

Augmenting Validation on the SCF

April 14, 2021

Shannon Nelson
Cathy Haggerty
Nella Coleman
Kate Bachtell
Micah Sjoblom
Steven Pedlow
Jesse Bricker



Overview

01 Significance of validation

02 Challenges

03 SCF validation protocol

04 Looking ahead



Our obligation as survey researchers: data integrity

- Must be truthful in data collection and reporting results
- Commitment to truthfulness must extend beyond researcher to all survey staff
- Interviewers must strictly adhere to protocols and we must strictly enforce protocols
- *Researchers have a responsibility to prevent and detect falsification*



Intentional departure from the designed interviewer guidelines or instructions, unreported by the interviewer, which could result in the contamination of data. Intentional means that the interviewer is aware that the action deviates from the guidelines and instructions.

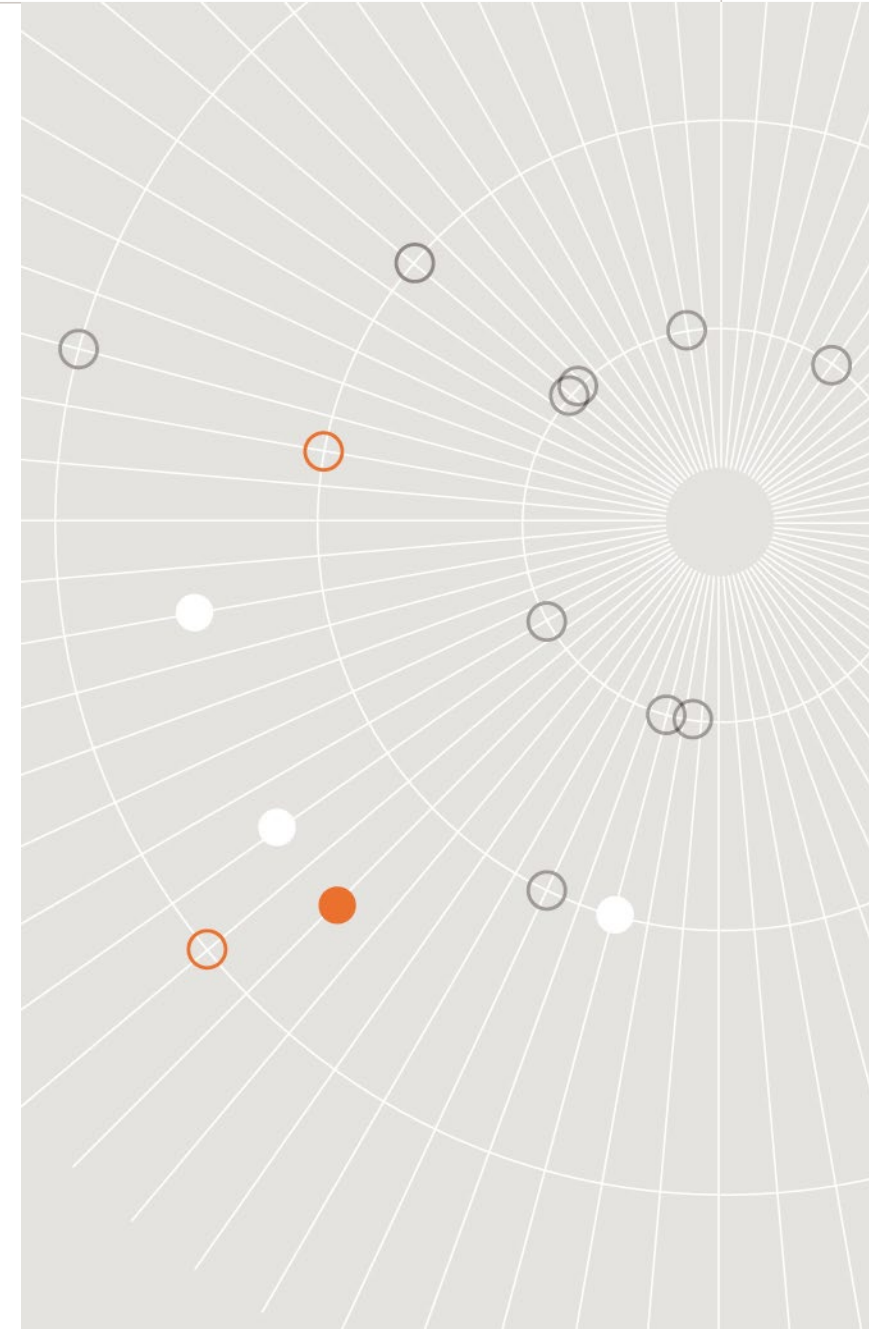
– AAPOR 2003

Collective responsibility to identify and mitigate falsification

Sharing methods and strategies for detecting and preventing falsification helps to establish best practices and benefits us all.

Preventing falsification

- Articulate values, goals, and rules as well as consequences
- Carefully select and train interviewers
- Provide adequate compensation, supervision, evaluation and rewards
- Design and execute programs of deterrence and detection
- Look for data anomalies
- Respond appropriately to proven instances of falsification



Challenges

The challenges are real

- Schedule
- Budget
- Resources

Study-level factors associated with a higher likelihood of falsification

- Long and burdensome questions and questionnaires
- Complex survey logic and study protocols
- Survey modes that are more difficult to monitor (e.g., field vs. telephone)
- Clustering of target population in areas with low resources or chronic difficulty w/completing interviews
- Presenting large cash incentives or other materials that may create a temptation for interviewer

SCF Validation Protocol

Standard and Expanded

Standard validation protocol included three key elements

Culture of continuous learning & high quality

Comprehensive training

- 3.5 days of IP training
- Expansion modules
- Weekly *In the Know*

FRB data quality review

- 50+ data quality checks applied to every case
- FI data quality reports

Recontact

Phone calls by independent validation team

- 1st 2 completes
- 10% completes thereafter

Mail and in-person efforts as necessary

Data review

Paradata

Interview duration, section timings, percent missing data

Benfords' Law

Distribution of leading digits

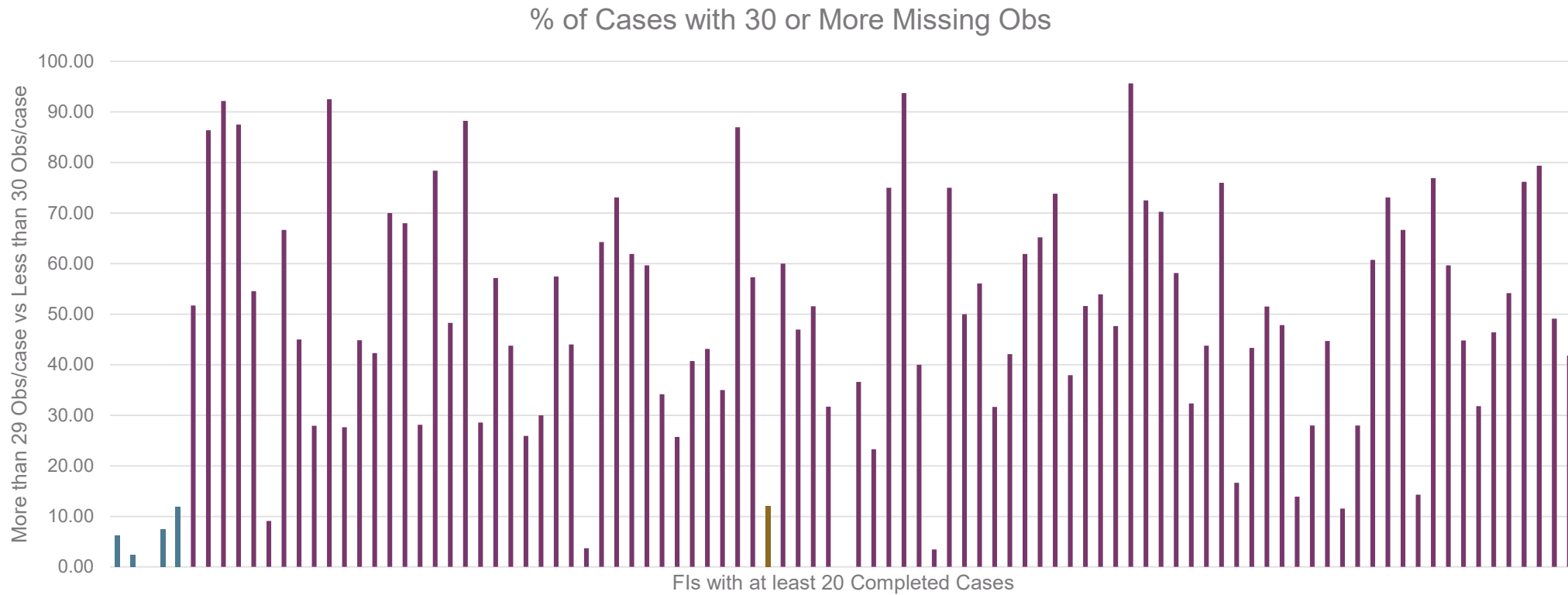
Standard validation methods work but small n is a problem

Every round a falsifier is identified...but not soon enough...

It takes time to gather enough data about interviewer behaviors to identify patterns that point to falsification.

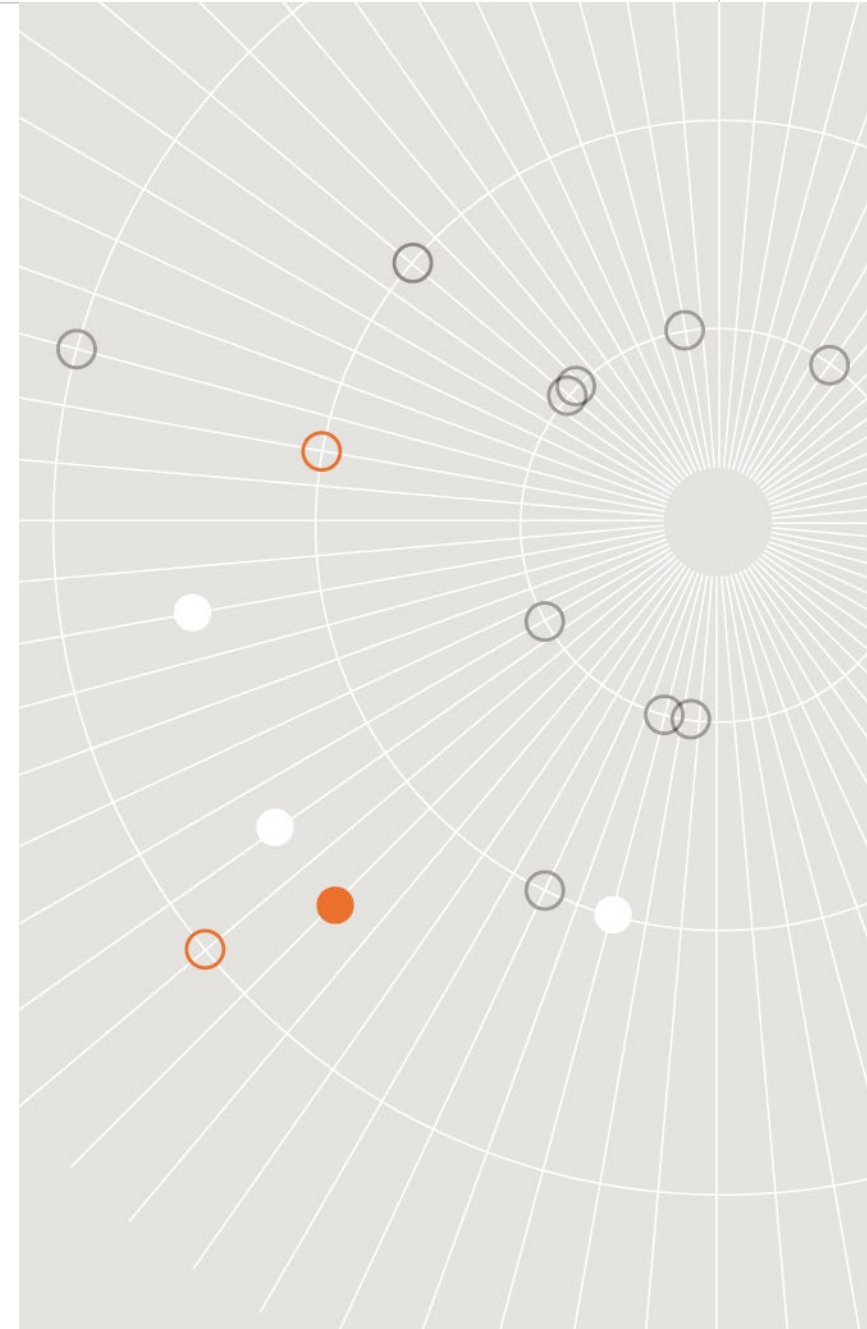


Deliberate falsifiers may attempt to thwart detection methods



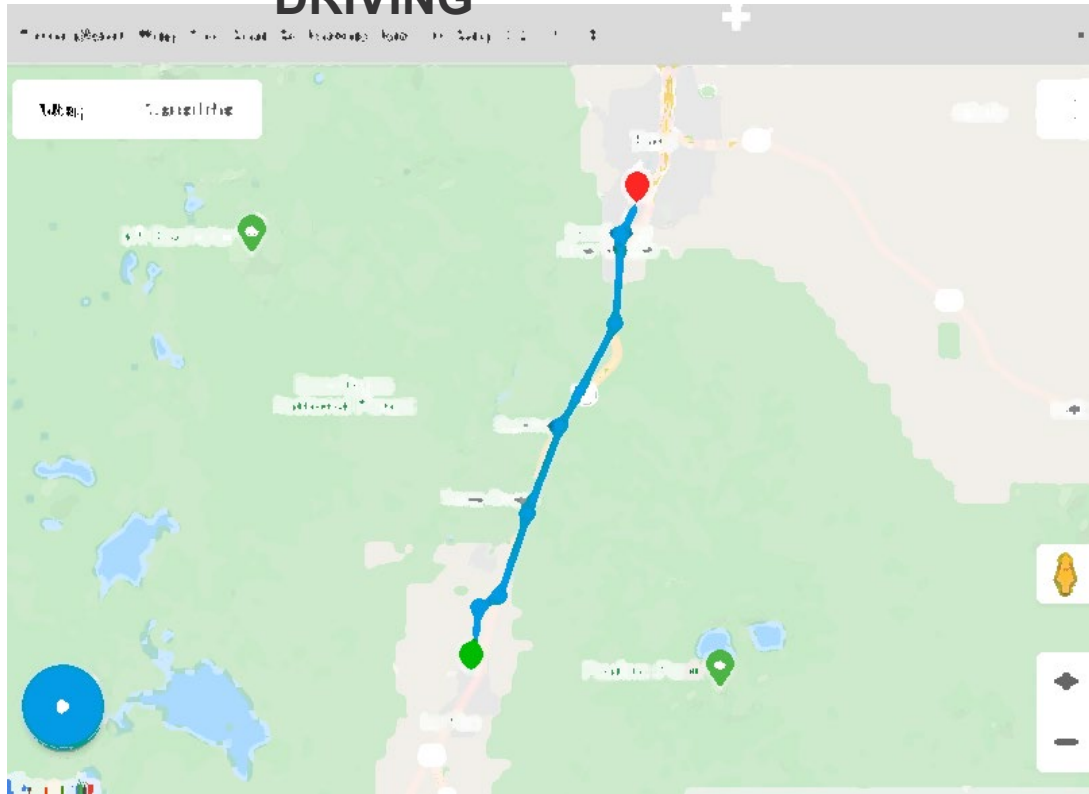
Technology speeds up the process: GPS

- Tablets
- TSheets (timetracking app)
- Dooblo (mobile survey software)

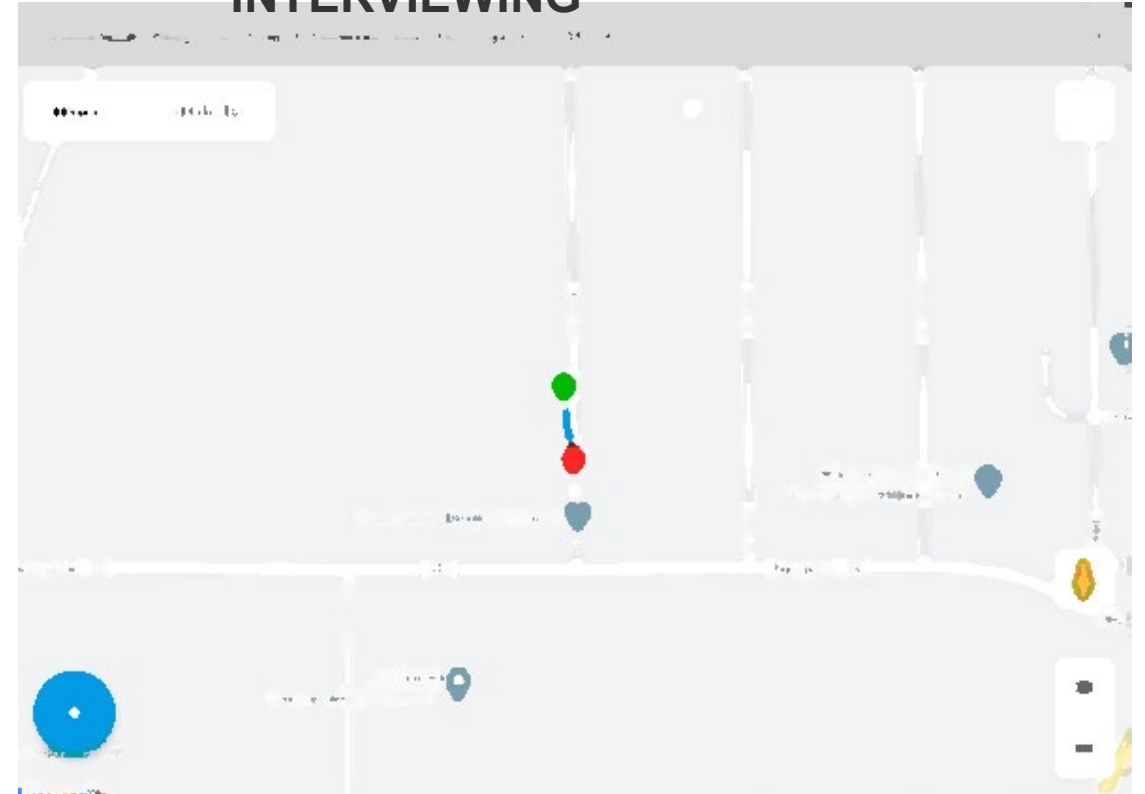


TSheets includes real time monitoring by task: what you expect to see

FI LOGGED IN AS DRIVING

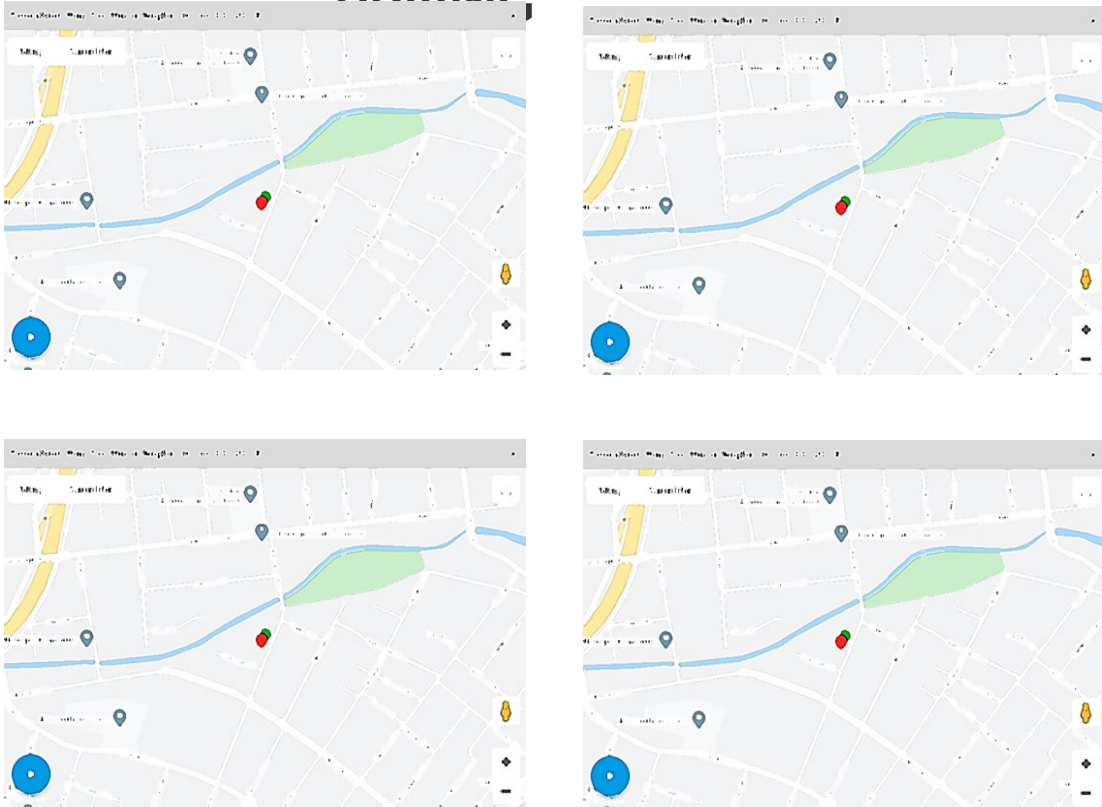


FI LOGGED IN AS INTERVIEWING

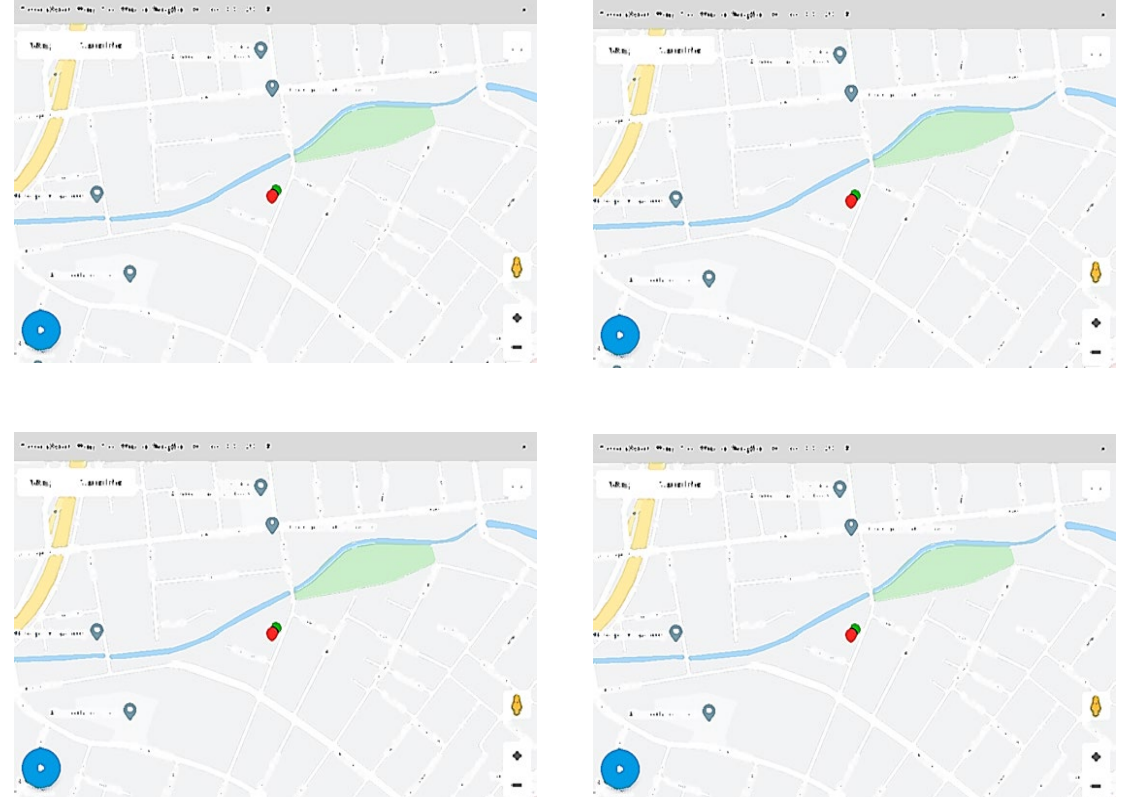


TSheets includes real time monitoring by task: Red Flag!

FI LOGGED IN AS
DRIVING



FI LOGGED IN AS
INTERVIEWING



Dooblo allows almost instant review and retrieval of signatures



CASE 1



CASE 2



CASE 3



CASE 4



CASE 5



CASE 1



CASE 2



CASE 3



CASE 4



CASE 5

Looking ahead

Expand, automate, and harmonize data used to validate completed cases

Training & culture of high quality

Focus on importance of high quality data
FRB data quality review and feedback
Ethics, values, and monitoring methods

Recontact

First two completed cases
10% of completed cases thereafter

Paradata and Benford's Law

Interview and subsection duration,
capture of contact information, missing
data, ROC data



Empowered field management

Regular contact and coaching
Access to validation dashboard & metrics

GPS (Tsheets and Dooblo)

Location at time of interview
Location at time of signature
Real time tracking by task

Data science techniques

Data mining/machine learning, cluster
analysis, data reduction

Other potential uses of technology for validation

- Electronic incentives
- Quick turnaround web/mobile confirmation surveys
- Voice detection software
- Geofencing



Data fabrication and falsification pose serious threats to the credibility of survey research. Falsified or fabricated data yield biased estimates, affect the precision of estimates, and impact multivariate relationships. The Office of Research Integrity (ORI) declared falsification as a form of scientific misconduct. In an era of tight budgets and increasing challenges for surveys, resources still need to be allocated to prevent, detect, and mitigate data falsification and fabrication.

– AAPOR Task Force on Data Falsification, September 2020

Questions?



Thank you.

Shannon Nelson
Senior Survey Director
nelson-shannon@norc.org

 Research You Can Trust™

 **NORC** at the
University of
Chicago