NATIONAL CENTER FOR UNDERSTANDING FUTURE
TRAVEL BEHAVIOR AND DEMAND

Final Project Report

Investigating Travel Survey Representativeness

BY

Amy Z. Fong
Email: amyzfong@umich.edu

F. Atiyya Shaw
Email: atiyya@umich.edu

University of Michigan, Ann Arbor
2350 Hayward St., Ann Arbor, MI 48105

October 2025

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
N/A	N/A	N/A	
4. Title and Subtitle Investigating Travel Survey Representativeness		**5. Report Date** October 7, 2025 **6. Performing Organization Code** N/A	
7. Author(s) Amy Z. Fong, https://orcid.org/0000-0002-5837-5003 F. Atiyya Shaw, https://orcid.org/0000-0001-8717-5118		**8. Performing Organization Report No.** N/A	
9. Performing Organization Name and Address Civil and Environmental Engineering Department University of Michigan, Ann Arbor 2350 Hayward St., Ann Arbor, MI 48105		**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, 2023-2024 **14. Sponsoring Agency Code** USDOT OST-R	
15. Supplementary Notes N/A			
16. Abstract Household travel surveys are a critical data source for transportation planning and forecasting, offering insights into traveler behavior such as trip purpose, mode choice, travel time, and temporal patterns. However, declining response rates, increasing response biases, and measurement errors pose growing challenges to survey data quality. This project develops and applies a methodological framework to evaluate the representativeness and behavioral bias of household travel surveys. Using 2021–2022 data from the Minneapolis–St. Paul metropolitan region, we examine how oversampling, the inclusion of convenience samples, and calibration weighting affect the bias and precision of travel behavior estimates. The framework builds upon a review of recruitment practices, sampling methods, and survey evaluation techniques, including findings from multiple metropolitan regions and recent literature. The resulting approach offers an accessible, theoretically grounded, and practically relevant tool for assessing survey design and improving data quality.			
17. Key Words travel survey data quality; targeted oversampling; convenience and multimethod sampling; nonresponse bias; survey weighting		**18. Distribution Statement** No restrictions.	
19. Security Classif.(of this report) Unclassified	**20. Security Classif.(of this page)** Unclassified	**21. No. of Pages** 30	**22. Price** N/A

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

DISCLAIMER

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 the University of Michigan Civil and Environmental Engineering Department for their support of university-based research in transportation, particularly for the funding provided for this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The core source of data for transportation planning and forecasting comes from household travel surveys. Travel surveys are used to obtain insight into the behavioral decisions of travelers; for example: trip purposes such as work or shopping; means/mode of transport such as car, walk, bus, etc.; travel time; and time of day/week. However, there are increasing challenges to survey data collection and its resultant data quality. Most critically, survey response rates continue to decline, introducing potential representation errors and behavioral biases in the resultant data. To compound this, it is increasingly difficult to obtain high quality data from those who *do* respond, introducing another class of errors that we do not address here (e.g., response biases, measurement errors). In this project, we focus on the former challenge to develop and apply an evaluation framework for assessing representativeness of travel survey data.

We begin by providing a brief overview of evolving survey design, sampling methods, and post-processing methods across travel surveys. To the goal of the paper, we focus on methods to improve and assess representativeness. Building on this synthesis, we develop and apply a methodological framework for evaluating travel survey representativeness and behavioral bias. We investigated travel survey data from several regions for this project, namely: Detroit, Seattle, San Francisco, and Minneapolis-St. Paul metropolitan regions. Due to data limitations, in this effort, we report only on data from the Detroit and Minneapolis-St. Paul regions. Specifically, we use 2021-2022 household travel survey data from the Minneapolis-St. Paul, Minnesota region to evaluate how oversampling, introducing a convenience sample, and calibration weighting bear on the bias and precision of travel behavior statistics.

Taken together, this work seeks to illustrate an accessible, theoretically informed, and modeling-relevant framework for evaluating travel survey representativeness. This effort is being developed into a shareable document for practitioners and policymakers in collaboration with the Graham Institute at the University of Michigan.

1. INTRODUCTION

Household travel survey (HTS) data enable many practical and research outputs. Central among them are travel demand models and related behavioral analyses, key tools used to inform transportation planning and infrastructure investments. Researchers and policymakers operate on the understanding that HTSs should be representative of all socioeconomic and demographic (SED) groups to produce reasonable recommendations from resultant analyses (Lubitow et al. 2021; Brown 2017). For instance, it is intuitive that an analysis based on data which under-represent SED groups more likely to use public transit (e.g., lower income individuals in the U.S.), might understate the usage and value of improved transit routes. But say that a hypothetical MPO oversampled lower income individuals, and so find that they *are* well represented in our data – that alone doesn't guarantee their *travel behavior* is well-represented. After all, we know that behavior is heterogenous: it may be the case that lower income individuals who are more likely to respond to travel surveys are less likely to take transit, suggesting that the travel behaviors captured in the data are biased. This produces nonresponse biases on travel behavior. We often find that people of color, people with lower socioeconomic status, immigrants, non-English speakers, and disabled people are both transportation disadvantaged and "hard-to-reach" in surveys, making these groups particularly vulnerable to potential representation errors (Tourangeau 2014).

Consequently, transportation analysts, planners, modelers, and policy makers are each concerned with travel survey representativeness, a key dimension of data quality. In this paper, we use the term representativeness to broadly refer to a survey dataset's ability to produce unbiased representations of the population and their travel behaviors, either weighted or unweighted. To address growing challenges to survey data collection, transportation agencies responsible for large-scale household travel data collection (e.g., state departments of transportation, metropolitan planning organizations -- MPOs) are exploring numerous means of improving data representativeness. Examples include: increased and varied incentives, follow-ups with nonrespondents, sending more invitations (i.e., increased sampling rates and oversampling), purchasing online samples or collecting convenience samples, or purchasing auxiliary data that can refine the sampling or weighting process. Faced with many (costly) options, agencies are eager to know how effective these potential survey design tools can be. However, fundamentally incomparable sampling and weighting approaches, time periods, and regional contexts across travel surveys limit generalizable insights. This motivates the aim of this effort to illustrate an accessible, theoretically informed, and modeling-relevant framework for evaluating travel survey representativeness

This report is organized as follows: First, we review the state of travel survey design and methods with an emphasis on approaches to measuring and improving travel survey representativeness. We then describe the data and analytic methods used to evaluate the respective effects of nonresponse, weighting, and random error on bias in travel survey statistics. Finally, we apply these methods to 2021-2022 household travel survey (HTS) data from the Minneapolis-St. Paul, Minnesota region, illustrating means of evaluating representativeness trade-offs.

2. LITERATURE REVIEW

2.1 Overview of travel survey practice

Here, we provide an overview of evolving HTS practices across the U.S., with an emphasis on representativeness. Findings in this section are sourced from the literature, as well as from our independent, detailed review of several HTS reports from differing agencies.

2.1.1 Emerging survey design considerations

There is variation as to the exact method of HTS data collection efforts in the U.S. For example, while most HTSs use "recruitment" or initial sign-up surveys to obtain demographic and household data, some regions collect this information initially and others collect aspects of these data at various timepoints throughout the survey (to reduce the upfront burden). With respect to the survey questions themselves, some agencies are making efforts to ask non-traditional questions (e.g., psychological questions like attitudes, context-relevant questions related to work from home conditions, e-commerce use, etc.) in addition to consistently collected travel diary data. These types of data have been shown in research to have the potential to improve travel behavior models and forecasts; however, respondent burden considerations typically restrict them from being included in large scale regional and national surveys.

With regard to method or mode of data collection, most regions no longer use paper surveys, some offer phone-based data collection[1], and all use web-based (online) data collection platforms. To supplement online data collection, passive smartphone-enabled GPS tracking for trips is becoming more widespread. The state of practice has recognized the increased accuracy of app-based GPS tracking for trips, although it has been found that this can exclude some categories respondents (e.g., older individuals) and introduce bias or underreporting for certain types of trips. Moreover, when this type of data collection occurs in two stages (recruitment or sign-up survey with transfer to app-based travel diary collection) attrition after the first stage can affect data representativeness. In particular, households with more adults and those which need to transfer from an online recruitment survey to an app-based travel diary (a response mode change mid-survey) are less likely to complete the second stage of data collection (Chestnut et al. 2025; Armoogum et al. 2024).

Survey frequency is also evolving as agencies respond to rapidly changing conditions and technologies. For example, more agencies are conducting surveys biennially with some even moving to annual data collection. In contrast, in the past, decennial data collection was common practice. In addition to survey frequency, agencies are also experimenting with the optimal day/duration for data collection. While all HTSs prioritize weekday data collection first, an increasing number are attempting to obtain data across multi-day or even up to seven-day periods (Khoury et al. 2025). Other variations include requiring at least consecutive two consecutive weekdays, requiring data across a 24-hour period, and so on (Selby et al. 2024).

2.1.2 Survey sampling and recruitment practices

As noted, regional HTS response rates in the U.S. are decreasing, and as of 2025 are often in the low single digits (~1-3%) (Chestnut et al. 2025). Certain SED groups tend to be under-represented in surveys across North American regions and contexts. Renters, racial and ethnic minorities, lower-educated individuals, lower-income households, and households with four or more members were under-represented or had lower response rates in numerous regional household travel surveys from 2016 to 2022 (regions encompassing Washington, DC; Phoenix, Arizona; Seattle, Washington; San Diego, California; and Greater Toronto) and the 2022 National Household Travel Survey (NHTS) (San Diego Regional Transportation Study, 2017; 2017 MAG Household Travel Survey Report, 2018; 2017/2018 (*San Diego Regional Transportation Study*

[1] Due to costs, agencies typically offer phone interview data collection only for respondents who prefer this method and/or are not able to utilize online platforms.

2017; *2017 MAG Household Travel Survey Report* 2018; *2017/2018 Regional Travel Survey* 2021; *2017-2022 PSRC Household Travel Survey Program: Lossons and Recommendations* 2022; *NHTS NextGen Study* 2022; Lo et al. 2020). In light of this, transportation agencies and departments are making deliberate efforts to recruit subgroups with relatively low response rates or with rare travel behaviors (i.e., ridehailing). These strategies, detailed in this section, each affect the representativeness of the final dataset.

Here, we discuss emerging and widely adopted practices taken to increase travel survey representativeness and/or improve recruitment efficiency. Many agencies now use targeted oversampling (typically through geographic stratification) to deliberately recruit from subgroups of interest. This aims to improve subgroup representation but may also magnify self-selection bias if these efforts are focused on low-response groups (Joh and McCall 2023; Selby et al. 2018). Other relevant advancements include using response rate predictive models and auxiliary (third-party) data to define strata (Anderson et al. 2018; Bradley et al. 2015; Coy et al. 2019). For instance, one survey used predictive behavioral models to oversample likely ridehailing users (Mark Bradley and Coy, n.d.).

Other recruitment strategies target response behavior, or in other words, seek to reduce unit nonresponse). Incentives increase willingness to respond (Joh and McCall 2023; Coy et al. 2019). However, it is unclear whether incentives resolve or merely reshape the self-selection biases present. Agencies also sometimes invest in public awareness campaigns to increase the likelihood that a household will respond to a travel survey invitation. More tailored recruitment strategies, or even reforms to travel survey development processes, may be on the horizon as practitioners grow more sensitive to the broader socio-political issues underlying low response willingness among marginalized groups (Bricka et al. 2013).

This is also increasing interest in combining convenience and probability samples. Currently, MPOs cautiously regard this as an opportunity to increase the representativeness of their overall dataset (Joh and McCall 2023). However, convenience samples lack well-understood statistical properties and have unknown self-selection biases (Federal Highway Administration 2023). One MPO found online panel data imposed weighting complications that outweighed the value of additional data (Joh and McCall 2023). However, another used multiple channels to recruit convenience respondents and deemed the effort successful in improving representativeness while lowering recruitment costs (Selby et al. 2018). Thus, evidence and present attitudes on convenience samples are mixed.

2.1.3 Survey data cleaning, processing, and expansion

Once survey data has been collected, it must go through quality control and cleaning to remove incomplete or inaccurate records. Some agencies may attempt to recontact households with incomplete data to amend or complete responses (sometimes with the offer of additional incentives). Overall, data cleaning procedures are often not clearly reported and there is a lack of standardization and transparency across agencies.

After cleaning, HTS data undergoes processing to address nonresponse and demographic discrepancies. To this end, agencies apply various weighting schemes to expand or redistribute the data along selected demographic targets. Weights can be produced for household or person-level targets, or increasingly, may be developed across these to produce a "joint" weight that removes complexity in analysis. Weights may also be developed differently for weekdays, weekends, etc.

The most common weighting approach is iterative proportional fitting (IPF; also known as raking) that attempts to develop weights based on the distributions of multiple variables (i.e., overlapping targets). Other agencies use a population synthesizer that samples from the survey households and persons until the final dataset reaches the targets for pre-selected variables. Common examples of variables used as targets for either approach include household demographics (size, income, number of vehicles) and person demographics (age, gender, race). Less common examples include area type, dwelling type, auto sufficiency, county, etc. Multiday and biennial surveys must grapple with how to develop weights that account for data collection across various days and/or how to fuse data across years.

Because travel surveys need to describe numerous behaviors and outcomes, a single set of weights is unlikely to address biases on all. It may even introduce new ones (further discussed in Section 2.2.2). Further, higher weight variation worsens the precision of weighted estimates. The efficacy of mitigating measures like weight trimming is circumstantial (Battaglia et al. 2009). For example, the quality of final weighted estimates is also a product of the baseline level of bias present in the unweighted survey dataset. Therefore, weighting methods also affect (final) travel survey data representativeness.

There is also increasing interest in supplementing HTS data with external data sources to improve representativeness and potentially expand coverage to visitors, long distance travel, etc. Common sources of passively collected, external data include data products from Replica, LOCUS, Streetlight, and Inrix. Some agencies also purchase an expanded sample that accompanies National Household Travel Survey (NHTS) data collection (i.e., NHTS add-on) (Selby et al. 2024).

2.2 Representativeness analysis of travel surveys

Survey data representativeness can refer to many things (Ochsner 2021). In the household travel survey context, the typical meaning refers to whether the dataset is a *miniature of the population*, or whether sample distributions match population distributions. In this paper, we follow this convention while noting that the term can have various meanings within the literature. For example, it is also common to treat response rates as an indicator of travel survey representativeness. We summarize several approaches to assessing representativeness here.

2.2.1 Representativeness evaluation based on sample distribution

To assess whether a sample achieves *miniature of the population*, analysts compare sample distributions to known population values (Ochsner 2021). A dataset may be representative of some variables but not others. Most commonly, these comparisons are performed on sociodemographic characteristics, for which censuses or reliable official statistics are often available to serve as a known value or benchmark. The more deviation the sample estimate from the known value, the "worse" the representativeness. Statistical tests such as t-tests (for continuous variables) and chi-square tests (categorical variables) can assess whether there are meaningful differences between the two. Qualitative assessments (i.e., without statistical tests) are often reported in travel survey methodology reports and are common prefaces to travel behavior studies.

HTS representativeness analyses often apply this analytic lens. This reveals one major limitation of the *miniature of the population* approach because trustworthy external data sources seldom include travel behavior estimates. As a result, there are more opportunities to generate evidence on the representativeness of sociodemographic characteristics than of travel behaviors in North American probability travel surveys (Lo et al. 2020; Selby et al. 2018). The National Household Travel Survey (NHTS) serves as the best external reference for travel behavior estimates, but there can be minor inconsistencies in time period of data collection, geographic scope, and measurement of travel behaviors. As a result, the presence of any travel behavior differences between a regional travel survey and the NHTS may be due to numerous factors beyond differential nonresponse. Nonetheless it provides analysts in the U.S. one of the few opportunities to examine travel behavior representativeness (Meister et al. 2025; Son et al. 2013).

2.2.2 Representativeness evaluation based on differential non-response

Sociodemographic comparisons with census data are not the only approach to representativeness analysis. Others are based on indicators of nonresponse bias, which is a prominent (but not the only) influence on sociodemographic representativeness. This type of bias is also commonly referred to a self-selection or self-selection bias. Nonresponse bias occurs when individuals who do not respond are systematically different from those who do (See Section 4.1 for more detail). This is particularly important for travel behavior analyses when *differential* nonresponse occurs across populations with distinct travel. Thus, it might be argued that the key measure of interest in travel behavior analyses may not be representativeness, but rather **nonresponse bias**.

One nonresponse bias analysis approach is to conduct follow up surveys with nonrespondents (Wittwer et al. 2024). By examining initial nonrespondents, this approach provides an estimate of how and whether nonrespondents and respondents have systematically different travel behaviors. It may be less frequently performed on travel surveys due to the additional cost and because it requires the follow-up survey to achieve a high response rate.

Another approach is to examine response rates. Travel survey analysts commonly regard survey response rates as an indicator of data quality and a means of assessing representativeness (Korimilli et al. 1998; Hubrich et al. 2018). The response rate represents the proportion of eligible units that responded to a survey invitation. It is often assumed a higher response rate yields more accurate survey estimates. Yet, substantial increases in response rate are only weakly associated with lower nonresponse bias, even when response rates are segmented by subgroup (Groves and Peytcheva 2008; Brick and Tourangeau 2017; Peytcheva and Groves 2009). Nonetheless response rates can also shed light on the representativeness of certain subgroups, for example by examining response rates across geographies (e.g., Census block groups) with distinct sociodemographic profiles (Bradley et al. 2015).

As mentioned previously, nonresponse is not the only source of error that influences representativeness. Overall *unweighted* sample representativeness can also be affected by coverage error in the sampling frame, random sampling error, because of the sample design itself, or nonresponse (differential nonresponse among different population groups). Further, *weighted* sample representativeness is also affected by post-survey adjustments to the data, like nonresponse adjustments, raking, and poststratification. For example, a weighted sample could be un-representative of educational attainment if it was not used explicitly when calibrating the sample to a known population.

To sum it up, unrepresentativeness introduces biases into ***travel behavior analyses*** when it is caused by *differential* nonresponse across populations with distinct travel patterns (i.e., termed nonresponse bias, to be further discussed in the next subsection). Thus, it might be argued that the key measure of interest in travel behavior analyses may not be representativeness, but rather **nonresponse bias**. However, regardless of the presence of nonresponse bias, if there are few observations of a certain groups of interest, then the survey dataset is also unrepresentative in that it cannot be used to prepare precise estimates. In this way, if a survey is to be used for subgroup analysis, severe under-representation and thus too few observations of those groups worsens travel data quality by prohibiting statistical inference. Because of this, we also see evaluation of **precision** as an important aspect of representativeness analysis.

2.3 Summary

This review identifies numerous takeaways that shape our study motivation and design. First, many see **nonresponse bias** (or self-selection bias) on SED characteristics associated with travel behavior as a central threat to representativeness. This means we need to elucidate the link between SED representativeness and travel behavior-related nonresponse bias. Second, travel surveys need to represent distinct travel behaviors across population subgroups. Thus, bias *within* distinct population segments is of critical concern to travel survey data applications. This further implies that **precision** of travel behavior estimates matters because it affects the ability to detect differences between subgroups (including when represented as regression coefficients). Third, all recruitment strategies have direct implications for the weighting process, which introduces further complexity to post-survey evaluations of bias and precision. Finally, the efficacy of emerging strategies to improve travel survey representativeness remains unclear for two main reasons. The first is that travel survey context and agency needs range, so different strategies are applied in different places and periods and results are incomparable. The second is that there are limited opportunities to evaluate representativeness of travel behaviors. This underscores the need for an analytic framework capable of (1) evaluating outcomes of these survey design choices *within* a single survey rather than across surveys and (2) evaluating modeling-relevant properties of travel surveys such that we understand not only representativeness of sociodemographic characteristics but also representativeness of travel behaviors and their *relationship* to sociodemographic characteristics.

3. DATA

For the remainder of this paper, we focus on the Twin Cities region in the U.S., although we first present some insights into the 2015 Detroit-area survey as well, as it motivates the next steps. In this report, we also rely on American Community Survey data to benchmark the travel survey (i.e., serve as the reference for true population distributions). The U.S. Census Bureau develops this product for purposes such as these.

3.1 Southeast Michigan Council of Governments Household Travel Survey (2015)

We report briefly on investigations into the 2015 Southeast Michigan Council of Governments (SEMCOG) HTS. This initial collaboration motivated much of the remainder of this project by highlighting the pressing need that many regional agencies are currently experiencing to understand past and current representativeness outcomes in order to inform current and ongoing data collection. It also revealed the importance of obtaining specific information on sampling characteristics from the consulting agencies responsible for data collection.

In 2015, SEMCOG collected a stratified, address-based sample. They oversampled block groups with higher prevalence of hard-to-survey, modeling-relevant populations including zero vehicle households, households with four or more people, and transit commuters. SEMCOG collected data in two phases, Spring and Fall 2015. The overall response rate was 5% and the final delivered dataset included 12,395 households. 90% of recruited households chose whether to complete the travel diary online or by telephone, and 10% were selected to participate in a GPS travel diary. The former group received $20 for each complete household while the latter received

$25 for each complete respondent in the household.

The final weights were developed using post-stratification, that is, to match totals of key subpopulations in the region. These subpopulations, or cells, were defined by varying combinations (i.e., joint) of household size, number of workers, and number of vehicles in the household.

In early meetings with the SEMCOG modeling and data team, we presented sample comparisons between the unweighted final SEMCOG dataset and weighted ACS estimates. The differences, summarized below, may be caused **by both different response propensities among subgroups and the stratified sampling design**. The relative contribution of each factor cannot be estimated without more information about the sample design, namely, the invitation rate in and geographic bounds of each sampling stratum. This was a driving factor in not moving forward with this data as we were not able to obtain additional sampling information needed to be able to disentangle these effects.

Here, for completeness, we discuss the results of our initial investigations in the SEMCOG 2015 HTS data. Unweighted comparisons of the 2015 SEMCOG data with ACS data showed that in every county, higher educated individuals are overrepresented relative to population estimates from the 2015 ACS, while those with high school degrees are underrepresented. We also find that individuals who commute by private vehicle are overrepresented, compared with individuals who walk/bike, take transit, or work from home. On the other hand, individuals who do not travel to work (work from home) are underrepresented in every region, and in St. Clair and Monroe counties no such individuals were observed at all. In every SEMCOG region (except Washtenaw and Monroe), individuals belonging to households with income below $25,000 a year are underrepresented compared with 2015 ACS. Likewise, in every region except Washtenaw individuals belonging to households with income greater than $75,000 a year are overrepresented in the SEMCOG sample. In every SEMCOG region (except Livingston and Monroe), individuals belonging to households with no household vehicles are overrepresented relative to population estimates from 2015 ACS. In every SEMCOG region (except Eastern Wayne) those belonging to households which rent are underrepresented while those belonging to households which own their dwelling are overrepresented.

An outcome of this investigation has been continued collaboration with SEMCOG on the current HTS data collection effort.

3.2 Twin Cities Regional Household Travel Survey (2021-2022)

The Metropolitan Council (Met Council) is the MPO for the Twin Cities region, which includes Minneapolis-St. Paul. We analyze their 2021-2022 regional household travel survey, hereafter referred to as the Travel Behavior Inventory (TBI). Met Council serves a seven-county planning region. TBI's sample included households in the region (core) and the twelve surrounding counties (ring). Figure 1 presents these strata delineations.

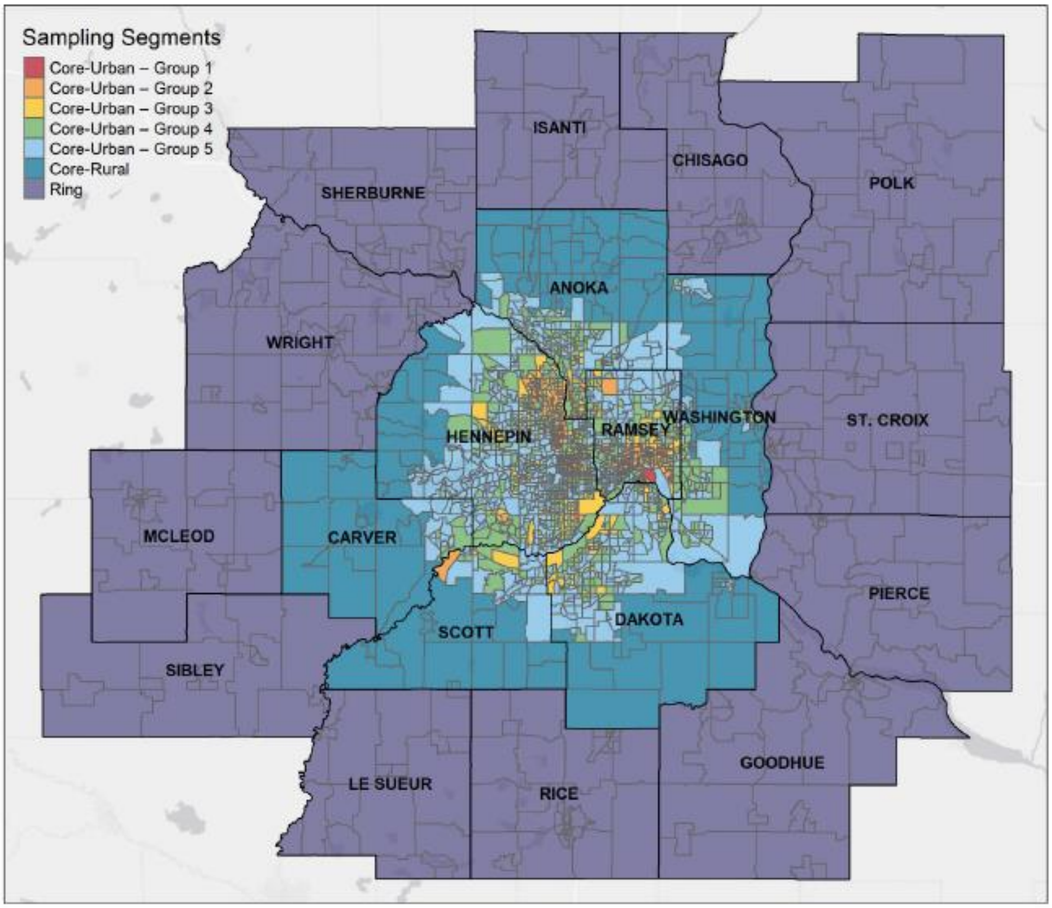

Figure 1: TBI study region by strata (sampling segments); sourced from survey documentation

TBI was fielded from June 2021 to January 2022 using address-based sampling (probability sample) and convenience sampling. Respondents could provide data through a smartphone app (seven-day travel diary, self-administered), online (one-day travel diary, self-administered), or telephone (one-day travel diary, with interviewer). The stratified sample design divided the core into six strata and treated the ring as one. Our analysis is limited to the six core strata. The core strata were delineated by block group-level race and income compositions, sourced from 2015-2019 American Community Survey (ACS) 5-year detailed tables. TBI over-sampled strata (disproportionate stratification) with higher proportions of BIPOC residents; it sampled nearly 100% of households in the highest-BIPOC stratum (composed of block groups where at least 80% of residents were BIPOC) and sampled 10% to 40% of households in the remaining strata.[2] The two urban strata with lowest BIPOC proportions had nearly twice the response rate as the stratum with the highest BIPOC proportion (3.1% and 3.2% compared with 1.6%).

Participants were offered incentives, varying based on recruitment mode, data collection mode, and whether the participant or household was considered hard-to-reach. Though we lack the requisite data to analyze incentive efficacy, a pilot early in data collection found that doubling or tripling the base incentive of $10 increased response rates.

[2] Estimated using response rates by strata reported in the documentation, household population by strata reported in the documentation, and number of completed households per stratum in the final dataset.

TBI's weights are calibrated jointly to household and person level targets using PopulationSim. Joint household-person weights are necessary for population synthesis inputs, though ACS microdata more often serve as seed data. The household-level calibration targets, sourced from the ACS, were based on household size, income, household workers, household vehicles, age of head of household, and presence of children. The person-level targets were based on sex, age, worker status, student status, educational attainment, race, and ethnicity. They also account for the proportion of convenience sample respondents in each strata.

We limit the analyses which follow to the "Core" strata, which encompass the seven-county metropolitan area. We exclude the "Ring" stratum which covers the twelve surrounding counties because their populations are too small for their residents to be identified in the ACS microdata.

3.3 American Community Survey (ACS)

This study uses weighted American Community Survey (ACS) microdata to benchmark TBI estimates. ACS is considered the most reliable and accurate source of population estimates for sub-state geographies in the U.S. because of its high response rate and large sample size.

However, like any other survey, ACS can have measurement and representation biases. Futher, differences between ACS and TBI in question wording and context (i.e., response mode, question order) may cause different estimates of the same construct. We assume this impact is minor for SED characteristics and vehicle ownership, which were measured similarly.

For temporal comparability, this analysis uses 2021 and 2022 ACS microdata restricted to the seven core counties in TBI. These data include 37,487 individuals and 16,942 households. ACS had 85% response rates in both years.

4. ANALYSIS

In this section, we present the analytical methods used to examine representativeness errors for TBI data. Specifically, we first evaluate the components of representation error – bias and variance; and then turn to approaches that can be used to evaluate the real-world implications of representation error.

4.1 Evaluating bias

Survey errors have a bias and a variance component (Groves 1989). Bias occurs when a survey statistic systematically deviates from the true value. Differential nonresponse patterns and adjustment methods (e.g., weights) can both introduce bias. We now elaborate on how weights can facilitate the interpretation of observed biases.

Base weights facilitate *nonresponse bias* analysis. All observations from a probability sample have a base weight, which represents the unit's inverse probability of selection (to be contacted to participate in the survey). Thus, base weights encode differences in sampling rates from oversampling and can be used to account for how oversampling may distort the representativeness of unweighted data. When statistics prepared with base weights differ from the true value, that difference may be attributable to nonresponse bias, which can also be expressed as (Haziza and Lesage 2016):

$$NR\ Bias(\hat{Y}) = \left(\frac{n_{NR}}{n_{sample}}\right)(\bar{Y}_R - \bar{Y}_{NR}) \quad (1)$$

where $\hat{Y}$ is the value of the travel behavior statistic of interest estimated from the dataset, n_{NR} is the number of nonrespondents, n_{sample} is the number of all contacted households, and thus the first term in the left-hand side of the Equation 1 is the nonresponse rate. $\bar{Y}_R$ is the value among survey respondents and $\bar{Y}_{NR}$ is the value among survey nonrespondents. $\bar{Y}_R - \bar{Y}_{NR}$ is non-zero if respondents and nonrespondents differ in their travel behavior. Equation 1 implies that the magnitude of nonresponse bias does not necessarily scale with response rate because it also depends on how different respondents and nonrespondents are on the measure of interest.

Estimates from final weights confer insight about both nonresponse and *adjustment bias* (e.g., from weighting). Weighting may not resolve all nonresponse biases and can meanwhile introduce adjustment bias. While we can expect that variables used as weighting targets will be largely unbiased after weighting, variables not included as targets can still have bias. Thus, when estimates prepared using final weights deviate from the true value, this bias can now be from both or either nonresponse or adjustment.

4.1.1 Evaluating bias in this study

ACS is a common benchmark for other surveys' statistics because of its rigorous design and high response rate. As such, in this analysis, if a TBI estimate deviates from that of the ACS, it implies bias in TBI.

To assess transportation-related bias, we examine TBI's estimates of household vehicle ownership. Of all variables common to ACS and TBI (and thus, benchmarking candidates), household vehicle ownership is most relevant to other travel behavior outcomes such as trip generation, destination choice, vehicle miles traveled, and mode choices. Next, we outline the process developed to examine vehicle ownership biases.

We first divide the population into 14 segments based on combinations of SED characteristics. We chose to do this after initially examining vehicle ownership biases across groups segmented by individual SED characteristics such as income and race. This did not produce interpretable patterns due to uncontrolled differences in the distributions of other vehicle ownership predictors (for example, a segment of middle-income individuals is diverse with respect to race, home ownership, and other factors that influence vehicle ownership). We thus instead apply a data-driven (in this case, decision tree) approach to identify key combinations of segmentation variables that minimize variation in vehicle ownership. This yields segments of combined SED characteristics that are homogenous with regard to vehicle ownership. Doing so refines our ability to observe modeling-relevant and systematic biases in vehicle ownership (compared with individual SED segments). These segments are based on combinations of residence county, household size, income, whether residence in single-family home (SFH) or multi-unit dwelling (MUD), and presence of children (see Table 1).

Table 1: Household Vehicle Ownership Segments

		HH Size			**Dwelling**		**HH Income**		**Children**		
Counties	**Segment ID**	1	2	3+	Multi-unit	Single-family	<25k	>25k	Yes	No	**TBI n**
Hennepin, Ramsey	1	✓				✓	✓	✓	n/a	n/a	728
	2	✓			✓		✓		n/a	n/a	488
	3	✓			✓			✓	n/a	n/a	904
	4		✓			✓	✓	✓	✓	✓	1090
	5		✓		✓		✓	✓	✓	✓	709

	6			✓		✓	✓	✓	✓		565
	7			✓		✓	✓	✓		✓	214
	8			✓	✓		✓	✓	✓	✓	287
Anoka, Carver, Dakota, Scott, Washington	9	✓				✓	✓	✓	n/a	n/a	234
	10	✓			✓		✓	✓	n/a	n/a	348
	11		✓			✓	✓	✓	✓	✓	468
	12		✓		✓		✓	✓	✓	✓	158
	13			✓	✓	✓	✓	✓	✓		331
	14			✓	✓	✓	✓	✓		✓	92

For each household segment, we use t-tests at the 95% confidence level to compare TBI and ACS estimates of subgroup proportion. All weighted estimates are prepared using the *survey* library in R to account for the stratified sample design when estimating standard errors. We present the bias in units of percentage difference:

$$\%\Delta = \frac{\hat{P}_{TBI} - \hat{P}_{ACS}}{\hat{P}_{ACS}} * 100 \tag{3}$$

where $\%\Delta$ is the percentage difference between the TBI population and corresponding ACS estimate, $\hat{P}_{TBI}$ is the proportion estimated from TBI data, and $\hat{P}_{ACS}$ is the proportion estimated from ACS data. The percentage difference itself also has a standard error, which we approximate per Census Bureau's technical documentation:

$$SE(\%\Delta) = \left|\frac{\hat{P}_{ACS}}{\hat{P}_{TBI}}\right| \sqrt{\frac{SE(\hat{P}_{ACS})^2}{\hat{P}_{ACS}^2} + \frac{SE(\hat{P}_{TBI})^2}{\hat{P}_{TBI}^2}} * 100 \tag{4}$$

where $SE(\%\Delta)$ is the standard error of the estimated percentage difference between the TBI and ACS estimates, $SE(\hat{P}_{ACS})$ is the standard error of the proportion estimated from ACS data, and $SE(\hat{P}_{TBI})$ is the standard error of proportion estimated from TBI data.

We then use t-tests to evaluate vehicle ownership biases by household segment in TBI. Because both ACS and TBI topcoded household vehicle count, we treat all households with six vehicles or more (the least restrictive common topcode possible) as having six vehicles.[3]

4.2 Evaluating precision

As discussed, U.S. travel surveys often use disproportionately stratified sampling, often referred to as oversampling. This design can either worsen or improve precision, depending on how relatively prevalent the targeted population is within the oversampled strata (Kalton 2009). Survey statisticians quantify this precision loss, or increased sampling variance, using design effects. Like nonresponse bias, the design effect is a property of a statistic, not a dataset. It can be expressed as the ratio between the sampling variance of an estimator under the complex sample design (CSD) and its variance under a simple random sample (SRS):

[3] Topcoding is the practice of, in this case, representing all households with six or more vehicles as a single category of household vehicle ownership. Of the nearly 8,000 TBI households, 24 had six or more vehicles.

$$DEFF = \frac{Var_{CSD}(\bar{y})}{Var_{SRS}(\bar{y})} \quad (2)$$

Thus, design effects over 1 indicate that a particular sample design requires more observations to achieve the same precision as a simple random sample, lower the survey's cost efficiency. For instance, the design effect can be combined with known data collection costs to evaluate the trade-off between information gain and additional cost for every additional day of travel diary data collection (Erhardt and Rizzo 2018).

4.2.1 Evaluating precision in this study

Specifically, in this study, we use design effects associated with each household segment's estimated mean vehicle ownership to evaluate how the sample design and adjustment stages influence these statistics' precision. We compare design effects in two ways: first, between TBI and ACS and second, between household segments within TBI. ACS and TBI design effects are not directly comparable because they used different sample designs and weighting methods. We instead treat the ACS design effects as a reference point and make only qualitative comparisons. We also examine how TBI design effects vary by household segment composition.

4.3 Evaluating policy implications

To illustrate implications of bias for policy analysis, we examine how survey design and weights impact estimates of vehicle miles travelled (VMT). Using TBI's trip diaries, we estimate the average VMT per household within each household segment. Because we cannot benchmark VMT estimates against ACS or any other reliable survey data source, we instead evaluate the efficacy of the weighting approach on bias qualitatively, based on whether they adjust VMT in the expected direction. Namely, if a set of weights increases average vehicle ownership among a population segment, then we should expect average VMT to increase as well, and vice versa. We reason this because vehicle ownership is homogenous within each household segment and vehicle ownership is a primary determinant of VMT. This is also based on similarly designed bias evaluations in other domains. When lacking a benchmark, survey methodologists instead examine how the statistics of interest change between the application of base and final weights (Miller et al. 2020).

We then represent and evaluate the precision of the VMT estimates using the minimum detectable effect (MDE). We calculate the MDE of VMT for each household segment based on sample size and under a statistical power of 80%. Assuming the VMT estimates from TBI represent base year conditions, this MDE represents the minimum amount of VMT reduction that would be necessary to detect a statistical difference under an alternative scenario. For reference, Met Council's Imagine 2050 Transportation Policy Plan states a regional objective to reduce per capita VMT by 20% by 2050 (*Imagine 2050: Transportation Policy Plan* 2025, 20). Therefore, MDEs larger than 20% of VMT are prohibitive to nuanced comparison of VMT across different transportation investment scenarios.

5. RESULTS

5.1 Baseline nonresponse bias

Table 1 summarizes which SED groups in TBI exhibit statistically significant nonresponse bias and their relative magnitude. Because convenience sample respondents do not have a known probability of selection and thus do not have base weights, these estimates are prepared using only the probability sample respondents. Positive percent differences in population proportion indicate that segment is over-represented *because they were systematically more likely to respond than other household segments*, and positive differences in vehicle ownership indicate that vehicle ownership is over-estimated *because households with more vehicles were more likely to respond than households with fewer vehicles in that segment*. Had we used final weights instead of base weights, we would observe biases nonetheless, but we would not be able to attribute them to systematic nonresponse.

Smaller households were more likely to respond to TBI, with the exception of low-income households in multi-unit dwellings in Hennepin-Ramsey counties (which contain the central cities of Minneapolis and St. Paul). In Hennepin-Ramsey counties, 1-person households in single-family homes were most likely to respond while 3+ person households in multi-unit dwellings were least likely. In Anoka-Carver-Dakota-Scott-Washington counties ('other' counties in Table 2), 1-person households in single-family homes were again most likely to respond while 3+ person households were least likely.

Where there exists nonresponse bias, TBI tends to underestimate mean vehicle ownership levels (see Figure 2). TBI only *over*estimates vehicle ownership in one segment, 1-person households in Anoka-Carver-Dakota-Scott-Washington counties. Self-selection of households with fewer vehicles may also lead to underestimates of VMT and private vehicle-based trips. On the other hand, groups that were relatively more likely to respond tended not to display any vehicle ownership bias, even though they also had single-digit response rates. Thus, low response rates are not necessarily indicative of travel behavior-related nonresponse biases, which is consistent with previous studies (Groves and Peytcheva 2008).

Main takeaway 1: Households with relatively lower automobility were most likely to self-select into this travel survey. Yet, low response rates alone do not imply that nonresponse bias is present.

Table 2: Nonresponse biases in TBI (TBI estimates prepared with base weights)

		Population Proportion (%)			Mean Vehicle Ownership		
Region[1]	SED[2]	ACS	TBI	% Diff[3]	ACS	TBI	Diff[3]
H-R	1-person SFH	6.1 (0.2)	9.8 (0.4)	**62.3 (9.2)**	1.16 (0.03)	1.19 (0.03)	0.02 (0.04)
H-R	1-person, MUD, income <25k	4.8 (0.2)	4.0 (0.3)	**-15.7 (6.8)**	0.54 (0.03)	0.59 (0.04)	0.06 (0.05)
H-R	1-person, MUD, income >25k	9.6 (0.3)	13.6 (0.5)	**42.0 (7.1)**	0.92 (0.01)	0.96 (0.02)	0.04 (0.02)
H-R	2-person, SFH	12.2 (0.3)	17.8 (0.6)	**46.3 (5.9)**	1.96 (0.02)	1.95 (0.03)	-0.02 (0.03)
H-R	2-person, MUD	7.7 (0.3)	9.3 (0.4)	**21.3 (7.2)**	1.50 (0.03)	1.46 (0.04)	-0.04 (0.05)
H-R	3+ person, SFH, w/ kids	10.6 (0.3)	8.4 (0.4)	**-21.4 (4.5)**	2.27 (0.03)	1.97 (0.03)	**-0.30 (0.04)**

Region[1]	SED[2]	Population Proportion (%) ACS	TBI	% Diff[3]	Mean Vehicle Ownership ACS	TBI	Diff[3]
H-R	3+ person, SFH, w/o kids	3.9 (0.2)	2.8 (0.2)	**-27.9 (7.2)**	2.94 (0.05)	2.41 (0.09)	**-0.53 (0.11)**
H-R	3+ person, MUD	5.0 (0.3)	2.3 (0.2)	**-53.5 (4.7)**	1.77 (0.06)	1.58 (0.08)	**-0.19 (0.10)**
Other	1-person, SFH	3.7 (0.2)	4.6 (0.3)	**23.7 (10.5)**	1.34 (0.04)	1.39 (0.06)	0.05 (0.07)
Other	1-person, MUD	5.5 (0.2)	6.1 (0.4)	11.2 (8.0)	0.89 (0.02)	1.01 (0.02)	**0.12 (0.03)**
Other	2-person, SFH	10.0 (0.3)	10.0 (0.5)	-0.6 (5.4)	2.20 (0.02)	2.09 (0.03)	**-0.11 (0.04)**
Other	2-person, MUD	4.1 (0.2)	2.9 (0.2)	**-29.8 (7.0)**	1.65 (0.03)	1.70 (0.06)	0.05 (0.07)
Other	3+ person, w/ kids	13.1 (0.3)	6.5 (0.4)	**-50.3 (3.2)**	2.30 (0.03)	2.15 (0.04)	**-0.15 (0.05)**
Other	3+ person, w/o kids	3.8 (0.2)	1.9 (0.2)	**-50.2 (6.1)**	3.04 (0.05)	2.97 (0.12)	-0.07 (0.13)

[1] H-R: Hennepin & Ramsey Counties; Other: Anoka, Carver, Dakota, Scott, & Washington Counties

[2] SED characteristics in order of: household size, dwelling type, and presence of children; SFH: single-family home; MUD: multi-unit dwelling

[3] Differences between ACS and TBI bolded if statistically different from zero (α=0.05)

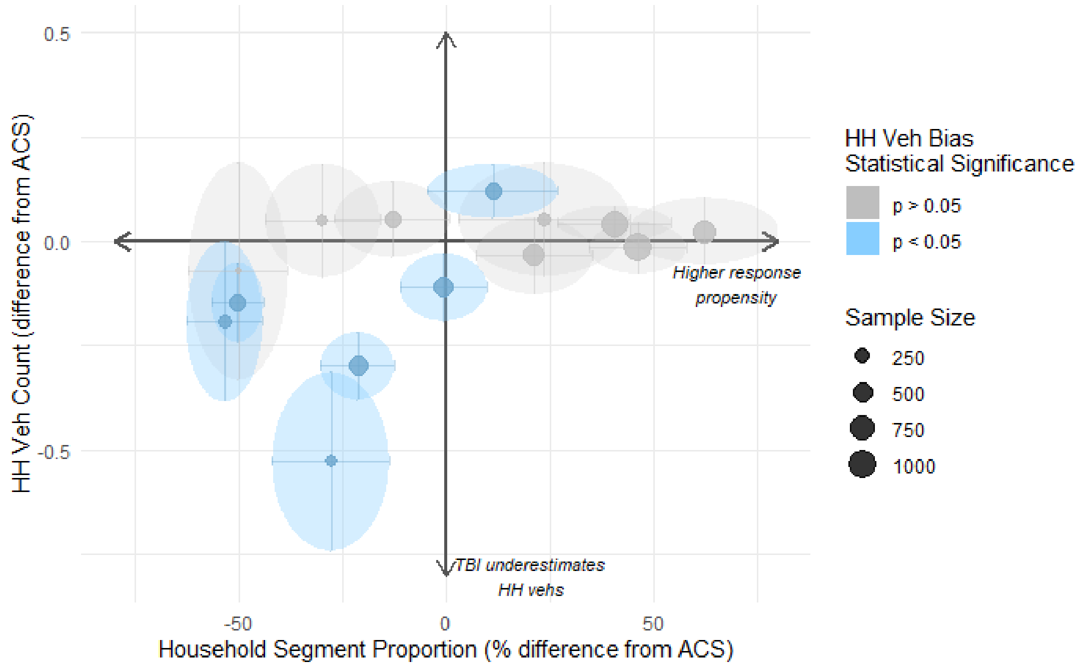

Figure 2: Vehicle ownership nonresponse bias vs. relative response propensity

We also evaluated SED-related nonresponse biases within each stratum (excised for brevity, full table available upon request). Figure 3 summarizes the types of households and persons with relatively low response propensity in each stratum.

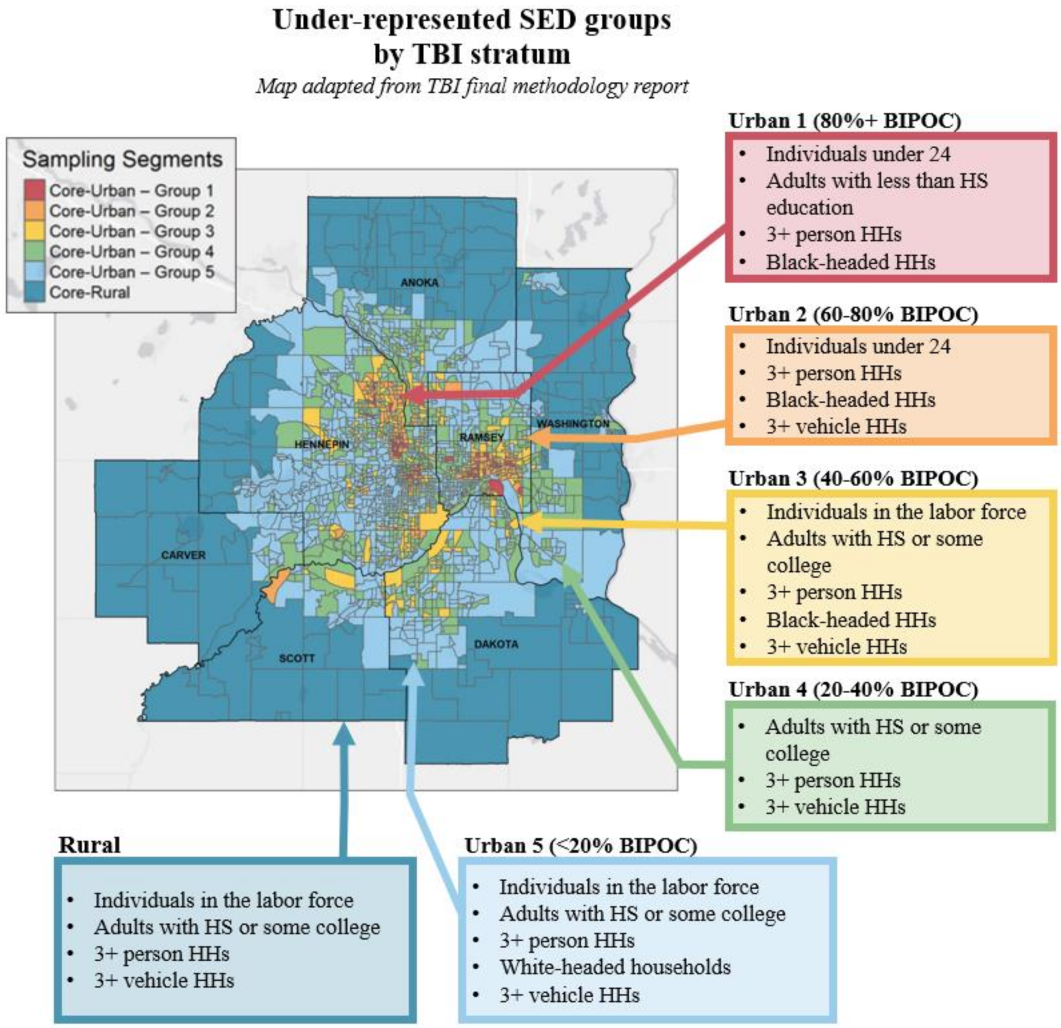

Figure 3: SED representativeness by stratum (sampling segments)

5.2 Impacts from convenience sample and weights

To prepare estimates adjusted to population totals and proportions, we develop calibrated weights. In so doing, we apply a consistent process to develop weights with and without inclusion of the convenience sample. While the final weights provided with the travel survey were developed using PopulationSim's entropy-maximizing list balancing algorithm, we use an iterative proportional updating algorithm implemented in R's *ipfr* library (Ward and Macfarlane, Greg 2025). Both are calibration methods designed to develop joint household-person weights for population synthesis. To further maintain consistency, we use the same method as in the Met Council data to develop initial expansion weights to initialize our weighting algorithm. We update the base weights such that the weighted sum of households in each stratum is consistent with the ACS population estimate for the same period. This is in effect a simple nonresponse adjustment that assumes every household in a stratum had the same response propensity, regardless of SED attributes.

Our weighting targets thus included household size, household income, number of workers, number of vehicles, presence of children, householder age, householder race and ethnicity, tenure (own vs. rent), and dwelling type. Even though the development of person weights is beyond the scope of this analysis, we also specify simultaneous person-level weight targets for age, gender, employment status, student status, educational attainment, race, and ethnicity to maintain consistency with how the official final weights were developed. Each of these targets are specified at a sub-region level. We apply these targets within each of 8 sub-regions (Anoka, Washington, Dakota, Scott & Carver counties, St. Paul in Ramsey County, the remainder of Ramsey county, Minneapolis in Hennepin County, and the remained of Hennepin county) to produce realistic weights that enable sub-region level analysis, as necessary in practice.

Compared with the original final weights, we also use the same weight inflation/deflation ratios (weights are floored at 0.125x and capped at 5.5x the initial input weight).

We next apply these calibrated weights to TBI data to re-estimate mean vehicle ownership by household segment and average daily VMT by household segment. Table 3 summarizes how estimates of household segment proportion and mean vehicle ownership differ between ACS and TBI, both with and without the convenience sample and corresponding weights. For brevity, we only report the differences between ACS and TBI for these. Again, positive differences indicate that TBI overestimates that particular value compared with ACS. We also report design effects associated with household vehicle ownership estimates from TBI data.

Table 3: Bias and Precision of Final Weighted Estimates

		With Convenience Sample			Without Convenience Sample		
Region[1]	SED[2]	SED % diff[3]	Veh Diff[3]	DEff[4]	SED % diff[3]	Veh Diff[3]	DEff
H-R	1-person HH, SFH	-8.6 (9.8)	0.08 (0.09)	7.14	-7.8 (10.5)	0.15 (0.09)	7.31
H-R	1-person HH, MUD, <$25k	**36.4** **(15.1)**	**-0.24** **(0.05)**	4.18	23.5 (14.9)	**-0.18** **(0.06)**	3.88
H-R	1-person HH, MUD, income >$25k	-13.4 (7.6)	0.01 (0.05)	5.25	-7.6 (8.1)	-0.01 (0.05)	5.14
H-R	2-person HH, SFH	0.6 (7.5)	**0.16** **(0.06)**	5.35	-0.7 (7.8)	**0.15** **(0.07)**	6.25
H-R	2-person HH, MUD	-0.6 (9.5)	-0.05 (0.09)	6.24	1.5 (10.2)	-0.07 (0.10)	6.04
H-R	3+ person HH, SFH, w/ children	12.3 (9.2)	**-0.18** **(0.06)**	3.12	13.3 (9.6)	**-0.13** **(0.06)**	3.05
H-R	3+ person HH, SFH, w/o children	-4.1 (14.1)	**-0.38** **(0.16)**	3.52	-0.7 (15.0)	**-0.42** **(0.15)**	3.06
H-R	3+ person HH, MUD	**-22.3** **(10.9)**	-0.13 (0.14)	4.67	**-27.1** **(11.1)**	-0.14 (0.13)	3.45
Other	1-person HH, SFH	-4.1 (16.5)	0.07 (0.15)	6.2	-4.2 (17.2)	0.03 (0.15)	6.66
Other	1-person HH, MUD	2.2 (14.5)	0.05 (0.07)	5.6	2.3 (14.7)	0.02 (0.06)	6.77
Other	2-person HH, SFH	-2.0 (10.2)	0.08 (0.08)	4.57	-4.9 (10.3)	0.06 (0.09)	5.21
Other	2-person HH, MUD	4.8 (17.7)	0.17 (0.12)	4.16	12.3 (19.0)	0.19 (0.12)	3.95
Other	3+ person HH, w/ children	1.4 (10.0)	**-0.15** **(0.07)**	2.68	1.2 (10.1)	-0.13 (0.07)	2.58
Other	3+ person HH, w/o children	-4.6 (17.9)	-0.13 (0.22)	3.1	-3.6 (18.1)	-0.01 (0.21)	2.77

		With Convenience Sample			Without Convenience Sample		
Region[1]	SED[2]	SED % diff[3]	Veh Diff[3]	DEff[4]	SED % diff[3]	Veh Diff[3]	DEff

[1] H-R: Hennepin & Ramsey Counties; Other: Anoka, Carver, Dakota, Scott, & Washington Counties
[2] SED characteristics in order of: household size, dwelling type, and presence of children; SFH: single-family home; MUD: multi-unit dwelling
[3] Differences between ACS and TBI bolded if statistically different from zero (α=0.05)
[4] DEff: Design Effect

Under the same weighting method, the dataset with the convenience sample has slightly more bias in SED and vehicle ownership estimates compared with the dataset without the convenience sample. Regardless, the weighting method undertaken appears more effective at resolving household SED biases than travel behavior biases within household SED segments.

Further, 9 of 14 household segments have larger design effects associated with vehicle ownership estimates with the convenience sample relative to the sample without. For reference, corresponding design effects using ACS data range from 1.4 to 1.8 (not depicted). Thus, differences in design effects between the datasets with and without the convenience sample are practically negligible compared to the impacts from oversampling and weighting.

Main takeaway 2: Although the convenience sample increased the *unweighted* representation of low-income or otherwise socially disadvantaged populations in the travel survey, it had no or minor negative consequences for the bias and precision of *weighted* SED and vehicle ownership estimates.

We next detail the potential influence of oversampling and weighting on design effects. These are illustrated in Figure 4 and Figure 5. In these figures, the highly oversampled stratum included block groups in which 80% or more of residents were Hispanic or non-white. The sampling rate was approximately 100% in this stratum, about 10 times higher than that of the stratum sampled at the lowest rate.

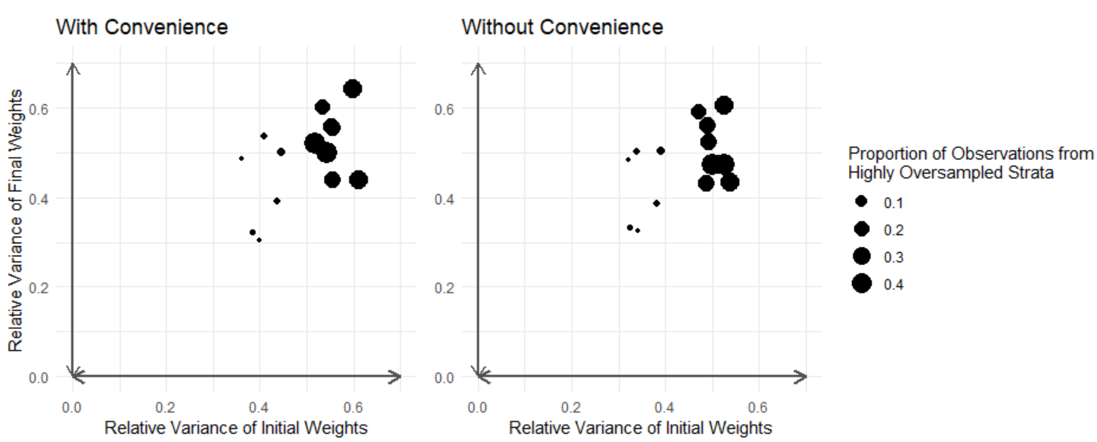

Figure 4: Final weight variance vs. initial weight variance by strata composition

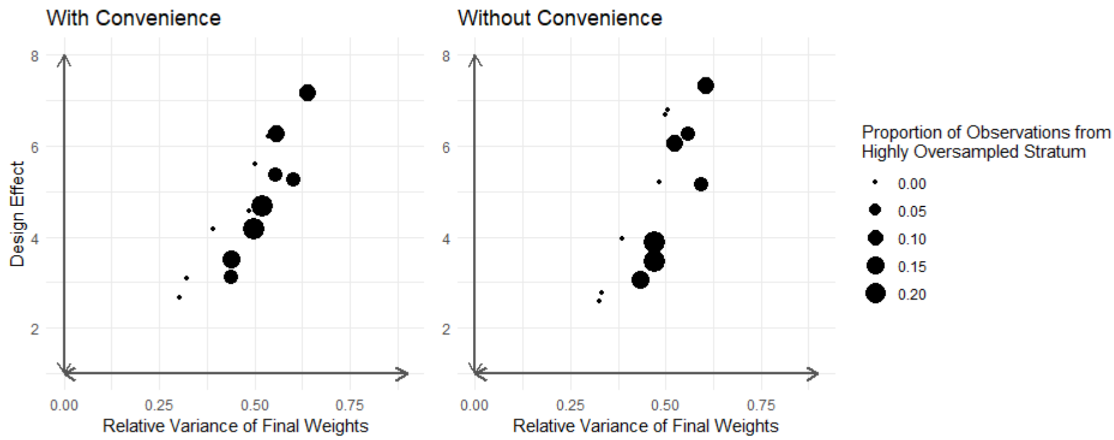

Figure 5: Vehicle ownership design effect vs. final weight variance by strata composition

Oversampling worsens final estimate precision by way of increased initial and final weights. First, disproportionate stratification (oversampling) increases the variance in base weights, and in turn, initial expansion weights (Figure 4). Then, higher relative variance in initial expansion weights provided to the weighting algorithm increases the relative variance of final weights, regardless of weight expansion minimums and caps (Figure 4). Finally, higher relative variance in final weights increases design effects and worsens the precision of vehicle ownership survey estimates (Figure 5).

Main takeaway 3: Extreme oversampling (or large disparities in selection probabilities) worsens the precision of vehicle ownership estimates (and likely other travel behaviors). The consequence is slightly more severe among household segments prevalent in oversampled neighborhoods.

5.3 Consequences for VMT

Sample design, weighting, and the convenience sample also have consequences for the bias and precision of downstream analyses – shown in this paper with VMT estimates. Table 4 presents VMT estimates by household segment, with and without the convenience sample, by both base weights and the corresponding final weights (Base VMT, Final VMT). It also presents how the final weights and convenience sampling increased or decreased vehicle ownership means (ΔVeh) and VMT (ΔVMT) estimates. The final column of each presents the minimum detectable effect (MDE) of VMT, the latter represented as the percent of VMT. The shaded green rows indicate those with statistically significant vehicle ownership non-response bias identified in Table **1**: Household Vehicle Ownership Segments

Table 4: VMT estimates after weighting

			With Convenience				Without Convenience			
Region[1]	SED[2]	Base VMT	Final VMT	ΔVeh	ΔVMT	MDE	Final VMT	ΔVeh	ΔVMT	MDE
H-R	1-person HH, SFH	20.7 (1.0)	18.5 (2.0)	0.06	-2.2	-27	19.5 (2.2)	0.12	-1.2	-27

Region[1]	SED[2]	Base VMT	With Convenience: Final VMT	ΔVeh	ΔVMT	MDE	Without Convenience: Final VMT	ΔVeh	ΔVMT	MDE
H-R	1-person HH, MUD, <$25k	14.4 (1.3)	8.1 (1.9)	-0.30	-6.3	-59	9.4 (2.0)	-0.23	-5.0	-52
H-R	1-person HH, MUD, >$25k	24.7 (1.3)	24.1 (2.7)	-0.03	-0.6	-28	24.2 (2.9)	-0.05	-0.5	-30
H-R	2-person HH, SFH	28.7 (1.6)	29.5 (2.3)	0.17	0.9	-19	30.8 (2.4)	0.17	2.1	-20
H-R	2-person HH, MUD	24.4 (1.5)	22.7 (2.1)	-0.02	-1.8	-23	24.2 (2.4)	-0.03	-0.2	-25
H-R	3+ person HH , SFH, w/ kids	34.0 (2.0)	33.3 (2.6)	0.12	-0.6	-20	35.1 (2.9)	0.16	1.2	-20
H-R	3+ person HH , SFH, w/o kids	30.3 (2.5)	34.0 (4.2)	0.15	3.7	-31	34.7 (4.7)	0.10	4.4	-34
H-R	3+ person HH , MUD	27.5 (3.3)	30.4 (7.2)	0.06	2.8	-61	28.6 (5.2)	0.06	1.1	-40
Other	1-person HH, SFH	29.2 (2.2)	25.2 (3.7)	0.02	-4.0	-36	27.5 (3.8)	-0.03	-1.7	-34
Other	1-person HH, MUD	27.1 (1.7)	23.7 (3.7)	-0.07	-3.4	-39	28.4 (6.3)	-0.10	1.2	-56
Other	2-person HH, SFH	37.7 (2.2)	39.3 (5.2)	0.19	1.6	-33	36.4 (3.1)	0.17	-1.4	-22
Other	2-person HH, MUD	34.3 (4.0)	33.2 (4.6)	0.12	-1.1	-34	34.1 (4.7)	0.14	-0.2	-34
Other	3+ person HH , w/ kids	42.3 (2.2)	43.5 (4.1)	0.00	1.2	-24	45.1 (4.2)	0.02	2.8	-24
Other	3+ person HH , w/o kids	43.6 (4.9)	45.3 (7.9)	-0.05	1.7	-40	47.8 (8.0)	0.06	4.2	-39

[1] H-R: Hennepin & Ramsey Counties; Other: Anoka, Carver, Dakota, Scott, & Washington Counties
[2] SED characteristics in order of: household size, dwelling type, and presence of children; SFH: single-family home; MUD: multi-unit dwelling

With two exceptions, the convenience sample decreased VMT estimates. This may be because it recruited individuals from the local transit provider's transit assistance program and community-based organizations that serve socially disadvantaged groups. The weights, regardless of the convenience sample, deflated vehicle ownership and VMT the most among 1-person, low-income, multi-unit dwelling households. This occurred despite a lack of vehicle ownership nonresponse bias among this segment to begin with Table 1.

There are also some unintuitive impacts to VMT among household segments that had vehicle ownership nonresponse bias (shaded green in Table 4). We expect that weights should affect mean vehicle ownership and VMT in the same direction (see ΔVeh and ΔVMT columns).

The dataset without the convenience sample shifts VMT intuitively among all 3+ person household segments. Yet, by increasing vehicle ownership while decreasing VMT, the weighted dataset with the convenience sample produces an unintuitive result among households with 3+ persons with kids in single family homes in Hennepin-Ramsey counties. On the other hand, the dataset with the convenience sample affects VMT in the expected direction while the dataset without the convenience sample does not among 1- and 2-person households with baseline nonresponse bias. Consistent with the previously described vehicle ownership adjustment biases, these mixed implications for VMT bias underscore how calibrated weights can include multiple SED margins but produce unintuitive results at their intersections.

Main takeaway 4: Calibration weighting (based on marginal SED totals) may have had unintended and unintuitive impacts on travel behavior estimates because they may not reflect *interactions between* SED attributes that influence travel behavior.

Moreover, the convenience sample likely had mixed consequences for VMT bias due to somewhat random interactions between its own composition of biases, baseline nonresponse biases in the probability sample, and the weight calibration process.

Finally, both datasets with and without the convenience sample estimate within-segment vehicle ownership and VMT with moderate to poor precision. To contextualize these VMT precision losses, consider that Met Council has a goal of reducing per capita VMT by at least 20% by 2050. Hypothetically, a projected future scenario would need to reduce VMT by 19 to 61% to detect a statistical difference from the base year VMT. Thus, extreme weights in combination with extremely disproportionate stratification may threaten the strength of inferences drawn from policy-relevant travel behavior analyses.

6. CONCLUSIONS AND POLICY IMPLICATIONS

In this paper, we use the Metropolitan Council's 2021-2022 Travel Behavior Inventory to examine measures taken to increase dataset representativeness. We find that there exists moderate self-selection or nonresponse bias on vehicle ownership, particularly among larger households. While we see that calibration weights address biases among target SED variables handily, they less consistently resolve vehicle ownership biases and potential VMT biases among household segments that are not explicitly used as weighting targets. We also observe precision losses attributable to weighting that are significant enough to undermine policy analyses based on travel behavior outcomes.

We next translate the four main takeaways highlighted in the results section into practical implications.

6.1 Tentative implications for travel demand modeling

The first implication concerns bias. Travel demand models increasingly use households as units of analysis due to the shift from aggregate trip-based models to disaggregate activity-based ones. This has increased the number of SED dimensions for which representativeness is feasible and desirable. As the number of marginal SED targets used to calibrate weights increase, adjustment biases may arise at the *intersections* of various SED characteristics. Such biases may then be encoded in travel demand models, which also typically treat SED characteristics as independent covariates. Other weighting approaches can explicitly mitigate these biases using joint

SED targets, but at the cost of the number of SED variables that can be included.

The second implication concerns precision. We find that joint household-person weight calibration magnified high variation in weights due to disproportionate stratification (oversampling), resulting in significant precision losses for final vehicle ownership estimates. Extreme weights will also worsen the precision of coefficients in regression-based travel demand models. This could affect model development by causing important covariates to become statistically insignificant. Accordingly, we elaborate on sample design remedies in our discussion of data collection implications, to which we now turn.

6.2 Tentative implications for travel survey data collection

First, response rates, subgroup sample sizes, or unweighted SED distributions do not consistently indicate travel ***behavior*** representativeness. Data collectors need not immediately respond to low response rates in a stratum by increasing sampling (invitation) rates. They should first look for indications of travel behavior bias in areas or subgroups with low response rates, as we use base weights to illustrate. Otherwise, they may take on unnecessary expenses that may not reduce relevant biases, only to oversample more heavily and face avoidable precision losses.

Relatedly, we should take into account downstream consequences for weighting and precision when considering a disproportionate stratified sample design. The stratification used to collect the present survey data was based on race; that it performs poorly with respect to the precision of travel behavior estimates merely indicates that race itself does not segment certain travel behaviors well in Met Council's region. Modelers and data collectors should together determine which key measures should inform any stratification.

Should these key measures be travel behaviors, we offer several sampling considerations. First, in auto-oriented places, disproportionate stratification (oversampling) may not be warranted over proportionate stratification if travel behaviors of interest are spatially dispersed. There exist methods to determine optimal oversampling rates (including whether it is necessary) based primarily on the prevalence of a targeted population in each stratum, the proportion of that population in each stratum, and relative recruitment costs (Kalton 2009). Regardless, strata should be based on the distribution of travel behavior indicators and determinants across the study region. This has limits because ACS is the most common data source and it measures few travel behaviors. Thus, we should explore whether measures of urban form and destination accessibility, each with extensively evidenced links to travel behavior, can serve as efficient stratification variables. Our review includes examples to this effect, and our analysis thus contributes an empirical reference point for MPOs as they weigh various reasonable approaches to oversampling.

MPOs should still consider using oversampling to deliberately reach certain SED groups. By doing so in 2021 and 2022, Met Council provided disproportionate opportunities to participate in a key civic process among those who are most likely to experience social exclusion. Considering modelers sometimes use expert judgment to modify parameters estimated from travel survey data anyways, other MPOs may also decide deliberate SED representation is worth the likely modest differences in demand forecasts caused by sample design and adjustment choices.

Our third data collection implication concerns convenience sampling. We find no basis to either support or oppose this emerging practice. However, there are valuable reasons to collect convenience samples beyond generating more data for modeling, such as if an MPO is also using the survey to gauge public opinion on policy issues. MPOs with interest and means should continue to collect convenience samples from any practicable source. Then, we should continue to systematically evaluate their impacts on data quality as done in this study.

6.3 Limitations and ongoing work

This analysis is ongoing, and next we plan to examine how other forms of weighting may yield similar or different findings. For example, post-stratification followed by calibration, rather than calibration largely in isolation, is an equally valid and feasible weighting approach. Accordingly, we are next evaluating household-only (as opposed to joint household-person) weights derived from post-stratification and calibration. These results will be compared with the joint calibration weights used in this paper.

Future work should advance upon the specificity and depth of the preceding practical implications. We acknowledge that practitioners seek more concrete guidelines for sample design and weighting. Doing so will require more information about survey costs, but currently few MPOs possess cost breakdowns with enough detail to support analyses that consider both the financial and data quality costs and benefits of various survey design choices (Khoury et al. 2025). Future research should also be sensitive to the diverse agency priorities across the U.S. Thus, future research could build on other recent, interview-based efforts to take stock of the existing travel survey landscape (Selby et al. 2024). Future interviews with travel survey data program managers could delve into how institutional factors such as policy and planning emphases, modeling needs, political considerations, risk tolerance, and time, budget, and human resources limitations necessarily constrain survey design. Advancement in these avenues would enable future work to weigh design trade-offs quantitatively (as discussed in our comments about optimal oversampling rates), and to further ground consequent survey design recommendations in practical realities.

7. REFERENCES

2017 MAG Household Travel Survey Report. 2018. Final Report. Maricopa Association of Governments.

2017/2018 Regional Travel Survey. 2021. Technical Documentation. National Capital Region Transportation Planning Board.

2017-2022 PSRC Household Travel Survey Program: Lessons and Recommendations. 2022. Supplemental Report. Puget Sound Regional Council (PSRC). https://www.psrc.org/media/6321.

Anderson, Rebekah Straub, Gregory Giaimo, Elizabeth Greene, Chloe Ritter, Leah Flake, and Megan Kelly. 2018. "Lessons from the Smartphone-Based Household Travel Survey in Ohio." Preprint, August.

Armoogum, Jimmy, Rico Wittwer, Jean-Loup Madre, and Yeshtabye Gopal. 2024. "Sampling, Nonresponse, and Quality in Travel Surveys." *Transportation Research Procedia* 76: 638–43. https://doi.org/10.1016/j.trpro.2023.12.085.

Battaglia, Michael P., David C. Hoaglin, and Martin R. Frankel. 2009. "Practical Considerations in Raking Survey Data." *Survey Practice* 2 (5): 1–10. https://doi.org/10.29115/SP-2009-0019.

Bradley, Mark, Asa Bergman, Michelle Lee, Elizabeth Greene, and Suzanne Childress. 2015. "Predicting and Applying Differential Response Rates in Address-Based Sampling for a Household Travel Survey." *Transportation Research Record: Journal of the Transportation Research Board* 2526 (1): 119–25. https://doi.org/10.3141/2526-13.

Brick, J. Michael, and Roger Tourangeau. 2017. "Responsive Survey Designs for Reducing Nonresponse Bias." *Journal of Official Statistics* 33 (3): 735–52. https://doi.org/10.1515/jos-2017-0034.

Bricka, Stacey, Ben Ettelman, and Chris Simek. 2013. "2012 Household Travel Survey Symposium: Conference Summary and Final Report." Southwest Region University Transportation Center, December. https://rosap.ntl.bts.gov/view/dot/27382.

Chestnut, Joe, Ilona Craigie, and Claire Goldhammer. 2025. "Household Travel Survey Response Rate Modeling in the Post-Pandemic Era." Poster. Annual Meeting of the Transportation Research Board, Washington DC, January 7.

Coy, Christopher, Mark Bradley, and Elizabeth Greene. 2019. "Innovations in Address-Based Sampling (ABS) for Travel Surveys." Paper presented at TRB Planning Applications Conference, Portland, Oregon. June 4.

Erhardt, Gregory D., and Louis Rizzo. 2018. "Evaluating the Biases and Sample Size Implications of Multi-Day GPS-Enabled Household Travel Surveys." *Transportation Research Procedia* 32: 279–90. https://doi.org/10.1016/j.trpro.2018.10.051.

Federal Highway Administration. 2023. *Travel Survey State of the Practice*. https://nhts.ornl.gov/assets/NextGen%20NHTS_State%20of%20Practice_032423.pdf.

Groves, R. M., and E. Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis." *Public Opinion Quarterly* 72 (2): 167–89. https://doi.org/10.1093/poq/nfn011.

Groves, Robert M. 1989. *Survey Errors and Survey Costs*. 1st ed. Wiley Series in Probability and Statistics. Wiley. https://doi.org/10.1002/0471725277.

Haziza, David, and Éric Lesage. 2016. "A Discussion of Weighting Procedures for Unit Nonresponse." *Journal of Official Statistics* 32 (1): 129–45. https://doi.org/10.1515/jos-2016-0006.

Hubrich, Stefan, Rico Wittwer, and Regine Gerike. 2018. "Quality Indicator Set for Household Travel Surveys." *Transportation Research Procedia* 33: 219–26. https://doi.org/10.1016/j.trpro.2018.10.098.

Imagine 2050: Transportation Policy Plan. 2025. Metropolitan Council. https://metrocouncil.org/Planning/Imagine-2050/Policy-Plan-Index/Imagine-2050-Accessible-PDFs/Transportation-Policy-Plan/Overview.aspx.

Joh, Kenneth, and Nicole McCall. 2023. "MPO Household Travel Survey State of the Practice 2022." National Capital Region Transportation Planning Board, February 22.

Kalton, Graham. 2009. "Methods for Oversampling Rare Subpopulations in Social Surveys." *Survey Methodology* 35 (2). https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2009002/article/11036-eng.pdf?st=lXaB_eIB.

Khoury, Issam, Deborah McAvoy, and Gregory D. Erhardt. 2025. "Synthesis Study of Costs and Trip Rates of Recent Household Travel Surveys." Ohio Department of Transportation, Office of Statewide Planning & Research, February 28.

Korimilli, Madhuri S., Ram M. Pendyala, and Elaine Murakami. 1998. "Metaanalysis of Travel Survey Methods." *Transportation Research Record: Journal of the Transportation Research Board* 1625 (1): 72–78. https://doi.org/10.3141/1625-09.

Lo, Albert, Siva Srikukenthiran, Eric J. Miller, and Khandker N. Habib. 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): 553–70. https://doi.org/10.1080/03081060.2020.1780707.

Mark Bradley, and Christopher Coy. n.d. *Exploring the SANDAG Survey Data on TNC Trips for Effective Geographic Oversampling Methods*. Memo. RSG.

Meister, Adrian, Nail F. Bashan, Rounaq Basu, Xianglu Shen, Ryan Q. Wang, and Andres Sevtsuk. 2025. "The BostonWalks Study: A Longitudinal Travel Survey Using Smartphone Tracking." *Transportation*, ahead of print, June 28. https://doi.org/10.1007/s11116-025-

10637-2.

Miller, Peter, Tala H. Fakhouri, Morgan Earp, et al. 2020. *A Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys*. FCSM 20-02. Federal Committee on Statistical Methodology. https://www.fcsm.gov/assets/files/docs/A_Systematic_Review_of_Nonresponse_Bias_Studies_Federally_Sponsored_SurveysFCSM_20_02_032920.pdf.

NHTS NextGen Study. 2022. 2022 NHTS Address-Based Sample Weighting Plan. U.S. Department of Transportation. https://nhts.ornl.gov/assets/2022/doc/2022%20NextGen%20NHTS%20Weighting%20Memo.pdf.

Ochsner, Michael. 2021. *Representativeness of Surveys and Its Analysis*. https://doi.org/10.24449/FG-2021-00015.

Peytcheva, Emilia, and Robert M Groves. 2009. "Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates." *Journal of Official Statistics*. https://www.rti.org/publication/using-variation-response-rates-demographic-subgroups-evidence-nonresponse-bias-survey-estimates.

San Diego Regional Transportation Study. 2017. Technical Report. San Diego Association of Governments.

Selby, Brent, Rachel Copperman, Kathryn DeBoer, et al. 2018. "Use of Convenience Sampling: Benefits and Challenges from the Phoenix, Arizona, Area Household Travel Survey." Paper presented at Transportation Research Board 97th Annual Meeting, Washington DC, United States. https://trid.trb.org/view/1496473.

Selby, Brent, Kate Dannemiller, and Rachel Copperman. 2024. "Memorandum of Task 1: Current Practices and Trends of HTS." Delaware Valley Regional Planning Commission as delivered under contract by Cambridge Systematics, August 3.

Son, Sanghoon, Asad Khattak, Xin Wang, Paul Agnello, and Ju-Yin Chen. 2013. "Quantifying Key Errors in Household Travel Surveys: Comparison of Random-Digit-Dial Survey and Address-Based Survey." *Transportation Research Record: Journal of the Transportation Research Board* 2354 (1): 9–18. https://doi.org/10.3141/2354-02.

Tourangeau, Roger. 2014. "Defining Hard-to-Survey Populations." In *Hard-to-Survey Populations*, 1st ed., edited by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates. Cambridge University Press. https://doi.org/10.1017/CBO9781139381635.003.

Ward, Kyle, and Macfarlane, Greg. 2025. *Ipfr: List Balancing for Reweighting and Population Synthesis*. V. 1.0.2. Released. R. https://cran.rediris.es/web/packages/ipfr/ipfr.pdf.

Wittwer, Rico, Stefan Hubrich, and Regine Gerike. 2024. "New Evidence on Nonresponse in Household Travel Surveys." *Transportation Research Procedia* 76: 233–45.

https://doi.org/10.1016/j.trpro.2023.12.051.