NATIONAL CENTER FOR UNDERSTANDING FUTURE
TRAVEL BEHAVIOR AND DEMAND

Final Project Report

Investigating the Influence of Alternative Survey Participant Recruitment Strategies on Measurement and Inference of Mobility Patterns

BY

Ram M. Pendyala
Email: ram.pendyala@asu.edu

Irfan Batur
Email: ibatur@asu.edu

Victor O. Alhassan
Email: valhassan@asu.edu

Fan Yu
Email: fanyu4@asu.edu

School of Sustainable Engineering and the Built Environment
Arizona State University
660 S. College Avenue, Tempe, AZ 85287-3005

May 2025

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. N/A	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle Investigating the Influence of Alternative Survey Participant Recruitment Strategies on Measurement and Inference of Mobility Patterns		**5. Report Date** May 14, 2025 **6. Performing Organization Code** N/A
7. Author(s) Ram M. Pendyala, https://orcid.org/0000-0002-1552-9447 Irfan Batur, https://orcid.org/0000-0002-8058-2578 Victor O. Alhassan, https://orcid.org/0009-0007-6408-7036 Fan Yu, https://orcid.org/0000-0001-9767-4156		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address School of Sustainable Engineering and the Built Environment Arizona State University 660 S. College Avenue, Tempe, AZ 85287-3005		**10. Work Unit No. (TRAIS)** N/A **11. Contract or Grant No.** 69A3552344815 and 69A3552348320
12. Sponsoring Agency Name and Address U.S. Department of Transportation, University Transportation Centers Program, 1200 New Jersey Ave, SE, Washington, DC 20590		**13. Type of Report and Period Covered** Final Report, 2024-2025 **14. Sponsoring Agency Code** USDOT OST-R
15. Supplementary Notes N/A		
16. Abstract There are growing concerns about the representativeness of survey data in an era of rapidly emerging and evolving technology, low response rates, and increasingly diverse and heterogeneous populations. Because of the complexities and costs associated with conducting surveys using traditional mail and phone methods, researchers and practitioners are adopting new methods to sample respondents. This project aims to provide a comprehensive assessment of the representativeness of the samples obtained from three survey sampling strategies utilized in the nationwide COVID Future Panel Survey: convenience sampling, email sampling, and online panel sampling. The three subsamples were statistically different from each other for all socio-economic and demographic variables except race, ethnicity, household size, and gender. However, these differences were ameliorated with the application of weights and the three subsamples converged to census distributions on many variables except educational attainment. Weighting was also able to reduce the differences between the subsamples for a variety of mobility variables except transit use frequency. Modeling the influence of survey sample recruitment strategy on measures of mobility shows that it is significant even after controlling for socio-economic and demographic variables in the model specification. It is likely that the survey sample recruitment strategy variable is accounting for unobserved traits such as attitudes and lifestyle preferences. It is therefore recommended to include attitudinal and lifestyle preference questions in transportation surveys so that these traits can be explicitly included in travel model specifications to enhance explanatory power and reduce bias.		
17. Key Words Survey methods, Sampling strategies, Sample recruitment, Online panels, Survey weighting, Mobility measurement		**18. Distribution Statement** No restrictions.
19. Security Classif.(of this report) Unclassified	**20. Security Classif.(of this page)** Unclassified	**21. No. of Pages** 22 / **22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

ACKNOWLEDGMENTS

This research was partially supported by the National Center for Understanding Future Travel Behavior and Demand (TBD), a National University Transportation Center sponsored by the U.S. Department of Transportation (USDOT) under grant numbers 69A3552344815 and 69A3552348320. The authors would like to thank the TBD National Center, USDOT, and TOMNET for their support of university-based research in transportation, particularly for the funding provided for this project. The authors also extend their thanks to Jose Roberto Dimas Valle, Tassio B. Magassy, Deborah Salon, and Chandra R. Bhat for their valuable contributions to the work presented in this report.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1

1. INTRODUCTION 2

2. SURVEY DESCRIPTION 4

3. COMPARISON OF SOCIO-ECONOMIC AND DEMOGRAPHIC VARIABLES 5

4. COMPARISON OF MOBILITY AND ACTIVITY PARTICIPATION VARIABLES 9

5. MODELS OF ACTIVITY-MOBILITY CHARACTERISTICS 12

6. STUDY IMPLICATIONS AND CONCLUSIONS 17

REFERENCES 19

LIST OF TABLES

Table 1 Distributions of Socioeconomic and Demographic Variables 8
Table 2 Pre-COVID Mobility and Activity Participation Characteristics 11
Table 3 Model Estimation Results for Select Activity-Mobility Characteristics 14
Table 4 Change in Commute Mode for Stayer Sample of COVID Future Panel Survey 16

EXECUTIVE SUMMARY

There are growing concerns about the representativeness of survey data in an era of rapidly emerging and evolving technology, low response rates, and increasingly diverse and heterogeneous populations. Because of the complexities and costs associated with conducting surveys using traditional mail and phone methods, researchers and practitioners are adopting new methods to sample respondents. This project aims to provide a comprehensive assessment of the representativeness of the samples obtained from three survey sampling strategies utilized in the nationwide COVID Future Panel Survey: convenience sampling, email sampling, and online panel sampling. The three subsamples were statistically different from each other for all socio-economic and demographic variables except race, ethnicity, household size, and gender. However, these differences were ameliorated with the application of weights and the three subsamples converged to census distributions on many variables except educational attainment. Weighting was also able to reduce the differences between the subsamples for a variety of mobility variables except transit use frequency. Modeling the influence of survey sample recruitment strategy on measures of mobility shows that it is significant even after controlling for socio-economic and demographic variables in the model specification. It is likely that the survey sample recruitment strategy variable is accounting for unobserved traits such as attitudes and lifestyle preferences. It is therefore recommended to include attitudinal and lifestyle preference questions in transportation surveys so that these traits can be explicitly included in travel model specifications to enhance explanatory power and reduce bias.

1. INTRODUCTION

In many fields, surveys of households and individuals serve as the source of information to obtain insights about behaviors, choices, preferences, attitudes, and trends over time. Transportation planning, policymaking, and modeling has long relied on transportation surveys of various types to understand, measure, and quantify activity-travel demand, time use patterns, mobility choices under a variety of scenarios, and attitudes and preferences related to transportation and emerging technologies. Survey data is used to estimate increasingly complex transportation demand forecasting models, understand how travelers feel about different transportation options, modes, policies, and technologies, and gain insights about trends in mobility patterns. Many metropolitan planning agencies as well as national governments conduct travel surveys on a periodic basis to obtain up-to-date information about travel behavior and values. In a few places, data is gathered on a more continuous basis to help monitor trends in mobility patterns on a more frequent basis. In the United States, the National Household Travel Survey (NHTS) that had historically been conducted about every half-dozen years or so has now transitioned to a more frequent data collection protocol, with surveys taking place every other year (Bricka et al. 2024). The recent COVID-19 pandemic motivated researchers and practitioners around the world to conduct surveys to gather critical information about changes in mobility patterns, activity-travel demand, and mode choices during and after the pandemic (Chauhan et al. 2021). This data has proven crucial to post-pandemic transportation planning and model calibration.

With surveys continuing to serve as a key source of information for transportation planning and modeling, there are growing concerns about the ability to collect representative data in an era of rapidly emerging and evolving technology, low response rates, and increasingly diverse and heterogeneous populations (Silvano et al. 2020; Stedman et al. 2019). The cost of conducting surveys using traditional methods (e.g., mail-out/mail-back surveys, paper-and-pencil surveys, random digit dialing based telephone surveys) has risen sharply due to low response rates, technological tools that allow prospective respondents to screen calls and ignore solicitations and increases in labor and material costs (Weber et al. 2020). With households and individuals constantly solicited for opinions, feedback, and input from many different entities, it is increasingly difficult to obtain the cooperation of an over-surveyed population that is incessantly experiencing survey fatigue.

Because of the complexities and costs associated with conducting surveys using traditional mail and phone methods (which yield dismally poor response rates in the current context), researchers and practitioners are adopting a variety of new methods to sample respondents and conduct surveys (Monzon et al. 2020; Silvano et al. 2020, Lo et al. 2020). While there are a number of survey methods that are being deployed, this project considers three specific methods that are gaining popularity – primarily due to affordable cost and convenience of administration. These methods are easy to execute, quite cost-effective, and often (but not always) considered adequate or good-enough for the purposes of collecting data.

The three methods of interest considered in this project are as follows. The first is the use of convenience samples for data collection. Convenience samples are generally comprised of individuals who can be recruited easily and conveniently without considerations of randomness or representativeness in the recruitment process. Convenience samples may include, among others, individuals who subscribe to certain channels, are easily reached because they belong to certain groups or lists, or acquaintances of the survey/research team who can be easily solicited and contacted via email or social media channels. The second is the use of email lists purchased from a commercial vendor. Using new messaging platforms and tools, it is quite easy to blast out hundreds of thousands of email solicitations to individuals, requesting their participation in a survey. Commercial entities have assembled lists of email addresses and sell them for marketing and survey research purposes. The lists are quite affordable and the effort involved in sending out large numbers of requests for participation in a survey is quite modest and cost-effective (Lohr 1999). The third is the use of online survey panels that are assembled by survey research companies. It is now possible to enter into an agreement with a survey research company to deploy a survey and obtain a requisite number of respondents. The survey research company is responsible for administering the survey, gathering the data from the agreed-upon number of respondents, and furnishing a reasonably clean data set to the survey research team or agency that has purchased their services. Depending on the complexity of the sampling quotas and geographic coverage, the cost per respondent varies – with higher costs associated with more complex sample quotas and geographically restrictive specifications. Survey research companies that offer this service have essentially compiled a large pool of professional survey-takers who complete surveys for a modest financial remuneration (Chandler et al. 2019).

As mentioned earlier, these three methods have gained considerable traction because of their ease of implementation and cost-effectiveness. However, all three methods raise critical questions of representativeness of the resulting survey respondent samples. Convenience samples are likely to be biased as they are not drawn from the general population of interest; rather, they are drawn from specific lists of individuals who are easily and conveniently contacted (Landers and Behrend 2015). Commercially available e-mail lists are likely to comprise of individuals who have larger online activity footprints; they subscribe to various online services and channels, purchase goods and services online, and use a variety of mobile apps to conduct activities, communicate, and travel. These email lists are likely to be biased in favor of technologically inclined (tech-savvy) individuals who are generally young, well educated, and have higher income. Finally, professional survey panels that have been assembled by survey research companies may be biased as well, with an over-representation of individuals who are interested in survey research, otherwise unemployed and need the financial compensation that comes with completing surveys, and free of other obligations and tasks thus allowing them the time and freedom to serve as professional survey-takers (Chandler et al. 2019, Griffin et al. 2020).

Despite the concerns of representativeness of associated with the adoption of these three survey modalities, they are here to stay in an era of low response rates and escalating survey administration costs. Within this project, these three survey modalities will be referred to as alternative survey sampling strategies even though they are not all strictly survey sampling strategies per se. While convenience sampling is a sampling strategy, the other two methods do not necessarily constitute established survey sampling strategies (rather, they are more like sample recruitment strategies). Nevertheless, for ease of presentation and articulation in this report, they will all be referred to as survey sampling strategies.

This project aims to provide a comprehensive assessment of the representativeness of the samples obtained from each of these survey sampling strategies. Many surveys are unlikely to adopt a multitude of strategies for recruiting survey respondents; rather, only one or two strategies will be used to obtain a survey sample. It is therefore of critical importance to understand the nature of the biases and the level of representativeness associated with each of these increasingly popular survey sampling strategies. Armed with such information, it will be possible for survey researchers and agencies to determine the most appropriate survey sampling strategy (or, strategies) that should be deployed in different survey data collection contexts to meet the objectives of the survey effort. In some instances, it may indeed be desirable to adopt multiple channels for survey data collection;

through such a composite survey sampling approach, it may be possible to obtain a survey sample that is more representative of the population as a whole.

The analysis in this project utilizes data from the COVID Future Survey, a national panel survey conducted in 2020 and 2021 to measure the impacts of the COVID-19 pandemic on traveler behaviors, mobility choices, activity modalities, and perceptions and preferences. The COVID Future Survey adopted all three survey sampling strategies described above, yielding a composite sample that includes respondents recruited via a convenience sampling approach, email messages sent to a list purchased from a commercial vendor, and an online survey panel commissioned by a survey research company. The source from which each respondent is drawn is recorded in the data set, thus allowing a comparison among the subsamples recruited via different means and a comparison of each subsample against census data to assess biases and level of representativeness. Each of the subsamples is then weighted to control for a series of socio-economic and demographic variables and ensure population representativeness (with respect to those variables). The weighted subsamples are then compared with respect to measures of mobility and activity participation (and transitions over time) with a view to assess the extent to which inferences about mobility drawn from different subsamples differ (or not) after weighting is performed.

The rest of this report is organized as follows. The next section offers a detailed description of the COVID Future Survey data set and weighting methodology. The third section presents a detailed comparison of unweighted and weighted survey subsamples with respect to socio-economic and demographic variables, while the fourth section offers a similar comparison with respect to activity-mobility characteristics. The fifth section presents econometric models of select activity-mobility variables with a view to identify the significance of the effect of sampling strategy on measures of activity-mobility after controlling for all other socio-economic, demographic, and contextual variables. It also includes a comparison of trends in mode choice over time, to examine differences and similarities in trends depicted by the different subsamples. Finally, the sixth section offers conclusions and implications for survey design and sampling strategies in different contexts and applications.

2. SURVEY DESCRIPTION

The COVID Future Survey was carried out during the pandemic years to study changes in mobility, attitudes and perceptions, and activity participation modalities. The survey is a panel survey of three waves with the first wave conducted in 2020. The first wave of the survey was comprised of two components. The first component was administered between April and June of 2020, at the height of the pandemic when many jurisdictions in the United States and around the world locked down and implemented stay-at-home orders to limit the spread of contagion. This component of the data set is referred to as Wave 1A. Participants in Wave 1A were recruited via a convenience sample recruitment scheme. The survey research team reached out to friends, family, colleagues, and acquaintances via email messaging, social media channels/contacts, and professional listservs of the transportation and urban planning fields. As the respondents were recruited via these convenient mechanisms, this sample may be considered a convenience sample. Through Wave 1A, a total of 1,127 survey responses were obtained.

A slightly modified and larger scale version of the survey was conducted between June and October 2020. The pandemic was still of considerable concern worldwide and people and businesses were largely operating in a virtual modality for many activities and services to minimize travel and in-person interactions. Work, school, shopping, dining, and medical appointments were taking place on a variety of virtual platforms, with stores and other establishments implementing strict distancing and masking protocols. This wave of the survey, dubbed Wave 1B, employed a dual sampling strategy to recruit respondents. First, a large email database of 350,000 email addresses was purchased from a commercial vendor. An additional 100,000 email addresses were randomly drawn from the rest of the United States, and a sample of 39,000 email addresses from the Greater Phoenix metropolitan region of Arizona was also contacted for potential participation in the survey. This recruitment strategy yielded a total of 2,946 survey responses. The email sample also contained individuals who were recruited through other means such as media adverts, the COVID Future web page and forwarded links. Second, Wave 1B also involved recruiting respondents through an online survey panel aggregated by an established survey research company, yielding an additional 5,262 responses that largely aligned with sampling quotas conveyed by the research team. Thus, the Wave 1B subsample comprises a total of 8,208 respondents (2,946 + 5,262). The number of respondents stated above refers to the numbers obtained after cleaning the dataset for missing entries for the variables of interest (variables shown in Table 1).

Following the administration of the first wave of the COVID Future Survey, subsequent waves were administered to all Wave 1 respondents through email communications. In order to ensure appropriate spacing between Wave 1 and Wave 2, the second wave was administered between November 2020 and April 2021. The third wave was administered during October – November 2021, by which time everybody who wished to obtain the COVID-19 vaccination could do so without any restrictions. All respondents from Wave 1 (i.e., Wave 1A and Wave 1B) received invitations for responding to Waves 2 and 3. Thus, all Wave 2 and Wave 3 respondents also responded to Wave 1, but not all Wave 3 respondents necessarily responded to Wave 2. A total of 3,093 individuals responded to Wave 2 while a total of 2,860 individuals responded to Wave 3 (Chauhan et al. 2021; Conway et al. 2020). The stayer sample numbered 1,933 individuals who responded to all three waves of the survey.

Upon compilation of the different waves of the survey data sets, a robust weighting methodology was employed to ensure that the survey samples were representative of the general population on a host of socio-economic and demographic variables at the census division level. These samples were weighted to provide geographic representativeness across a broad array of socio-economic and demographic characteristics for each of the nine census divisions in the United States. Sample size limitations prevented controlling for socio-economic and demographic variables at a finer geographic resolution. The weighting was done using the PopGen iterative proportional updating (IPU) algorithm that has proven to be reliable in producing weighted samples that replicate known population distributions (Pendyala et al. 2011). To meet the objectives of this study, three survey subsamples recruited through different means – convenience sample, email sample, and online panel – were each weighted separately to represent the population. The selected control variables included gender, age, education, race/ethnicity, household size, household income, and vehicle ownership. The American Community Survey (ACS) 2021 summary files of the US Census Bureau served as the source of information for marginal control distributions on the variables of interest.

3. COMPARISON OF SOCIO-ECONOMIC AND DEMOGRAPHIC VARIABLES

This section focuses on a comparison of socio-economic and demographic variables across the three subsamples. The three subsamples are convenience sample (CS), email sample (ES), and online panel (OP). These three subsamples are compared both for unweighted and weighted statistics and distributions. The summary of this comparison is presented in Table 1.

The table shows a number of socio-economic and demographic attributes with several categories for each of the variables. In some cases, only a few categories could be represented because of small sample sizes in disaggregate categories. The first section of the table presents the comparison for unweighted statistics and the second section presents the comparison for weighted statistics. The chi-square (□2) p-value is indicative of the level of statistical significance that may be associated with differences in statistical distributions among the survey subsamples. The last column of the table presents the American Community Survey 2021 statistics for the variables shown in the table.

The comparison of the unweighted statistical distributions yields a number of interesting insights. Although the gender distributions are not statistically different from one another, it can be said that the online panel is clearly comprised of a larger percent of females (from a qualitative standpoint). A review of the age distributions shows that they statistically differ across the three subsamples at the 0.05 significance level. As expected, the convenience sample is largely comprised of employed individuals who are of working age. The online panel subsample is rather uniformly distributed across all age groups. The email sample, on the other hand, is clearly skewed towards the older age groups with more than one-half of the subsample aged 55 years or over. This is rather consistent with the general tendency for older individuals to respond at a higher rate to surveys than individuals in other age brackets (Gigliotti and Dietsch 2014). Older individuals generally exhibit a higher level of civic participation and are able to spare the time necessary to respond to surveys, thus contributing to the unweighted age distribution seen in the table for the email sample.

As mentioned earlier, the convenience sample (CS) is drawn largely from the social and professional networks of the survey research team. As such, it is not surprising that this subsample exhibits a high level of educational attainment, a very high employment rate, and a substantially higher income profile than the other subsamples. The subsamples differ statistically significantly at the 0.05 level on education, employment, and income. The online panel respondents tend to be less educated, lower income, and less employed than the other two subsamples. This suggests that individuals who sign up to be professional survey takers are generally of a lower socio-economic status and become members of the online survey panel to derive some income (Hays et al. 2015; Ternovski 2022). With respect to education, employment, and income, the email sample generally falls in between the convenience sample and the online panel, largely consistent with expectations and the age profile of the subsample (Peytcheva and Groves 2009).

All three subsamples are predominantly white, more so than the general population, with the online panel depicting a greater proportion of nonwhite individuals. However, the three subsamples do not statistically significantly differ with respect to race. It is quite notable that the unweighted subsamples do not differ significantly with respect to household size distribution. The household size distributions are quite similar, with the online panel depicting fewer two-person households and a larger percent of 4+ person households. In terms of home ownership, the email sample depicts the highest rate of home ownership with the convenience sample and the online panel depicting similar patterns of home ownership. Given the age distribution of the respondents in the email sample (older), it is not surprising that this group depicts a higher rate of home ownership (Angelini et al. 2014; Hochstenbach and Arundel 2021; Feijten et al. 2003). Vehicle ownership distributions differ at the 0.10 significance level, suggesting that the patterns are rather consistent with one another. The email subsample, comprising older individuals to a greater degree, depicts a greater proportion falling into the category of owning three or more vehicles (Heinonen et al. 2021).

Overall, unweighted subsamples differ significantly from one another on all the socio-economic and demographic variables considered in this table, with the exception of ethnicity, race, and household size. In other words, each of these sample recruitment strategies results in biased respondent samples that need to be weighted appropriately to draw statistically valid inferences about the population as a whole. Weighting of the survey subsamples was done in this study on a host of socio-economic and demographic variables (marked with an asterisk in the table). The weighted statistical distributions generally mirror the census distributions (shown in the last column), suggesting that weighting is capable of compensating for socio-economic and demographic biases that may arise from using a specific survey recruitment and sampling strategy.

A few noteworthy findings may be discerned from the table in the context of weighted distributions. First, for those variables that were not controlled in the weighting process, the distributions still differ from one another (across the different recruitment strategies), but the differences diminish – even to the extent of becoming statistically insignificant in several instances. This happens for employment and housing tenure, both of which were significantly different for the unweighted subsamples, but not so for the weighted samples. All the controlled variables show no statistically significant difference across the survey subsamples, with the exception of education. This particular variable is still statistically different across the survey subsamples simply because the convenience sample included a very tiny number of individuals at the low end of the educational spectrum. Despite the application of a robust weighting methodology, the sample sizes in those categories for the convenience sample were simply too small for the weighting process to produce weights that could replicate the census distribution. As such, while the weighting process did improve the percent of individuals in the weighted convenience sample falling into these lower education categories, it was not able to fully correct the large bias in the unweighted sample. This suggests that the bias associated with a convenience subsample that is largely drawn from professional and social networks may not be fully overcome even through a robust weighting scheme. Nevertheless, the results in this table suggest that weighting is a reasonably effective way of correcting for socio-economic and demographic biases associated with different survey recruitment and sampling strategies (Miratrix et al. 2018).

Table 1 Distributions of Socioeconomic and Demographic Variables

Variable	Attribute	Unweighted					Weighted					ACS 2021 (%)
		χ2 p - value`	CS n = 1,127 (%)	ES n = 2,946 (%)	OP n = 5,262 (%)	All n = 9,335 (%)	χ2 p - value	CS n = 1,127 (%)	ES n = 2,946 (%)	OP n = 5,262 (%)	All n = 9,335 (%)	
Gender*	Male	0.55	41.1	41.8	34.9	37.8	0.83	52.6	48.8	48.8	49.1	49.0
	Female		58.9	58.2	65.1	62.2		47.4	51.2	51.2	50.9	51.0
Age* (years)	18 - 24	0.00	6.1	2.3	12.1	8.3	0.92	14.2	11.7	11.8	11.7	11.7
	25 - 34		28.7	9.5	19.0	17.2		12.9	17.7	17.4	17.6	17.4
	35 - 44		22.5	12.9	18.6	17.3		19.1	17.2	17.0	17.1	17.0
	45 - 54		18.9	16.0	14.1	15.3		20.4	15.4	15.7	15.6	15.7
	55 - 64		16.7	23.4	16.5	18.7		9.8	16.7	16.4	16.4	16.6
	65+		7.1	35.8	19.8	23.3		23.6	21.3	21.7	21.5	21.6
Education*	< High school	0.00	0.3	0.9	2.7	1.8	0.03	7.5	10.3	10.5	10.5	10.7
	High school or GED		0.4	6.0	21.2	13.9		9.9	27.2	26.9	26.9	27.3
	Some college		8.6	25.8	34.7	28.7		38.0	29.3	29.7	29.6	29.5
	≥ Bachelor's degree		90.8	67.3	41.4	55.6		44.6	33.3	32.9	33.0	32.5
Employment	Employed	0.00	86.4	55.9	49.3	55.9	0.15	61.2	62.1	49.8	54.1	64.2
	Not employed		13.6	44.1	50.7	44.1		38.8	37.9	50.2	45.9	35.8
Ethnicity	Hispanic	0.29	6.7	7.8	12.7	10.4	0.46	10.9	16.9	15.0	14.7	16.9
	Not Hispanic		93.3	92.2	87.3	89.6		89.1	83.1	85.0	85.3	83.1
Race*	White	0.22	84.5	85.6	77.0	80.6	0.98	64.5	63.4	63.6	63.7	63.6
	Nonwhite		15.5	14.4	23.0	19.4		35.5	36.6	36.4	36.3	36.4
Household Income*	Less than $50,000	0.00	13.0	22.2	46.6	34.8	1.00	29.9	29.9	29.9	29.9	31.4
	$50,000 to $149,999		54.4	54.5	45.2	49.3		48.9	48.9	48.9	48.9	47.8
	$150,000 or more		32.7	23.4	8.1	15.9		21.2	21.2	21.2	21.2	20.8
Household Size*	1	0.74	19.3	18.1	18.9	18.7	1.00	16.9	16.9	16.9	16.9	16.9
	2		42.8	45.1	33.7	38.4		33.0	33.0	33.0	33.0	33.0
	3		15.0	15.0	20.0	17.8		18.7	18.7	18.7	18.7	18.7
	4 or larger		23.0	21.8	27.4	25.1		31.4	31.4	31.4	31.4	31.4
Housing Tenure	Own	0.01	57.1	78.3	57.3	63.9	0.07	55.0	64.5	62.8	62.7	67.2
	Rent		35.0	18.2	37.2	30.9		27.7	28.0	31.3	30.8	28.5
	Other		7.9	3.4	5.5	5.1		17.3	7.5	5.8	6.5	4.4
Vehicles Available*	0	0.10	9.5	3.5	8.7	7.1	1.00	8.9	8.9	8.8	8.9	8.9
	1		31.8	28.7	42.9	37.1		23.1	23.2	23.1	23.1	23.1
	2		41.5	43.4	34.9	38.4		37.8	37.8	37.8	37.8	37.8
	3 or more		17.2	24.4	13.5	17.4		30.2	30.2	30.2	30.2	30.2

Note: * Variables controlled during weighting; CS=Convenience Sample; ES= Email Sample; OP= Online Panel; ACS = American Community Survey

4. COMPARISON OF MOBILITY AND ACTIVITY PARTICIPATION VARIABLES

While it is possible to correct for survey biases by weighting subsamples with respect to census distributions on socio-economic and demographic variables, the same cannot be said of mobility and activity participation variables (Ampt and Ortuzar 2004). It is generally not feasible to include measures of mobility and activity participation in survey weighting processes because there are no census distributions for such variables. As such, weighting based on socio-economic and demographic variables should help compensate for biases in mobility measures, resulting in a weighted sample that provides reliable population-wide estimates of mobility characteristics (Forsman et al. 2007). Table 2 presents a comparison of distributions for several mobility and activity participation variables across the three subsamples. In order to ensure that COVID effects do not impact the comparisons presented in this analysis, all mobility and activity participation variables are depicting pre-COVID patterns of behaviors. It should also be noted that there are no readily available census distributions for these variables. However, due to the socio-economic weighting process, one would expect the mobility and activity participation metrics across the three subsamples to resemble one another quite closely. If such "convergence" occurs with weighting, then it can be said with a reasonable degree of confidence that the mobility metrics derived from any weighted sample are likely to be representative of the overall population characteristics.

A comparison of unweighted statistical distributions shows that the subsamples differ substantially with respect to travel and activity participation variables. The online panel (which depicts lower levels of educational attainment, employment, and income) shows a lower level of driver's license holding. They also have significantly less access to a bike, are more likely to choose the private automobile for commuting, and least likely to choose transit for commuting. The convenience sample, which is largely comprised of individuals derived from professional and social networks of transportation professionals, exhibits a higher level of transit use, the lowest level of private vehicle use, and the highest level of "other mode" (walking, bicycling, ridehailing, and micromobility modes) use for commuting. While these findings are not surprising (Dingil and Esztergár-Kiss 2022; De Vos and Alemi 2020), they do convey a sense of the strong biases inherent in different sample recruitment strategies. The private vehicle use frequency and the transit use frequency variables generally depict patterns that are similar to those seen for commute mode choice, with convenience sample individuals more prone to using transit more frequently and less prone to using the private vehicle frequently, when compared with email sample and online panel respondents.

Although not statistically significantly different at the 0.05 level (but significantly different at the 0.11 level), there are key differences in work-from-home frequency across the subsamples who have the option to telecommute. The convenience sample shows the highest level of hybrid work modality (frequent or occasional work from home), consistent with the nature of the social and professional networks from which the convenient sample respondents were recruited. As these individuals are largely transportation professionals and individuals within their networks, it is likely that they are largely office workers who have some flexibility with respect to work location even in the pre-COVID era (Sanchez et al. 2021; Holgersen et al. 2021). The email sample and the online panel tend to be more similar to one another in terms of distributions of pre-COVID work-from-home frequency. The unweighted distributions for online grocery shopping frequency are not statistically significantly different from one another. In all cases, the percent who never shopped for groceries online in the pre-COVID era is quite high, with the online panel subsample depicting higher levels of online grocery shopping frequency than the other two subsamples (from a qualitative numeric perspective). The three subsamples are found to differ significantly with respect to their frequency of air travel for business. Consistent with the employment and education profile of the convenience sample, they depict higher levels of air travel for business when compared with the other two subsamples. The online panel sample depicts the lowest level of air travel for business, consistent with the lower income and education profile of the respondents in the online panel (Enzler 2017; Buchs and Mattioli 2021).

Overall, it can be seen that unweighted subsamples recruited through different strategies show significant differences with respect to mobility and activity participation variables. Virtually all the measures included in Table 2 are statistically different across the three subsamples (at the ~0.1 level), with the exception of online grocery shopping frequency. However, what is quite noteworthy and encouraging is that the weighted samples resemble one another more closely with a number of statistically significant differences becoming insignificant after weighting. The percent of individuals with driver's licenses is no longer significantly different across subsamples. Differences in bike access, commute mode choice, and private vehicle use frequency are all statistically insignificant across the three subsamples, suggesting that weighting to compensate for biases in a number of socio-economic and demographic variables may also correct for biases in mobility and activity participation variables. That is, regardless of the sampling recruitment strategy that is adopted, the weighted sample yields statistically identical measures of mobility and activity participation characteristics. The only variable that remained statistically different across the three subsamples is that of transit use frequency. This may be because transit use is heavily influenced by service quality and coverage, attitudes and perceptions related to transit and the environment, and the geographic and cultural context in which the travelers are located (Aston et al. 2020; Legrain et al. 2019). As such, weighting for socio-economic and demographic variables is unlikely to be sufficient in bringing about convergence in estimates of transit use frequency across subsamples recruited via different sampling strategies. However, even in the case of transit use frequency, it is clear that the convenience sample is the one contributing to the statistical difference. Both the email sample and the online panel depict transit use frequency patterns that are quite similar, suggesting that the use of a convenience sampling approach may be quite detrimental to obtaining robust estimates of mobility choices that are not strongly correlated with and explained by socio-economic and demographic variables (that are controlled in the weighting process).

Table 2 Pre-COVID Mobility and Activity Participation Characteristics

		Unweighted					Weighted				
Variable	**Attribute**	**χ2** p-value	**CS** n = 1,127 (%)	**ES** n = 2,946 (%)	**OP** n = 5,262 (%)	**All** n = 9,335 (%)	**χ2** p-value	**CS** n = 1,127 (%)	**ES** n = 2,946 (%)	**OP** n = 5,262 (%)	**All** n = 9,335 (%)
Drivers' License	Yes	0.00	98.0	96.6	87.5	91.6	0.55	92.1	90.8	87.6	88.6
	No		2.0	3.4	12.5	8.4		7.9	9.2	12.4	11.4
Regular Bike Access	Yes	0.00	67.3	53.7	42.2	48.9	0.11	60.9	50.4	46.6	49.4
	No		32.7	46.3	57.8	51.1		39.1	49.6	53.4	50.6
Private Vehicle Use Frequency	Every day	0.06	43.9	64.6	55.0	56.7	0.51	45.7	59.3	55.8	55.4
	Frequent		43.9	29.2	32.4	32.8		39.4	29.2	31.4	31.4
	Occasional		6.0	2.6	3.1	3.3		6.5	4.8	3.2	3.6
	Never		6.2	3.7	9.5	7.3		8.4	6.7	9.6	9.5
Transit Use Frequency	Every day	0.00	15.6	4.9	4.0	5.7	0.01	16.5	5.9	4.1	5.6
	Frequent		26.4	9.6	15.3	14.9		22.1	12.3	15.3	14.3
	Occasional		26.6	18.7	15.8	18.0		17.4	15.7	17.0	16.7
	Never		31.4	66.8	64.9	61.4		44.0	66.2	63.6	63.4
Online Grocery Shopping Frequency	Every day	0.23	0.1	0.0	0.7	0.4	0.53	0.0	0.0	0.7	0.4
	Frequent		6.5	7.0	15.6	11.8		12.6	10.4	15.6	12.6
	Occasional		10.3	7.6	10.5	9.5		4.7	9.7	9.9	9.3
	Never		83.1	85.4	73.3	78.3		82.7	79.8	73.8	77.7
Commute Mode Choice (*)	Private vehicle	0.00	54.1	74.2	77.6	72.4	0.52	65.6	74.0	78.4	74.2
	Transit		22.3	10.3	8.7	11.6		13.4	11.8	8.8	10.8
	Work-from-home		7.5	10.3	8.8	9.1		11.5	7.1	8.5	8.7
	Other mode		16.0	5.2	4.8	6.9		9.6	7.1	4.4	6.3
Airplane Travel for Business (*)	Very frequent	0.01	0.7	1.1	1.3	1.1	0.94	0.5	0.9	1.5	1.2
	Frequent		38.3	29.2	20.9	26.5		23.8	25.6	23.6	24.4
	Occasional		24.9	16.2	12.8	16.0		18.3	15.1	12.8	14.5
	Never		36.1	53.5	64.9	56.4		57.4	58.4	62.1	59.8
Work-From-Home Frequency (**)	Every day	0.11	10.5	21.4	21.8	18.8	0.83	21.7	17.2	21.1	20.1
	Frequent		58.0	48.0	49.0	51.0		42.7	50.4	48.0	47.2
	Occasional		26.2	21.5	16.9	20.8		26.4	21.7	19.0	21.4
	Never		5.3	9.0	12.2	9.4		9.2	10.8	12.0	11.2

Note: CS: Convenience Sample; ES: Email Sample; OP: Online Panel. Frequent = a few times per month *or* a few times per week; Occasional = a few times per year. (*): n = 942, 1743, 2679 and 5364 for CS, ES, OP & All respectively. (**): n = 676, 840, 1081 and 2597 for CS, ES, OP & All respectively.

5. MODELS OF ACTIVITY-MOBILITY CHARACTERISTICS

The analysis in this report has so far relied on simple descriptive comparisons to assess the extent to which sample recruitment strategy may play a role in shaping inferences regarding activity-mobility characteristics. To further investigate the influence of sampling strategy on metrics of activity-mobility characteristics derived from sample surveys, this section presents a series of statistical and econometric models in which the influence of the sampling strategy is assessed while controlling for socio-economic and demographic variables that significantly explain activity-travel choices. The report presents four distinct models of endogenous activity-travel variables to help conduct this assessment. The four models are: (1) Ordered probit model of vehicle ownership; (2) Ordered probit model of transit use; (3) Ordered probit model of online grocery shopping; (4) Multinomial logit model of commute mode choice. The entire Wave 1 dataset was used to estimate the three ordered probit models while a workers-only dataset was used to estimate the multinomial logit model of mode choice. All the variables represent pre-COVID activity-travel choices to control for any effects that the COVID-19 pandemic may have had in shaping these measures of behavior.

Model estimation results are presented in Table 3. In general, the models offer reasonable interpretations with respect to the influence of socio-economic and demographic variables on the endogenous variables of interest. For the sake of brevity, an exhaustive description of the influence of socio-economic and demographic variables is not provided here. A few key highlights are offered for illustrative purposes. Vehicle ownership is higher for those in the oldest age group (>55 years), as evidenced by the negative coefficients for other age groups. The youngest group depicts a positive coefficient for vehicle ownership, primarily because they continue to reside at home with individuals in the oldest age group (which serves as the base alternative), thus increasing overall household vehicle ownership. The likelihood of frequent transit use decreases with age, while the frequency of online grocery shopping appears to be highest for those in the middle age groups of 25-34 and 35-44 years. Older workers tend to favor the car for their commute and are found to work from home more so than their younger counterparts. Larger household sizes are associated with higher levels of vehicle ownership and higher levels of online grocery shopping, presumably because larger households are pressed for time and find the use of such services convenient (Karjalainen et al. 2021). As household vehicle ownership increases, the frequency of transit use decreases as does the frequency of online grocery shopping; these findings are entirely consistent with expectations as higher levels of vehicle availability will naturally be associated with greater levels of automobile use and in-person engagement in activities (Liu and Cirillo 2015; Kim and Kim 2005). Owning at least one vehicle is also associated with the use of car as the preferred commute mode; at the same time, zero vehicle ownership is likely indicative of lower income, education, and occupational status that is less conducive to work from home. Those in a single-family house exhibit lower levels of transit use, presumably because they are in lower density areas less served by transit (Mattson 2020). The presence of home internet is associated with higher levels of vehicle ownership, lower levels of transit use, and greater use of online grocery services, all of which are findings consistent with expectations (Pérez-Hernández and Sánchez-Mangas 2011). Higher transit level of service (referring to residing in cities with extensive transit coverage and service, including New York City, Boston, Chicago, Seattle, and San Francisco) is associated with greater frequency of transit use and lower levels of vehicle ownership, along with a higher probability of using transit for commuting. Residing in areas with higher population density is associated with greater levels of transit use, lower levels of vehicle ownership, and a lower probability of using the car for commuting. Once again, these findings are entirely consistent with expectations (Polzin et al. 2000; Costa et al. 2023; Merlin et al. 2021; Taylor and Fink 2013).

The key variables of interest in this context are those shown at the end of Table 3. These variables represent the sampling strategy employed, with convenience sampling serving as the base alternative and email recruitment and online panel explicitly included in the model specification. It is found that recruited via email sampling exhibit lower levels of transit usage frequency than those in the convenience sample even after controlling for a host of socio-economic and demographic variables. Arguably, this finding is not surprising given that the email sample is older, lives in single family detached houses, and owns their homes at higher rates than the convenience sample and the online panel. They also exhibit the highest levels of vehicle ownership as seen earlier in Table 1. However, all these socio-economic and demographic characteristics are now controlled in the ordered probit model of transit use frequency. It is therefore quite interesting to note that the email sample exhibits lower rates of transit use frequency even after controlling for these socio-economic and demographic characteristics, suggesting that the email recruitment variable is capturing the effects of unobserved variables (such as attitudes and lifestyle preferences) that vary across three subsamples. In this instance, It appears that the email sample is more inclined towards a car-oriented lower-density lifestyle, and it is this unobserved lifestyle preference/inclination that is reflected and captured via the effect of the email sampling variable, suggesting that an email sampling strategy is likely to yield a respondent sample that is more automobile-oriented.

The online panel variable is found to be statistically significant in several models of endogenous variables. The model estimation results show that online panel respondents exhibit lower levels of vehicle ownership and transit use frequency and higher rate of online grocery shopping frequency, presumably due to their comfort with navigating online services and applications (Lee et al. 2015). Online panel members were found to exhibit lower levels of income, education, and employment. These characteristics are generally associated with lower levels of vehicle ownership (King et al. 2022). Because of the special nature of the convenience sample (transportation professionals who use transit and other modes at a relatively higher rate than the general population), both the email sample and the online panel depict lower levels of transit usage frequency as evidenced by the negative coefficients for these variables in the model corresponding to transit usage frequency. However, the model specifications already include a host of socio-economic and demographic variables, suggesting that the online panel variable is reflecting the influence of unobserved variables and attitudes and preferences towards a more home-oriented online modality of activity participation. As online survey takers are likely to be tech-savvy and very adept at using online platforms and services, it is not surprising that their online grocery shopping frequency tends to be higher than groups recruited via other means (Hernandez et al. 2011). Thus, the online panel variable is capturing the effect of "being tech-savvy", which is an unobserved trait not captured by any of the other observed variables in the model specification. For mode choice , the online panel depicts a higher rate of work from home, once again reflecting a penchant for a more home-based online activity participation modality. The model specification also includes a couple of interaction effects to reflect that males in the convenience sample are less likely to shop online frequently for groceries and those who are college graduates in the convenience sample are less likely to use car for their commute or work from home compared to the other modes. The latter group is also less likely to reside in households with higher vehicle ownership and to shop online frequently for groceries.

Table 3 Model Estimation Results for Select Activity-Mobility Characteristics

Variable	Attribute	Model 1		Model 2		Model 3		Model 4					
		Ordered Probit		Ordered Probit		Ordered Probit		Multinomial Logit					
		Vehicle Ownership		Transit Usage Frequency		Online Grocery Shopping Freq		Commute Mode Choice (base = Other)					
								Car		Transit		WFH	
		Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Constant		--	--	--	--	--	--	0.85	5.33	-2.08	-11.08	-0.34	-1.32
Gender	Female	0.10	3.90	-0.25	-9.23	-0.21	-6.45	0.32	3.84	0.21	1.88	0.22	1.83
Age	18 to 24 years	0.20	3.85	0.57	10.13	0.39	6.35	-2.34	-18.35	-1.57	-7.13	-2.87	-10.93
	25 to 34 years	-0.23	-6.00	0.38	9.09	0.49	10.26	-0.83	-8.49	-0.39	-3.12	-1.71	-9.71
	35 to 44 years	-0.36	-9.48	0.29	7.08	0.50	10.61	--	--	--	--	-0.28	-2.37
	45 to 55 years	-0.12	-3.26	0.21	4.99	0.22	4.54	--	--	--	--	--	--
Education	Some college	0.11	3.95	0.15	3.36	--	--	--	--	--	--	--	--
	Bachelor's or higher	--	--	0.43	9.45	--	--	--	--	0.37	3.19	--	--
Employment	Employed	0.27	8.86	0.12	3.52	0.09	2.23	--	--	--	--	--	--
	Full-time	--	--	--	--	--	--	1.06	13.48	1.24	8.75	--	--
Student Status	Student	0.15	2.22	0.25	3.57	--	--	--	--	--	--	--	--
Race	White	0.12	4.06	-0.09	-2.81	-0.13	-3.50	--	--	--	--	0.27	2.07
	Black	--	--	--	--	--	--	--	--	0.56	3.86	--	--
Ethnicity	Hispanic	0.12	2.96	--	--	--	--	--	--	--	--	--	--
Household Size	One	--	--	--	--	--	--	0.34	3.21	--	--	0.36	2.50
	Two	0.84	23.64	--	--	--	--	0.21	3.09	--	--	--	--
	Three	1.21	28.24	--	--	0.18	4.34	--	--	--	--	--	--
	Four or more	1.42	33.57	0.07	2.17	0.32	8.13	--	--	--	--	--	--
Household Vehicle Ownership	Zero	--	--	--	--	--	--	-3.32	-12.67	0.48	3.08	-0.48	-2.21
	One	--	--	-0.89	-17.40	-0.30	-5.38	--	--	--	--	--	--
	Two	--	--	-1.14	-20.33	-0.57	-9.31	--	--	--	--	--	--
	Three or more	--	--	-1.19	-18.80	-0.87	-12.09	--	--	--	--	--	--
Household Income	$50,000 or less	--	--	--	--	--	--	--	--	--	--	-0.22	-1.78
	$50,000 to $149,999	0.55	18.69	0.11	3.12	0.10	2.78			--	--	--	--
	$150,000 or more	0.79	18.58	0.35	7.33	0.31	5.80	-0.31	-3.99	--	--	--	--
Housing Type	Single family house	0.43	13.23	-0.12	-3.39	--	--	--	--	--	--	--	--
	Apartment	--	--	--	--	--	--	-0.32	-3.67	0.38	3.35		
Work modality	Work-from-home	-0.26	-8.42	0.28	8.49	0.34	9.02	--	--	--	--	--	--
Home Internet	Yes	0.17	3.34	-0.17	-2.94	0.27	3.92	--	--	--	--	--	--
Housing Tenure	Own	0.28	8.98	-0.11	-3.21	0.12	3.22	--	--	--	--	--	--
Transit Service Level	High	-0.24	-7.75	0.48	14.61	--	--	--	--	1.31	13.61	--	--

Table 3 (continued)

Variable	Attribute	Model 1		Model 2		Model 3		Model 4					
		Ordered Probit		Ordered Probit		Ordered Probit		Multinomial Logit					
		Vehicle Ownership		Transit Usage Freq		Online Grocery Shopping Freq		Commute Mode Choice (base = Other)					
								Car		Transit		WFH	
		Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
Population Density	High (>2900/km^2)	-0.40	-13.10	0.56	17.74	0.30	8.59	-0.75	-9.78	--	--	-0.62	-4.87
Sample	Email	--	--	-0.38	-8.72	--	--	0.50	3.75	--	--	0.42	1.91
	Online panel	-0.30	-10.93	-0.43	-10.23	0.36	9.88	0.85	6.62	--	--	0.62	2.86
Interaction Effects	Male & CS	--	--	--	--	-0.38	-3.95	--	--	--	--	--	--
	Graduate & CS	-0.17	-3.86	--	--	-0.22	-3.17	-0.54	-3.61	--	--	-0.54	-2.06

Goodness of fit statistics

Model 1; sample size = 9,335; final log likelihood = -9236.15; initial log likelihood = -11465.2; r-square = 0.19

Model 2; sample size = 9,335; final log likelihood = -8180.06; initial log likelihood = -9790.41; r-square = 0.16

Model 3; sample size = 9,335; final log likelihood = -5774.76; initial log likelihood = -6443.47; r-square = 0.10

Model 4; sample size = 6170 (workers only); final log likelihood = -5131.49; initial log likelihood = -8553.44; r-square = 0.40; AIC = 10336.97; BIC= 10585.89

Note: Coef = coefficient; t-stat = t-statistic; Freq. = Frequency; WFH = work from home; "--" = not applicable. Base category corresponds to all complementary/omitted categories in each set of attributes.

As mentioned earlier, the COVID Future Survey was a panel survey that collected information from the same respondents at multiple points in time. The panel survey data set allows the examination of transitions in behavior for the same set of individuals through the period covered by the panel. To further compare the three subsamples recruited through different means, changes in the distributions of commute mode choice are examined specifically for the worker subsamples of each recruitment method (who responded to all three waves of the survey). This transition is shown in Table 4. While it is feasible to examine transition matrices (either in tabular form or via Sankey diagram), such transition matrices and diagrams are not included for the sake of brevity. Table 4 depicts the univariate distribution of commute mode choice for each of the three subsamples in each of the three periods. The commute mode choice variable is selected for this examination because of the widespread interest in this variable in transportation modeling and planning processes (Kane et al. 2020) and because of the effect that COVID-19 had on commuting (Magassy et al. 2023, Sinha and Chakraborty 2023, Edward et al. 2024). It is found that the convenience sample is especially different from the other two samples. The email sample and the online panel depict similar patterns of change; for example, both samples exhibit a drop in private vehicle mode share of about 25-30 percentage points. The corresponding drop for the convenience sample is nearly 40 percentage points. Similarly, the shift to work from home for the convenience sample is dramatically larger than that for the email sample and the online panel. While the percent of those working from home increases by 30-40 percent for the email sample and the online panel, the corresponding increase for the convenience sample (of mostly transportation professionals) is nearly 85 percentage points – reflecting the professional office nature of their occupation (Asmussen et al. 2024). In all cases, the subsamples expect to rebound to some degree – but not entirely back to pre-COVID percentages – in the post-COVID period (Magassy et al. 2023) (note that these percentages reflect what respondents stated that they expected to do in a post-COVID era, since actual post-COVID era behaviors could not be measured or observed within the duration covered by the panel survey).

Table 4 Change in Commute Mode for Stayer Sample of COVID Future Panel Survey

	Pre-COVID	**During COVID**	**Post-COVID**
		Convenience Sample (n = 166)	
Private Vehicle	45.8	6.6	37.4
Transit	31.3	0.0	22.8
Work-from-home	5.4	90.4	21.7
Other	17.5	3	18.1
		Email Sample (n = 212)	
Private Vehicle	69.4	40.1	64.6
Transit	9.9	1.4	7.1
Work-from-home	16.0	56.1	22.6
Other	4.7	2.4	5.7
		Online Panel (n = 400)	
Private Vehicle	76.0	52.6	73.4
Transit	9.7	2.8	6.3
Work-from-home	11.0	41.3	16.3
Other	3.3	3.3	4.0
		All (n = 778)	
Private Vehicle	67.7	39.5	63.3
Transit	14.4	1.8	10.0
Work-from-home	11.2	55.7	19.2
Other	6.7	3.0	7.5

6. STUDY IMPLICATIONS AND CONCLUSIONS

In an era of low survey response rates and high survey administration costs, transportation surveys are increasingly adopting a variety of sample recruitment strategies to boost respondent sample sizes. Among the variety of methods being deployed, three methods are of particular interest in the current context as they are being increasingly adopted for transportation surveys. This includes the use of convenience samples, the use of commercially available email lists (for administering surveys via email), and the use of online survey panels aggregated by survey research companies. All three methods are considered efficient, cost effective, and potentially beneficial from the standpoint of realizing desired/large respondent sample sizes.

This project aims to assess and compare these three methods with respect to differences and biases in sample characteristics that result from the adoption of each of the methods. In general, it is very difficult to perform such a comparison because any single survey will generally adopt just one single administrative modality to conduct the survey. It is therefore difficult to perform a controlled comparison of these three survey methods while controlling for other survey features (such as content and length of survey). A unique opportunity to perform such a comparison presented itself in the context of the COVID Future Panel Survey, a multi-wave nationwide longitudinal survey conducted in 2020 and 2021. Respondents were recruited via all three methods noted previously, i.e., convenience sampling, email messaging to a large database of email addresses purchased from a commercial vendor, and use of an online survey panel assembled by a commercial survey research company. The survey sample has a total of 9,335 respondents, with 1,127 in the convenience sample, 2,946 in the email sample, and 5,262 in the online panel. The same survey was administered to all three survey subsamples.

The assessment is conducted through three primary types of analyses. First, the comparison of socio-economic and demographic characteristics shows that the convenience sample tends to be of working age and depict higher levels of employment, income, and education. This is primarily because the convenience sample is largely a professional and social network of transportation professionals who are well educated and employed in the transportation profession. The online panel, on the other hand, is lower income, less educated, and exhibits lower levels of employment. The only variables for which the three subsamples were not statistically different from one another included race, ethnicity, household size, and gender (although the online panel clearly had a numerically larger percent of female respondents).

Second, the analysis involved comparing pre-COVID activity-mobility characteristics across the three subsamples (to eliminate any COVID effects). It is found that the subsamples differ significantly from one another on a host of mobility and activity-travel participation variables considered in this study. Online panel members show lower levels of driver's license holding, higher levels of regular access to a bicycle, and higher levels of online grocery shopping. On the other hand, the convenience sample of largely transportation professionals depicts the highest level of transit use frequency, driver's license holding, regular access to a bicycle, commuting by transit and other modes, and airplane travel for business. These findings are entirely consistent with expectations, given the nature and source of the convenience sample. The bottom line is that the three subsamples differed substantially with respect to measures of mobility and activity-travel modality, suggesting that the method of sample recruitment does influence measurement of travel behavior. However, what is particularly encouraging is that weighting the subsamples on socio-economic and demographic variables does compensate and overcome these differences quite substantially. When weighted distributions of mobility characteristics are compared across survey subsamples, it is found that statistical differences fade away for all mobility measures except for transit use frequency; this is because the convenience sample is so unique in comparison to the email sample and the online panel that transit use patterns in the weighted convenience sample continue to show a different pattern. Given that the convenience sample is largely comprised of transportation professionals, this finding is not surprising.

Finally, econometric models of vehicle ownership, transit use frequency, online grocery shopping frequency, and commute mode choice show that, even after controlling for a host of socio-economic and demographic variables in the model specification, the survey sample recruitment strategy has a significant effect on measures of mobility that serve commonly as endogenous variables of interest in the field of travel behavior research. Model estimation results show that the email sample is less likely to use transit frequently when compared with the convenience sample and online panel respondents. The online panel is more likely to engage in online grocery shopping frequently and work from home, and less likely to own a larger number of vehicles or use transit frequently. In general, these individuals appear to be more home-bound (less mobile) and more tech-savvy, thus enabling greater activity engagement (work and shopping) through online modalities.

This project has helped uncover the unique characteristics of survey respondent samples recruited through different means. The findings suggest that extreme care should be exercised in the recruitment and use of convenience samples as they may exhibit substantial biases with respect to socio-economic, demographic, and mobility characteristics. While weighting the survey sample to match census distributions compensates to some degree, a few significant differences may remain for variables where the convenience sample exhibits very uniquely different patterns. When it comes to email sample recruitment, the respondent sample appears to be older, higher income individuals, with a lifestyle that is more automobile-oriented, residing in single family detached houses in lower density areas. This is not necessarily a serious concern as weighting the sample appears to be able to correct for these biases and provide a more representative distribution of sample characteristics. The online panel is biased in the other direction, comprised of individuals who are younger and exhibit lower levels of income, educational attainment, vehicle ownership, and employment status. Once again, however, weighting the sample to match census socio-economic and demographic distributions is found to compensate for these biases for the most part, enabling the drawing of population-wide inferences regarding activity-travel characteristics. In other words, the use of online panels for travel surveys appears to be a cost-effective robust approach for survey sample recruitment, especially given the dismally low response rates associated with mail and email-based solicitations.

The study shows that the survey respondent sampling strategy variable is significant in explaining a host of mobility characteristics and choices even after controlling for socio-economic and demographic variables. This means that the survey method variable is capturing the effects of unobserved traits including attitudes and lifestyle preferences that are not adequately captured by socio-economic and demographic variables. Weighting survey samples compensates for biases in socio-economic and demographic variables, but does not necessarily compensate for biases in unobserved traits such as attitudes and lifestyle preferences (because there are no census distributions for such variables). It would therefore be of considerable value to include attitudinal and lifestyle preference questions in transportation surveys on a consistent and routine basis so that these traits can be explicitly included in travel model specifications, helping to enhance explanatory power, reduce omitted variable bias, and account for biases in unobserved traits that may arise from the choice of sample recruitment strategy.

REFERENCES

Ampt ES, Ortúzar JDD (2004) On best practice in continuous large-scale mobility surveys. Transport Reviews 24(3):337-363. http://dx.doi.org/10.1080/0144164032000140708

Angelini V, Brugiavini A, Weber G (2014) The dynamics of home ownership among the 50+ in Europe. Journal of Population Economics 27**:**797-823. https://doi.org/10.1007/s00148-013-0477-5

Asmussen KE, Mondal A, Batur I, Dirks A, Pendyala RM, Bhat CR (2024) An investigation of individual-level telework arrangements in the covid-era. Transportation Research Part A Policy and Practice 179:103888. https://doi.org/10.1016/j.tra.2023.103888

Aston L, Currie G, Delbosc A et al (2020) Exploring built environment impacts on transit use – an updated meta-analysis. Transport Reviews 41(1):73-96. https://doi.org/10.1080/01441647.2020.1806941

Bricka S, Reuscher T, Schroeder P et al (2024) Summary of Travel Trends: 2022 National Household Travel Survey. FHWA, U.S. Department of Transportation. FHWA-HPL-24-009. https://nhts.ornl.gov/assets/2022/pub/2022_NHTS_Summary_Travel_Trends.pdf

Buchs M, Mattioli G (2021) Trends in air travel inequality in the UK: from the few to the many? Travel Behaviour and Society 25:92-101. https://doi.org/10.1016/j.tbs.2021.05.008

Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L (2019) Online panels in social science research: expanding sampling methods beyond mechanical turk. Behavior Research Methods 51:2022-2038. doi: 10.3758/s13428-019-01273-7

Chauhan RS, Bhagat-Conway MW, Da Silva DC et al (2021) A database of travel-related behaviors and attitudes before, during, and after Covid-19 in the United States. Scientific Data 8:245. https://doi.org/10.6084/m9.figshare.15141945

Conway MW, Salon D, Da Silva DC, Mirtich L (2020) How will the COVID-19 pandemic affect the future of urban life? Early evidence from highly educated respondents in the United States. Urban Science 4:50. https://doi.org/10.3390/urbansci4040050

Costa AM, Marra AD, Corman F (2023) Public transport commuting analytics: a longitudinal study based on GPS tracking and unsupervised learning. Data Science for Transportation 5(15). https://doi.org/10.1007/s42421-023-00077-8

De Vos J, Alemi F (2020) Are young adults car-loving urbanites? Comparing young and older adults' residential location choice, travel behavior and attitudes. Transportation Research Part A Policy and Practice 132:986-998. https://doi.org/10.1016/j.tra.2020.01.004

Dingil AE, Esztergár-Kiss D (2022) The influence of education level on urban travel decision-making. Periodica Polytechnica Transportation Engineering 50(1):49-57. https://doi.org/10.3311/PPtr.16871

Edward D, Soria J. Stathopoulos A (2024) Multi-modal machine learning investigation of telework and transit connections. Data Science for Transportation 6(14). https://doi.org/10.1007/s42421-024-00100-6

Enzler HB (2017) Air travel for private purposes. An analysis of airport access, income and environmental concern in Switzerland. Journal of Transport Geography 61:1-8. DOI: 10.1016/j.jtrangeo.2017.03.014

Feijten P, Mulder CH, Baizán P (2003) Age differentiation in the effect of household situation on first-time homeownership. Journal of Housing and the Built Environment 18:233-255. https://www.jstor.org/stable/41107094

Forsman A, Gustafsson S, Vadeby A (2007) Impact of nonresponse and weighting in Swedish travel survey. Transportation Research Record 1993:80-88. https://doi.org/10.3141/1993-11

Gigliotti L, Dietsch A (2014) Does age matter? the influence of age on response rates in a mixed-mode survey. Human Dimensions of Wildlife 19(3):280-287. https://doi.org/10.1080/10871209.2014.880137

Griffin GP, Mulhall M, Simek C, Rigs WW (2020) Mitigating bias in big data for transportation. Journal of Big Data Analytics in Transportation 2: 49–59. https://doi.org/10.1007/s42421-020-00013-0

Hays RD, Liu H, Kapteyn A (2015) Use of Internet Panels to Conduct Surveys. Behavior Research Methods 47:685-690. https://doi.org/10.3758%2Fs13428-015-0617-9

Heinonen J, Czepkiewicz M, Árnadóttir A, Ottelin J (2021) Drivers of car ownership in a car-oriented city: a mixed-method study. Sustainability 3(2):619. https://doi.org/10.3390/su13020619

Hernandez B, Jiminez J, Martin MJ (2011) Age, gender and income: do they really moderate online shopping behaviour? Online Information Review 35(1):113-133. http://dx.doi.org/10.1108/14684521111113614

Hochstenbach C, Arundel R (2021) The unequal geography of declining young adult homeownership: divides across age, class, and space. Transactions of the Institute of British Geographers 46:4. https://doi.org/10.1111/tran.12466

Holgersen H, Jia Z, Svenkerud S (2021) Who and how many can work from home? evidence from task descriptions. Journal of Labour Market Research 55(4). https://doi.org/10.1186/s12651-021-00287-z

Kane K, Cryer J, Hsu J, Anderson M (2020) Affecting commute mode choice in southern California: which employer-based strategies work? Journal of Transport and Land Use 13(1):255-272. https://doi.org/10.5198/jtlu.2020.1558

Karjalainen LE, Tiitu M, Lyytimäki J et al (2023) Going carless in different urban fabrics: socio-demographics of household car ownership. Transportation 50**:**107-142. https://doi.org/10.1007/s11116-021-10239-8

Kim SH, Kim E (2005) Effects of public transit on automobile ownership and use in households of the USA. Review of Urban & Regional Development Studies 16(3):245-262. http://dx.doi.org/10.1111/j.1467-940X.2005.00090.x

King DA, Smart MJ, Manville M (2022) The poverty of the carless: toward universal auto access. Journal of Planning Education and Research 42(3):464-481. https://doi.org/10.1177/0739456X18823252

Landers RN, Behrend TS (2015) An inconvenient truth: arbitrary distinctions between organizational, mechanical turk, and other convenience samples. Industrial and Organizational Psychology 8(2):142-164. http://dx.doi.org/10.1017/iop.2015.13

Lee RJ, Sener IN, Handy SL (2015) Picture of online shoppers: specific focus on Davis, California. Transportation Research Record 2496(1):55-63. http://dx.doi.org/10.3141/2496-07

Legrain A, Buliung R, El-Geneidy AM (2015) Who, what, when, and where: revisiting the influences of transit mode share. Transportation Research Record 2537(1):42-51. https://doi.org/10.3141/2537-05

Liu Y, Cirillo C (2015) Measuring transit service impacts on vehicle ownership and use. Public Transport 7:203-222. http://dx.doi.org/10.1007/s12469-014-0098-8

Lo A, Srikukenthiran S, Miller EJ, Habib KN (2020) Impact of multiple sample frames on data quality of household travel surveys: the case of the 2016 Transportation Tomorrow Survey. Transportation Planning and Technology 43(6), 553–570. https://doi.org/10.1080/03081060.2020.1780707

Lohr SL (1999) Sampling: Design and Analysis. Duxbury Press. New York.

Magassy TB, Batur I, Mondal A et al (2023) Evolution of mode use during the covid-19 pandemic in the United States: implications for the future of transit. Transportation Research Record 00:1-13. https://doi.org/10.1177%2F03611981231166942

Mattson J (2020) Relationships between density, transit, and household expenditures in small urban areas. Transportation Research Interdisciplinary Perspectives 8(100260). https://doi.org/10.1016/j.trip.2020.100260

Merlin LA, Singer M, Levine J (2021) Influences on transit ridership and transit accessibility in US urban areas. Transportation Research Part A Policy and Practice 150:63-73. https://doi.org/10.1016/j.tra.2021.04.014

Miratrix LW, Sekhon JS, Theodoridis AG, Campos LF (2018) Worth weighting? how to think about and use weights in survey experiments. Political Analysis 26(3):275-291. doi:10.1017/pan.2018.1

Monzon A, Juliob R, Garcia-Martinez A (2020) Hybrid methodology for improving response rates and data quality in mobility surveys. Travel Behaviour and Society 20:155-164. https://doi.org/10.1016/j.tbs.2020.03.012

Pendyala RM, Konduri KC, Christian KP (2011) PopGen 1.1 User's Manual. Lulu.com Publishers, USA.

Pérez-Hernández J, Sánchez-Mangas R (2011) To have or not to have internet at home: implications for online shopping. Information Economics and Policy 23:213-226. DOI: 10.1016/j.infoecopol.2011.03.003

Peytcheva E, Groves RM (2009) Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. Journal of Official Statistics 25(2):193-201.

Polzin SE, Chu X, Rey JR (2000) Density and captivity in public transit success: observations from the 1995 Nationwide Personal Transportation Study. Transportation Research Record 1735(1):10-18. https://doi.org/10.3141/1735-02

Sanchez DG, Parra NG, Ozden C et al (2021) Who on earth can work from home? The World Bank Research Observer 36(1):67-100. https://doi.org/10.1093%2Fwbro%2Flkab002

Silvano AP, Eriksson J, Henriksson P (2020) Comparing respondent characteristics based on different travel survey data collection and respondent recruitment methods. Case Studies on Transport Policy 8:870-877. https://doi.org/10.1016/j.cstp.2020.05.015

Sinha S, Chakraborty M (2023) Causal analysis-based reduced order modeling of human mobility in the U.S. during COVID-19 pandemic. Data Science for Transportation 5(18). https://doi.org/10.1007/s42421-023-00081-y

Stedman RC, Connelly NA, Heberlein TA et al (2019) The end of the (research) world as we know it? understanding and coping with declining response rates to mail surveys. Society & Natural Resources 32 (10):1139-1154. http://dx.doi.org/10.1080/08941920.2019.1587127

Taylor BD, Fink CNY (2013) Explaining transit ridership: what has the evidence shown? Transportation Letters 5(1):15-26. https://doi.org/10.1179/1942786712Z.0000000003

Ternovski JA (2022) Note on increases in inattentive online survey-takers since 2020. Journal of Quantitative Description: Digital Media 2:2022. https://doi.org/10.51685/jqd.2022.002

Weber J, Hubrich S, Wittwer R, Gerike R (2020) Non-probability recruitment strategies for innovative smartphone-based travel surveys. Internationales Verkehrswesen 72:58-61.