and processing, errors in mode detection and trip purpose persist (Nitsche et al., 2013). Mode detection errors, for instance, occur more frequently when trying to discriminate between modes that share similar speed and acceleration profiles and in high density urban settings that lack accurate GIS contextual information (Rasmussen et al., 2015).

In order to address the limitations of inference, the FMS employs a validation structure. The validation structure of the FMS system trades a small increase in participant burden (relative to a completely passive inferential system) for increases in data accuracy (Cottrill et al., 2013). For each travel day, participants are presented with a map of their travel and activities and asked to validate the system's inferences on mode selection, route selection, start/stop times, and activity. Such validation decreases inference errors and adds further contextual information for data processing to improve upon future inferences (Cottrill et al., 2013).

Moreover, relative to a traditional pen-and-paper travel survey, the validation structure employed in smartphone apps improves data accuracy while reducing the number of questions a participant must actively fill out. Time-consuming travel diary prompts requiring the participant to fill out start/stop times

Figure 5 -- System Architecture

Source: Ghorpade et al., 2015, An integrated stop-mode detection algorithm for real world smartphone-based travel survey. Conference paper, 3.

and locations, mode choice, and trip purpose for each and every trip are removed when using smartphone. Instead, users simply assess whether the algorithmic inferences made in data processing are accurate and are then asked to validate them. This system results in an overall reduction in the number of questions asked by the survey form and time spent filling out the form while improving data quality and accuracy (Zhao et al., 2015a).

4.2.2.1 Smartphone Applications

The following is a list of smartphone applications that have been developed and employed to capture travel data. Some, like the FMS and rMove, were designed with the intention of supplementing or replacing the predominant large-scale regional HTS methodologies. Others, such as CycleTracks, were designed to capture travel data from specific groups or for a narrower research purpose.

- Future Mobility Survey (FMS)
- SITSS
- rMove
- CycleTracks
- GPS-ATD
- Florida Trip Tracker
- Mobile Market Monitor
- Motion X GPS
- DVMobile

4.3 PASSIVE DATA COLLECTION STRATEGIES

The ability to capture, store and retrieve device-generated data makes it possible to infer behaviors that previously required direct responses from household members. For example, through the collection and visualization of GPS devices moving on the network, it is possible to infer a household member left their home and traveled to work as the start location of the device was their residential location and the next long period location was their work location. While more aggressive processing is required to identify the patterns that are used in the inference, most, it not all, of the processing can be automated through machine-learning algorithms. Additionally, characterization of household members may be possible to construct using marketing data. Taken together, algorithmic inference and fusion of demographic characteristics can allow researchers and practitioners to generate synthetic populations without needing to directly survey individuals.

4.3.1 Mobile Phone Trace

Mobile phone traces are generated every time a mobile device attempts to connect to the communication network. Traces are timestamped and offer an estimation of the mobile devices location at the time of the trace. Mobile phone penetration, as defined by the percentage of adults who own a cell phone, is over 80% in the United States and devices connect to the cellular network with increasing regularity due to shifts in communication norms and web-based application availability. As a result, mobile phone traces are huge datasets that hold valuable insights into human mobility. Moreover, location data is gathered without placing any burden on the mobile phone user. These characteristics make mobile phone traces an attractive source for travel behavior information. Much research in this field focuses on developing methods for using mobile phone trace in transportation modeling.

In 2013, Jiang et al., analyzed and processed mobile phone traces from a 2-month period in 2010 in the Boston metropolitan area to identify stays (activities) and pass-by's (travel). Their goal was "to develop methods to reproduce previous findings on road usage with a platform that is computationally integrated and publicly available." The authors developed methods for tying land-use data to geographic coordinates sourced from the mobile phone trace to identify likely activities at the stay locations.

Figure 6 -- Cell Phone Ownership, 2000-2014

[Online image]. (2014) Retrieved December 3, 2015 from http://www.pewinternet.org/2014/02/27/part-1-how-the-internet-has-woven-itself-into-american-life/

Another study looked at the effectiveness of using mobile phone traces for urban modeling. Along with the mobile trace data, the study aggregates supplemental data from sources like census tracts and odometer readings from safety inspections to develop algorithms that reveal variables like trip and home locations at a 500m by 500m grid cell level (Calabrese et al., 2013). Without processing, raw trace cellular data presents obstacles to researchers; it is often without "socio-economic and demographic attributes,

might not represent a random sample of the population, and the data-sets will not originally be packaged with modeling purposes in mind" (Calabrese et al., 2013). The study proposes a methodology for processing mobile trace data for optimal use towards transportation modeling. The proposal was developed after a study of one million mobile phone users within the Boston metro area over the course of three months.

Liu et al. (2013) used a set of machine-learning algorithms to analyze mobile phone traces from 80 users over one year. With post-processing algorithm, they achieved a 76.6% activity prediction accuracy for activities. Their work illustrates the potential of annotating mobile phone trace data with activities. However, the sociodemographic requirements of activity-based travel demand models pose a major challenge for replacing traditional survey methodologies with mobile phone trace data, which fails to capture and associate demographic variables with accompanying location information.

4.3.1.1 Location-based Social Networks (LBSN)

Location-based social networks (LBSN) are an area of interest for transportation researchers and practitioners. Geotagged posts to a social network can provide a record, albeit incomplete, of a user's mobility.

Lee et al. (2015) analyzed large-scale mobility data from Twitter against the output of a travel demand model from a regional MPO in southern California. The authors used the following five-step approach: "1) Harvest geotagged tweets and extract OD-pairs; 2) Examine spatial zoning levels for aggregating OD pairs for the comparison; 3) Test the correlation between twitter based trips, trip-based and activity-based travel demand model trips using multivariate methods; 4) Create a conversion method that builds OD matrices from tweets; and 5) Develop spatio-temporal distributions of travel to different types of activities for the entire region at different days of the week" (Lee et al., 2015).

In another study, researchers constructed origin-destination tables from Foursquare check-in data and compared the results to model output from a regional MPO. They conclude that LBSN data "have better spatial and temporal coverage, built-in user verification, real-time updating capability, and much lower data collection cost" than do traditional O-D estimation methods (Jin et al., 2013).

Nevertheless, a review of the literature suggests that no LBSN survey methodology has been developed that can effectively and consistently capture the complex and diverse trip and demographic variables captured by more traditional HTS methods.

4.3.2   Passive Data Model

Travel demand models can be constructed using third party data sources obviating the need for travel survey data. By purchasing or acquiring O-D data, O-D travel time data, and consumer marketing data (which includes demographic data) and applying data science techniques, it is possible to create a synthetic population and synthetic travel diaries for use in a travel demand model. Kressner et al. (2016) developed one such model using household-level data, firm-level data, O-D data, O-D travel time data, and the NHTS for the French Broad River MPO (FBRMPO) region in North Carolina. Model output was compared with the FBRMPO aggregate trip-based model using standard validation measures. At least 90 percent of the passive data model counts were less than the maximum desirable deviation. Model development times were compared (452 hours for FBRMPO model; 75 hours for passive data model).

Further research is required to determine the capacity of the tour-based model to handle multimodal trips or a more complex modeling environment. Additionally, future research could address many interesting questions. For instance, could this approach be used to assist in forecasting and scenario analysis? How might different data sets be used in the NYBPM modeling process and what assumptions might be tested (e.g. can O-D travel time data provide insight into travel time reliability over days of the week or months of the year)? What are the advantages and disadvantages of household and firm data from third party marketing companies and those in the ACS PUMS? What other data sets (e.g. LBSN data) might be used to create a more accurate and more predictive model?

Using passive data in modeling has the potential to eliminate household travel surveys from the travel demand modeling process thereby drastically reducing data collection expenditures for NYMTC. Although initial research suggests the passive data model can reduce cost and model development times, further research will be required to understand whether a passive data approach is viable in a region as large and complex as the New York metropolitan area.

5 COMPARATIVE ANALYSIS

5.1 METHODOLOGY USED TO COMPARE DATA COLLECTION OPTIONS

In order to compare different data collection strategies, the research team created a cost-benefit analysis spreadsheet. The spreadsheet contained all potential data collection strategies, including survey methodologies and "Big Data" approaches that might be executed in the NYMTC region. After consulting with NYMTC, AVAIL researchers chose to differentiate between survey instrument technologies (e.g. paper questionnaires) and location technologies (e.g. GPS or cellphone trace) to create cost-benefit analyses for each of these technology categories. The types of survey instruments are Smartphone, Online Diary, Paper Travel Diary with Mail Retrieval and Paper Travel Diary with Telephone Retrieval. The types of location technologies are Smartphone, Bluetooth, GPS, and Cellphone Trace.

The AVAIL research team then used the ratings of the survey and location technologies to construct a cost-benefit analysis of survey methodologies. Every survey methodology includes at least one survey component. Many of the feasible methodologies include a supplemental location technology.

Survey instruments, location technologies, and travel survey methodologies are rated on attributes. The primary categories under which these attributes fall are: (1) Survey Data, which includes trip characteristic and demographic data; (2) Survey Instrument Characteristics, which covers attributes related to survey material versatility, data formats, commercial availability and outlook; (3) Respondent Experience, which addresses attributes like respondent burden and equity; (4) Administrative Experience, which deals with data management and cleaning, managing survey materials, retention rates, and comparisons; and (5) Pricing, which we estimate based on a market scan and documentation of previous surveys. Definitions and the importance of each attribute in the five primary categories are provided.

5.1.1 Comparative Analysis Inclusion Parameters

To focus on the appropriate new sources amongst a seemingly endless stream of candidate technologies, this cost/benefit study will be limited to those that could be deployed in the NYMTC region within the next five years.

The recent attention to potential uses of "Big Data" being produced from a variety of new technologies has captured the imagination of transportation researchers and professionals. These innovative, "reality reflective," accurate, and inexpensive sources of data could offer innovative and novel solutions to the challenge of collecting data for travel demand modeling. Given that this approach is based on modern data science techniques and available passively collected datasets, it may be ready for deployment in the next five years. The bulk of research on "Big Data" solutions for data collection conclude that these large data sets, typically containing location data used to build origin-destination matrices, are helpful for travel demand model calibration and/or validation, but cannot yet replace survey data (Ozbay et al., 2014). A "proof-of-concept" study in Asheville, NC, however, challenges this assumption (Kressner et al., 2016). The fusion of behavior, sociodemographic, and location data may be sufficient to revise the relationships that govern the core choice models of the NYBPM and eliminate the need to collect survey data.

5.1.2 Comparative Analysis Structure

In the comparative analysis, AVAIL compared each methodology on its current features and, where necessary, its potential and ideal features. This includes cases where the technology is permissive of further development.

AVAIL deployed a pyramid structure to the comparative analysis. At the base of this pyramid are survey instrument technology and location technology comparisons rated on an interval scale (0-100). These scores serve as the foundation for comparing methodologies on that same interval scale (0-100). The ordinal rankings that serve as the top of the pyramid, are the rankings of each of the nineteen survey methodologies from best interval scale score to worst (1-19).
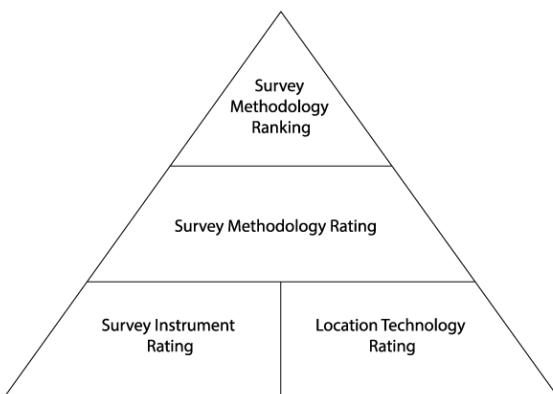
Figure 7 -- Organization of the Comparative Analysis Scoring

The research team organized the comparative analysis as a spreadsheet where columns house a single technology or survey methodology and rows house a single travel survey attribute. Figure 8 is an excerpt from the spreadsheet. It shows these attributes as well as one of the survey methodologies, "Smartphone Survey."

5.1.3 Scoring

In order to rank each survey from best to worst, AVAIL first implemented an interval scale of 0 – 100. For example, a survey methodology will score a 100 in the Survey Data attribute if it generates the most accurate and complete data. In Survey Instrument Characteristic, Respondent Experience, and Administrative Experience a score of 100 is the most ideal, comprising the most versatile technology and/or highest reduction of respondent and/or administrative burden.

	Smartphone Survey
Total Score	**79**
Survey Data	**78**
Origin Location	80
Destination Location	80
Trip Purpose	80
Mode	80
Parking Info	75
Toll Info	75
Route Info	55
Departure Time	75
Arrival Time	85
Travel Party	80
Demographic Data	**90**
Survey Instrument Characteristics	**86**
Data Structure	85
Data Collection Method	80
Extended Collection Period	75
Longitudinal	95
Machine Readable Data	90
Customizability	95
Language	95
Presentation	95
Branching Logic	95
Commercial Availability	70
Future Outlook	95
Respondent Experience	74
Respondent Burden	70
Ease of Participation	65
Recall Burden	60
Time Required	65
Particpant Responsibility	90
Equity	78
Language	95
Age	55
Gender	95
Underrepresented Race/Low Income	60
Ability	85
Administrative Experience	**80**
Respondent Recruitment	**83**
Initial Recruitment	90
Respondent Retention	75
Survey Completion	85
Historical Comparison	40
Data Collection Burden	100
Data Processing Burden	95
Hardware Dispersal/Collection	75
Pricing	

Figure 8 -- Excerpt from Comparative Analysis showing travel survey methodology attributes.

Each of the attribute category headings are an average of the subheading scores. Some subheadings themselves are averages of their sub-attributes. The attribute category heading scores are then averaged together to generate the Total Score. See Figure 7 for an example of the category organization.

5.1.3.1 Weighting

Based on NYMTC's input, AVAIL identified Survey Data as the most important attribute category. This is reflected in the Total Score. In order to capture the significance of the Survey Data category, AVAIL applied a weighting procedure that triples the Survey Data Category score in generating the Total Score. Additionally, AVAIL weighs the Primary Model Input Data three times in determining the Survey Data category score.

5.1.3.2 Attributes, Definitions, Importance

Each survey technology and methodology attribute is listed and defined with brief descriptions of the importance of each attribute to the travel survey process, including, where possible, their capacity to provide necessary NYBPM inputs and their capacity to simplify and/or enhance the travel survey process.

5.1.3.3 Survey Data

5.1.3.3.1 Trip Characteristic Data

Trip characteristics are vital elements for an activity-based model. The important trip characteristics are listed below:

1. **Primary Model Input Data**

   - **Origin Location**. Origin location refers to the location recorded at the beginning of a trip.

   - **Destination Location**. Destination location refers to the location recorded at the conclusion of a trip.

   - **Trip Purpose**. Trip purpose refers to the reason the survey-taker made a trip.

   - **Mode**. Mode refers to the transportation modality the survey-taker used to make their trip.

   - **Departure Time**. Departure time refers to the time the survey-taker begins a trip from their origin location.

   - **Arrival Time**. Arrival time refers to the time the survey-taker concludes a trip at their destination location.

   - **Travel Party**. Travel party refers to the number and characteristics of people who journeyed with a survey-taker from their origin to their destination.

2. **Secondary Model Input Data**

- o **Parking Information**. Parking information refers to the location, cost, and characteristics of the place a survey-taker parked their vehicle.

- o **Toll Information**. Toll information refers to the cost and payment method associated with highway or bridge tolls.

- o **Route Information**. Route information refers to the path taken on the transportation network from the origin location to the destination location.

5.1.3.3.2   Demographic Data

Detailed demographic data is typically captured in telephone recruitment interviews. However, bulk collection of demographic information from respondents produces negative outcomes in perceived burden. Participants may refuse to answer questions or answer incorrectly which reduces data quality (Cottrill et al. 2013). Though evidence is limited, distribution of demographic questions across the survey timeline—during the recruitment interview, as part of the survey instrument, and as part of the reporting and exit interview—may reduce negative perception of survey burden (Cottrill et al., 2013). In turn, the accuracy and completeness of data is higher when perceived burden is low (Yuan, 2001).

5.1.4   Survey Instrument Characteristics

5.1.4.1   Data Structure

5.1.4.1.1   Data Collection Method

This field refers to methods used to collect the data by the particular survey methodology. The primary distinction here is whether or not the data is collected automatically or manually.

5.1.4.1.2   Extended Collection Period

Due to limitations in respondent retention and completion, traditional travel surveys collect data for a single 24-hour period, usually during a weekday. Extended collection period refers to the ability of the survey methodology to collect multiple days of travel data from survey participants.

5.1.4.1.3   Longitudinal (Panel Survey)

This attribute refers to the ability of the survey methodology to collect longitudinal data in the form of a panel survey. Although, currently the NYBM does not take panel surveys into account, it is possible that future models will require it.

5.1.4.1.4   Machine-Readable Data

Traditional travel surveys require human input to translate survey results into machine-readable data formats. New survey technologies (e.g. smartphone, online) are inherently machine-readable data generating technologies.

5.1.4.2   Customizability

The ease with which a survey methodology and its instruments can be adapted to meet the objectives of the MPO is a valuable metric for comparing the usefulness of survey methodologies. This attribute refers

to how easily and completely technologies and methodologies can be adapted to meet the needs of NYMTC.

5.1.4.2.1   Language
This attribute refers to the ability to present survey questions in the primary languages of the survey universe. Surveys that cannot be easily adapted to meet the needs of those under study will incur greater costs and/or achieve lower representativeness and data quality. The ideal methodology could be translated into every language present within a survey region.

5.1.4.2.2   Presentation
As survey technology advances and people's expectations of survey content and presentation change, it becomes important to be able to structure the survey interface in a manner that encourages participation, accurate response, and completion. Paper surveys are static, though they offer freedom to structure survey questions in any order, and use icons and legends, they cannot respond or change once administered. Smartphone screens are smaller, limiting the amount of information that can be displayed at one time. However, they are also dynamic and interactive, allowing the user to touch, rotate, and zoom content. Moreover, you can embed video, gifs, and create places for expanded content. Online surveys have many of the same features as smartphones. Telephone interviews have the advantage of a human interface that can address the specific concerns and questions of participants. However, survey methodologies that employ telephone interviews lack an information-laden visual display to aid in answering survey questions completely and accurately (Tang & Waters, 2005).

5.1.4.2.3   Branching Logic
Three of the four reviewed survey instruments have the capacity to question in an increasingly focused manner based on previous response thereby limiting the amount of information the participant needs to comprehend. Paper surveys can approximate this by saying, for example, "if you do not own a vehicle, please skip to question x." Nevertheless, valuable visual space is taken up by information that will not be relevant to every participant. Online surveys and smartphones are capable of choosing the next question using answers to previous questions thereby limiting the amount of time a participant must spend engaging with the survey, as well as reducing participant confusion. Moreover, clear feedback can be provided by these instruments as to whether or not the survey is complete—an assessment that a participant must make on their own when completing a paper survey. Telephone interviews offer a similar type of branching logic that minimizes the total information a participant must comprehend.

5.1.4.3   Commercial Availability
Commercial availability is important because competition drives quality and reduces expense to the MPO. There are numerous existing organizations that work with MPOs throughout the country to develop paper travel diaries (e.g., NuStats, Westats, RSG, SRBI). Some of these organizations are developing online surveys and smartphone applications to meet the needs of the travel demand modeling communities. Still, at present, there are few commercially available online and smartphone app surveys.

5.1.4.4   Survey Instrument Lifecycle (Future Outlook)
This attribute points toward the future potential of the survey instrument. For instance, it is difficult to imagine the paper surveying community developing a new approach that drastically changes the physical medium or reduces the cost of deployment. On the other hand, GPS devices will likely become more

ubiquitous, fundamentally changing the survey methodologies that surround the GPS as a survey instrument accessory. The potential of the smartphone is even greater. The graphic user interface allows a surveyor to engage with the participant in ways that scientists are only beginning to understand. Additionally, through language translation software, smartphones are capable of supporting surveys for different populations. Smartphones may also facilitate the capture of travel of the visually impaired population. Moreover, the device contains a host of different sensors-- accelerometer, GPS, gyroscope, Wi-Fi, Bluetooth-- a list that continues to grow with new technologies and applications.

5.1.5   Respondent Experience

5.1.6   Respondent Burden
Respondent burden can be defined as, "the perceived 'difficulty', dissonance, or intrusion that individuals associate with a survey that they are being asked to do" (Bricka, 2008). It is related to how easy it is for a respondent to give you accurate information (Richardson et al., 1995: 30).

Respondent burden is both an objective measure--amount of time required to complete survey questions--and a subjective measure--opportunity cost, perceived strain. Respondent burden can be affected by amount of survey questions, length of survey (in days), recall expectation, and reporting methods (Richardson et al., 1995). Decreasing the amount of necessary self-reporting is a way to decrease burden (e.g., use of a survey technology that deduces trip purpose through machine-learning).

There are many ways to reduce participant burden without compromising the amount or quality of data collected including: the type of survey conducted (panel, cross-sectional, other); changes to the survey instrument (e.g., in a smartphone application, changing the way the user inputs trip characteristics or demographic variables from typed entry to selection from a drop down menu); offering monetary incentives or informational incentives (e.g., connecting the survey results to infrastructure changes that may occur as a result); gamification; and others (Bricka, 2008; Cottrill et al., 2013).

For the sake of this comparative analysis, respondent burden scores take into consideration both objective and subjective components. High scores in participant burden indicate either minimal time or effort expenditure on the survey and/or effective methods for controlling the participant's perception of that strain, while still gathering accurate information from that participant.

5.1.6.1.1   Recall Burden
This attribute refers to the amount of information a traveler is requested to remember, multiplied by the length of time elapsed since the trip occurred. Recall Burden negatively effects travel survey accuracy. If a respondent reports only half of their trips, the data is inaccurate no matter how well that half is reported. Smartphone survey applications offer solutions that reduce recall burden. Smartphone applications are usually carried on-person and can include notification prompting recall directly after trips conclude. Smartphones are also capable of employing machine-learning to deduce trip purpose, mode, etc. which can reduce the respondent recall responsibility by offering trips information that only require a simpler validation process. For the purposes of this comparative analysis, a higher Recall Burden score corresponds to a less recall burden.

5.1.6.1.2   Ease of Participation

This attribute refers to the ease by which a respondent can participate in the survey. Ease of participation includes clarity of definitions and terms (e.g., household, trip, etc.), user-interface design, user support, general number or perceived amount of survey requirements. Clear, succinct instructions as well as definition of terms create consistency between different respondents. A streamlined, visually appealing interface can improve a user's interest and therefore improve completion compliance.

5.1.6.1.3   Time Required

Time requirements can deter participation or hinder completion compliance. Paper travel surveys with telephone validation require more a participant's time than a smartphone application. Additionally, an online survey requires more time than a handheld smartphone.

5.1.6.1.4   Participant Responsibility

Responsibility for the tangible survey method or technology is included in the measurement of burden. Concerns include: Do they have to carry a device all day? Is it easy to lose or forget about? Are their mechanisms to help a respondent remember to record or validate a trip? Do they have to charge a device?

5.1.6.2   Equity

Equity can be defined as the distribution of impacts (benefits and cost) as well as the survey's ability to capture vital (and often underemphasized) data with the purpose of mitigating inequitable patterns in transit. In broad terms, the equity of a survey can be analyzed in two ways: the effectiveness of outreach and recruitment processes used to secure the appropriate representation of all traveler demographics within a random sample, and the survey's ability to capture data that will result in more equitable transportation modeling. In this analysis, highly ranked survey technologies or methodologies encourage equitable response from underrepresented populations in travel surveys.

"The most common demographic characteristics associated with non-response in traditional household surveys include being low income, young (under age 25), having a lower education, being a part of a larger household, being of minority descent, and living in an urban area" (Bricka, 2008).

Beyond the legal equity requirements for MPO travel plans to comply with Title VI of the 1964 Civil rights act and to secure Fixing America's Surface Transportation (FAST) Act federal funding, equity considerations are a vital component to accurate transit models (Karner & Niemeier, 2013: 1). In current form, an increased burden of substandard travel outcomes is felt by historically marginalized communities. If travel surveys do not equitably capture habits from all demographics, inaccurate modeling is inevitable.

For years, paper surveys had difficulty capturing responses from hard to reach demographics. There is mounting evidence to suggest that smartphone based travel surveys can increase participation from this cohort (Anderson et al., 2016; Buskirk & Andres, 2013; CISCO, 2016). The increased ubiquity of smartphones provides an opportunity to improve equitable representation in travel surveying due to the ability of smartphones to employ language technology, the increased automation of the survey process, survey notification reminders and the reduced respondent requirement (e.g. the removal of the requirement to return the survey or even have a consistent residence or telephone number). Data plan costs can be mitigated by smartphone survey applications that cache participant travel data for upload

later, when wifi access is available. Smartphones may need to be deployed to low income households as a means to participate which could have the secondary effect of incentivizing participation.

5.1.6.2.1 Language

This attribute refers to the ability of a survey technology or methodology to present survey questions in various languages. Electronic surveys, such as online or smartphone can employ modern language translation software, including accessibility features accommodating visual and aural disabilities.

5.1.6.2.2 Age

Survey technology that is comfortable to use in traditionally low responding age groups should be given favor in an equity analyses. Survey methods that acknowledge potential transit dependency or that do not unnecessarily emphasize motorized travel will also be more equitable in gathering data from younger and older populations.

5.1.6.2.3 Gender

Early studies indicated that women were more traditionally likely to answer paper surveys than men. Furthermore, there was a gender gap in early technology use for telecommunications. It is unclear whether these trends have persisted and what impact they would have on survey deployment. Beyond technological comfort and willingness to use, the survey itself needs to be pre-emptively aware of care related travel. Gendered division of labor makes women much more statistically likely to complete travel related to caring for children and elders, and to trip chain (Bianco & Lawson, 1996). Questions that prompt respondents to remember and accurately report trip chains in a consistent manner, remember passengers, and the various reasons that may exist for what they perceive as one trip will be analyzed as more equitable. Survey methods that encourage respondents to see non-motorized care trips (e.g., walk to a playground) as trips worthy of reporting will be considered more equitable. Because asking these questions could often add significant respondent burden in length of survey, surveys that can create logic chains will be seen as the best balance for an equity analysis.

5.1.6.2.4 Underrepresented Race/Low Income

"Transportation outcomes still show disparities on the basis of race and class" (Karner & Niemeier, 2013: 127). An equitable survey actively seeks out the participation of communities of color and is affordable. In certain cases, transportation policies may perpetuate spatial mismatch between low and moderate income housing and employment, lack of transit options, and disparate economic hurdles. Therefore, survey methods with the potential to increase participation of underrepresented race and low income demographics are ranked as more equitable.

5.1.6.2.5 Ability

This attribute refers to the ability of survey technology or methodology to increase participation and decrease burden for people with disabilities. Electronic survey technologies, and smartphone applications in particular, score highly in this attribute sub-category due their ability to offer software innovations that target various disabilities (Li, 2015).

5.1.7   Administrative Experience

5.1.7.1   Respondent Recruitment

Respondent recruitment refers to a survey methodology's means of reaching out to potential participants and ensuring participation in the surveying effort. Respondent recruitment is a major challenge due to many of the same cultural factors that cause perception of burden to be high -- a distrust of government from a conspiracy perspective and a government effectiveness perspective, have had negative experiences with previous surveys. Additionally, failure of participants to see the usefulness of a survey, or feel oversaturated by survey requests. Monetary incentives were not considered when scoring survey technologies and methodologies on respondent recruitment and retention. Obtaining participants is a two-fold challenge: first, you must contact and obtain interest; second, you must get the participant to complete the survey.

5.1.7.1.1   Initial Recruitment

Initial recruitment is a very large administrative project. In NYMTC's 2010-2011 Travel Survey, 711,551 households were invited, of which 31,156 households participated or - 4% of the total invited. Survey methodologies such as online surveys and smartphone applications provide new recruitment opportunities due to the novelty. Resource Systems Group (RSG) deployed a smartphone survey application with online survey supplement in Ohio where twice as many participants were recruited as were anticipated based on previous travel survey recruitment results (Anderson et al., 2016). They allow for new types of recruitment, emails and other social media posts can contain links to survey participation for easier access (though email response may suffer due to hesitance surrounding opening unknown links).

5.1.7.1.2   Respondent Retention

Being able to follow up with respondents may actually place less of a burden on them in this moment by distributing questions across several months/years that otherwise may have gone unanswered or been answered less accurately (Schillewaert & Meulemeester, 2005: 164). Being able to associate demographics to travel behavior and observe how that behavior changes over time is valuable to better modeling.

5.1.7.1.3   Survey Completion/Respondent Compliance

From the perspective of both data collection and administrative effort, survey completion is the most important feature of the administrative experience of the various survey technologies and methodologies. The administrative effort taken to invite and on-board survey participants is costly. In NYMTC's 2010-2011 Travel Survey, 31,156 households participated, of which, 18,965 households completed the survey - or 61% of the participants. In the smartphone application survey with implemented by RSG in Ohio, 76% of participants completed at least one full day 100% answered trip surveys. Variables affecting completion include reporting and validating trips and returning surveys. In this analysis, electronic surveys score higher as they require less hardware retrieval. Additionally, online and smartphone surveys simplify the trip reporting and validation and offer the possibility of gamification.

5.1.7.1.4   Historical Comparison

One key feature of collecting travel data is comparing it to travel data collected in previous years to ask "how is the data we collected this time around different than in the past" and "what does this mean about

both our methods and travel in our region?" Collecting new data can also permit validation of previous future forecasts of a travel demand model to determine what may need to be tweaked in order to generate more accurate results.

5.1.7.1.5   Data Collection Burden

This attribute refers to the administrative burden of collecting the survey data. Traditional travel surveys require a telephone validation as well as paper survey transcription. Electronic survey technologies and methodologies offer the feature of automated collected data, which reduces administrative burden.

5.1.7.2   Data Processing Burden

This attribute category is a measure of the amount of data processing required after the data has been collected. Traditional paper travel surveys require paper survey transcription and geo-coding. Paper Travel Survey with GPS supplement requires GPS correction, weighting, and expansion. Most commercially available electronic surveys (Online and Smartphone survey only) require little post processing. Smartphone applications with GPS technology embedded require little, to no, additional processing. Smartphone location history is promising but currently requires participant retrieval (download) and submittal (upload).

5.1.7.3   Hardware dispersal and collection

This attribute refers to the distribution of physical survey components (e.g., wearable GPS units, paper surveys, smartphones, and smartphone location history files). Survey technologies and methodologies with fewer dispersal and collection requirements score higher on this attribute.

5.1.7.4   Future Outlook

This attribute refers to the growth trajectory of the survey or location technology or combination survey methodology. "While the diary has undergone significant evolutionary change over the past 20 years, mainly in an attempt to improve its reporting capability, and to reduce respondent burden, non-response rates have continued to rise" (Stopher & Greaves, 2007). More recent technologies, such as the online diary, smartphone application and smartphone location history, show promise.

5.2   SUMMARY

This section details the findings of the Cost/Benefit Comparative Analysis. The Comparative Analysis methodology scores survey technologies (e.g. paper, online, smartphone) and location technologies (e.g. GPS, Smartphone GPS, Bluetooth) as foundational technologies. These scores are then used to rate the deployable survey methodologies, which include at least one survey technology component and are often supplemented by location technologies. The complete set of Comparative Analysis tables are located in the Appendix of this report.

What follows are a series of visualizations built upon the Comparative Analysis scores and the rankings of Survey Methodologies. The first is a table of all of the deployable survey methodologies with their rankings by category. The next set of visualizations are one page reports on single survey methodologies. Each survey methodology page contains a description, strengths and weaknesses analysis, primary pros and cons, and outlook. Only the methodologies relevant to this report are included here. All of the thirteen survey methodologies scrutinized in the Comparative Analysis are presented in Appendix C, in order of their ranking from most to least effective.

5.3 FINDINGS

Below are the findings of the travel data collection methodology comparative analysis that ranked thirteen data collection methodologies. Pricing comparisons are discussed separately in section 5.3.1.
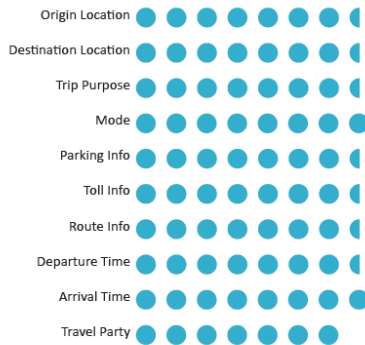
Ranking of Deployable Travel Survey Methodologies													
	Smartphone Survey	Smartphone App	Smartphone App + Bluetooth	Smartphone App + Online Diary	Online Diary	Online Diary + Smartphone Location History	Online Diary + Wearable GPS	Paper Travel Diary+GPS	Paper Travel Diary + Smartphone GPS	Paper Travel Diary + Online Diary + Smartphone GPS	Paper Travel Diary + Online Diary + Smartphone App	Passive Data 1 - Cellphone Trace + ACS	Passive Data 2 - O/D Data + ACS + Marketing Data
Total Score	**6**	**1**	**2**	**3**	**9**	**5**	**8**	**12**	**11**	**10**	**7**	**13**	**4**
Survey Data	**8**	**2**	**1**	**4**	**12**	**2**	**4**	**9**	**9**	**7**	**6**	**13**	**11**
Origin Location	7	1	1	4	12	1	4	10	10	9	7	13	4
Destination Location	7	1	1	4	12	1	4	10	10	9	7	13	4
Trip Purpose	6	1	1	4	9	1	4	9	9	8	6	13	12
Mode	7	2	1	4	12	2	4	7	7	10	4	13	11
Parking Info	8	1	1	4	11	1	4	8	8	7	6	13	12
Toll Info	8	1	1	1	8	1	1	8	8	7	6	13	8
Route Info	11	2	1	4	11	2	4	9	9	8	7	13	4
Departure Time	8	1	1	4	12	1	4	8	8	7	6	13	11
Arrival Time	6	1	1	4	11	1	4	7	7	10	7	13	12
Travel Party	6	1	1	4	9	1	4	9	9	8	6	13	9
Demographic Data	1	1	1	1	1	1	1	1	1	1	1	13	1
Characteristics	**3**	**2**	**6**	**4**	**5**	**7**	**11**	**13**	**12**	**10**	**9**	**8**	**1**
Data Structure	6	2	3	5	8	7	9	13	12	11	10	3	1
Data Collection Method	5	3	3	6	7	9	8	13	12	11	9	1	1
Extended Collection Period	7	4	4	6	7	3	9	12	12	11	10	1	1
Longitudinal	2	2	2	2	6	6	10	13	12	11	8	9	1
Machine Readable Data	7	1	4	4	8	4	9	13	12	11	10	1	1
Customizability	2	2	6	5	2	6	8	13	12	10	9	11	1
Language	2	2	2	2	2	2	10	13	12	10	9	2	1
Presentation	2	2	7	6	2	7	12	12	11	10	9	2	1
Branching Logic	2	2	2	2	2	2	2	11	11	10	9	13	1
Commercial Availability	4	6	12	6	4	11	13	2	3	9	6	10	1
Future Outlook	2	2	2	2	8	6	11	13	12	8	8	6	1
Respondent Experience	**6**	**2**	**3**	**1**	**9**	**7**	**10**	**13**	**12**	**11**	**8**	**4**	**4**
Respondent Burden	6	3	4	5	9	8	10	13	12	11	7	1	1
Ease of Participation	6	3	4	5	7	8	11	13	11	10	9	1	1
Recall Burden	9	4	3	7	12	4	4	13	10	10	8	1	1
Time Required	7	3	4	5	8	10	10	13	12	9	6	1	1
Particpant Responsibility	3	3	6	3	7	9	12	13	11	10	8	1	1
Equity	2	2	2	1	6	5	6	13	12	9	8	10	10
Language	1	1	1	1	1	1	9	13	12	11	9	7	7
Age	2	2	2	1	2	8	2	13	10	8	2	10	10
Gender	1	1	1	1	1	1	1	10	10	9	8	12	12
Underrepresented Race/Low Income	2	2	2	1	12	5	9	12	9	11	5	7	7
Ability	2	2	2	1	9	2	2	13	12	11	10	2	2
Administrative Experience	**4**	**2**	**3**	**1**	**8**	**8**	**10**	**13**	**11**	**12**	**7**	**5**	**5**
Respondent Recruitment	4	1	3	1	6	6	13	12	8	8	5	10	10
Initial Recruitment	4	4	4	3	10	10	13	10	8	8	7	1	1
Respondent Retention	4	1	3	2	4	4	11	10	8	8	4	12	12
Survey Completion	6	4	4	3	10	10	9	13	10	8	7	1	1
Historical Comparison	11	6	6	6	9	9	5	1	4	2	2	12	12
Data Collection Burden	1	1	1	1	7	7	7	12	12	11	10	5	5
Data Processing Burden	1	1	1	1	7	7	7	12	12	11	10	5	5
Hardware Dispersal/Collection	6	6	6	6	1	1	12	13	10	11	5	1	1

Figure 9 -- Heat-map Ranking of Deployable Travel Survey Methodologies

Paper Travel Diary+GPS

Description: After agreeing to participate in a survey respondents are mailed a paper travel diary and instructed to complete it for their assigned day of travel. The paper travel diary contains codes for activities and travel modalities as well as spaces to enter these and other trip characteristics. Respondents may mail back these travel diaries or report their contents via telephone. Additionally, a subsample of survey participants is given a GPS unit to be worn on the day of their travel. Data from this GPS subsample is used to correct for trip underreporting in the full sample.

Survey Data

Survey Instrument Characteristics

Respondent Experience

Administrative Experience

Total Score	Rank
56%	12

Pros	Cons
• Widely accepted as the state-of-the-art for collecting travel data from survey participants.	• Wearable GPS unit may be forgotten by participant on their day of travel. Wearable unit introduces a greater respondent burden as the participant must remember to keep the unit with them in addition to completing the paper travel diary.
• Wearable GPS provides valuable location data without interfering with the participant's normal use of their smartphone device.	• Purchasing, distributing, and collecting GPS units introduces great cost to the MPO.

Outlook: If GPS units have the capacity to transmit data wirelessly the GPS data could be uploaded to MPO/consultant server, processed, and displayed for an Online Diary. This facilitates the process of conflating GPS traces to self-reported travel data.

Figure 10 -- Analysis and Visualization of Paper Travel Diary with GPS Data Collection

Paper Travel Diary + Online Diary + Smartphone App

Description: Participants are given three options for survey completion: paper travel diary, online diary, and smartphone app.

Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection
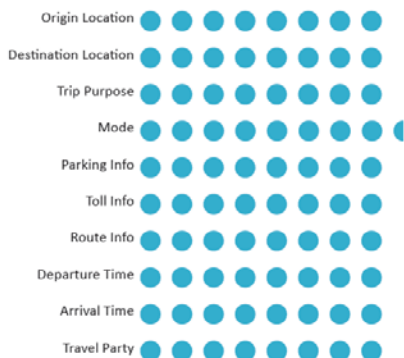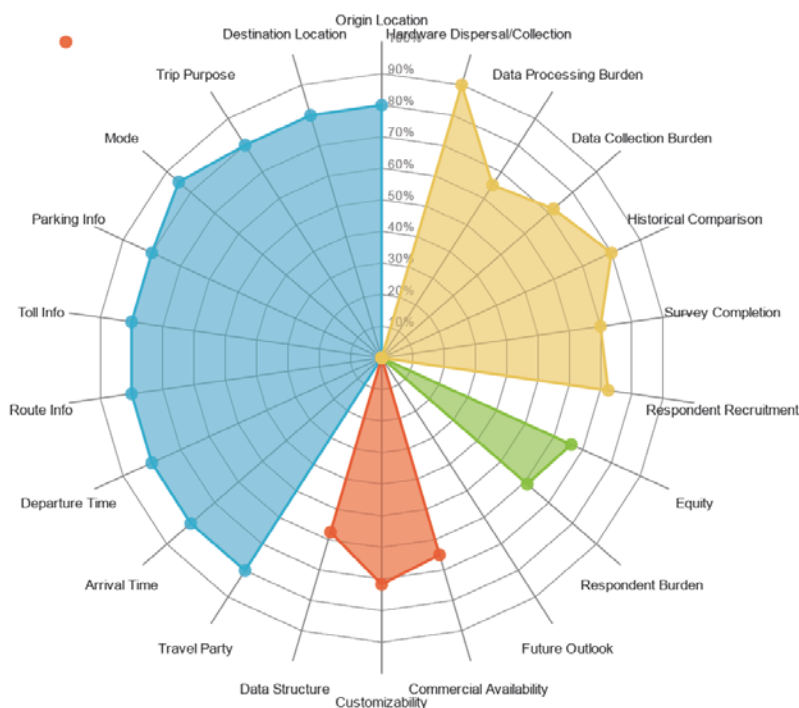
Total Score

75%

Rank

7

Pros

- Integrates new technology and approaches with the traditional approach.

- GPS data collected by the smartphone app can be used to correct the self-report data collected in the paper travel diary and online diary.

Cons

- Expense of developing and deploying all three survey instruments is high.

- Data management challenges that occur when using multiple survey instruments.

Outlook: Strong. Variations on this methodology are likely to emerge in the coming years as MPOs strive for new approaches to travel surveys while remaining attached to the tried and true. The broad net cast by employing all three methods means you are likely to achieve high equity and representativeness in your survey.

Figure 11 -- Analysis and Visualization of Paper Travel Diary with Online and Smartphone Travel Diary Data Collection

Smartphone App (Survey + Smartphone GPS)

Description: Collects travel data from participants automatically through smartphone sensing technologies and uses algorithms and computing power to present those trips and their characteristics to a participant. A participant then validates or adjusts those trip characteristics.

Survey Data
- Origin Location
- Destination Location
- Trip Purpose
- Mode
- Parking Info
- Toll Info
- Route Info
- Departure Time
- Arrival Time
- Travel Party

Survey Instrument Characteristics
- Data Structure
- Customizability
- Commercial Availability
- Future Outlook

Respondent Experience
- Respondent Burden
- Equity

Administrative Experience
- Respondent Recruitment
- Survey Completion
- Historical Comparison
- Data Collection Burden
- Data Processing Burden
- Hardware Dispersal/Collection
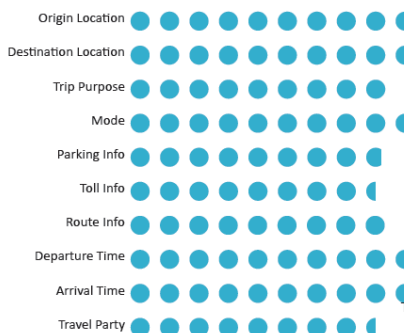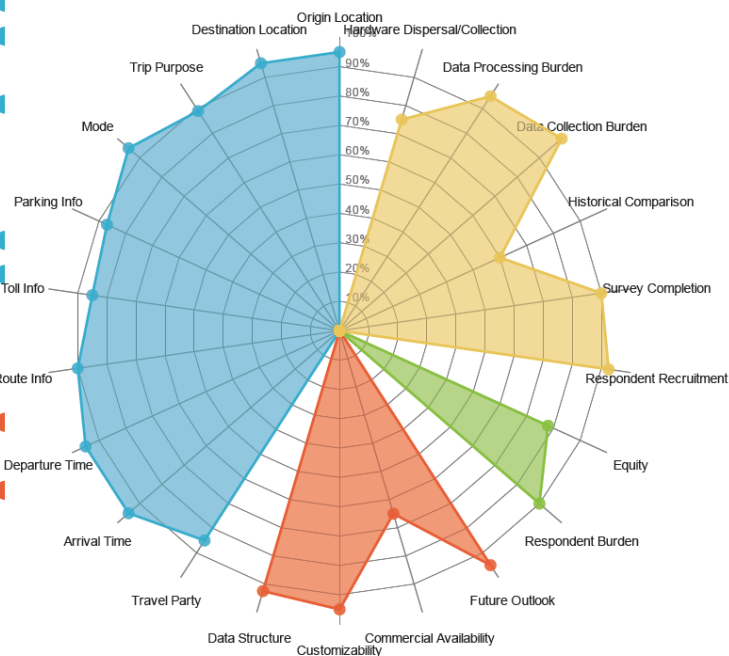
Total Score	Rank
88%	1

Pros

- Passively gathering travel information then presenting it to the user for validation represents the state-of-the-art in balancing participant burden reduction and trip data quality.

- The smartphone eliminates the need to distribute and collect GPS devices from participants, reduces the likelihood of forgetting the GPS device on the travel day, and has a user interface for viewing and adjusting trip characteristics. As a result it is considered by many in the industry to be the natural evolution of GPS use in travel surveys.

Cons

- GPS traces remain inaccurate underground and in areas known as urban canyons.

- The algorithms for identifying route, mode, and trip purpose need further refinement and are often inaccurate.

- Though supporting data is somewhat lacking there is some evidence to suggest that using 'high tech' to gather travel information could lead to oversampling of populations that are younger and more highly educated.

Outlook: The outlook for smartphone applications that use sensing technologies and user interface to gather accurate travel data is promising. A number of different companies are developing these applications and using feedback and rapid improvements in sensing and computing to refine them.

Figure 12 -- Analysis and Visualization of Smartphone Travel Diary App Data Collection

Passive Data 2 - O/D Data + ACS + Marketing Data

Description: Gather (purchase) O/D probe data and marketing data (or PUMS Data) for MPO region, run through algorithm to create synthetic population and generate O/D matix trip table. Run trip table through NYBest Practices Model or use resulting O/D table and socieoeconomic demographic data as regional transportation model.

Total Score
80%

Rank
4

Pros

- Reduces or removes the need for survey participants.

- Can be used as model input or as a model by itself.

- Can be used to calibrate forecast models by validating previous forecasts.

Cons

- Accuracy of O/D probe data is still a question.

- No precedent for employing this methodology in large metropolitan areas comparable to NYMTC.

Outlook: Although there is little precedent for this methodology, it is highly likely that this passive data method will replace both the current active survey and the statistical modeling process.

Figure 13 -- Analysis and Visualization of Passive Data Collection and Modeling

5.3.1 Pricing

Pricing is critical in the process of comparing different survey methodologies. Due to the competitive and proprietary nature of vendors and the complexity of cost estimation for the New York metropolitan region, it is difficult to ascertain exact price estimates for each survey methodology. This challenge is compounded by the nascence of innovative strategies and technologies whose prices are likely to shift dramatically as they are developed further. The following table represents best cost estimates for the methodologies reviewed in this comparative analysis. Sources include vendor interviews and published sources. For methodologies where no cost sources were available the research team made estimates based on best available information.

Table 2 -- Travel Data Collection Methods Cost Estimate

Cost Estimate Table				
Methodologies	**Price Estimate (per HH)**	**Total Cost**	**Source**	**Confidence**
Smartphone Survey	~$220	~$4.18m	Vendor	Medium
Smartphone App	~$220-$250	~$4.18m to ~$4.75m	Vendor	High
Smartphone App + Bluetooth	~$250+	~$4.75m+	None	Low
Smartphone App + Online Diary	~$220-$250	~$4.18m to ~$4.75m	Vendor	High
Online Diary	~$150-$180	~$2.85m to ~$3.42m	Vendor	High
Online Diary + Smartphone Location History	~$220-$250+	~$4.18m to ~$4.75m	None	Low
Online Diary + Wearable GPS	~$200+	~$3.8m+	None	Medium
Paper Travel Diary + GPS	~$220	~$4.18m	Vendor, NYMTC, Hartgen, D. T., & San Jose, E. (2009)	High
Paper Travel Diary + Smartphone GPS	~$220	~$4.18m	None	High
Paper Travel Diary + GPS + Smartphone GPS	~$220	~$4.18m	None	High
Paper Travel Diary + Online Diary + Smartphone GPS	~$220-$250	~$4.18m to ~$4.75m	None	High
Passive Data 1 - cellphone trace and ACS	N/A	~$500k to ~$1.5m	Data Science Vendor, Data Vendor	Medium
Passive Data 2 O/D Data + ACS + Marketing Data	N/A	~$1m to ~$2m	Data Science Vendor, Data Vendor	Medium-High

6 PATHWAYS
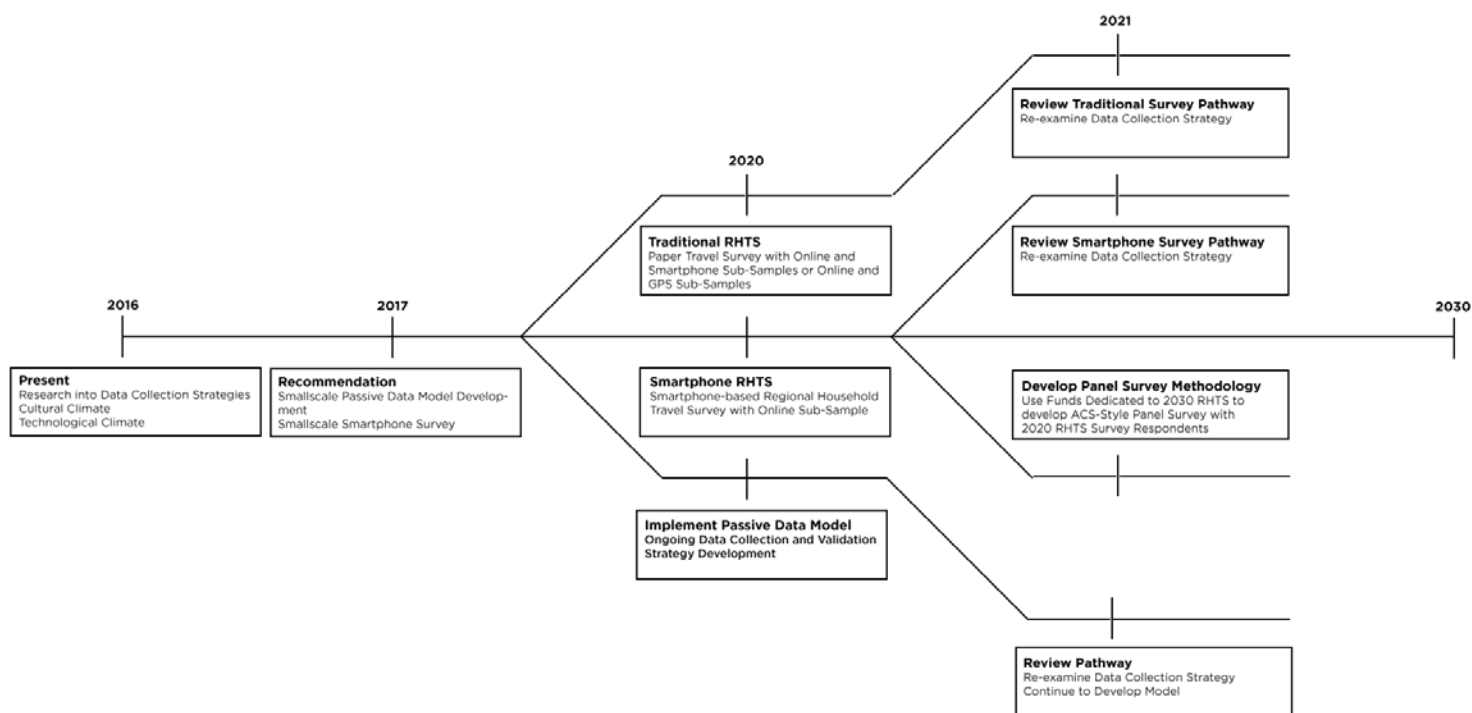
Figure 14 -- Three Pathways to the Future of Data Collection

As illustrated in Figure 14, three pathways (Paper/Online Travel Survey, Smartphone Survey, and Passive Data) have been identified. The opportunities and challenges faced for each pathway will now be discussed, as well as a basic timeline. Emerging possibilities are also discussed for continued future consideration as is the potential use of passive data in model validation and calibration. A recommendation is then made for each of the pathways and these various recommendations are then collated into a final set of recommendations for NYTMC.

Within the Active Data Collection Paradigm, AVAIL offers two potential pathways. The first, the Paper and Online Diary Travel Survey, couples the paper travel diary with an online diary for a large-scale RHTS. A GPS subsample, using distributed standalone GPS devices or leveraging smartphone GPS, is recommended. The second approach is a full-scale Smartphone Diary Travel Survey. Both of these approaches are simplified in order to easily understand the core differences in the two. There are, however, gray areas between the two pathways where survey technologies can be combined and recombined, where percentages of demographic cohorts using various survey instruments can be adjusted up or down.

The third pathway, Passive Data Collection, uses origin-destination data, demographic data, and consumer behavior data purchased or sourced from third party companies. Data scientists then fuse the data sets into a synthetic population. An activity-based microsimulation model (like the NYBPM) then distributes and conducts that synthetic population through the transportation network.

Statistical Extrapolation

Regional Household
Travel Survey
0.36% Of Population
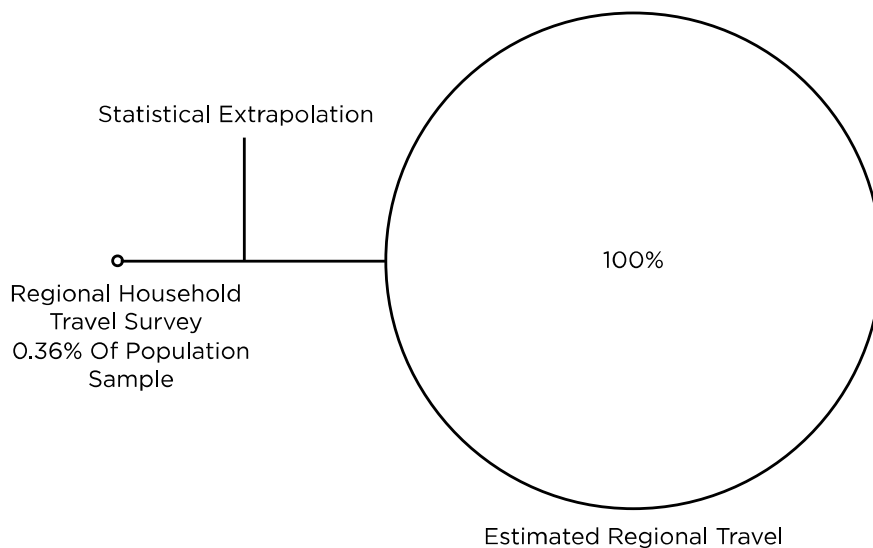Sample

100%

Estimated Regional Travel

Figure 15 -- Visualization of Active Data Collection

6.1 PAPER AND ONLINE DIARY TRAVEL SURVEY

Major Event(s): Early 2020s RHTS.

Estimated Cost Savings (through 2025): Savings of appx. $50-75/HH for cohort that participates using online survey.

This pathway requires that NYMTC orient data collection efforts around an early 2020's RHTS using paper and online diaries and collecting GPS traces from a subsample (possibly by smartphone). It is the least risky and logistically challenging approach requiring the fewest operational changes in data collection and modeling.

One strategy that may provide value is to change the manner in which GPS traces are collected. In the 2010-11 RHTS, a subsample of nearly 2,000 households were given standalone GPS units to use on their travel day in addition to filling out a paper travel diary. In the next RHTS it may be possible to leverage widespread smartphone ownership to gather GPS traces in lieu of distributing and collecting GPS devices. Or, it may be possible to turn this subsample into a smartphone application pilot study.

The costs of travel surveying have remained relatively stable after adjusting for inflation (Hartgen & San Jose, 2009) and costs for surveys using newer survey instruments (e.g. smartphone) are, initially, at least, comparable to paper travel diaries. One potential area for savings is in utilizing the online diary. It is estimated by one industry expert interviewed for this study that the online diary provides a potential for savings that should start to bear out in the marketplace in the next few years. As vendors continue to develop and deploy online diaries, the cost of development ceases to weigh into the cost of deployment. It is estimated that the cost of conducting online surveys ranging from $150 to $180 per household. As a

comparison, the cost of conducting a paper travel survey ranging from $220 to $250 per household. Additionally, all vendors with the logistical capacity to conduct a regional household travel survey in the NYMTC region have the capacity to deploy online surveys (though there is a high-risk of low return rates as people receive many online survey invitations).

6.2 SMARTPHONE SURVEY APPROACH

Major Event(s): 2017-2019 Smartphone Pilot Study OR 2017-2019 Simultaneous Deployment Pilot Study; Early 2020s RHTS.

Estimated Cost Savings (through 2025): None.

Using a smartphone as a survey instrument appears far more promising than a paper travel diary (pending verification with localized pilot study deployments). Using a smartphone survey instrument offers numerous improvements:

- It eliminates distribution and retrieval challenges,
- It is constantly present on the person of most owners,
- It has built-in reminder capabilities,
- It can format data to reduce data processing time and associated costs,
- It may be perceived as less intimidating and easier to use (more intuitive) for a respondent,
- It contains built-in sensing technologies coupled with inferential algorithms which shift the respondent responsibility to confirmation or validation as opposed to manual entry,
- It is capable of producing visual prompts to aid in recall.

In many ways the smartphone is the marriage of the travel diary to the location sensing device (GPS), except the whole is greater than the sum of its parts. The smartphone attenuates the weaknesses of both. By coupling GPS traces to the diary, fewer trips are missed during reporting and, because smartphones are nearly always with their owners, the GPS is rarely forgotten.

In terms of cost, app-makers who have conducted actual travel surveys using smartphones have indicated that the costs of administering a travel survey are comparable between smartphones and traditional paper surveys.

Figure 16 -- Screenshots from Smartphone Travel Diary Apps

While it is true that there are concerns about representativeness, and reasonably so, these concerns seem manageable. Smartphone ownership skews to younger populations (and perhaps other demographics). Perhaps a small subsample, probably older adults, should receive a paper travel diary and a GPS. Or it may even be possible to distribute smartphones to this group, similar to other smartphone pilot studies. Regardless, it is reasonable to expect increased costs associated with this additional sampling. However, that increased cost may not be prohibitive. Furthermore, a shift to a smartphone data collection strategy increases numerous possibilities that may offset those costs.
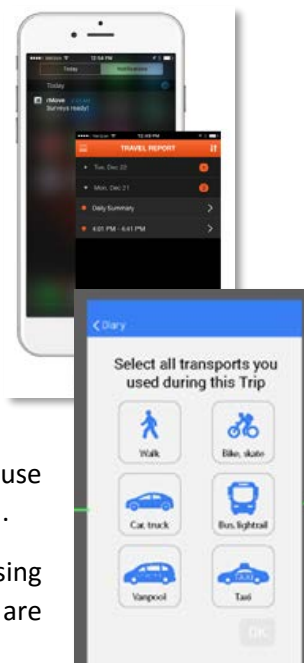
Planners have expressed an interest in collecting tourist travel activity. Attempts to develop a tourist travel survey deployment strategy would require asking hotel patrons to voluntarily participate using traditional travel survey techniques. A smartphone tourist survey, however, may be more successful if tourists are invited via smartphone invitation. Nevertheless, using hotels as the origin may require coordinating hotel staff to ask their guests to participate (or grant surveyors permission to ask them to participate) which could prove problematic. Information or monetary incentives may induce high rates of app download and use. Nearly 60 million people visited New York City in 2015. Even if those 60 million visitors only visited for a single day that represents an additional 160,000 people in New York using the transportation network that are difficult to account for in the RHTS.

Similarly, a smartphone app permits surveying of freight that moves through the region. At least one app-maker is considering a dashboard design that would allow their clients (MPOs) to adjust the survey questions that appear in the app. This could manifest in many interesting ways. Perhaps the app could house several different surveys (household, tourist, freight, bike/ped, emergency response survey, validation surveys that identify past users who might be affected by a road closure or new policy and re-surveys them etc.) The initial investment in the application may be costly but the opportunities afforded moving forward may warrant every penny spent.
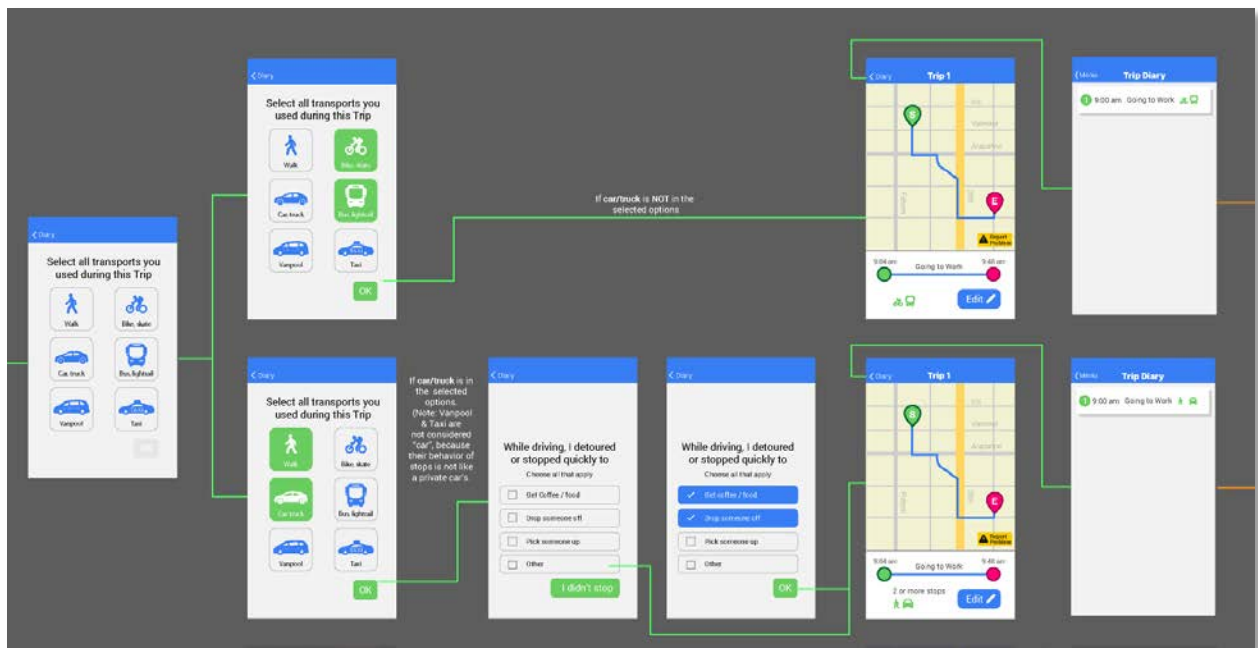
Figure 17 -- Wireframe of a Recent Version of a DVMobile Smartphone Travel Diary App

Moreover, consider the trajectory of computing and data storage. Advances in data computing are made rapidly. Just ten years ago there was no such thing as a smartphone. Now, it is estimated that over sixty percent of Americans carry one. The US Census, a valuable source of demographic information, now has an API that allows users to rapidly gather and integrate demographic data into applications, models, and research, a process that was previously prohibitive to all except the most well-funded and persistent. The entire transportation network of the NYMTC region is digitally represented and available to anybody who

is interested on numerous platforms, both proprietary and open source. More than 95% of all data ever generated was generated in the last five years.

In summary, the smartphone is a survey instrument that is ready to be tested in the NYMTC region. The preliminary research suggests higher response rates and lower respondent burden, and the cultural climate continues to shift toward greater ubiquity of sensing devices, a reduced reliance on landlines (important in the past for sampling), and a growing distaste for lengthy and complicated surveys. When we project forward the choice is even more obvious. The adoption of the smartphone for travel data collection will lead to greater R & D investment, a larger supply of apps (and by extension approaches to data collection) in the marketplace, and innovative uses explored and shared within the surveying community. The potential to examine new vectors, collect data on previously unexplored components of travel, policy impacts, construction impacts, and, in general, uses of the network that were previously impossible to study.

6.3 PASSIVE DATA COLLECTION
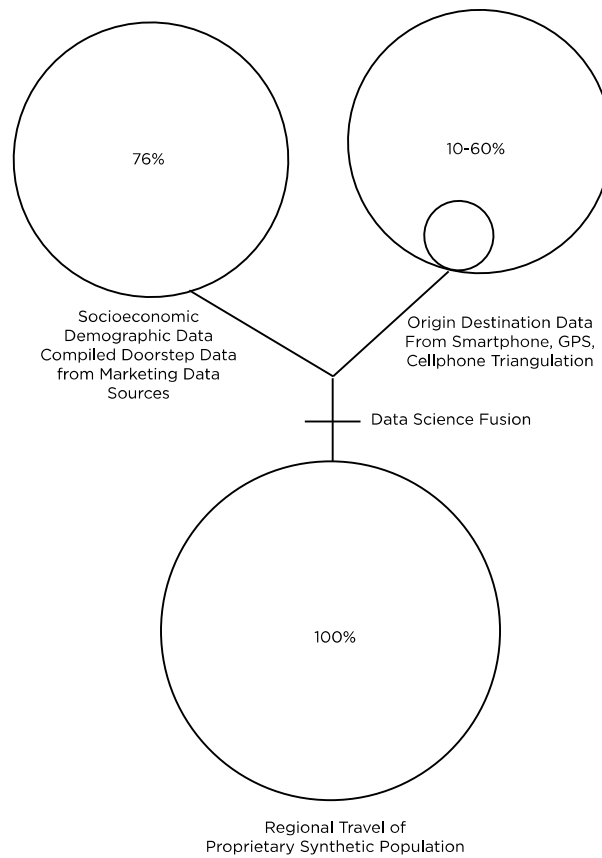
Figure 18 – Visualization of Passive Data Collection

Major Event(s): 2017-2019 Simultaneous Deployment Pilot Study

Estimated Cost Savings (through 2025): Savings unknown, but possibly upwards of $1,000,000.

Admittedly, this "Passive Data" or "Data Science" approach has never been conducted in a modeling environment as complex or massive as the NYMTC jurisdiction. But based on research by Kressner et al. (2016), scaling this approach to meet NYMTC's needs seems possible at this time. Investing in this approach may save NYMTC hundreds of thousands, perhaps millions, of dollars on data collection, data processing, and travel modeling.

Fully comprehending the possibilities associated with the Passive Data Approach requires reexamining why MPOs conduct household travel surveys in the first place. MPOs use the data generated by household surveys as inputs into travel models. The survey data is processed and expanded to create synthetic populations whose travel flows are modeled. The 40,000+ people surveyed in the 2010-11 RHTS represents approximately 0.3-0.4% of the NYMTC population. Therefore, ~99.6% of the population that moves through the model is synthetic, though based on "ground truth."

The Passive Data Approach also generates a synthetic population. The basis for this synthetic population, however, is massive datasets from different sources that are joined together. Origin-Destination data from companies like HERE, Airsage, or INRIX, are used to show aggregated and disaggregated travel flows.

Demographic data, from the American Community Survey (ACS), is combined with marketing data about households and the behaviors of those households, from companies such as Nielsen, Epsilon, or InfoUSA (or any other number of consumer behavior/marketing companies) in order to create trip purpose and frequency. These datasets provide large capture percentages that cover nearly the entire population, and in the case of the O/D and marketing datasets, are updated on continuously.

This abundance of data can now be integrated using data science techniques. Algorithms can link household demographics and consumer behavior, essential to marketing companies, to the origin destination data provided by probe tracking companies, an approach that has been successfully tested in Asheville, North Carolina (Kressner et al., 2016).

How We Compile Our Data

We use 29 billion records from over 100 different sources to aggregate our consumer database every year. We gather raw data from real estate and tax assessments, voter registration files, utility connects, bill processors, behavioral data, and other sources before we integrate dozens of proprietary enrichment sources.

infoUSA®

Targeted Data

To help you refine your consumer and residential mailing lists, we offer dozens of search selections, including:

✔ Age
✔ Estimated Household Income
✔ Marital Status
✔ Single vs. multi-family home
✔ Gender
✔ Home Value
✔ Presence of children
✔ Ethnicity
✔ Hobbies & Interests

Figure 19 -- List of Marketing Data Options from InfoUSA

This Passive Data / Data Science Approach to travel data collection provides the added benefit of real-time updates and improved forecasting models. The traditional regional household survey generates a day of data that is then run through models to create nowcasts and forecasts. The Passive Data Approach can offer now-cast insights as regularly as is affordable and can use machine-learning to analyze months, and years, worth of data to generate increasingly accurate forecasts. Strategies to use back-casting (e.g., taking data from a previous time period and forecast to the present and compare with actual data from the present) can be used to evaluate the accuracy of the approach. Although it is not possible to validate future forecasts, finding high correlation with model outputs and actual data provides a strategy for moving the concept forward. Back-casting using machine learning has the potential to be a more powerful forecasting tool than the current practice.

Consider the unique travel habits of the millennial generation, many of whom do not own a vehicle, or the emergence of semi-autonomous vehicles and the seeming inevitability of self-driving cars - these shifts, cultural and technical, change the pattern of activity on the transportation network. Profound changes can impact transportation choices in less than a ten year period (less time than that between the two most recent NYMTC RHTS) amplified by the paradigm shift regarding car ownership. Certainly, Uber, Lyft, and other ridesharing services are changing travel patterns. There is reason to believe these changes will permanently impact society and that it is unlikely to revert back to increased individual car ownership. To accurately capture emerging changes requires a strategy to collect data about travel habits and transportation network activity with greater regularity to increase confidence in the model results, ensuring they reflect contemporary reality, not the travel behaviors of ten years ago.

It may be possible for NYMTC to eventually eliminate travel surveys from the travel demand modeling process or at the very least, to mitigate their costly implementation to the realm of sub-sampling, as a control measure for testing the validity of the Passive Data Approach.

The future of Travel Data Collection may well be the Passive Data Approach. Large passive consumer and location data sets can be combined to create person-level synthetic travel data (Kressner et al., 2016; Wolf et al., 2014). Using statistical techniques on these data sets, it is possible to recreate and revise the relationships that comprise the core choice models of the New York Best Practices Model (NYBPM).

The approach trades ground truth data of <1% of the population for aggregate data on large portions of the population from various sources. In Asheville, Kressner et al. (2016) created a tour-based model using passive data for comparison with the regional MPOs trip-based model and concluded, "The passive data model produces effectively equivalent levels of accuracy when compared with the aggregate trip-based model using standard validation measures." Notably, though, this model did not include multimodal capacity. Further, Asheville does not have the NYMTC region's complicated transportation system and travel patterns.

Nevertheless, the Passive Data Approach, with its backbone of data science, could play an increasingly larger role in travel modeling. It has the potential to save large sums of money and keep travel models updated indefinitely with regular data acquisition. As with the smartphone, the potential here is not fully known. Companies are gathering more and more data about people. InfoUSA (*see* Figure 19), for instance, collects data on age, estimated HH income, marital status, presence of children, and many other variables immediately relevant to travel demand modeling. As online behavior is tracked more closely and the Internet of Things (IoT) expands and produces new data vectors on our real-world behavior, modeling activity choices and related travel decisions will expand to take advantage of this new data.

Additionally, passive data can be used to back-cast (e.g., taking data from a previous time period and forecast to the present and compare with actual data from the present) to evaluate the accuracy of the data science approach. Although it is not possible to validate future forecasts, finding high correlation with model outputs and actual data provides a strategy for at least moving the concept forward.

Furthermore, there are additional opportunities. For instance, Alphabet, the Google parent company has under their purview a company called Sidewalk Labs. Sidewalk Labs is a "new type of company that works with cities to build products addressing big urban problems." One of these products is a transportation coordination platform called Flow.

7 PILOT STUDIES

This section outlines pilot studies that NYMTC should consider exploring in advance of their next regional household travel survey.

7.1 SIMULTANEOUS DEPLOYMENT PILOT STUDY

To maximize the effectiveness of conducting pilot studies to prepare for future data needs, a multi-location approach is recommended that compares the active data collection with smartphone apps and the passive data approach in four geographically diverse settings.

Four geographies approach include:

- Manhattan
- High density/low transit
- Suburban medium density
- Suburban low density

The first location would be in downtown Manhattan, representing the highest level of density and most saturated transit network, to test the viability of the smartphone use in this environment and the fidelity of the passive data approach. The second location should represent high urban density with sufficient transit network available (e.g., Brooklyn, Bronx). The third locations should represent medium suburban density (little or no transit service). The final pilot test location should be a medium to low density suburban area.

To achieve statistical representativeness, it is recommended each of the smartphone pilot test produce at least 300 completed surveys, for a total of 1200 completed surveys. However, in low-density suburban areas, it may be challenging to achieve such a statistical sample. Realistically, a smaller number of completed surveys would be sufficient to demonstrate the effectiveness of the methodology. If the low density area is indeed problematic, it will require further attention before recommending full deployment.

Each of the passive data experiments would need to "create" 300 synthetic households with their associated travel patterns, for a total of 12 synthetic samples. Conducting both methodologies in the same locations generates the maximum opportunity for comparability. These pilot studies only test the ability to use these methods, however, it may be possible to compare the characteristics of the data collected with outputs from the NY Best Practices Model (NYBPM), using overall aggregate metrics.

7.2 SMARTPHONE TRAVEL SURVEY PILOT STUDY

To evaluate the emerging marketplace of smartphone travel survey applications, AVAIL conducted a scan of available services and tested a sample of available applications to gauge real-world readiness of the technology as a whole and to describe best practices within the marketplace. Although a thorough market analysis was beyond the scope of this research, the testing procedure was a sufficient analysis to understand best practices.

To test smartphone survey applications in the New York City geography, AVAIL recommends NYMTC work with a consultant to conduct a small-scale Smartphone-based Household Travel Survey (HTS) Pilot Study.

The results of the smartphone pilot study should provide favorable results and can inform which pathway to choose going forward. A pilot study allows NYMTC to gather valuable data and simultaneously test a new approach before completely overhauling their data collection strategies.

A Smartphone Survey Pilot Study would provide NYMTC with the evidence it needs to invest in the Smartphone Survey Approach. This includes the possibility of using the smartphone as a sub-sample in the Paper Travel Survey Approach, in place of the GPS device, as a sub-sample in the Data Science Approach, or as a stand-alone survey methodology.

Of the smartphone vendors interviewed for this study, all were in the process of revising and upgrading their mobile interface. This suggests improvements in the user experience will be on-going, while the functionality of the smartphone has reached maturity. As a result, in the near term, it is possible to conduct a pilot study to determine the "use" of a smartphone app as a travel survey methodology, but not necessarily needing to modify the interface to the exact specifications for NYMTC's needs. The concern expressed most often was whether a smartphone approach would work in the urban canyon environment in Manhattan. This suggests the need to first test the ability for the system to function compared to a medium dense environment (e.g., suburban location). If the pilot study produces sufficient robust results in Manhattan (suburban applications have many examples of successful deployment in other states), then the next step would be to request modifications of the web interface to specifically suit the Manhattan experience (e.g., multi-modal transfers, high taxi/Uber use). It is possible that an off-the-shelf app will be available in the near term that has the needed elements to meet NYMTC's high-density urban environment needs. There are companies that have experience in other international cities with similar characteristics. For instance, the Mobile Market Monitor (MMM) application is currently being employed in large-scale surveys in densely populated cities such as Tel Aviv and Singapore, where multi-modal transfers are critical. A key feature of the MMM system is the treatment of these change mode activities. MMM incorporates the local transit networks into its machine learning backend for each survey in order to detect change mode stops.

Based on a literature review and product testing, AVAIL recommends the following application features for a Smartphone-based HTS Pilot Study: reducing respondent burden; battery life; data plans; equity; data processing and validation; and privacy.
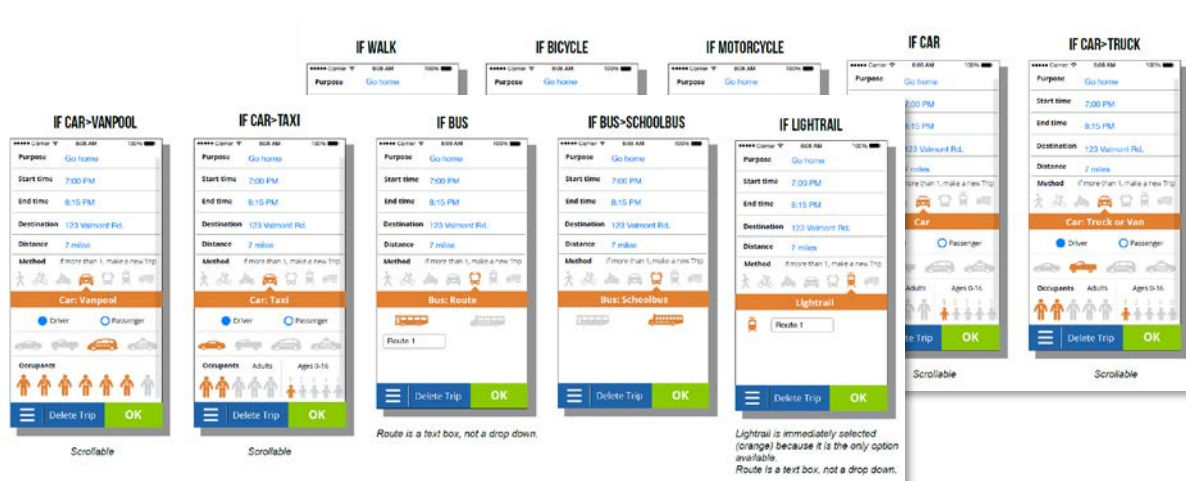

Figure 20 -- Branching Logic and Iconography in App

7.2.1 Reducing Respondent Burden

User Interface is a key feature for reducing respondent burden in smartphone travel diary apps. Shown above (Figure 20) is an excellent example of branching logic and iconography from DVMobile. In the Figure 20 example, DV Mobile shows how a series of questions about mode and travel party can be asked in one screen, using iconography. Another example of this is Mobile Market Monitor (Figure 21). This type of approach to user interface design should be explored by all smartphone app vendors.

Another feature that reduces respondent burden is mode and activity inference. By inferring aspects of trips, apps can reduce the time it takes respondents to reply to surveys. Inference transfers respondent responsibility from recording trip details to validating trip details. Apps such as Mobile Market Monitor and services such as Google Location History are using machine-learning technology to infer mode and activity. This technology should continue to improve and adoption of machine learning inference technology is expected across the marketplace.

Figure 21 -- Mobile Market Monitor, Iconography and Validation

7.2.2 Battery Life

Proposals should contain details on how the app-maker balances the trade-off of data quality and battery consumption. The most popular technique for preserving participant's battery is "phased sampling." This process involves periods of "sleep" and "wake" cycles for the smartphone GPS. In the "sleep" cycle, GPS pings with much lower frequency. In the "wake" cycle the GPS pings with much higher frequency. "Sleep" cycles are used when the participant is stationary. "Wake" cycles are used when the participant is moving from one location to another. Phased sampling is a trade-off between battery life and data quality. If a participant begins movement during a "sleep" cycle there is a possibility of missing exact origin location and departure times. However, if no "sleep" cycle is included in the GPS units' activity then users' devices experience rapid battery drain.

Another tactic for preserving battery life is using the phone's other sensors in concert with a "phased sampling" approach. Accelerometers, for instance, can be used to trigger an increase in GPS sampling frequency. When accelerometer measurements reach a pre-defined threshold the application moves the GPS sampling into a "wake" cycle.

All smartphone survey applications should be tested for battery life. Improvements are continually being made on this aspect of the technology. The survey technologies sampled in 2015 and 2016 showed drastic differences in battery life outcomes. The highest performing technology in 2016 was the DVMobile application which employs an open source customizable plug-in to aid in battery life preservation.

7.2.3   Data Plans

Proposals should detail how the application transmits data collected during the travel day. For the sake of equity and preserving a participant's positive attitude toward the survey it is important to give the user several options for transmitting data collected by the application. Many app-makers, for instance, give users the option to transmit data only when a Wi-Fi connection is available.

7.2.4   Equity

Proposals should detail what steps will be taken to ensure that all users, regardless of technical competency, will be able to use the application as intended. Iconography aids in this, as does language translation software. Trip inference using machine learning is another means for achieving equity within the survey tool.

7.2.5   Data Processing & Validation

Proposals should detail how data gathered from the application will be processed and presented to the user. At present, machine-learning is capable of making modestly accurate inferences on a number of trip characteristics. Nevertheless, there are some areas where inference is challenging. Differentiating between private car and taxi via sensing technology, for instance, remains difficult. As a result, proposed smartphone applications that use machine learning to identify trip characteristics should also outline detail an approach to validating machine inferences.

7.2.6   Privacy

Proposals should detail how participant's data will be protected.

7.2.7   The 2016 Smartphone Travel Survey App Landscape

Table 2 lists the best practices for smartphone travel diary apps that were field test by the research team.

Table 3 -- Smartphone Best Practices

Smartphone Travel Survey Application Best Practices	
Best Practice	**Feature Details**
Reducing Respondent Burden	Use of icons, branching logic, user onboarding
	Place saving features that remove respondent validation steps
	Use of user notification and survey prompts
Battery Conservation	Use of customizable plug-in (bgGeo) to aid in battery conservation
Data Plans	Features that collect data offline for sync with wifi
Trip Recording	Automatically detects trip starts and stops, suggests trip mode and trip purpose for validation by user
Equity	Use of icons and branching logic improve usability by non-English speakers and illiterate respondents
	Language translation

	Offline data collection makes it possible for low-income respondents to participate without expensive data plans
Data Collection and Validation	Machine learning features for detecting places, modes, etc.,
Privacy	Data collected on secure servers and privacy policies that notify users on how their data will be used and how their privacy will be protected

7.3 PASSIVE DATA PILOT STUDY

A passive data approach relies on data acquired from various sources integrated using data science techniques to create travel diaries for a synthetic population. Unlike traditional data collection and modeling strategies, modeling with passive data does not require a travel survey to be conducted. A passive data model relies on the same data categories--O-D data (including O-D travel time) and population/demographic data (including household information)--but sources this information from third party providers, instead of individuals in the region. Further details can be found in Kressner et al. (2016).

AVAIL recommends a small-scale pilot study to examine the effectiveness of a passive data approach in the NYMTC region.

An agile passive data model was built for the French Broad River MPO (FBRMPO). The FBRMPO serves a population of more than 450,000 people. Automotive travel dominates the region; nearly 95% of all trips are made using automobiles. As a result, the model was built to accommodate only automotive travel. However, it would be possible to expand the model to include multimodal assignment capabilities.

Kressner et al. (2016) claim model development took 75 person-hours to complete, including "processing all the passive datasets, running the simulation model, processing the results into a format accepted by the assignment model, and validation." For the NYMTC region, however, it is an absolute necessity for modeling capabilities to include multimodal assignment, a feature absent in the Kressner et al. model. However, even after adjusting for the more complex urban environment of New York, it is reasonable to imagine the construction of a relatively inexpensive model.

The research team estimates that a simple passive data model that tests this methodology in a series of sub-regions, can be built for NYMTC at approximately $200,000-$450,000, assuming total data costs of approximately $100,000-$200,000 for the pilot study geographies and another $100,000-$250,000 for model development.

A pilot study to examine the feasibility of an agile passive data model in the NYMTC region would have numerous benefits for NYMTC. First, it is an inexpensive way to explore an innovative approach to data collection that could save the organization hundreds of thousands of dollars while augmenting the quantity of data and the "up-to-dateness" of data used in modeling. These expenses are further reduced if consideration is given to the reusability of the acquired data sets.

The most expensive of the needed data sets for passive data modeling is origin-destination data (O-D data). O-D data can be purchased to be suitable for passive data modeling and for validation or calibration of NYMTCs extant travel demand model, the NYBPM. Similarly, O-D travel time data can be used to answer questions about travel time reliability throughout different days and different times of the year.

Household data sourced from consumer marketing companies can be used to update the synthetic population created in the NYBPM modeling process. Further, this data can expand upon ACS PUMS data permitting richer lifestyle segmentation for the core choice models that underlie the NYBPM.

Importantly, the development of a small-scale passive data model does not mean replacing the NYBPM. As shown above, acquiring the data needed to construct the passive data model contributes to the efficacy of the NYBPM, as a form of validation. Further, data that comes out of the passive data model can be fed into the assignment model of the NYBPM (or any dynamic assignment model). Investing in a small-scale passive data model is one step toward adopting new data collection strategies that, in the long-term, will drive down costs and leverage novel techniques for integrating data and the rapid growth of sensing and consumer data to create a more accurate and up-to-date picture of travel.

7.3.1   Data sets
In the Data Science Pilot Study, the researcher/consultant would identify, acquire, and analyze large data sets that may be used to generate a synthetic population. Potential data sets include:

Sociodemographic Data

- American Community Survey (ACS)
- Census Transportation Planning Products (CTPP)
- National Household Travel Survey (NHTS)
- Nielsen consumer data
- Epsilon
- InfoUSA
- Axiom
- Experion
- Department of Motor Vehicle (DMV) vehicle registration data

Location Data

- Call Detail Records (CDRs) from Airsage
- INRIX movement data
- HERE
- Google Location History (GLH)
- Uber
- Transit Applications
- Taxi

7.4   GOOGLE LOCATION HISTORY (GLH) PILOT STUDY
Another potential component to the Data Science Pilot Study could be the construction of a Google Location History website to investigate the feasibility of crowdsourcing user's GLH as a supplement to or replacement of a Regional Household Travel Survey (RHTS). Researchers would conduct low-cost outreach efforts to recruit volunteers to upload their GLH for analysis.

When an Android device—smartphone or tablet—uses its GPS to obtain latitude and longitude coordinates the device logs these coordinates to a linked Google account. This historical record of location data is known as Google Location History. Google makes this record available to users for viewing through Google Maps (https://www.google.com/maps/timeline) and download through Google Takeout, a webpage for downloading personal data from Google products

 (https://takeout.google.com/settings/takeout).

Google uses machine learning and algorithms to infer activity and mode from a user's Location History. In the data Google attaches confidence intervals to inferences.

The GLH project would consist of the following key elements:

- Construction of Google Location History Crowdsource Website.
- Outreach to potential volunteers.
- Data analysis.

Volunteers will be directed to the GLH survey website where they will be given instructions for downloading and uploading their Google Location History data. A simple online survey could also be included in the GLH website for collecting socioeconomic demographic on respondents.

```
}, {
  "timestampMs" : "1458311046400",
  "latitudeE7" : 426861737,
  "longitudeE7" : -738226538,
  "accuracy" : 20,
  "activitys" : [ {
    "timestampMs" : "1458311044686",
    "activities" : [ {
      "type" : "tilting",
      "confidence" : 100
    } ]
  } ]
}, {
  "timestampMs" : "1458311018515",
  "latitudeE7" : 426857694,
  "longitudeE7" : -738219186,
  "accuracy" : 21

  "timestampMs" : "1458310946647",
  "latitudeE7" : 426846228,
  "longitudeE7" : -738227457,
  "accuracy" : 50

  "timestampMs" : "1458310893136",
  "latitudeE7" : 426836926,
  "longitudeE7" : -738214613,
  "accuracy" : 28,
  "activitys" : [ {
    "timestampMs" : "1458310893807",
    "activities" : [ {
      "type" : "onBicycle",
      "confidence" : 92
    {
      "type" : "unknown",
      "confidence" : 8

  "timestampMs" : "1458310811131",
  "latitudeE7" : 426820707,
  "longitudeE7" : -738159904,
  "accuracy" : 37

  "timestampMs" : "1458310728618",
  "latitudeE7" : 426786837,
```
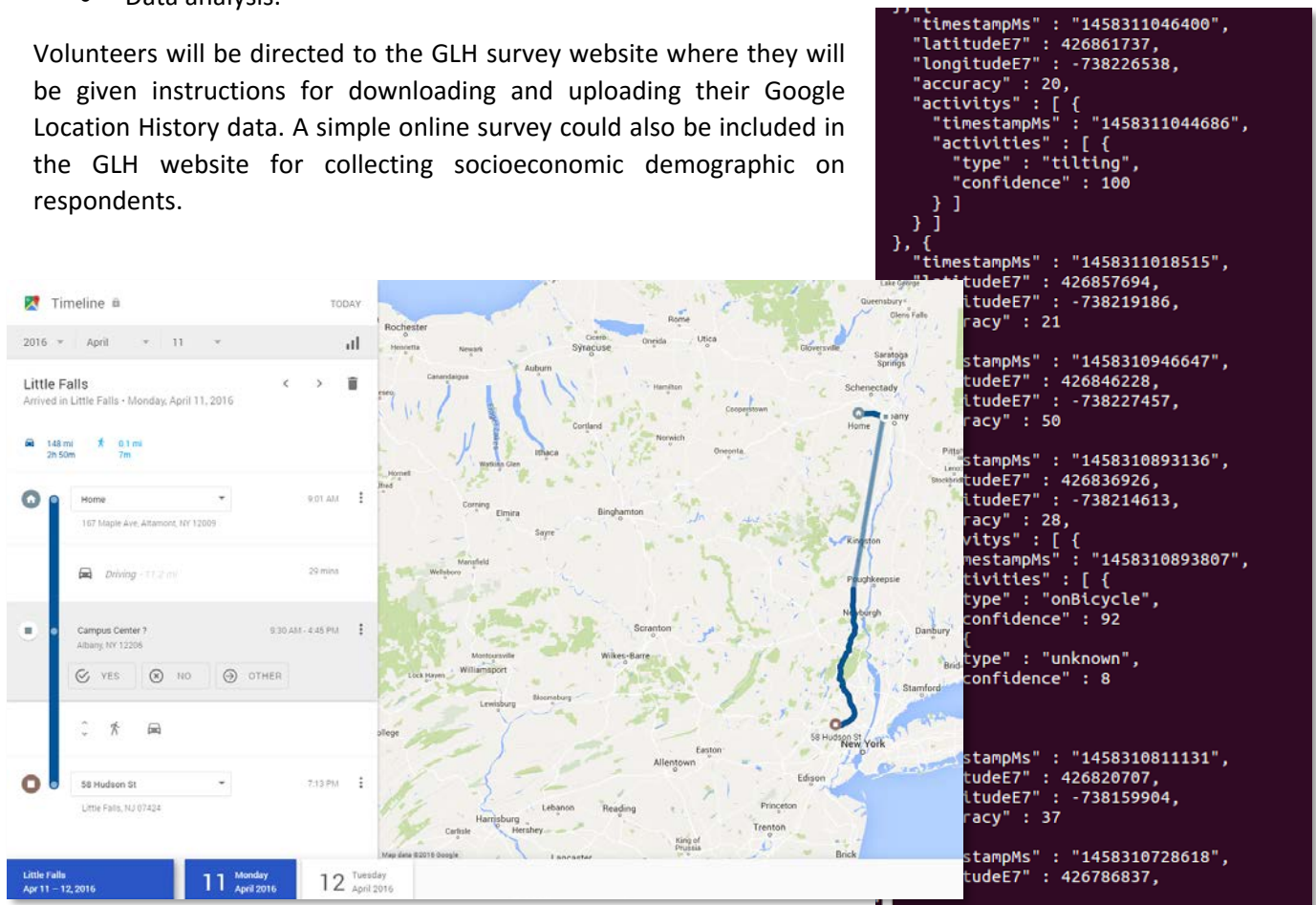
Figure 22 -- Google Location History Data and Visualization

7.4.1.1   Construction of Google Location History Crowdsource Website

The website will serve as the primary interface with the public. It would contain information on the Pilot Study, steps taken to ensure the privacy of the volunteer, clear instructions on how to download Google Location History and upload the .zip file to the website, a Frequently Asked Questions (FAQs) section, a portal for uploading Location History, a socioeconomic demographic survey, and analytics on respondent travel behavior.
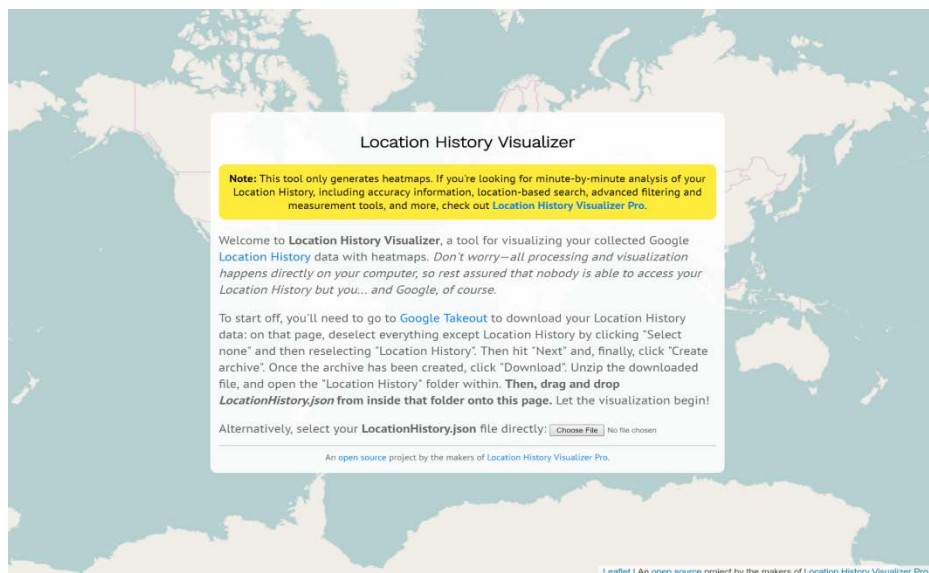
Figure 23 -- Open Source Google Location History Visualizer

Figure 22 shows a website (https://locationhistoryvisualizer.com/heatmap/) that allows users to upload and visualize their GLH.

7.4.1.2   Outreach to Potential Google Location History Volunteers

Outreach strategies will be left to the discretion of NYMTC and the consultant. Nevertheless, AVAIL recommends engaging in outreach through traditional and non-traditional media outlets.

Outreach strategies would be needed to increase the viability of this approach as some sections of the population may be less likely to volunteer their information.

Partnerships with popular travel or New York improvement Twitter accounts could be leveraged to bring awareness (e.g. 511 NY). Researchers may be able to partner with Waze and other travel apps to advertise on those platforms. Facebook, whose network includes millions of people in the NYMTC jurisdiction, offers low cost outreach options.

8 ADDITIONAL RECOMMENDATIONS AND FUTURE CONSIDERATIONS

8.1 USING PASSIVE DATA FOR CALIBRATION & VALIDATION

Passive Data has great potential as a tool for calibrating and validating travel models. NYMTC should consider increasing its use of passive data for validation and calibration. Below are three examples of the use of passive data technology in model validation.

8.1.1 Airsage

AirSage, cellphone trace data, is sometimes used for model calibration. The South Alabama Regional Planning Commission (SARPC) used AirSage data to calibrate their model. The Research Triangle Region of North Carolina used AirSage to validate their Triangle Regional Model (TRM). Other MPOs throughout the United States have used AirSage data to calibrate or validate their models.

AirSage collects and aggregates cell phone traces. When a mobile device connects to the cellular network it leaves a trace of its location, device ID, and timestamp. AirSage collects these traces and uses proprietary algorithms to identify trips, and infer origin, destination, and trip purpose for trips in the area under study.

Modelers use AirSage data to adjust trip distribution models. In the case of the SARPC, AirSage data was used to create values for average trip length by trip purpose and trip length frequency distributions and to calculate friction factors--mathematical interpretations of traveler's response to travel impedances (American Association of State Highway and Transportation Officials, 2014).

AirSage data is valuable for model calibration because it captures actual travel. Models calibrated with AirSage data may represent travel more accurately in their region. However, at present, AirSage is not a replacement for household travel survey data. AirSage data is necessarily aggregated to protect the privacy of mobile phone users. The sociodemographic characteristics of travelers captured by AirSage are currently unknowable. Though trip purpose can be algorithmically inferred, the level trip of purpose specificity is limited by location inaccuracies. Further, inferences have a small but significant error rate. Finally, it is not currently possible to map intra-household interactions that may affect travel decisions (e.g. rideshare or vehicle availability).

8.1.2 Probe Generated Origin/Destination Datasets

Passively collected, probe generated, Origin/Destination Data is comparable to the Airsage cellphone trace data. Vendors, such as HERE, INRIX, and TomTom, previously offered speed related probe data, but are now beginning to offer origin/destination data. These datasets will have similar capacity as the Airsage data and will offer potentially higher levels of granularity.

8.1.3 Skycomp

Skycomp produces traffic flow metrics using time-lapse aerial photography, video and direct observation. Their aerial traffic surveys deliver metrics for peak periods over study areas of up to twelve square miles. They have the ability to provide the following set of concurrent metrics:

- Origin-Destination (O-D) Matrices
- Route Information

- Traffic Volumes
- Vehicle Classification
- Turning Movement Counts (TMCs)
- Queue Lengths
- Travel Times
- Densities and LOS

8.1.4 "Big Data Exhaust"

Davidson (2016) outlines an approach to model validation that utilizes passive data from crowd-sourcing. Davidson used data harvested from a transit schedule/routing smartphone application, The Transit App, to validate NYMTC's Best Practices Model. The author built a table of origin-destination pairs from searches made by users of the app (the origin, destination, and device ID are logged in a server when a search is made).

Next, the author compared those O-D pairs to those taken from the 2010-11 NYMTC RHTS. Comparison occurred at the community board level geography (71 total in the NYMTC region). The Transit App data was not a record of actual trips but searches made to the app. The assumption is made that many of the trips that are searched are also made. The Transit App dataset showed trips made that were not represented in the RHTS. Although it is unclear which dataset more closely represents what trips actually occur in the region, it is possible to identify discrepancies between the two datasets. For instance, The Transit App identified a large number of trips from the airport into the city where the RHTS recorded no such trips. It is most likely that these trips represented travel by visitors to the region and would not have been possible to capture with a household travel surveying effort.

8.2 EMERGING POSSIBILITIES

This section describes emerging technologies or potential partnerships that are promising opportunities in the near future.

8.2.1 Partnerships

Future travel data collection efforts could include partnerships with other agencies, especially regional partners. For instance, the joint purchase of passive datasets or the cooperative development of a regional smartphone application with embedded travel survey technology could provide value to numerous transportation agencies in the region.

8.2.1.1 Transit Ticketing Application with Built In Survey Technology

NJ TRANSIT recently rolled out a smartphone application that allows users to view schedules and purchase and display tickets on a phone. Users are required to supply their credit card and billing information and allow the application to verify location, which means that a user's location and identifying information be connected to ticket search and purchase data.

There may be an opportunity for NYMTC to partner with NJ TRANSIT on an application update to include travel survey capabilities, GPS tracking, and travel behavior data outputs, with effective privacy protections in place. Such additional features, although costly, are relatively easy to implement in terms of software development and given the dedicated base of users, would allow for immediate deployment without the costs of recruitment.

Perhaps it is also possible to partner with MTA on a similar transit app for NYC that allows people to view transit schedules, buy tickets, etc. NYMTC could consider leveraging travel survey funding to help in building the application on the grounds that the application would include survey capabilities, GPS tracking, and travel behavior data outputs.

Partnerships with MTA and NJ TRANSIT may provide unprecedented data on transit trips and access to potential survey participants.

8.2.1.2   Sidewalk Labs

Sidewalk Labs is a company under the Alphabet/Google umbrella. Sidewalk Labs want to improve understanding of urban transportation by integrating emerging data sets into a single platform that can provide insights for traffic engineers and transportation planners as well as commuters and residents. Though Sidewalk Labs only real accomplishment to date is the installation of Wi-Fi kiosks in New York City, their ambitions are grand. Sidewalk Labs planned analytics program, Flow, is being built in partnership with US Department of Transportation (USDOT). The Sidewalk Labs webpage states that Flow analytics can "Help cities understand how roads are being used and how people may respond to policy, infrastructure, or technology changes." Though AVAIL was unable to test the Flow platform, it would appear that Flow analyzes many of the same phenomena as travel demand models. The Sidewalk Labs website claims that Flow can:

- Integrate aggregated, anonymized smartphone data from billions of miles of trips (starting with Google's Urban Mobility program) along with sensor data (for example, via LinkNYC Wi-Fi kiosks) to create a real-time view of road and curb usage;
- Select and analyze[s] specific road segments to understand what's driving congestion based on the type of trip being made and the neighborhoods where traffic originates; and
- Simulate the impact of new roads, transit routes, mobility services, and incentives on traffic by asking "what if" questions and sharing data across Flow cities.

The Flow platform would appear to integrate more fully with transportation planning and real-time use than traditional travel demand models. For instance, Flow claims to be able to integrate with third party mobile applications to enable things like dynamic parking, which routes drivers to available parking to reduce or eliminate circling and waiting times. Another purported feature of Flow is dynamic transit, which would adjust routes on the basis of ridership demand. Flow is currently being implemented in seven cities-- Austin, Denver, Kansas City, Pittsburgh, Portland, San Francisco, and Columbus. The timeline of its availability to other cities is unclear.
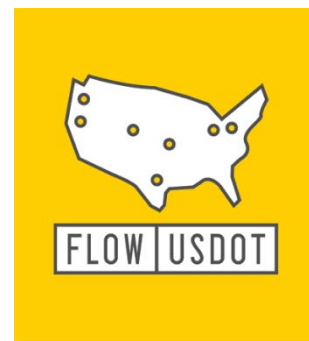
Interestingly, it seems highly likely that Flow leverages Google Location History (GLH). Google Location History is a record, kept by Google, of all the locations a phone has been while the GPS was gathering samples. This historical record of location data is available to users for viewing through Google Maps (*see* Section 7.4).

8.2.1.3   National Household Travel Survey Opportunity

One way to approach the traditional survey is through data acquisition from the New York State deployment of the National Household Travel Survey (NHTS). The NHTS is a national effort to gather data

on travel behavior trends and these trends are reported at the national level. The current methodology uses a web interface and a phone retrieval methodology, based on user preferences (https://www.nationalhouseholdtravelsurvey.com/). In addition to scoping a sample sufficient to develop a national understanding, states can also participate in the "add-on" program that uses a sampling strategy sufficient for a state-wide understanding.   New York State has taken advantage of the add-on program and is currently participating in the 2016 deployment.

The NHTS provides an immediate opportunity to expand travel survey data resources through acquisition of the 2016 NHTS data for the NYMTC region counties.   In the past, state DOTs (and state DOT partners) would be given access to the geocoded survey data just for their own add-on samples sufficient for state-wide aggregations. NYMTC has indicated participation in the 2016 add-on sample and should request the geocoded data and keep it secure for future use as a source of validation data. In addition, if these data were collected from a larger sample (e.g. recycling the survey instrument and conducting an expanded deployment) they could serve as a substitute for collecting local survey data if the data elements meet the needs of the NY Best Practices Model (NYBPM). If the core data elements are sufficient, a new strategy for developing a localized NHTS with a modified sampling frame is possible. A program similar to the American Community Survey with its continuous surveying effort could be developed for the web interface data collection effort. The data will be self-aggregating, and after the overall sampling frame is developed, the cost per completed survey would be relatively low.

The data collected for the NHTS may contain data elements similar enough to those required for the Best Practices Model to warrant an element-by-element comparison. The sampling frame used in NYS add-on program, however, would most likely not be sufficient for actual model estimation for the NYMTC region. A comparison of the sampling frame used for the NHTS 2016 deployment to the sampling frame developed for the most recent NYMTC household travel survey deployment would need to be conducted to identify areas of sampling weakness (e.g., specific household types). Based on the outcome of sampling frame comparison, it is feasible to then continue to gather additional surveys locally, using a modified sampling frame, and using the established methodology and procedures for the current deployment, in order to produce data with sufficient coverage for model estimation purposes.

8.2.1.4   *Consideration of Estimation-based "Local" Transferability Strategy*
Recent research using the 2009 NHTS tested the feasibility of estimating "utility coefficients based on observed choices from two regions, using the same exact model specification for both regions" (Bowman et al., 2014), and testing for statistically significant differences. The results of this research suggests the ability to develop model parameters that can be used in multiple locations, based on a major finding that "although estimation of models using a large local sample is best, it is better to transfer models that are based on a large sample from a comparable region than it is to estimate new models using a much smaller local sample" (Bowman et al., 2014). Taking this concept to the NYMTC region, it may be possible to estimate models for sub-areas and then transfer those parameters to other like-kind sub-areas, including using samples from non-NYMTC regions (e.g., medium dense suburban counties). This strategy would not be sufficient, however, for the five boroughs that are unique to the greater New York metropolitan region. While this research is in its infancy, it is worth following the success of other transferability studies and considering future research to serve the needs of the NYMTC region using the existing NHTS, and/or additional investments in a modified NHTS strategy for the NYMTC region.

9 CONCLUSION

Looking forward to the next two decades of travel data collection strategies to support the modeling efforts at NYMTC, there are a growing variety of approaches. These approaches include traditional survey methods, the use of GPS devices, smartphone apps, and emerging strategies that rely on administrative data and innovative methods for meeting the needs of travel demand modeling.

The analysis of these various approaches found two distinct paradigms for travel data collection—Active Data Collection (e.g. paper travel diary, online travel diary, and smartphone travel diary app) and Passive Data Collection (e.g. synthetic population and activity generation from a variety of third party data sources). In a comparative analysis three recommended pathways emerged. The first, a Paper and Online Diary Travel Survey (current practice), the second, a Smartphone Survey Approach, and the third, a Passive Data Collection, are presented with associated timeline trajectories.

The research revealed the rapidly advancing nature of travel data collection as a result of new computing capabilities, storage capacity, data science advancements, and generational cultural shifts. As a result, it will be necessary to conduct a set of sequenced pilot studies to correctly identify the travel data collection strategy that will provide the best outcomes for the NYMTC region. These pilot studies include deploying a crowdsourced strategy to gather Google Location History (GLH) from survey participants and a simultaneously deployed Smartphone and Passive Data pilot studies.

To meet the NYMTC data collection and modeling improvement schedule, it is recommended that consideration of these pilot studies to evaluate the effectiveness of these strategies begin as soon as possible. Correctly identifying the appropriate strategy will provide NYMTC with the required data elements for their travel demand modeling needs, at a potentially lower cost, and with data quality improvements.

10 BIBLIOGRAPHY

Abdulazim, T., Abdelgawad, H., Habib, K., & Abdulhai, B. (2013). Using smartphones and sensor technologies to automate collection of travel data. Transportation Research Record: Journal of the Transportation Research Board, 2383, 44-52. doi:10.3141/2383-06

American Association of State Highway and Transportation Officials (2014). Use of Mobile Phone Data for Planning Purposes. http://research.transportation.org/_layouts/AASHTORAC/FileDownload.aspx?SurveyFileID=357

Anderson, R. S., Gregory Giaimo, P. E., & Flake, L. (2016). First Large-Scale Smartphone-Based Household Travel Survey in USA – Pilot Results. http://tfresource.org/images/b/bb/ITM16_First_Large-Scale_Smartphone-Based_Household_Travel_Survey_in_USA_%E2%80%93_Pilot_Results.pdf

Bianco, M. & Lawson, C.T., (1996). "Trip Chaining, Childcare, and Personal Safety: Critical Issues in Women's Travel Behavior". *Proceedings of the Second National Conference on Women's Travel Issues* held in Baltimore, Maryland, on October 23-26, 1996.

Bowman, J. L., Bradley, M., Castiglione, J., & Yoder, S. L. (2014). Making advanced travel forecasting models affordable through model transferability. Presented at the 93rd Annual Meeting of Transportation Research Board, Washington, DC.

Bricka, S. (2008). Non-response challenges in GPS-based surveys. International Steering Committee on Travel Survey Conferences Workshop on Non-Response Challenges in GPS-based Surveys (pp. 2-5). Prepared for the 8TH International Conference on Survey Methods in Transport, Annecy, France. May 2008

Buskirk, T. D., & Andres, C. (2013). Smart Surveys for Smart Phones: Exploring Various Approaches for Conducting Online Mobile Surveys via Smartphones. Survey Practice, 5(1).

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr., J., & Ratti, C. (2013). Understanding Individual Mobility patterns from Urban sensing Data: A mobile trace example. Transportation Research Part C 26, 301-313.

Cell phone ownership, 2000-2014 (2014). Retrieved from http://www.pewinternet.org/2014/02/27/part-1-how-the-internet-has-woven-itself-into-american-life/

CISCO. Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015-2020 (2016). www.cisco.com%2Fc%2Fen%2Fus%2Fsolutions%2Fcollateral%2Fservice-provider%2Fvisual-networking-index-vni%2Fmobile-white-paper-c11-520862.html&sa=D&sntz=1&usg=AFQjCNHlwjAzjvh-CgPKUlgRpKrfitnLUg

Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M., & Zegras, P. (2013). Future mobility survey: Experience in developing a smartphone-based travel survey in Singapore. Transportation Research Record: Journal of the Transportation Research Board, (2354), 59-67.

Davidson, A. (2016). Big Data Exhaust for Origin-Destination Surveys: Using Mobile Trip-Planning Data for Simple Surveying. In Transportation Research Board 95th Annual Meeting (No. 16-2984).

Ghorpade, A., Pereira, F. C., Zhao, F., Zegras, C., & Ben-Akiva, M. (2015). An integrated stop-mode detection algorithm for real world smartphone-based travel survey. Presented at Transportation Research Board 94th Annual Meeting.

Gong, H., Chen, C., Bialostozky, E., & Lawson, C. T. (2012). A GPS/GIS method for travel mode detection in New York City. Computers, Environment and Urban Systems, 36(2), 131-139.

Greaves, S. P., Ellison, A. B., Ellison, R. B., Rance, D., Standen, C., Rissel, C., & Crane, M. (2014). A web-based diary and companion smartphone app for travel/activity surveys. In International Conference on Transport Survey Methods.

Greene, E. R., Wertman, R., Murakami, E., Hathaway, K., Geilich, M., Flake, L. (2015). "In the moment" travel study: A GPS smartphone application. Planning Applications Conference. Atlantic City, New Jersey.

Hartgen, D. T., & San Jose, E. (2009). Costs and Trip Rates of Recent Household Travel Surveys. http://www.hartgengroup.net/Projects/National/USA/household_travel_summary/2009-11-11_Final_Report_Revised.pdf

Huang, Y., & Wang, Q. (2015). Evaluation of existing smartphone applications and data needs for travel surveys (No. FHWA/TX-15/0-6767-1).

Jiang, S., Fiore, G. A., Yang, Y., Ferreira Jr, J., Frazzoli, E., & González, M. C. (2013). A review of urban computing for mobile phone traces: current methods, challenges and opportunities. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (p. 2). ACM.

Jin, P. J., Yang, F., Cebelak, M., Ran, B., & Walton, C. M. (2013). Urban travel demand analysis for Austin TX USA using location-based social networking data. In TRB 92nd Annual Meeting Compendium of Papers.

Karner, A., & Niemeier, D. (2013). Civil rights guidance and equity analysis methods for regional transportation plans: a critical review of literature and practice. Journal of Transport Geography, 33, 126-134.

Kim, Y., Pereira, F. C., Zhao, F., Ghorpade, A., Zegras, P. C., & Ben-Akiva, M. (2015). Activity recognition for a smartphone and web based travel survey.arXiv preprint arXiv:1502.03634.

Kressner, J. D., Macfarlane, G. S., Huntsinger, L., & Donnelly, R. (2016). Using Passive Data to Build an Agile Tour-Based Model.

Lee, J. H., Gao, S., Janowicz, K., & Goulias, K. G. (2015). Can Twitter data be used to validate travel demand models?. In IATBR 2015-WINDSOR.

Li, S. (2014). Innovative procedures for travel data collection and processing. Sydney Digital Theses (Open Access).

Liu, F., Janssens, D., Wets, G., & Cools, M. (2013). Annotating mobile phone location data with activity purposes using machine learning algorithms. Expert Systems with Applications, 40(8), 3299-3311.

Nitsche, P., Widhalm, P., Breuss, S., & Maurer, P. (2012). A strategy on how to utilize smartphones for automatically reconstructing trips in travel surveys. Procedia - Social and Behavioral Sciences, 48, 1033-1046. http://dx.doi.org/10.1016/j.sbspro.2012.06.1080

Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., & Maurer, P. (2013). Supporting large-scale travel surveys with smartphones – A practical approach. Transportation Research Part C: Emerging Technologies, 43, Part 2, 212-221. http://dx.doi.org/10.1016/j.trc.2013.11.005

NYMTC. (2014) 2010/2011 Regional Household Travel Survey: Final Report. https://www.nymtc.org/DATA-AND-MODELING/Travel-Surveys/2010-11-Travel-Survey

Ozbay, K., Yang, H., Morgul, E. F., Mudigonda, S., & Bartin, B. (2014). Big data and calibration and validation of traffic simulation models. Traffic and Transportation Simulation: Looking Back and Looking Ahead: Celebrating, 50, 92-122.

Rasmussen, T. K., Ingvardson, J. B., Halldórsdóttir, K., & Nielsen, O. A. (2015). Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area.Computers, Environment and Urban Systems.

Richardson, A. J., Ampt, E. S., & Meyburg, A. H. (1995). Survey methods for transport planning (p. 314). Melbourne: Eucalyptus Press.

Safi, H., Assemi, B., Mesbah, M., Ferreira, L., Hickman, M. (2015). Design and implementation of a smartphone-based system for person travel survey: Case study from New Zealand. Conference paper. http://www.researchgate.net/profile/Behrang_Assemi/publication/277249675_Design_and_Implementation_of_a_Smartphone-based_System_for_Personal_Travel_Survey_Case_Study_from_New_Zealand/links/5565163908ae89e758fd9bcf.pdf

Schillewaert, N., & Meulemeester, P. (2005). Comparing response distributions of offline and online data collection methods. International Journal of Market Research, 47(2), 163-178.

Smartphone ownership, over time (2014, February 27). Retrieved from http://www.pewinternet.org/2014/02/27/part-1-how-the-internet-has-woven-itself-into-american-life/

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

Tang, K. X., & Waters, N. M. (2005). The internet, GIS and public participation in transportation planning. Progress in Planning, 64(1), 7-62.

Wolf, J., Bachman, W., Oliveira, M., Auld, J., Mohammadian, A., & Vovsha, P. (2014). Applying GPS Data to Understand Travel Behavior (Vol. 1). Washington, DC: Transportation Research Board.

Yuan, M. (2001). Reoperationalizing economic data collection. Annals of Tourism Research, 28(3), 727-737.

Zhao, F., Ghorpade, A., Pereira, F. C., Zegras, C., & Ben-Akiva, M. (2015a). Stop detection in smartphone-based travel surveys. Transportation Research Procedia, 11, 218-226.

Zhao, F., Pereira, F. C., Ball, R., Kim, Y., Han, Y., Zegras, C., & Ben-Akiva, M. (2015b). Exploratory analysis of a smartphone-based travel survey in Singapore. In Transportation Research Board 94th Annual Meeting (No. 15-4343).

10.1 ADDITIONAL RESOURCES

Abbasi, A., Rashidi, T. H., Maghrebi, M., & Waller, S. T. (2015). Utilising Location Based Social Media in Travel Survey Methods: bringing Twitter data into the play. In Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (p. 1). ACM.

Berger, M., & Platzer, M. (2015). Field Evaluation of the Smartphone-based Travel Behaviour Data Collection App "SmartMo". Transportation Research Procedia, 11, 263-279.

Biler, S., Šenk, P., & Winklerová, L. (2013). Willingness of Individuals to Participate in a Travel Behavior Survey Using GPS Devices. NTTS.

Bonsall, P., Wolf, J., & Holroyd, S. (2006). Review of the potential role of 'new technologies' in the National Travel Survey. Contract no UG599. Institute for Transport Studies, University of Leeds.

Bourbonnais, P. L., & Morency, C. (2013). Web-based travel survey: a demo.Transport survey methods: best practice for decision making, 207-223.

Carlson, J. A., Jankowska, M. M., Meseck, K., Godbole, S., Natarajan, L., Raab, F, & Kerr, J. (2015). Validity of PALMS GPS scoring of active and passive travel compared with SenseCam. Med. Sci. Sports Exerc, 47(3), 662-667.

Clifton, K. J., & Gehrke, S. R. (2014). Wider Dissemination of Household Travel Survey Data Using Geographical Perturbation Methods.

De Abreu e Silva, J., & Davis, M. (2015). Workshop Synthesis: Respondent/Survey Interaction in a World of Web and Smartphone Apps. Transportation Research Procedia, 11, 289-296.

Deutsch, K., McKenzie, G., Janowicz, K., Li, W., Hu, Y., & Goulias, K. (2012) Examining the use of smartphones for travel behavior data collection during the FHWA webinar on Smartphones and travel behavior.

Duffy, B., Smith, K., Terhanian, G., & Bremer, J. (2005). Comparing data from online and face-to-face surveys. International Journal of Market Research, 47(6), 615.

Ellison, R. B., Ellison, A. B., & Greaves, S. P. (2015). Assessing spatial distribution and variability of destinations in inner-city Sydney from travel diary and smartphone location data.

Fan, Y., Chen, Q., Liao, C. F., & Douma, F. (2012). Smartphone-based travel experience sampling and behavior intervention among young adults.

Fan, Y., Wolfson, J., Adomavicius, G., Vardhan Das, K., Khandelwal, Y., & Kang, J. (2015). SmarTrAC: A Smartphone Solution for Context-Aware. Minneapolis: Volpe Center at the U.S. Department of Transportation & Intelligent Transportation Systems Program.

Gerike, R., & Lee-Gosselin, M. (2015). Workshop Synthesis: Improving Methods to Collect Data on Dynamic Behavior and Processes. Transportation Research Procedia, 11, 32-42.

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. Behavior research methods, 47(4), 1237-1259.

Gong, L., Morikawa, T., Yamamoto, T., & Sato, H. (2014). Deriving personal trip data from GPS data: A literature review on the existing methodologies. Procedia - Social and Behavioral Sciences, 138, 557-565. http://dx.doi.org/10.1016/j.sbspro.2014.07.239

Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C., & Crane, M. (2015). A web-based diary and companion smartphone app for travel/activity surveys. Transportation Research Procedia, 11, 297-310.

Hoogendoorn-Lanser, S., Schaap, N. T., & OldeKalter, M. J. (2015). The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. Transportation Research Procedia, 11, 311-329.

Huang, A., & Levinson, D. (2015). Axis of travel: Modeling non-work destination choice with GPS data. Transportation Research Part C: Emerging Technologies, 58, 208-223.

Körner, C. (2012). Modeling visit potential of geographic locations based on mobility data (Doctoral dissertation, Bonn, Rheinische Friedrich-Wilhelms-Universität Bonn, Diss.

Kurkcu, A., Ozbay, K., & Morgul, E. F. (2016). Evaluating the Usability of Geo-located Twitter as a Tool for Human Activity and Mobility Patterns: A Case Study for New York City. In Transportation Research Board 95th Annual Meeting (No. 16-3901).

Lane, S. J., Heddle, N. M., Arnold, E., & Walker, I. (2006). A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC medical informatics and decision making, 6(1), 1.

Liao, C. F., Fan, Y., Adomavicius, G., & Wolfson, J. (2016). Battery-Efficient Location Change Detection for Smartphone-Based Travel Data Collection: A Wi-Fi Fingerprint Approach. In Transportation Research Board 95th Annual Meeting (No. 16-2090).

Maruyama, T., Sato, Y., Nohara, K., & Imura, S. (2015). Increasing Smartphone-based Travel Survey Participants. Transportation Research Procedia, 11, 280-288.

McNally, M. G., & Lee, M. S. (2002). Putting behavior in household travel behavior data: An interactive gis-based survey via the internet. Center for Activity Systems Analysis.

Mobile Technology Fact Sheet. (2013). Retrieved from http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/

Mohammadian, K., Karimi, B., Pourabdollahi, Z., & Frignani, M. (2013). Modeling Seniors' Activity-Travel Data. FHWA-ICT-13-026.

Mooney, S. J., Sheehan, D. M., Zulaika, G., Rundle, A. G., McGill, K., Behrooz, M. R., & Lovasi, G. S. (2016). Quantifying Distance Overestimation From Global Positioning System in Urban Spaces. American journal of public health, 106(4), 651-653.

Morris, S., Gimblett, R., & Barnard, K. (2005). Probabilistic travel modeling using GPS. In International Congress on Modelling and Simulation, Melbourne (pp. 12-15).

Nour, A., Hellinga, B., & Casello, J. (2016). Classification of automobile and transit trips from Smartphone data: Enhancing accuracy using spatial statistics and GIS. Journal of Transport Geography, 51, 36-44.

Ribeiro, M. D., Larrañaga, A. M., Arellana, J., & Cybis, H. B. (2014). Influence of GPS and Self-reported Data in Travel Demand Models. Procedia-Social and Behavioral Sciences, 162, 467-476.

Safi, H., Mesbah, M., & Ferreira, L. (2013, October). ATLAS Project–Developing a mobile-based travel survey. In Australasian Transport Research Forum.

Taghipour Dizaji, R. (2013). Acquiring Multimodal Disaggregate Travel Behavior Data Using Smart Phones.

Wells, T., Bailey, J., & Link, M. (2013). Filling the void: gaining a better understanding of tablet-based surveys. Survey Practice, 6(1).

Xu, Y. (2011). Using Volunteer Tracking Information for Activity-Based Travel Demand Modeling and Finding Dynamic Interaction-Based Joint-Activity Opportunities. http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2033&context=utk_gradthes

11 APPENDICES

11.1 APPENDIX A: TRIAL PROCESS FOR ASSESSING SMARTPHONE SURVEY TECHNOLOGY

In December 2015, AVAIL began the process of interviewing companies and MPOs that developed or implemented a travel survey that utilized smartphone survey technology. Our initial leads were based on an extensive literature review on the current evolution of household travel surveys.

Our initial phase included contacting any large MPO that conducted a travel survey which utilized smartphones in any capacity and asking them which company they partnered with. AVAIL tried to maintain the broadest perspective on the most current capabilities of smartphone survey technology within a global assessment. We conferenced with companies currently in Singapore and New Zealand.

The initial outreach to every APP designer included an introductory survey.

- What is your overall intention for your app?
- What was the overall cost of your study?
- Is your app commercially available? If it isn't yet do you have an idea as to what the cost of expanding would be?
- Regardless, is it open source? And how much did it cost to make it? Do you have an interest and capacity in expanding to a market like NYC?
- What steps did you take to recruit and/or retain users? And how much was cost per respondent?
- In what format(s) is the data collected by app? Can you share a slug?
- What measures were taken to preserve the privacy of the user and data?
- What steps were taken to preserve the battery life of the smartphone?
- How did you ensure minimal consumption of user's data plan?
- What was the method for surveying the traveler on travel reasons? Other validation?
- Can we have a trial run/ test the app?

After receiving the responses, the second phase included a conference call meeting with each app companies where we discussed survey answers and did an initial walk through of their app technology. The third phase was an actual trial of the app. This included trials in NYC as well as New York's Capital Region. The final phase was a follow up meeting to discuss the collected data and any further inquiries post trial. Many app companies are quickly updating their apps as the technology evolves. Indeed, one company, DV Mobile, launched a new version of the app during our research program and AVAIL was able to complete two trials; one with each app version.

Parallel to the app trials, AVAIL interviewed MPOs that had an opportunity to work with the app. This allowed for the inclusion of not only a technical overview of the app, but also a perspective that includes the logistics of actually disseminating region wide surveys utilizing smartphones.

AVAIL encourages interested parties to be on the lookout for University of Toronto Transportation Research Institute's report on their "controlled experiment assessing the strengths and weaknesses of multiple location logging smartphone apps, as well as trace processing algorithms and software."

11.2 Appendix B: Key Elements for Consideration in Assessing Passive Data Approaches

11.2.1 Background

Based on the research conducted for this report, it may be possible to create trip tables for input into the NY Best Practices Model (NYBPM) without conducting a travel survey. Researchers used "household-level data, firm-level data, origin-destination data, travel time data in traffic, and the National Household Travel Survey" to construct synthetic trip tables for input into a regional MPO travel demand model in North Carolina and achieved promising results (Kressner et al., 2016). By using statistical techniques and other data science methods, researchers were able to create synthetic trip tables from these "passive" data sets, suggesting that it may be possible to eliminate travel surveys from the travel demand modeling process. However, it remains to be seen whether or not it is possible to adapt this, or some similar methodologies for use in a complex, multi-modal travel demand model such as the NYBPM.

11.2.2 Technical Elements

The following criteria are crucial for assessing passive data approaches:

How are origins and destinations generated? What data are used to generate those points? How are trip purposes generated? How are trip modes generated? How does the approach handle trip chaining? Approaches should communicate with clarity and conciseness the answers to these questions. Approaches are likely to generate origins and destinations in one of three ways (ordered by increasing specificity):

1. Randomized within a polygon,
2. Semi-randomized within a polygon (by incorporating land use data), or
3. Whole universe point-based O/D.

It is not necessary to use the same method to generate origins and destinations. Due to logistical constraints, for example, it may be possible to use the whole universe for origins, but a randomized polygon approach for destinations.

Approaches should present a sound reasoning on the number and type of trip purposes chosen for assignment. Further, it is important to analyze methods for assigning origins and destinations to households, methods for generating a trip purpose distribution by sociodemographic characteristics, and methods for mode selection.

11.2.3 Methodological considerations

Point-based Trip Purpose Allocation is the preferred method for developing Trip Tables using passive travel data. It has the potential to achieve higher trip table estimation accuracy than generating estimates based on aggregated geographies. The Point-Based Trip Purpose Allocation methodology described below is based on the prior modeling experience and information developed by Kressner et al., (2016). It should be noted, however, that the only known evidence of the accuracy of the passive data approach is the paper published by Kressner et al., (2016), and that their exact methodology was not delineated in their report and remains unavailable.

Below is a basic overview of two approaches to passive travel data modeling: Point-based and Geographic Flows. The following outline addresses a possible process for generating trip tables based on origins, destination and purpose. It covers the assignment of origins and destinations to households/landmarks and of generating trip purpose distribution at the point-based level and at the polygon level (methods of mode selection are not addressed here).

- **Acquire datasets:** Dataset Acquisition is a key point of consideration. The potential accuracy of a proposed passive data methodology is higher if all data being used is point-based and if the methodology utilizes all three of the following data types: Probe Travel Data, Socioeconomic Data, and Establishment Land-Use Data.
  - o **Probe Data:** A passive data approach methodology that utilizes Origin/Destination (O/D) data is able to develop trip tables based on point-based synthetic trip information. Aggregated geographic flows data will necessarily aggregate trip tables and corresponding purposes, modes, etc. to the polygon. A methodology that utilizes a point-based O/D dataset has potential to be more accurate than a geographic flow dataset. Examples of probe datasets:
    - ▪ Point-based O/D
      - • INRIX *Trips*
      - • HERE O/D
      - • Google Location History (GLH)
    - ▪ Aggregated geographic flows
      - • AirSage
      - • CTPP
  - o **Socioeconomic Data**
    - ▪ Publicly available
      - • American Community Survey (ACS)
      - • Public Use Microdata (PUMS)
    - ▪ Proprietary. The use of marketing data appears to be the easiest and most reliable method of assigning socioeconomic demographic data to households at the point-based level.
      - • Epsilon
      - • InfoUSA
  - o **Establishment Land Use Data**
    - ▪ InfoGroup
    - ▪ NETS
    - ▪ Dunn & Bradstreet
    - ▪ Google Places API
  - o What would be the estimated operational cost of purchasing/acquiring data sets needed to create synthetic trip tables?
- **Possible Methodology for Construction of Passive Data Trip Table**
  - o Creating Trip Tables with Trip Purpose Distribution by Destination Category
    - ▪ **Point-based Trip Purpose Allocation**
      - • Create HH IDs for SE data.

- Map Household IDs (HH ID) from SE data to origins in O/D data to create Synthetic Origins.
- Using synthetic origins, cross-reference work locations from SE data with corresponding destinations from O/D data to create Synthetic Destinations with trip purpose "work".
- Cross remaining non-work location destinations with establishment/landmark data to assign non-work trip purposes.
  - Assign Synthetic Destination IDs and Destination Categories to points using Landmark Data.
  - Assign Destination Category to each Synthetic Destination ID.
  - Assign trip purposes to HH ID by cross-referencing remaining (non-work) destinations in O/D data to Synthetic Destination IDs.
- Assign trip purpose distribution to each HH ID by Destination Category.
- Develop database that collects trip purposes by various socioeconomic attributes and geographic location.
  - Distance weighted random distribution formula for assigning other destinations from origins.
  - Attractiveness x distance = distance attractiveness weight (weighted random distribution).
  - Assign confidence intervals to estimates.
- Assign Synthetic Origins and Destinations with associated confidence intervals to each trip in the O/D data set.
- Aggregate trips by socioeconomic characteristic to TAZ level.
- Develop Trip Table for entry into New York Best Practices Model.
- Validate using current validation methodology.
  - Counts / farebox
- **Geographic Flows (TAZ-level) Trip Purpose Allocation Method**
  - Randomly distribute households across TAZs.
  - Generating trip purpose (algorithm behaves differently depending on trip purpose).
    - Trip Purpose Work:
      - Reference CTTP for estimate of TAZ to TAZ work flows.
      - Develop a distribution to other TAZs. Weighted random distribution assignment to one.
      - If whole universe of destinations is assumed then establishment data would be required.
        - InfoUSA, Google Places API
      - Assigned to destination in other TAZ.
    - Trip Purpose Other:
      - Distance weighted random distribution formula for assigning other destinations from origins.

- Attractiveness x distance = distance attractiveness weight (weighted random distribution).
  - o Assign Trip Purpose Distribution by Destination Category.
- Develop TAZ to TAZ Trip Table by assigning Origin and Destination distribution by destination category and socioeconomic variables.
- Run the trip table through the New York Best Practices Model.
- Validate using current validation methodology.
  - o Counts / farebox
  - o How well do the point-based trip tables match the like-kind time frame of the NYBPM trip table?

The methodological and cost considerations outlined above are intended to facilitate understanding of a passive data approach to travel demand modeling. It could be used as a starting point for the development of a more in-depth application of a passive data modeling approach.
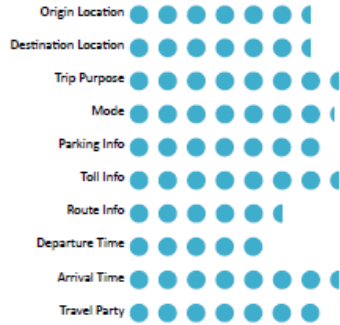
Reference:

Kressner, J. D., Macfarlane, G. S., Huntsinger, L., & Donnelly, R. (2016). Using Passive Data to Build an Agile Tour-Based Model.

Online Diary

Description: A survey created for the web. The MPO creates a webpage and directs survey participants to enter the code they received when agreeing to participate in the survey. Participants enter their travel information periodically throughout the day or at the end of their travel day based on their convenience. Online diary can mirror the layout of a paper travel diary or can employ different workflow methods to encourage accuracy and completion and minimize burden.
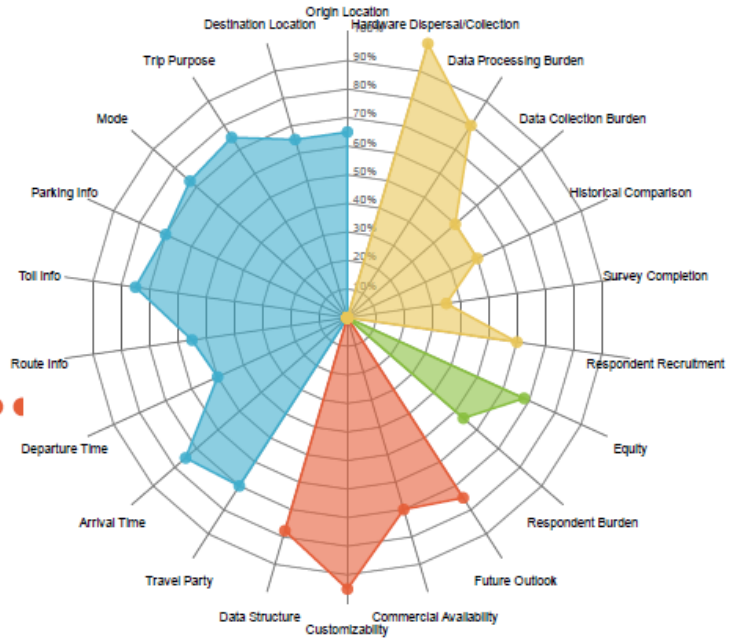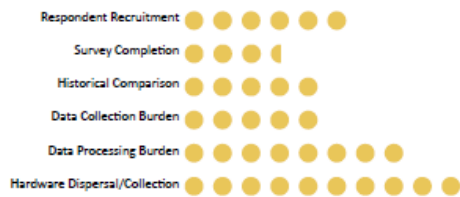

Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

68%

Rank

9

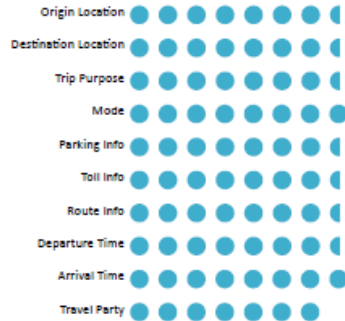| Pros	Cons
• Familiar medium for engaging with surveys and form completion.	• Limited trip characteristics accuracy due to self-reporting.
• Large display and high usage of Google Maps and other mapping services increases comfort with viewing and adjusting origin and destination locations.	• Requires manual entry of trip characteristics (as opposed to validation) limiting its ability to reduce respondent burden.
• Web-development may be straightforward and inexpensive while being easily customizable based on user demographics (age and language).	

Outlook: Without any location technology it is difficult to imagine the Online Diary becoming popular or particularly useful for NYMTC.

Paper Travel Diary + Smartphone GPS

Description: After agreeing to participate in a survey respondents are mailed a paper travel diary and instructed to complete it for their assigned day of travel. The paper travel diary contains codes for activities and travel modalities as well as spaces to enter these and other trip characteristics. Respondents may mail back these travel diaries or report their contents via telephone. Additionally, survey participants are asked to provide GPS information taken from their smartphones. In practice, this requires a basic application that must be downloaded by the participant. The only necessary functions for this application are a means of accessing the phone's GPS sensors and transmitting the output.
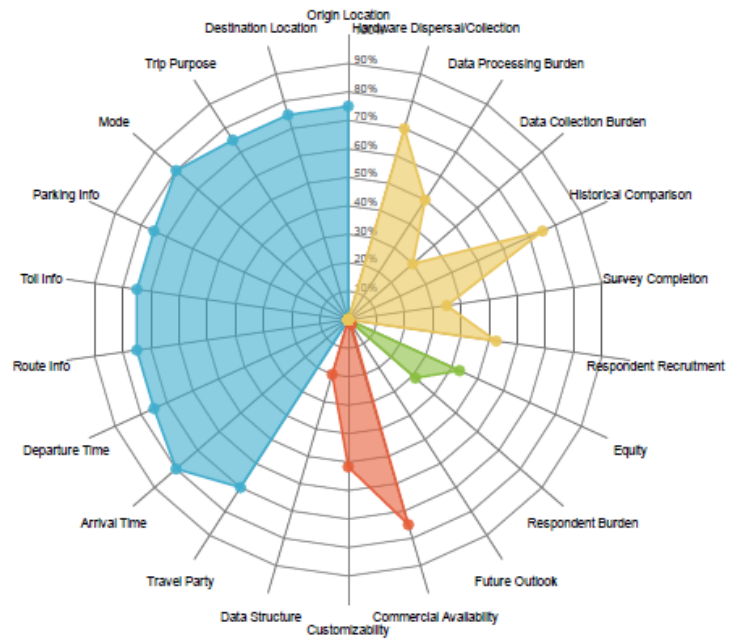
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score	Rank
60%	11

Pros

- Collects accurate location information from participants while removing the need to purchase, distribute, and collect GPS devices.

- Collection of paper travel diary data permits comparison to previous years of survey data.
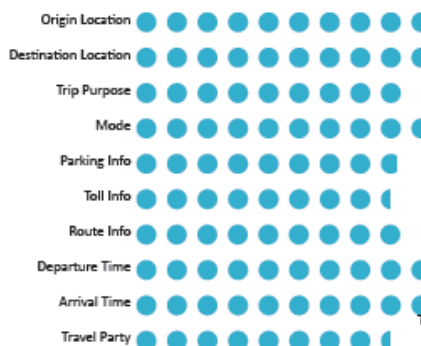
Cons

- Continual use of smartphone GPS wears on the battery creating burden for the participant.

- No precedent for replacing wearable GPS devices with smartphone GPS.

Outlook: The outlook for this methodology is not particularly strong considering the likelihood that in future years the paper travel diary will be re-created for phones, effectively uniting accurate location data with a simple survey in a single device.

Online Diary + Smartphone Location History

Description: A web portal that allows you to import Google Location History or Apple's Frequent Locations history for annotation and validation.
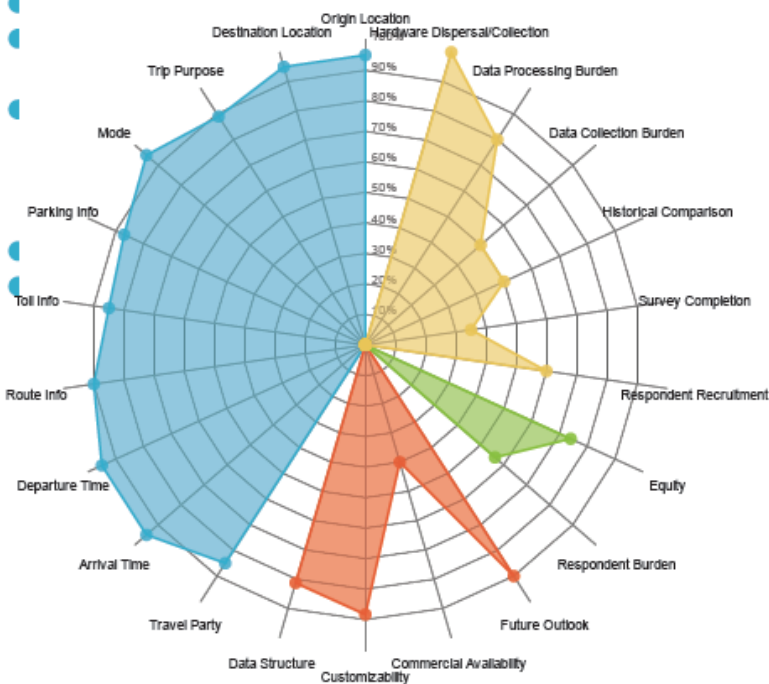
Survey Data

Survey Instrument Characteristics

Respondent Experience

Administrative Experience

Total Score

79%

Rank

5

Pros	Cons
• Minimal respondent burden for high quality data.	• May encounter difficulties developing an API that allows you to import this data.
• Can potentially annotate months or years worth of data.	• Possible incomplete data for short walking trips or others where a user might not have brought their phone or phone may have run out of battery. Since the user is reliant on passive collection of data they may not think to update this information to accurately a day's travel.

Outlook: Promising. Google Location History is already available for download in a zip format. A website could be constructed to manage this data and represent it to a participant in order for them to validate and adjust trip characteristics and demographic variables as necessary.

Online Diary + Wearable GPS

Description: A survey created for the web. Participants are given a code, a wearable GPS, and a travel day and asked to wear the GPS on the day of travel then validate and adjust the output displayed on the website.
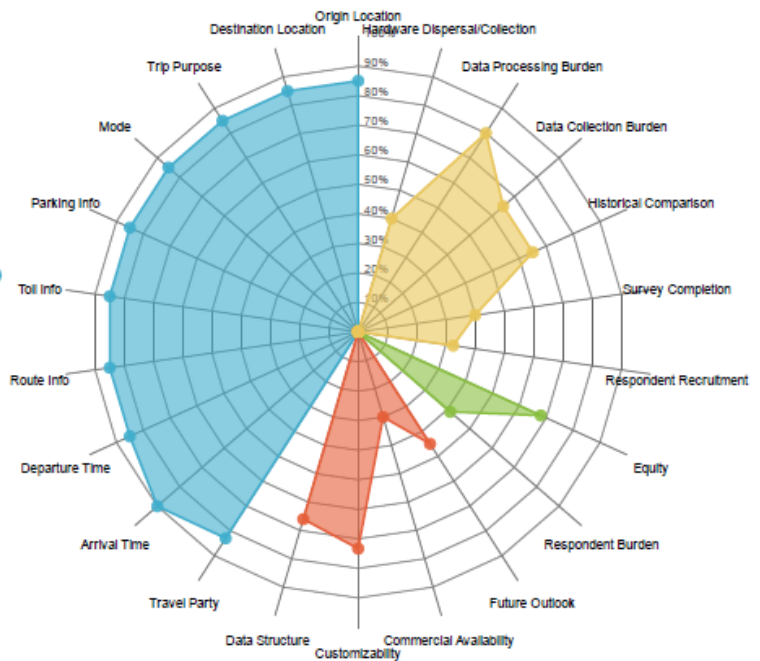
Survey Data
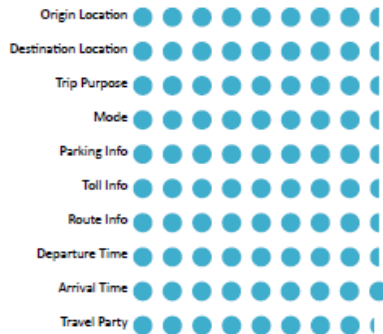
- Origin Location
- Destination Location
- Trip Purpose
- Mode
- Parking Info
- Toll Info
- Route Info
- Departure Time
- Arrival Time
- Travel Party

Survey Instrument Characteristics

- Data Structure
- Customizability
- Commercial Availability
- Future Outlook

Respondent Experience

- Respondent Burden
- Equity

Administrative Experience

- Respondent Recruitment
- Survey Completion
- Historical Comparison
- Data Collection Burden
- Data Processing Burden
- Hardware Dispersal/Collection

Total Score

70%

Rank

8

Pros

- Passively collects trip characteristics and presents those characteristics on a familiar medium for validation and adjustment.

- Web development relatively inexpensive.

- Data processing simplified by methodologies use of machine readable data.
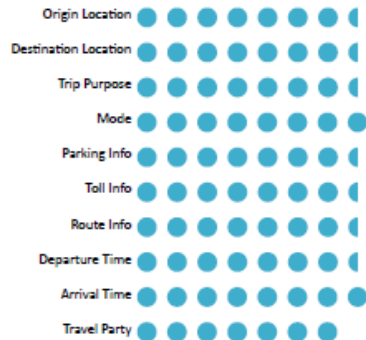
Cons

- Wearable GPS may be forgotten, may be uncomfortable, may run out of battery, etc.

- High costs associated with purchasing, distributing and collecting wearble GPS from participants.

- Potentially biased toward higher income and higher education individuals. (Documented bias in high tech for that SES)

Outlook: May become popular in coming years but unlikely as smartphone applications accomplish many of the same goals with reduced burden on participant and administrator alike.

Paper Travel Diary+GPS

Description: After agreeing to participate in a survey respondents are mailed a paper travel diary and instructed to complete it for their assigned day of travel. The paper travel diary contains codes for activities and travel modalities as well as spaces to enter these and other trip characteristics. Respondents may mail back these travel diaries or report their contents via telephone. Additionally, a subsample of survey participants is given a GPS unit to be worn on the day of their travel. Data from this GPS subsample is used to correct for trip underreporting in the full sample.
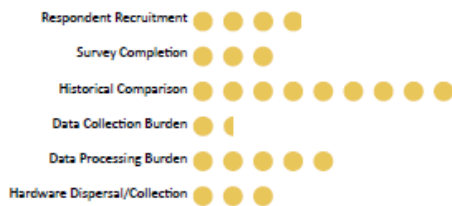
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

56%

Rank

12

Pros

- Widely accepted as the state-of-the-art for collecting travel data from survey participants.

- Wearable GPS provides valuable location data without interfering with the participant's normal use of their smartphone device.
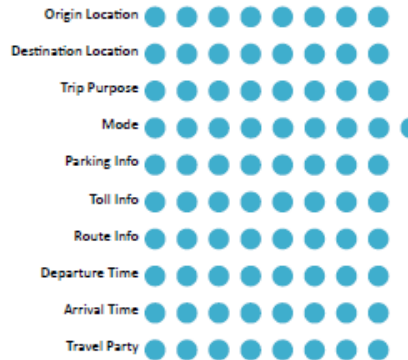
Cons

- Wearable GPS unit may be forgotten by participant on their day of travel. Wearable unit introduces a greater respondent burden as the participant must remember to keep the unit with them in addition to completing the paper travel diary.

- Purchasing, distributing, and collecting GPS units introduces great cost to the MPO.

Outlook: If GPS units have the capacity to transmit data wirelessly the GPS data could be uploaded to MPO/consultant server, processed, and displayed for an Online Diary. This facilitates the process of conflating GPS traces to self-reported travel data.

Paper Travel Diary + Online Diary + Smartphone App

Description: Participants are given three options for survey completion: paper travel diary, online diary, and smartphone app.
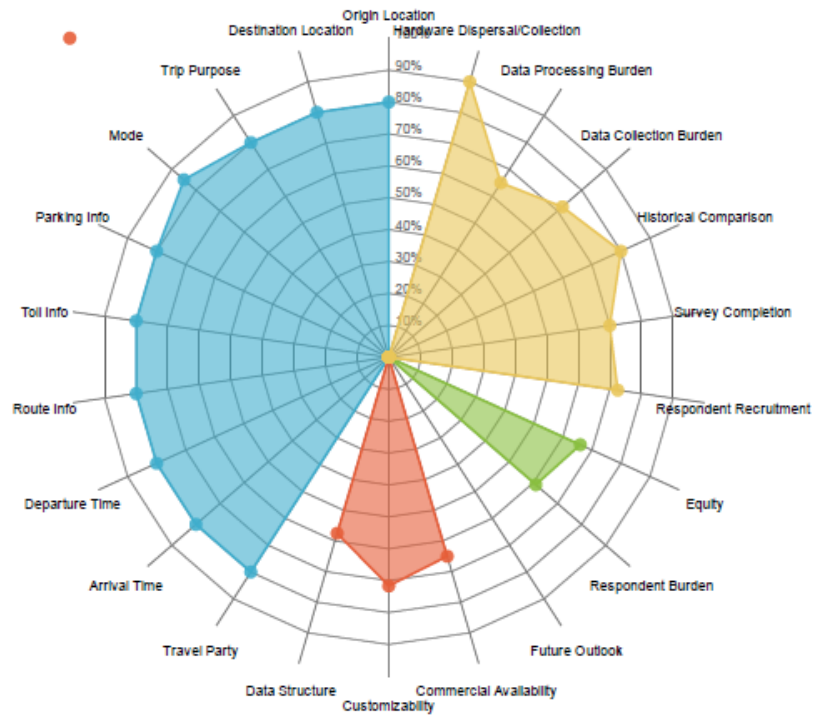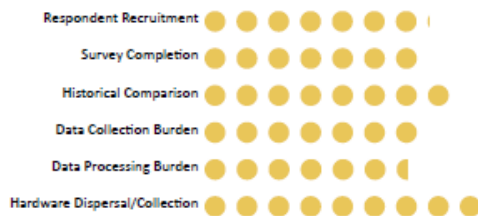
Survey Data

- Origin Location
- Destination Location
- Trip Purpose
- Mode
- Parking Info
- Toll Info
- Route Info
- Departure Time
- Arrival Time
- Travel Party

Survey Instrument Characteristics

- Data Structure
- Customizability
- Commercial Availability
- Future Outlook

Respondent Experience

- Respondent Burden
- Equity

Administrative Experience

- Respondent Recruitment
- Survey Completion
- Historical Comparison
- Data Collection Burden
- Data Processing Burden
- Hardware Dispersal/Collection

Total Score

75%

Rank

7

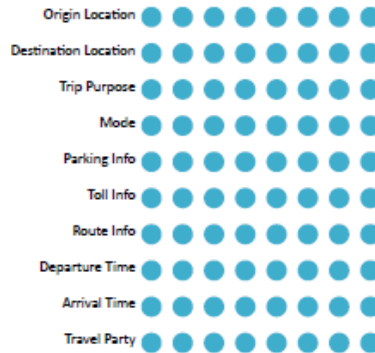| Pros	Cons
- Integrates new technology and approaches with the traditional approach.	- Expense of developing and deploying all three survey instruments is high.
- GPS data collected by the smartphone app can be used to correct the self-report data collected in the paper travel diary and online diary.	- Data management challenges that occur when using multiple survey instruments.

Outlook: Strong. Variations on this methodology are likely to emerge in the coming years as MPOs strive for new approaches to travel surveys while remaining attached to the tried and true. The broad net cast by employing all three methods means you are likely to achieve high equity and representativeness in your survey.
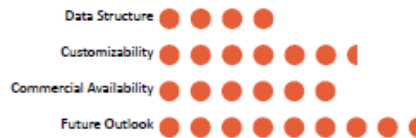
Paper Travel Diary + Online Diary + Smartphone GPS

Description: Participants are given the option to fill out a paper travel diary or complete a diary online. Additionally, a subsample is asked to download and use a basic smartphone application that records and transmits data on the day(s) of their travel.
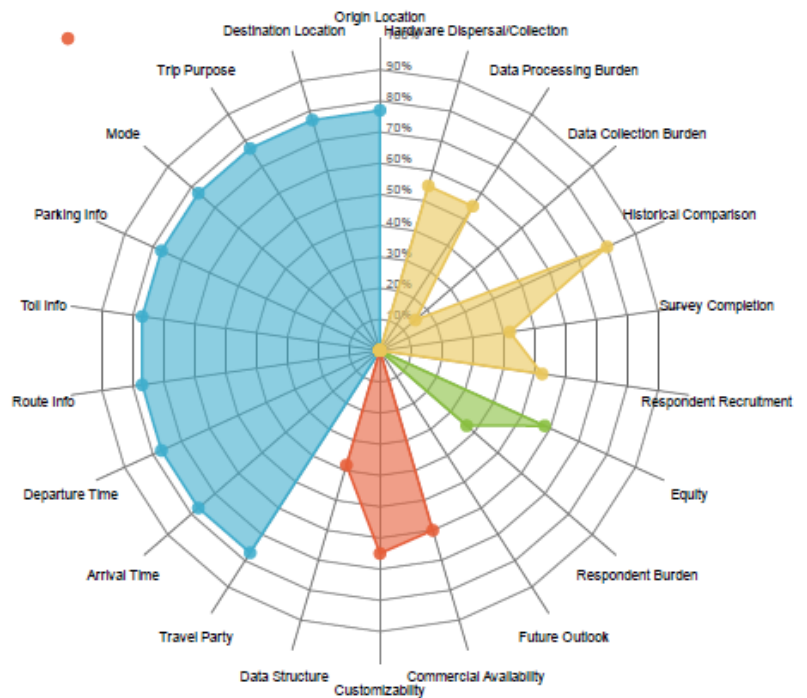
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

65%

Rank

10

Pros

- Using both paper and online diaries to survey regional travelers allows MPOs to reach a greater number and diversity of participants.

- Collects accurate location information from participants while removing the need to purchase, distribute, and collect GPS devices.
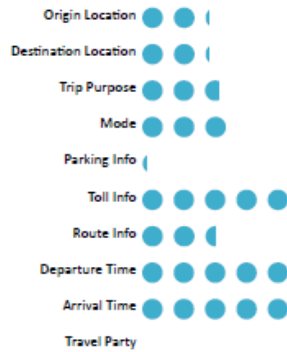
Cons

- Employing different survey instruments means creating different procedures for managing the data collection and processing of survey data.

- Continual use of smartphone GPS wears on the battery creating burden for the participant.

- No precedent for replacing wearable GPS devices with smartphone GPS.

Outlook: Paper travel diaries are likely to go by the wayside as rapid technological advances increase the accuracy and ease of collecting trip characteristic variables through 'high tech' survey instruments. Online diaries may persist and even grow in popularity as novel techniques for gathering travel data become more widely accepted (e.g. smartphone location history).

Passive Data 1 - Cellphone Trace + ACS

Description: Aggregated cellular traces are purchased from a company (e.g. Airsage) for the MPO jurisdiction. Aggregated cellular traces can be used to show regional mobility. Demographic variables are gathered through ACS and used to conflate the cell trace aggregations.
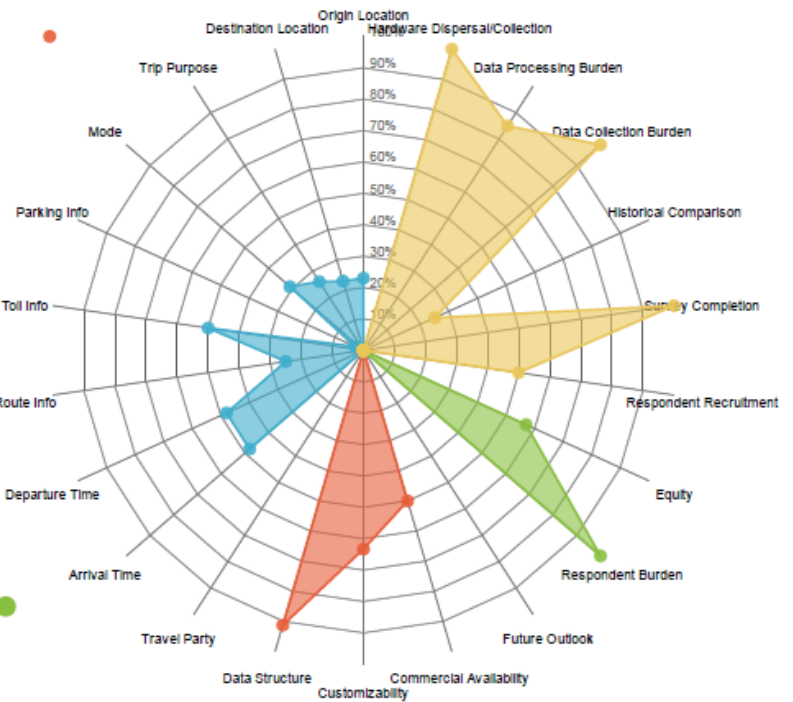

Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

54%

Rank

13

Pros

- Imposes zero burden on survey population.
- Valuable as a validation tool for mode ouput.

Cons

- Requires mathematical disaggregation to generate inputs for NYBPM. These disaggregated data points do not contain ground truth, limiting the value of the NYBPM output.
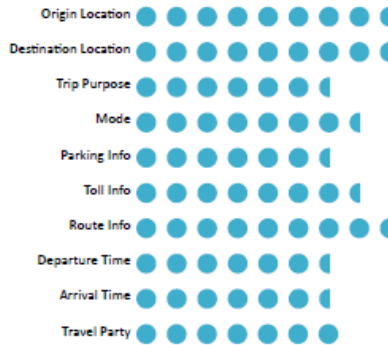
Outlook: As an input for the NYBPM, aggregated cell traces have limited value due to NYBPMs requirement for disaggregated inputs. There is value in using traces to validate models.

Passive Data 2 - O/D Data + ACS + Marketing Data

Description: Gather (purchase) O/D probe data and marketing data (or PUMS Data) for MPO region, run through algorithm to create synthetic population and generate O/D matix trip table. Run trip table through NYBest Practices Model or use resulting O/D table and socieoeconomic demographic data as regional transportation model.
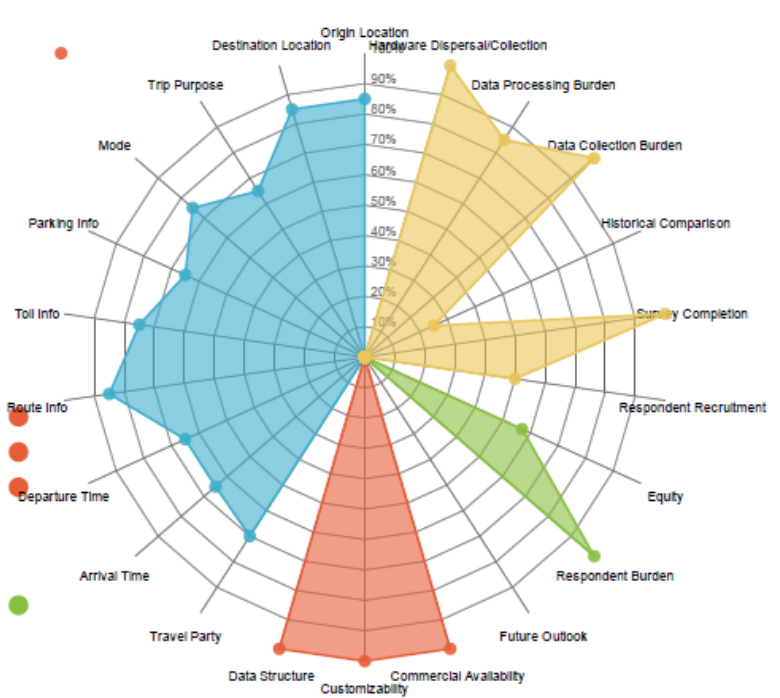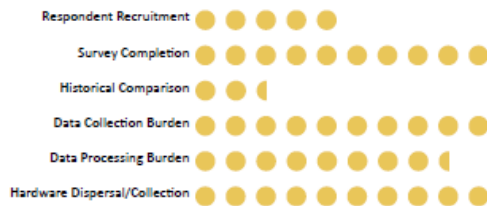
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

80%

Rank

4

Pros

- Reduces or removes the need for survey participants.
- Can be used as model input or as a model by itself.
- Can be used to calibrate forecast models by validating previous forecasts.
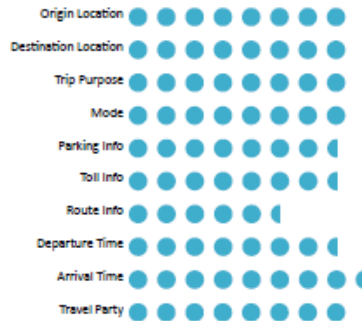
Cons

- Accuracy of O/D probe data is still a question.
- No precedent for employing this methodology in large metropolitan areas comparable to NYMTC.

Outlook: Although there is little precedent for this methodology, it is highly likely that this passive data method will replace both the current active survey and the statistical modeling process.

Smartphone Survey

Description: Collects travel data from participants strictly via self-report on a smartphone. Does not use phone's location/sensing technologies to supplement or enhance the accuracy of reporting.
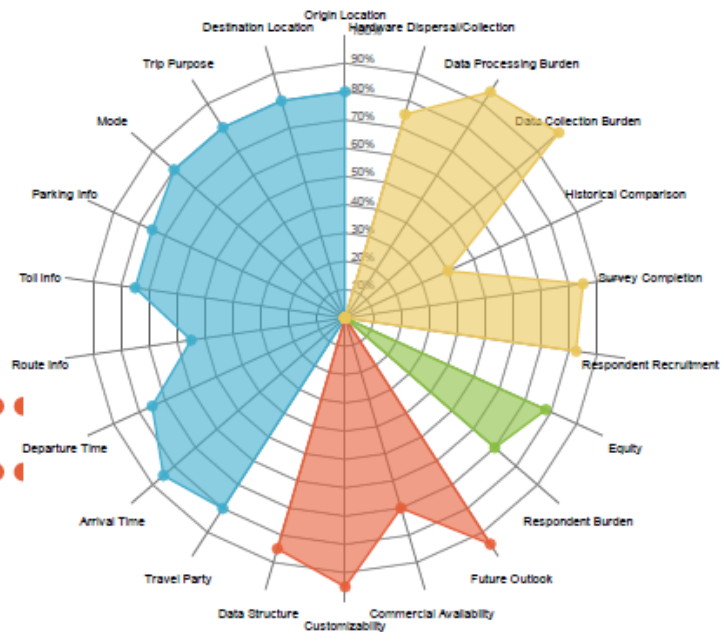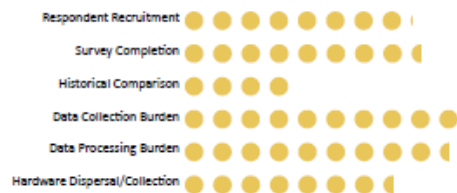
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

79%

Rank

6

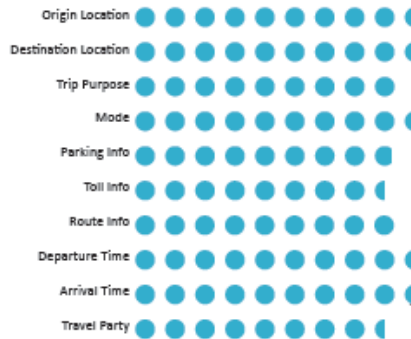| Pros	Cons
• Because it is not reliant on GPS and other phone sensors the drain on participants battery is greatly reduced.	• At present there are few travel survey smartphone applications available and even fewer than do not employ sensing technologies. As a result, an MPO that chooses the smartphone as a survey instrument must select from a limited vendor list, design the survey in-house, or contract the development of an app.
• A smartphone survey is easy to customize and present in many languages.	
• Smartphones are unlikely to be forgotten by the participant and can offer periodic reminders to input travel data. As a result, participants are more likely to enter trip data with reduced latency resulting in increased accuracy.	• By not employing the phone's sensing technologies a greater burden is placed on participant's recall. Even though there is reduced battery consumption relative to a smartphone app that uses GPS, the burden to the participant is greater when relying on participant recall especially when considering the quality of the data that is gathered.

Outlook: There may be value for using smartphone surveys to collect stated preference items from residents. For the purposes of collecting accurate travel data (esp. those items listed under 'Trip Characteristic Data'), however, the smartphone survey seems unlikely to achieve prominent use by MPOs. The large body of literature on self-report travel surveys shows underreporting and trip characteristic inaccuracies in participant recall. Model inputs collected from travel surveys without location/sensing technologies are less accurate than those that employ those technologies.

Smartphone App (Survey + Smartphone GPS)

Description: Collects travel data from participants automatically through smartphone sensing technologies and uses algorithms and computing power to present those trips and their characteristics to a participant. A participant then validates or adjusts those trip characteristics.
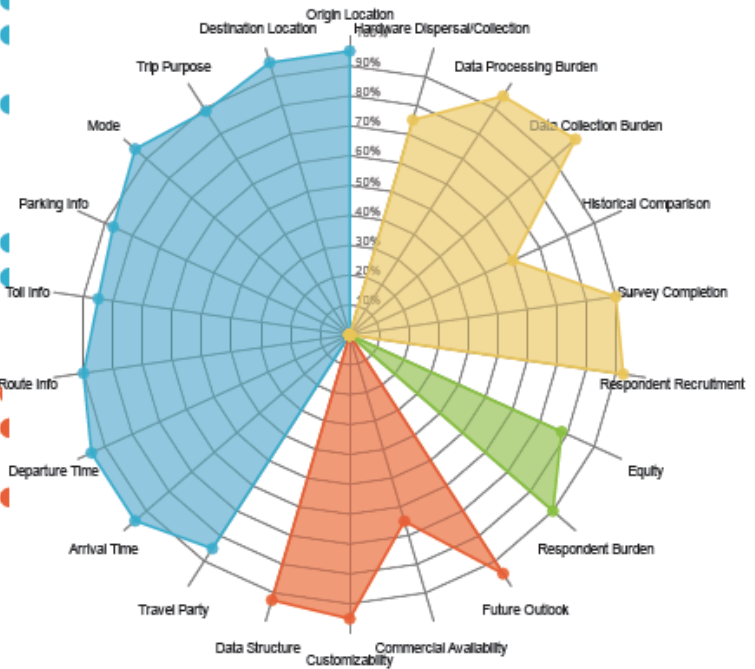
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection

Total Score

88%

Rank

1

| Pros	Cons
• Passively gathering travel information then presenting it to the user for validation represents the state-of-the-art in balancing participant burden reduction and trip data quality.	• GPS traces remain inaccurate underground and in areas known as urban canyons.
• The smartphone eliminates the need to distribute and collect GPS devices from participants, reduces the likelihood of forgetting the GPS device on the travel day, and has a user interface for viewing and adjusting trip characteristics. As a result it is considered by many in the industry to be the natural evolution of GPS use in travel surveys.	• The algorithms for identifying route, mode, and trip purpose need further refinement and are often inaccurate. • Though supporting data is somewhat lacking there is some evidence to suggest that using 'high tech' to gather travel information could lead to oversampling of populations that are younger and more highly educated.

Outlook: The outlook for smartphone applications that use sensing technologies and user interface to gather accurate travel data is promising. A number of different companies are developing these applications and using feedback and rapid improvements in sensing and computing to refine them.

Smartphone App + Bluetooth

Description: Collects travel data from participants automatically through smartphone sensing technologies and uses algorithms and computing power to present those trips and their characteristics to a participant. A participant then validates or adjusts those trip characteristics. This methodology attempts to address GPS signal loss underground by using existing bluetooth infrastructure in the subway system to track location.
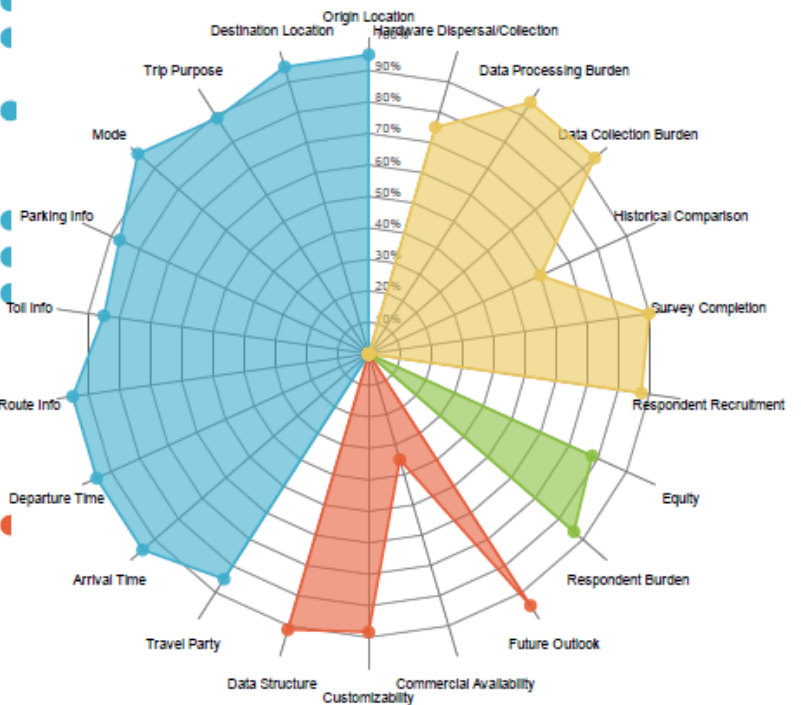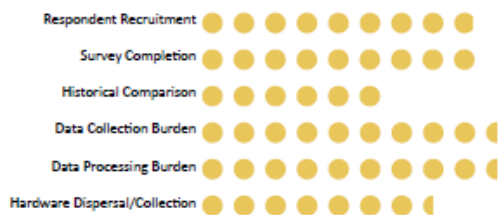
Survey Data
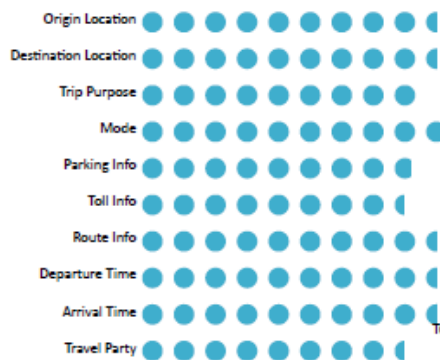
Survey Instrument Characteristics

Respondent Experience

Administrative Experience

Total Score	Rank
86%	2

Pros

- Bluetooth component fills in for GPS signal loss underground covering one of the more substantial limitations of GPS and/or smartphone use in New York City.

- Smartphone applications can encourage annotation of highly accurate spatiotemporal data. As a result, important trip characteristic variables can be directly reported instead of just inferred as is the case with GPS loggers.
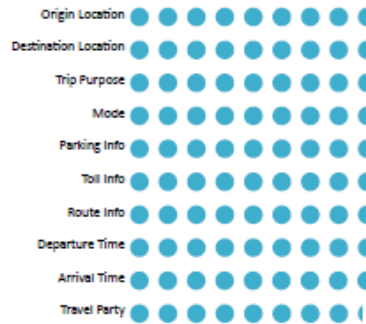
Cons

- Potential equity issues. Kiosks are not equally distributed throughout the subway system.

- Bluetooth infrastructure in New York City subway may not be sufficient to capture a significant number of trips.

Outlook: The outlook for this methodology is positive. Bluetooth, specifically, is becoming more prominent in New York City subways with more than 2,000 help kiosks with bluetooth capability installed in hundreds of subway stations (article).

smartphone app + online diary

Description: Survey methodology calls for the development of a smartphone application that uses the phone's sensing capabilities and an accompanying website for editing and validating. Use of the website is optional and participants may complete the survey exclusively on their phone. Participants may also limit smartphone use to passive data collection and complete the validation and adjustment of computed trip characteristics on the website.
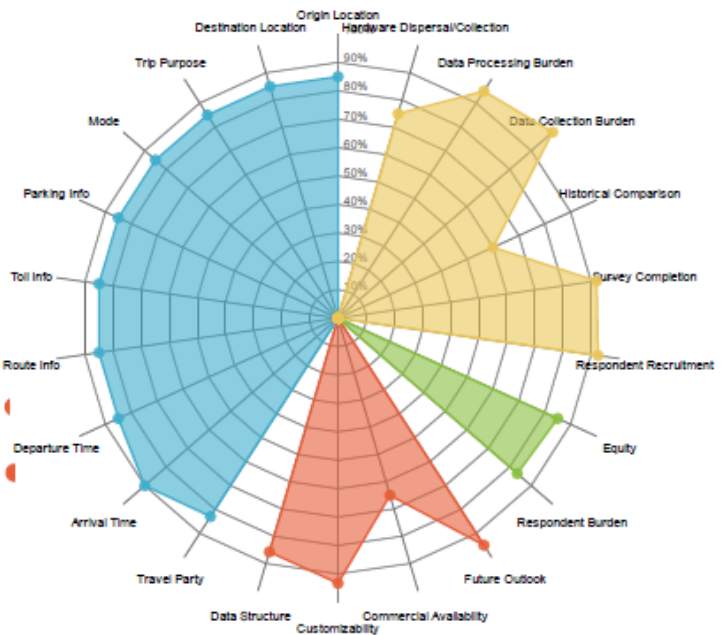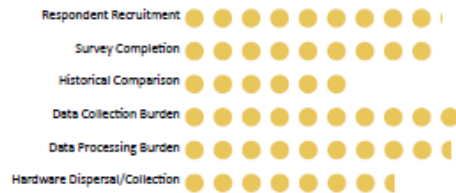
Survey Data

Origin Location
Destination Location
Trip Purpose
Mode
Parking Info
Toll Info
Route Info
Departure Time
Arrival Time
Travel Party

Survey Instrument Characteristics

Data Structure
Customizability
Commercial Availability
Future Outlook

Respondent Experience

Respondent Burden
Equity

Administrative Experience

Respondent Recruitment
Survey Completion
Historical Comparison
Data Collection Burden
Data Processing Burden
Hardware Dispersal/Collection
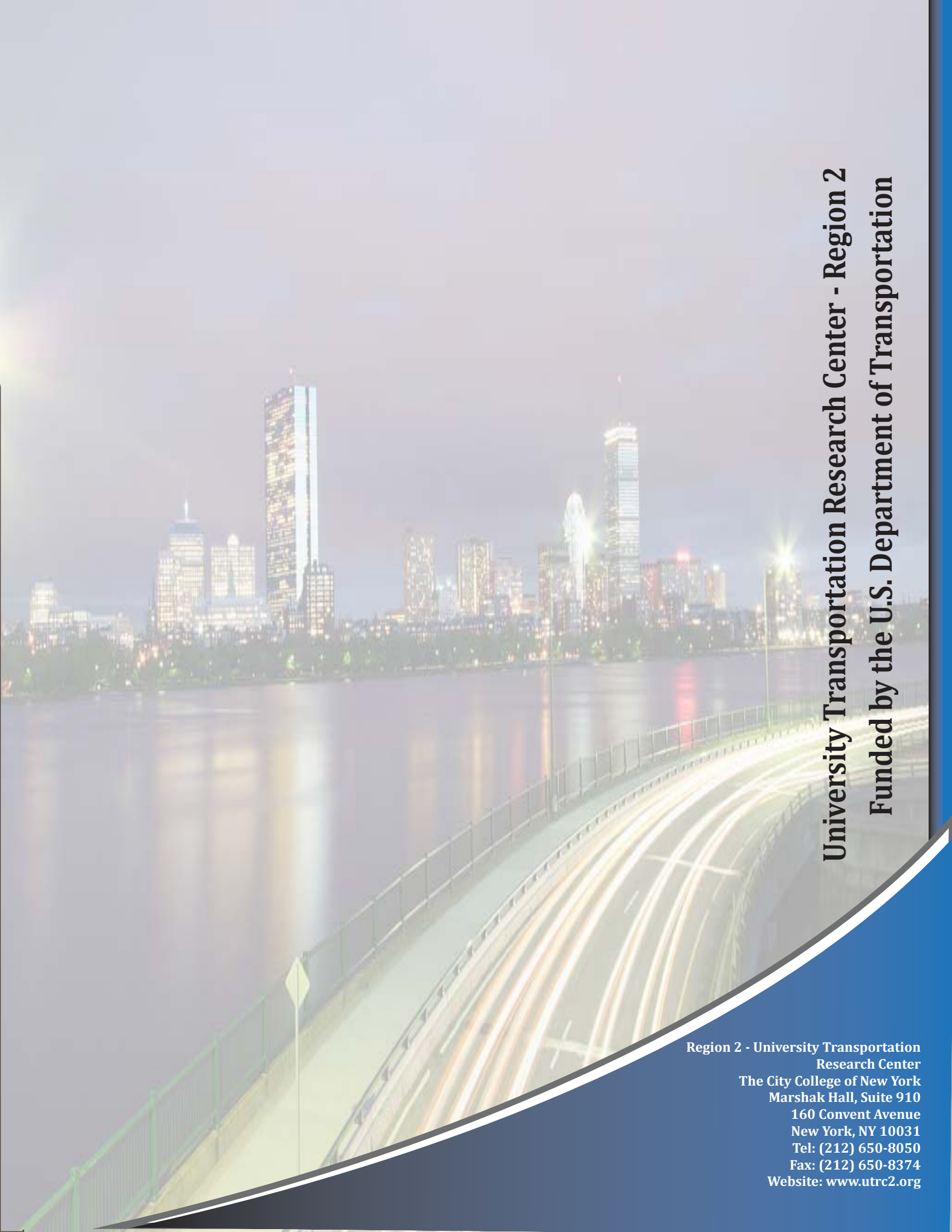
Total Score

85%

Rank

3

Pros

- Flexibility allows more people to engage with the survey in a way that seems best for them encouraging higher participation and completion rates.

- Online diary component allows self-reporting of trips through a familiar medium that can reduce respondent burden by employing branching logic, visualizations, video tutorials, and easily-updated FAQs. Additionally, the Online Diary can encourage accurate completion by employing a workflow that reminds a user that they haven't filled out a field or have used an invalid value.

Cons

- Higher costs associated with development of web content that may have low valuation. In other words, the extra costs spent developing a website in addition to and consistent with a smartphone application may not be justified by the mild upticks in participation and completion rates. However, this is largely conjecture. There is little evidence to examine to determine how much flexibility in survey participation options increases participation or completion.

- Does not administer a wearable GPS for those who do not own a smartphone. Therefore, study is limited to smartphone owners. If the MPO desires to make it possible for participants to complete the Online Diary without GPS component you must weight this different type of data which adds administrative burden and cost.

Outlook: Outlook for this methodology is promising. Users can complete the survey on their phone or online and this versatility is likely to encourage survey completion.