USDOT NATIONAL CENTER FOR UNDERSTANDING
TRAVEL BEHAVIOR AND DEMAND

Evaluating Strategies for Improving Travel Survey Representativeness

Amy Z. Fong and Atiyya Shaw
University of Michigan – Ann Arbor

Modeling Mobility Conference
September 2025

Acknowledgements

Met Council: Jonathan Erlich and Brandon Whited

RSG: Joel Freedman, Abigail Rosenson, and Ashley Asmus

University of Michigan: Ivan Shih, Joe Grengs, and Sunghee Lee

Funding from: the USDOT National Center for Understanding Travel Behavior and Demand, the U-Mich Graham Sustainability Institute, and the National Science Foundation

Travel survey recruitment is difficult

- Low **response rates** lead to
  - Uncertainty during survey sample design and data collection
  - Increased cost per recruited household
  - Potential for nonresponse bias
- Low response rate implies high **unit nonresponse**
- Unit nonresponse causes **nonresponse bias** when it coincides with **self-selection** along important variables

$$Nonresponse\ bias(\bar{X}) = (Nonresponse\ rate) * (\bar{X}_{respondents} - \bar{X}_{nonrespondents})$$

Strategies to improve representation

Design Feature	Mechanism	Challenges/Trade-offs
Disproportionate stratification (oversampling)	Oversample rare or 'hard-to-reach' groups	May increase weight variation -> possible precision loss
Nonprobability sampling	Add 'inexpensive' observations	Unknown statistical properties and response mechanism; self-selection
Weighting	Force SED representativeness	Increased constraints -> extreme weights (even with trimming) -> possible precision loss
Incentives	Increase motivation to respond	Cost; Unclear impacts on self-selection
Response mode flexibility	Prevent drop-outs from smartphone requirements	Confounds mode effects and self-selection

Joh & McCall, 2023; Federal Highway Administration, 2023

Motivating questions

How do travel survey representation strategies affect the quality of travel behavior estimation?

and

How and when can we answer this question in the face of practical data limitations?

(w/o experiments or access to data from numerous surveys)

Evaluation barriers

It's difficult to understand how travel survey design impacts representativeness because…

- Lack of resources and time to conduct experiments
- Incomparable survey designs across time and regions
- Common representativeness measures (e.g., response rate, sociodemographic distributions) are limited indicators of travel behavior nonresponse bias

Study objectives

1. Develop a survey data evaluation framework that is:
    - Applicable to probability travel surveys
    - Low cost (few add’l data collection requirements)
    - Germane to travel behavior inference
    - Based on the theoretical properties of survey statistics

2. Provide actionable insights about survey design features like…
    - Targeted oversampling
    - Convenience sampling
    - Weighting methods

3. Use these insights to propose efficiencies in sample design and weighting (ongoing work, not included in today’s talk)

Preliminary Results

Evaluation

- Response rates and sociodemographic representativeness may not always *indicate nonresponse biases* on travel behavior
- Data owners can *enable direct nonresponse bias analysis* by requesting key information about the sample design (i.e., selection probabilities, strata geographies) from data collectors

Insights (in order of specificity)

- *Trade-offs between civic participation and data collection efficiency*: sample design (geographic oversampling) can increase raw observations of targeted groups but may also lower precision of travel behavior estimates
- *Limiting data to one set of analysis weights (compared with unique household & person weights)* eases data analysis but may also worsen statistical precision
- *Calibration weighting to SED marginals* can inadvertently introduce new biases because it may not account for interaction effects that explain response propensity

Objective 1: Evaluation framework

1. **Identify** statistics and traveler subgroups with policy and/or modeling relevance

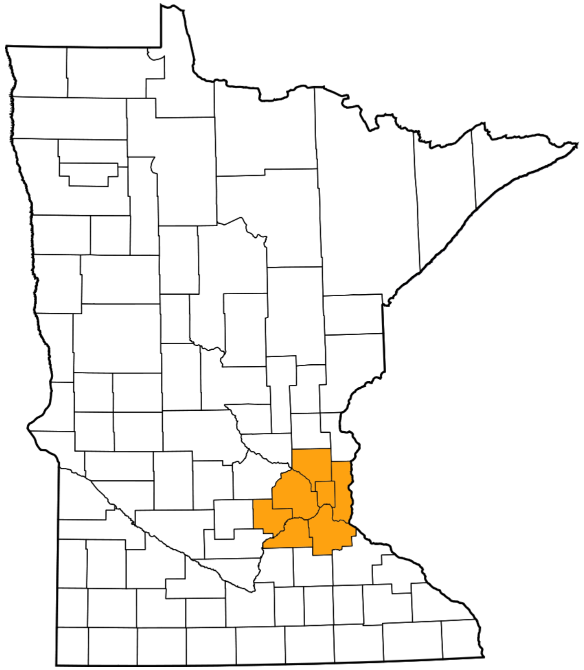
2. **Evaluate** data quality (bias and precision of travel behavior estimates) *without* survey design impacts (baseline)

3. **Compare** with the bias and precision of final estimates, which reflect impacts from survey design

Objective 2: Application to Met Council data

2021-2022 Travel Behavior Inventory (TBI):

- **Sampling unit**: households
- **Sample design**: disproportionate stratified sampling and convenience sampling
  - Oversampled high-BIPOC block groups
  - Recruited from transit users list and community orgs
- **Weights**: calibrated on American Community Survey (ACS)
  - *Household level targets*: household size, income, workers, vehicles, age of head of household, presence of children, total households
  - *Person level targets*: gender, age, worker status, university student status, educational attainment, race, ethnicity, total persons



Applying the framework to TBI

1. **Identify statistics and traveler subgroups with policy and/or modeling relevance**

2. Evaluate data quality without survey design impacts (baseline)

3. Compare with the bias and precision of final estimates, which reflect impacts from survey design and final weights

Which statistics and which subgroups?

- Proposed statistics:
  - Modeling-relevant: **mean household vehicle count** (HH veh)
    - Can be benchmarked directly using American Community Survey
    - Predicting HH veh well is critical because it is subsequently used to predict numerous other activity and travel choices in activity-based models
  - Policy-relevant: **mean vehicle miles traveled per household** (VMT)
    - Met Council aims to reduce VMT per capita by 20% by 2050
    - Predicting VMT well nuances evaluation of prospective infrastructure investments
- Proposed subgroups need two qualities:

$$Nonresponse\ bias(\bar{X}) = (Nonresponse\ rate) * (\bar{X}_{respondents} - \bar{X}_{nonrespondents})$$

1. Nonresponse-relevant: segmented by varying levels of nonresponse
2. Travel behavior-relevant: homogenous within, and heterogenous across, segments

Traveler subgroups and their measures of interest

Region	SED	Mean HH Vehs	Mean HH VMT
Urban	1-person household in SFH	1.2 (0.03)	18.5 (2.0)
Urban	1-person, MUD, income <25k	0.5 (0.03)	8.1 (1.9)
Urban	1-person, MUD, income >25k	0.9 (0.01)	24.1 (2.7)
Urban	2-person, SFH	2.0 (0.02)	29.5 (2.3)
Urban	2-person, MUD	1.5 (0.03)	22.7 (2.1)
Urban	3+ person, SFH, w/ kids	2.3 (0.03)	33.3 (2.6)
Urban	3+ person, SFH, w/o kids	3.0 (0.05)	34.0 (4.2)
Urban	3+ person, MUD	1.8 (0.06)	30.4 (7.2)
Other	1-person, SFH	1.3 (0.04)	25.2 (3.7)
Other	1-person, MUD	0.9 (0.02)	23.7 (3.7)
Other	2-person, SFH	2.2 (0.02)	39.3 (5.2)
Other	2-person, MUD	1.7 (0.03)	33.2 (4.6)
Other	3+ person, w/ kids	2.3 (0.03)	43.5 (4.1)
Other	3+ person, w/o kids	3.0 (0.05)	45.3 (7.9)

Urban: Hennepin & Ramsey Counties; **Other:** Anoka, Carver, Dakota, Scott, & Washington Counties
SFH: Single family home; **MUD**: multi-unit dwelling

Applying the framework to TBI

1. Identify statistics with policy and/or modeling relevance

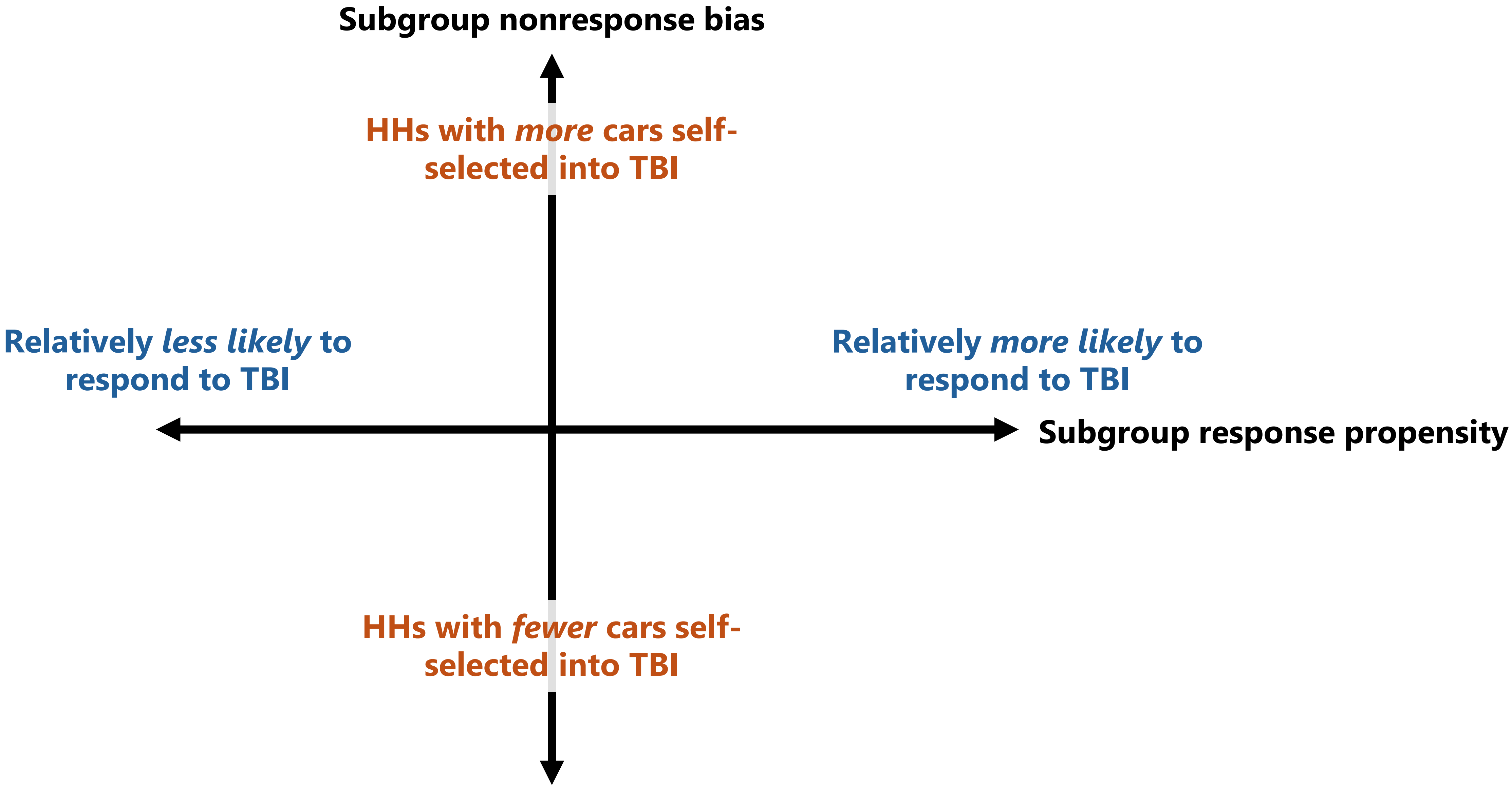
2. **Evaluate data quality without survey design impacts (baseline)**
   - <u>nonresponse bias</u> on SED, vehicle ownership
   - <u>margins of error (precision)</u> of vehicle ownership, VMT

3. Compare with the bias and precision of final estimates, which reflect impacts from survey design and final weights

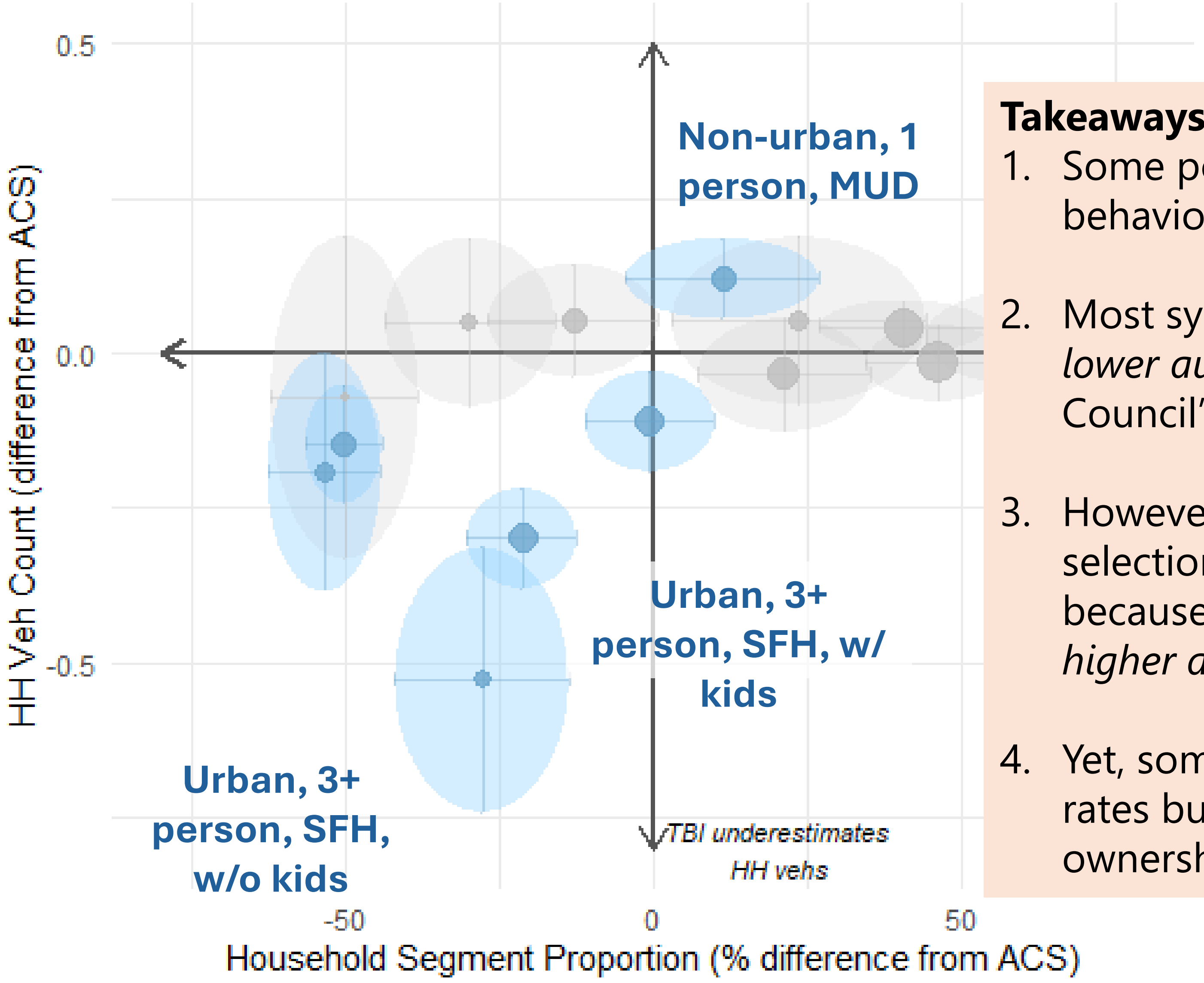
Nonresponse bias among traveler subgroups

Region	Traveler Segment	SED (Who was more/less likely to respond?)	Vehicle Ownership (Who self-selected by automobility?)
Urban	1-person household in SFH	+	
Urban	1-person, MUD, income <25k	-	
Urban	1-person, MUD, income >25k	+	
Urban	2-person, SFH	+	
Urban	2-person, MUD	+	
Urban	3+ person, SFH, w/ kids	-	-
Urban	3+ person, SFH, w/o kids	-	-
Urban	3+ person, MUD	-	-
Other	1-person, SFH	+	
Other	1-person, MUD		+
Other	2-person, SFH		-
Other	2-person, MUD	-	
Other	3+ person, w/ kids	+	-
Other	3+ person, w/o kids	+	

When response rate predicts bias… and when it does not



Vehicle ownership nonresponse bias in TBI



Takeaways

1. Some population segments display travel behavior-related self-selection into the survey.

2. Most systematically, larger households with *lower automobility* self-selected into Met Council's travel survey.

3. However, we see that the nature of self-selection can be different across subgroups, because non-urban smaller households with *higher automobility* also self-selected.

4. Yet, some groups had relatively lower response rates but no statistically detectable vehicle ownership nonresponse bias.

Applying the framework to TBI

1. Identify statistics with policy and/or modeling relevance

2. Evaluate data quality without survey design impacts (baseline)

3. **Compare with the bias and precision of final estimates, which reflect impacts from survey design and final weights**

    weighting and convenience sampling

How TBI's sample design & weights impact vehicle ownership estimates

	Calibration Weighting Calibrate subregion sociodemographic distributions to Census benchmarks	**Convenience Sampling** TBI specific: from transit assistance registry and relevant community-based organizations
Bias	+/-	+/-
Precision	–	+/-

Considerations for the Zephyr Community

Evaluating Strategies to Improve Travel Survey Representativeness

Feedback appreciated!
AmyZFong@umich.edu

Modeling	Data Collection

• **Weighting can *introduce* bias**, particularly at the *intersections* of various SED characteristics. These biases may then become encoded in models and forecasts.
 	
• **Extreme weights can worsen statistical precision,** including of of coefficients in travel demand models. This could affect model development by potentially:
  • Overlooking statistically insignificant, but important covariates
  • Limit model sensitivity testing
  • Limit our ability to distinguish outcomes between scenarios
	• Traditional *SED-based* data quality measures (response rates, subgroup sample sizes, or unweighted distributions) may not always indicate travel *behavior* biases. **We may be cautious when using these to determine mid-survey invitation rate adjustments.**
	
	• **When considering whether to increase incentives, follow-up effort, or invitation rates (and combinations)**, we may consider not only their respective costs but also their differing downstream consequences for weighting, bias, and precision.
	
	• **TBI's specific convenience sample did not significantly impact car ownership or VMT estimates,** but this doesn't apply to other regions or if Met Council samples through different means in the future. The practice is also valuable for gauging public opinion on regional policy issues or increasing the civic participation of hard-to-survey groups.

Appendix

Food for thought

- What are we trying to measure?
- Incentives vs oversampling
- Sample size determination

Key concept: survey weights

Base (design) weights

- Inverse selection probability
- Compensates for stratified sample design
- If no unit nonresponse, or respondents are *missing completely at random*, base weights should produce unbiased estimates
- Convenience sample observations have no base weight

Final (calibrated) weights

- "Expands" observations: a weight is number of people in the population represented
- Can be developed with non-response adjustments and/or calibration (raking, post-stratification)
- If nonresponse mechanism well captured, then respondents *missing at random* and final weights should produce unbiased estimates

Key concept: bias sources

Weight used on travel survey		**Source(s)** Random Sampling	Sample Design	Nonresponse	Weighting Process	Convenience Sample (if applicable)
Unweighted		X	X	X		X
Base Weight		X		X		N/A
Calibrated Weights	Variable used in weighting	X			X	X
	Variable **NOT** used in weighting	X		X	X	X

Key concept: statistical precision

- **Design effects** quantify precision loss, or increased sampling variance

  expressed as the ratio between the sampling variance of an estimator under the present sample design (CSD) and its variance under a simple random sample (SRS)

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

- The **effective sample size** of a statistic collected by a present sample design is the number of SRS observations necessary to obtain the same precision

$$n_{effective} = \frac{n}{DEFF}$$

How TBI's sample design & weights impact household VMT estimates

Challenge: there is no gold standard benchmark for household VMT in the Met Council region

Solution: use engineering judgment to examine magnitude of VMT weighting adjustments and correspondence with vehicle ownership adjustments (common approach in survey evaluation)

	Calibration Weighting
Bias	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.
Precision	Among certain segments, weighting worsened precision enough to make regional VMT reduction targets indetectable.