

# RSG the science of insight

# Innovations in address-based sampling (ABS) for travel surveys

Christopher Coy, Mark Bradley, and Elizabeth Greene June 4, 2019, TRB Planning Applications Conference

### **Presentation Overview**



Survey-based data collection provides valuable, comprehensive travel behavior data to support modeling and planning



Who you sample can significantly impact your results and representativeness



There are many opportunities to leverage third-party data to improve address-based sampling plans for travel surveys

### Better sampling allows for better modeling and analysis



How participants are recruited into surveys drives what sample frames are viable. Most household travel surveys use ABS.





## Let's review 3 recent case studies on sampling methods.





#### Oversampling for bike, walk, transit users

 To support modeling and analysis, we want more data for 'rare' or 'new' travel modes





#### Oversampling for **TNC users**





- Overcoming income-based non-response in surveys
  Low-income HHs are typically less likely to participate in
  - surveys



#### CASE STUDY 1: Oversampling for bike, walk, & transit users



## ABS travel surveys can oversample bike, walk, transit users.

#### Challenge

 There is a persistent need to get more data for 'rare' or 'new' travel behaviors (transit, bike, walk, TNC, etc.)

#### Approach

 Use American Community Survey (ACS) data to identify Census Block Groups where relatively more **bike**, walk, or transit commuters live

 $\rightarrow$  Oversample those areas





# Data from San Diego County shows how geographic concentration of bike, walk, and transit commuters.





This approach is effective to get more walk, bike, transit users. The degree of oversampling determines how much more.

#### **Benefits**

- **It works!** Can often get 2–3x the unweighted data for given populations
- 2x the data for rare behaviors can really reduce margins of error for small groups
- This is neutral (or helps) overall response rates and can save project budget
- Still a probability sampling approach, with all the statistical benefits

#### Limitations

- Aggressive oversampling leads to geographically lumpy data (but that also reflects reality)
- Oversampling leads to higher variance (less precision) in the overall weighted sample (but likely more precision for bike/walk/transit users)



### CASE STUDY 2: Oversampling for TNC Users

• —

Ē

Data Source 1: ACS data at the block group level

#### Data Source 2: San Diego Regional Transportation Study

- 2016–17 household travel survey for SANDAG
- Included TNC as a travel mode → Captured 1,700 TNC trips from 630 people
- Largest known TNC trip count in a household travel survey

#### **Predictive model:**

- The model estimates the average TNC trips per household in that block group
- Model variables include: income, presence of children, age, block group density, number of vehicles, and more





# After identifying areas where more TNC-using residents <u>live</u>, we invite more households from those neighborhoods.

Model where TNC users live  $\rightarrow$  Heavy Oversampling



Darker-shaded Block Groups had higher expected TNC use and received higher invitation rates. **Obtain more participants who live in those Block Groups** 



As expected, more participants are coming from oversampled areas.\*



Maps shown are for San Francisco. The location points are 'jittered' by adding random noise.

## ABS oversampling for TNC users is very effective. It's scalability/impact depends on the region's level of TNC use.

Sample Region	Sample Strata	% self-described 'TNC Users'	TNC mode share	'Shared mobility' mode share (car/bike/scooter)	
San Francisco County	Very High TNC	62%	7.2%	1.5%	
	High TNC	58%	5.1%	1.6%	T
	Medium TNC	45%	3.1%	0.3%	
	General Population	47%	1.9%	0.3%	
8-County Region	Subtotal	57%	5.5%	1.3%	
	Very High TNC	43%	2.0%	2.1%	
	High TNC	64%	2.7%	1.9%	
	Medium TNC	51%	1.9%	0.9%	
	General Population	31%	0.7%	0.4%	
	Subtotal	45.0%	1.6%	1.0%	
	Total	52.1%	3.9%	1.2%	



# This approach is effective at getting more 'local' TNC users.

#### **Benefits**

- It works! Not a lot of other (good) options for sampling TNC users
  - Captures more self-described and actual TNC users
- Slightly helps response rates, compared to the general population
- Still a probability sampling approach, with all the statistical benefits

#### Limitations

- Geographically lumpy data (but that also reflects reality)
- Most TNC users are infrequent users, meaning data collection/diary methods matter a lot for capturing TNC trips
- Doesn't capture TNC use from visitors, who likely use TNCs a lot



## CASE STUDY 3:

## **Overcoming income-based non-response**



# RSG has used third-party household income data to help overcome income-related non-response bias.

**Challenge**: Low-income households are less likely to respond to surveys, and income is a key determinant of travel behavior



Approach: Use 3<sup>rd</sup>-party estimated household income (e.g., credit agency data) to send relatively more invitations to low-income households

- RSG gets this income data from our ABS data provider (MSG)
- Typically 80-90% of households have an estimated income, 10-20% don't

HH ID	Address	City	State	Zip Code	<b>Estimated HH income</b>
1		WEST SACRAMENTO	CA	95691	\$50,000-\$74,999
2		SACRAMENTO	CA	95828	\$35,000-\$49,999
3		SACRAMENTO	CA	95823	\$75,000-\$99,999
4		SACRAMENTO	CA	95828	NA



To help overcome non-response bias, a project can samplebut-not-invite medium and higher income households.

#### **Benefits**

- It works!
  - Other options for this are more geographically lumpy
- This approach is scalable and changeable
  - Scalable: Degree of oversampling can be dialed up or down easily, based on budget
  - Changeable: If you need to scale back or undo part way during sampling, you can
- Still a probability sampling approach, with all the statistical benefits

#### Limitations

- By its nature, this approach invites more people who are 'hard-to-reach,' lowering survey response rates
- Lower response rates lead to higher sampling/recruitment budgets







## **Concluding Thoughts**

• Travel behavior is changing. We need to ensure that our data collection and sampling methods keep up



- The concepts of representativeness and data quality should apply to all data products (including passive/big data)
- There are **opportunities to use third-party data** (including other travel survey data) to improve travel survey sampling
- These methods are viable, effective, continually improving, and in use today

