Maintaining Data Integrity: The Evolution of the Census Bureau's Field Quality Monitoring Program

Presentation by Elizabeth Mahoney, U.S. Census Bureau

Contributors: Mary C. Davis, Richard A. Denby, Scott W. Glendye, Laura B. Hergert, Rachel Huang, Sadaf Rohani and Karen Pennie U.S. Census Bureau

2024 AAPOR Conference
May 16, 2024

Disclaimer: Any views expressed are those of the authors and not those of the U.S. Census Bureau

Data Quality begins with the Interviewer

- Survey interviewers are a significant source of potential measurement error in survey data collection (Groves, 1989)

    - Error introduced when interviewers are improperly trained, fail to follow procedures, or deliberately falsify data

- Detecting and mitigating interviewer error is challenging, especially since falsification is a relatively rare event
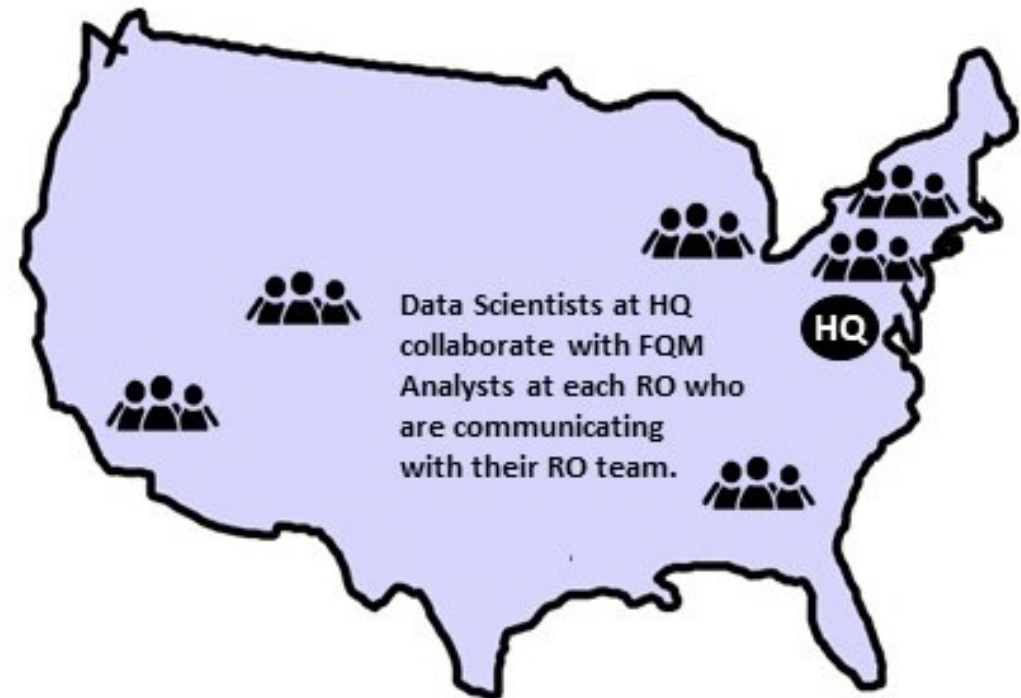
Why Field Quality Monitoring (FQM)?

- One Field Representative (FR) not following procedures can have a broad impact on estimates

- Existing processes (like reinterview) does not always catch these problems

- Need to rethink how we identify potential data quality issues more quickly

FQM Vision

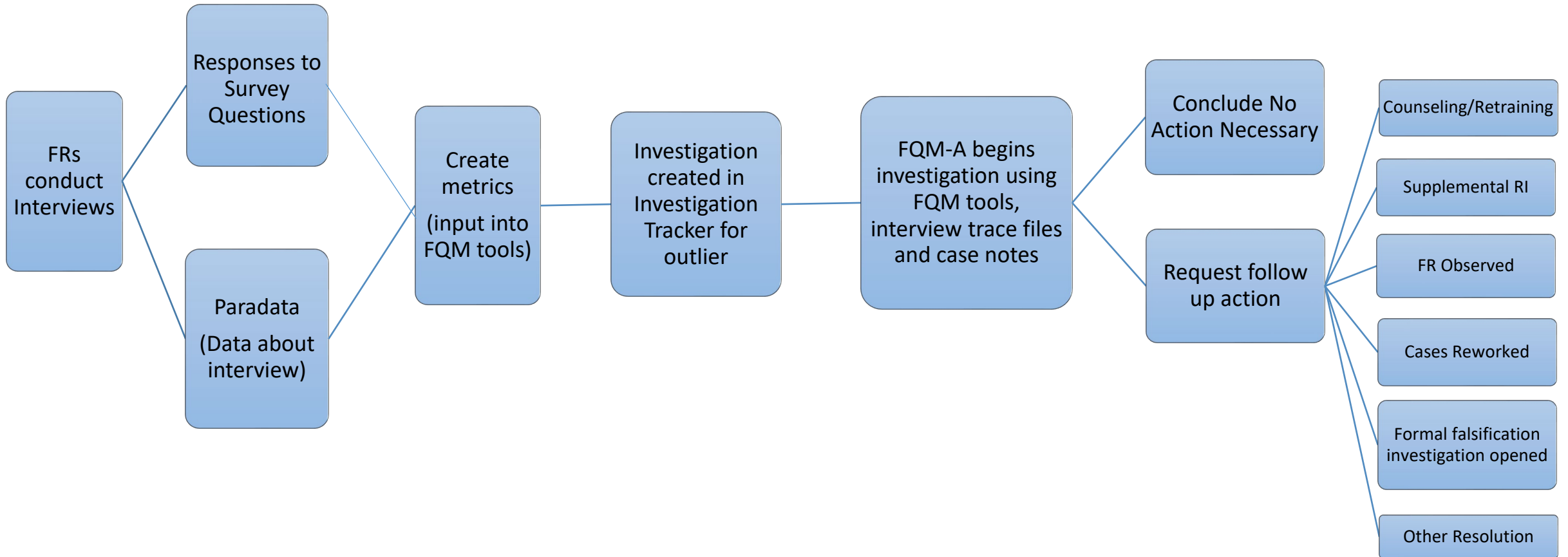- Create one unified system by which data quality for an FR can be monitored across surveys in near real-time.
  - Monitoring across surveys will allow us for identification of deviation from procedures earlier than by looking at one survey itself
  - Data-driven analysis using investigations to gain context
  - Identify and correct isolated quality problems before becoming widespread or requiring fixes during processing

What Makes FQM Successful

- FQM Analyst (FQM-As) in each of our 6 regional office

- HQ Liaisons work with FQM-As

- Weekly Team FQM Check-in meetings

- Identifying potential issues in near-real time

- Acting on potential issues in near-real time

Data Scientists at HQ collaborate with FQM Analysts at each RO who are communicating with their RO team.

HQ

From Interview to Outlier Investigation Resolution

FRs conduct Interviews

Responses to Survey Questions

Paradata (Data about interview)

Create metrics (input into FQM tools)

Investigation created in Investigation Tracker for outlier

FQM-A begins investigation using FQM tools, interview trace files and case notes

Conclude No Action Necessary

Request follow up action

Counseling/Retraining

Supplemental RI

FR Observed

Cases Reworked

Formal falsification investigation opened

Other Resolution

United States® Census Bureau

Office of Survey & Census Analytics
SCA

Original Quality Components and Metrics for 2021

FR Workload

- Response Rate
- Partial Rate
- Refusal Rate
- Type B/C rate (noneligible sample rate)
- Incomplete Callback Rate

Response Behavior

- Item Nonresponse Rate
- Time in Instrument
- Average Section Time Flags

Interviewer Behavior

- Days Since Last Contact

United States® Census Bureau

Office of Survey & Census Analytics
SCA

Original Methodology

- Monitoring 9 metrics for 1 survey

- Flagging Methodology

  - Interquartile range (IQR)

    - Lower threshold: Area median - 1.5*IQR

    - Upper threshold: Area median + 1.5*IQR

  - Areas must have 30 or more cases and at least 10% of cases complete

The FQM Program's Evolution

- Evolving metrics

- Improved flagging

- Sophisticated monitoring tools

- Reinterview Pilot Program

- FR Quality Score

Evolving Metrics

FR Workload

- Partial Rate
- Vacant/URE rate
- Type B/C rate (noneligible sample rate)
- Completion Rate

Response Behavior

- Item Nonresponse Rate
- Time in Instrument
- One Person Households
- No Child Households

Interviewer Behavior

- Time between Interviews (Back-to-back)
- Overnight Attempts
- Interviews per Attempt (IPA)
- Unable to Conduct Interview
- Phone Number Discrepancies
- Variability in Attempts
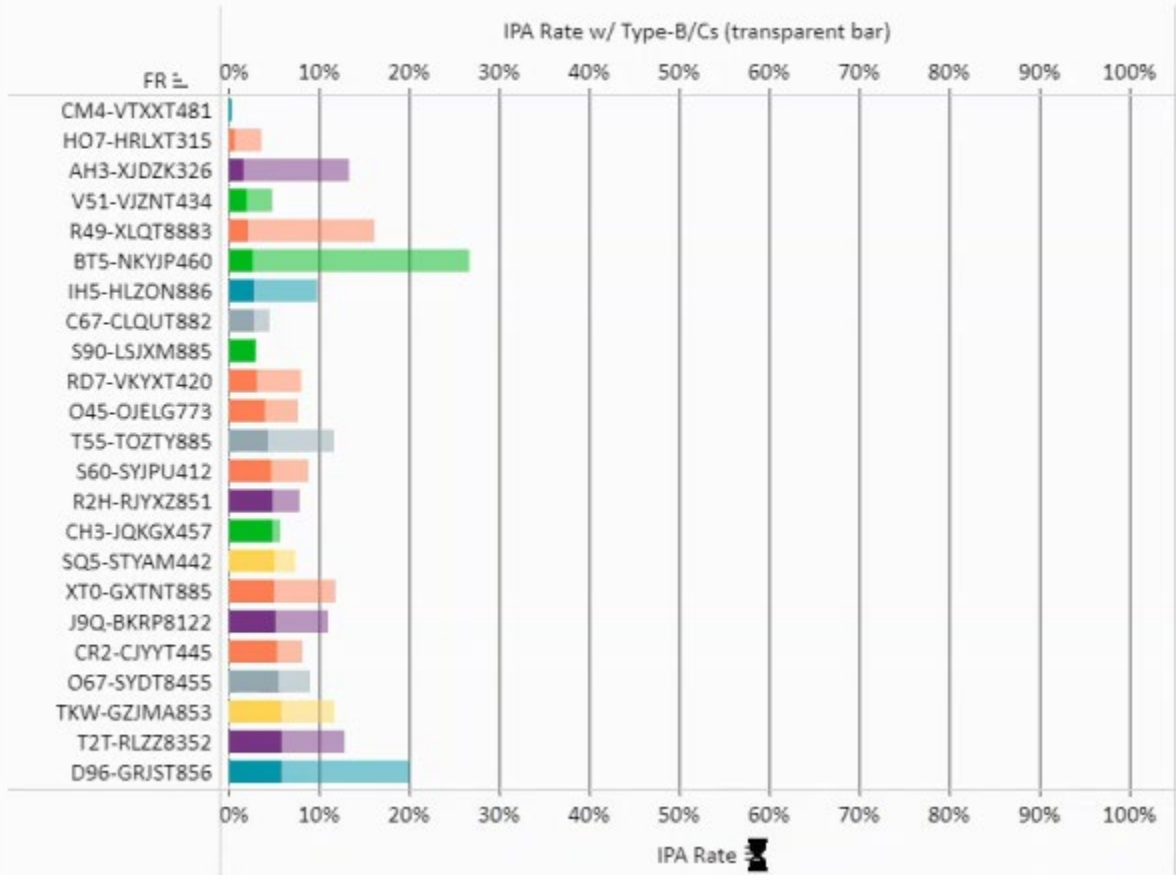- Completes by Text/Email

Cost Consideration

- Miles/Hours per Attempt
- Miles/Hours per Case
- PV Attempts per Day
- Geocoded miles between home and visits vs. Miles charged
- Instances with miles or hours charged with no attempts

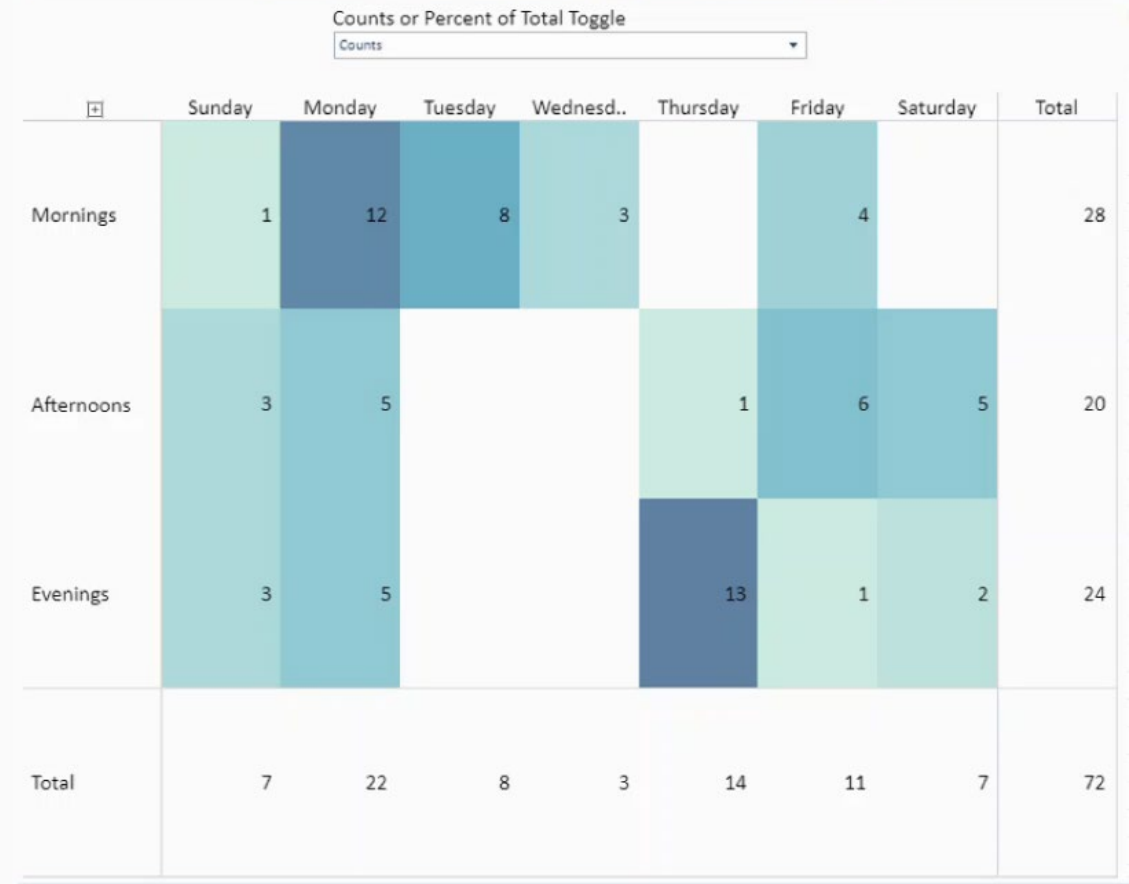# Sophisticated Cross Survey Monitoring Tools

- CHI Dashboard

- Miles and Hours Dashboard

- Item Nonresponse Dashboard

- Response Dashboard

- ArcGIS Mapper

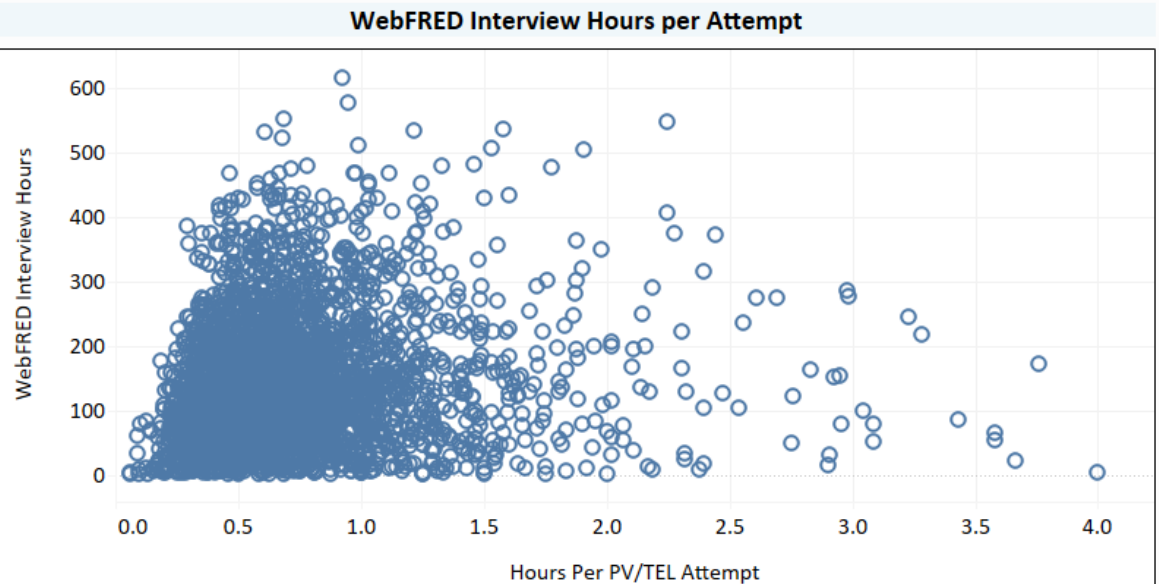# Sophisticated Monitoring Tools

NOTE: Data shown is fictitious

Sophisticated Monitoring Tools

WebFRED Interview Miles per PV Attempt

WebFRED Interview Hours per Attempt
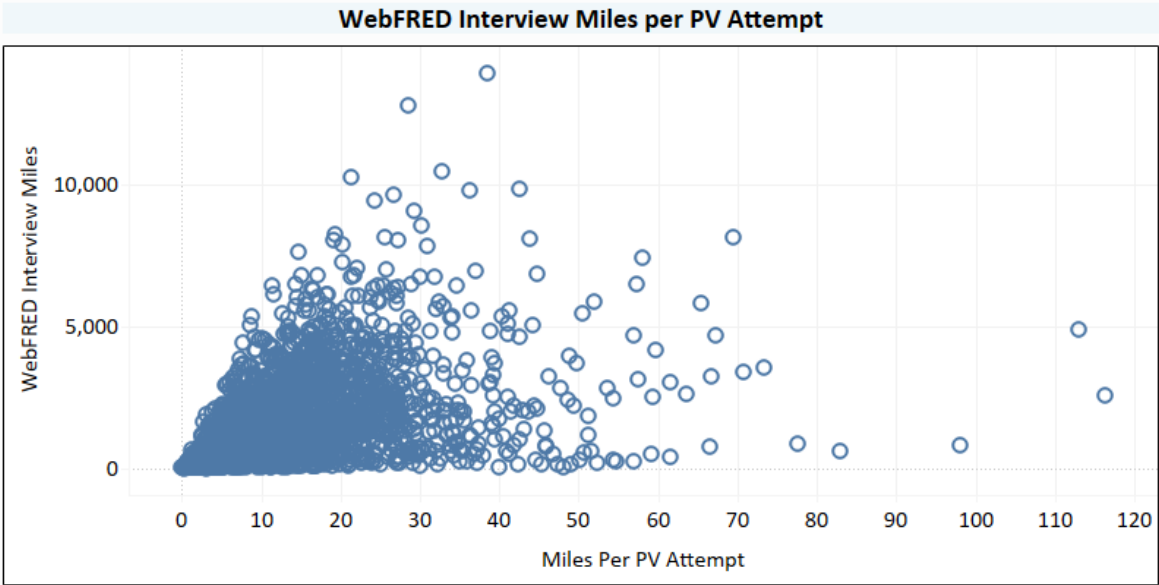
Sophisticated Monitoring Tools

16.2 Miles
: 2:34:27 PM
Survey 1

- Geocoded Miles from interviewer home
- Date and time of PV attempt
- Survey

2.4 Miles
3:57:28 PM
Survey 1

- Geocoded Miles from last visit
- Date and time of PV attempt
- Survey

NOTE: Data shown is fictitious

14

Sophisticated Monitoring Tools

NOTE: Data shown is fictitious

Sophisticated Monitoring Tools
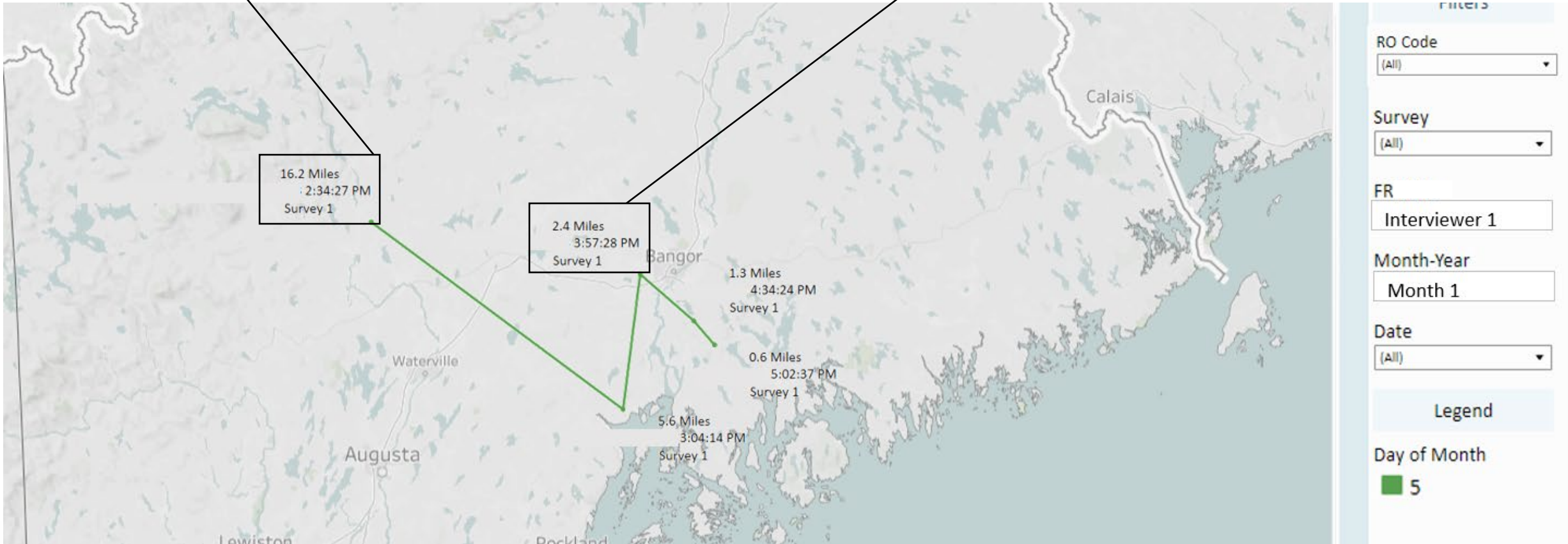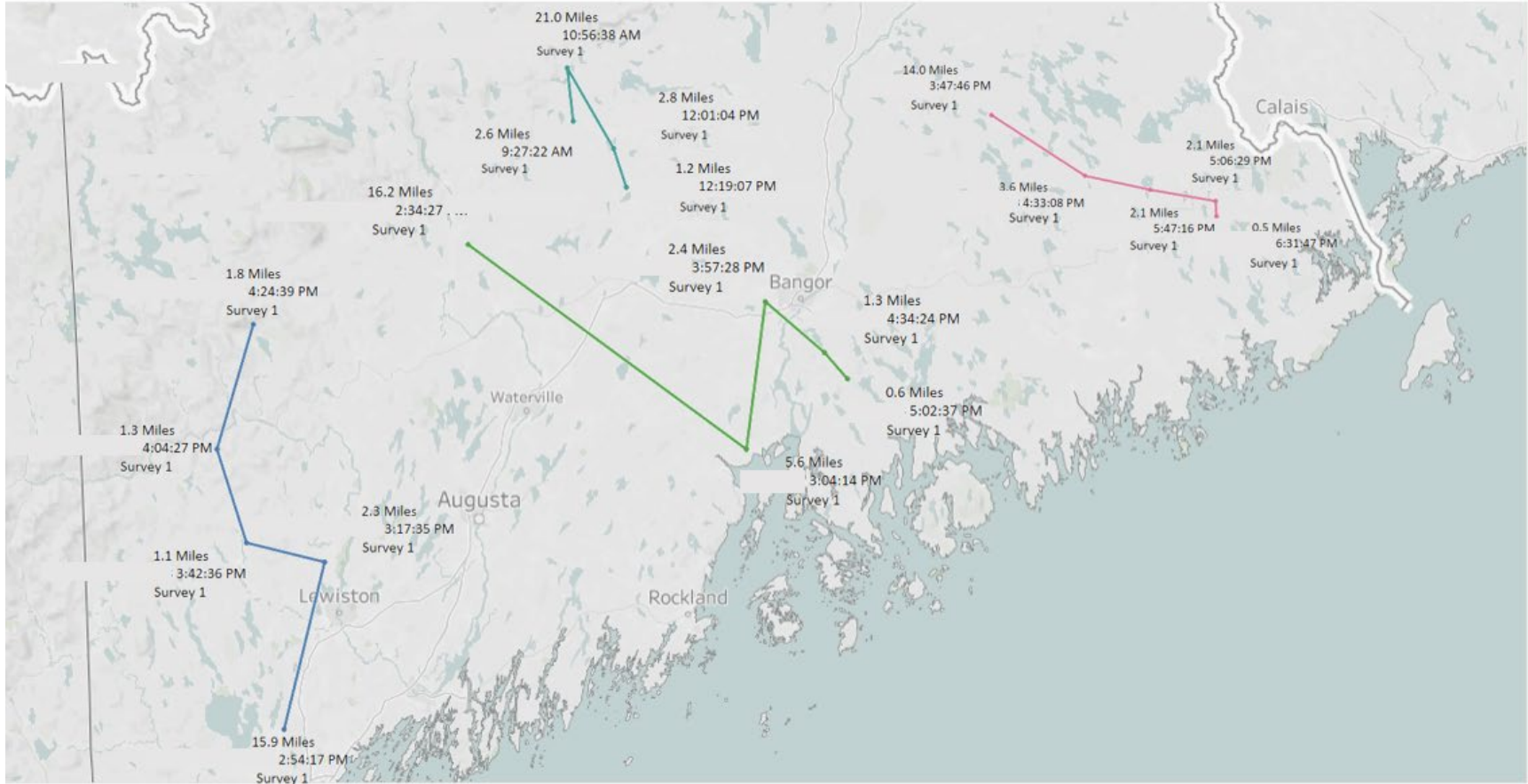
	DOBD	YRNAT	SCHL	MGW2	HICEMP..	MARYR	WKL	INW2	PWW1	JWLH	YBL	RMS	ELE	GAS	WAT	VALPROP	MRG	TAX	INS	EARN	SEM	INTR
											Variable Name											
FRAJ1	6.5%	23.8%	5.9%	0.0%	0.4%	11.4%	0.0%	20.9%	36.4%	10.2%	17.2%	28.2%	18.0%	22.2%	23.5%	18.8%	12.1%	27.4%	35.5%	51.6%	72.7%	75.0%
FRAJ2	11.7%	33.3%	3.6%	18.4%	3.5%	30.9%	3.2%	9.0%	6.8%	15.7%	44.3%	9.7%	12.9%	14.6%	16.1%	19.2%	23.3%	23.4%	40.4%	42.8%	36.0%	66.7%
FRAJ3	10.2%	3.8%	3.0%	2.3%	2.2%	11.2%	3.9%	16.0%	19.8%	3.8%	44.1%	19.0%	11.2%	14.4%	11.9%	31.9%	16.1%	40.0%	46.1%	42.7%	14.8%	66.7%
FRAJ7	11.7%	33.3%	5.3%	11.1%	5.3%	10.9%	7.7%	16.7%	24.0%	17.2%	18.8%	6.4%	15.4%	28.6%	25.0%	11.1%	0.0%	42.3%	38.5%	28.3%	16.7%	57.1%
FRAJ8	0.0%		0.0%	0.0%	0.0%	8.3%	0.0%	16.7%	87.5%	12.5%	11.8%	5.9%	22.2%	25.0%	40.0%	16.7%	0.0%	33.3%	50.0%	33.3%	60.0%	50.0%
FRAJD	8.2%	10.0%	3.0%	12.5%	2.4%	11.2%	3.6%	15.3%	50.0%	8.9%	16.7%	16.8%	6.2%	23.5%	11.3%	23.0%	0.0%	17.0%	26.4%	44.9%	26.7%	87.5%
FRAJF	0.0%	11.1%	0.0%		0.0%	2.9%	0.0%	5.4%	2.9%	0.0%	5.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	30.0%	3.2%	12.5%	
FRAJJ	9.4%	0.0%	9.5%	0.0%	7.5%	16.9%	10.7%	21.2%	58.1%	3.2%	27.3%	9.1%	10.0%	6.7%	7.4%	25.0%	6.3%	50.0%	68.8%	33.3%	75.0%	63.6%
FRAJO	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	14.3%	0.0%	0.0%	0.0%	0.0%		0.0%	14.3%	0.0%	16.7%	33.3%	14.3%	0.0%	
FRAJQ	7.0%	60.0%	7.3%	25.0%	3.5%	24.3%	3.8%	20.0%	56.3%	6.3%	3.0%	9.1%	16.7%	100.0%	18.2%	5.9%	0.0%	29.4%	29.4%	41.2%	0.0%	66.7%
FRAJY	9.5%	0.0%	0.0%		0.0%	7.1%	0.0%	0.0%	0.0%	0.0%	8.3%	0.0%	0.0%	0.0%	0.0%	0.0%	20.0%	20.0%	40.0%	14.3%	0.0%	
FRAK1	3.9%	26.3%	5.0%	7.7%	3.5%	12.7%	7.2%	7.0%	10.0%	3.4%	43.7%	16.1%	10.7%	16.7%	14.0%	23.7%	10.0%	37.7%	37.7%	53.0%	0.0%	75.0%
FRAK8	12.1%		13.3%	100.0%	13.3%	26.1%	3.0%	12.5%	55.0%	63.2%	45.2%	44.3%	7.1%	31.8%	25.0%	56.3%	0.0%	48.3%	75.9%	50.0%	100.0%	100.0%
FRAKI	4.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	7.0%	9.7%	0.0%	2.0%	2.1%	2.6%	0.0%	0.0%	3.1%	14.3%	3.4%	10.3%	0.0%	0.0%	
FRAKQ	7.7%		2.9%	0.0%	2.6%	13.6%	2.2%	5.7%	12.5%	2.5%	10.0%	2.5%	9.4%	0.0%	8.3%	8.3%	12.5%	41.7%	50.0%	10.3%	37.5%	0.0%
FRAKT	1.9%	0.0%	0.0%	66.7%	0.0%	4.2%	0.0%	43.3%	77.3%	0.0%	45.2%	32.3%	4.8%	0.0%	0.0%	16.7%	0.0%	25.0%	50.0%	29.2%	0.0%	
FRAL1	14.3%	28.6%	6.3%	40.0%	1.6%	20.8%	0.0%	44.1%	56.5%	10.5%	45.7%	34.1%	36.0%	16.7%	20.0%	21.4%	0.0%	40.0%	60.0%	75.9%	75.0%	100.0%
FRAL2	0.0%	100.0%	5.3%		0.0%	0.0%	20.0%	20.0%	75.0%	0.0%	44.4%	0.0%	0.0%	0.0%	0.0%	0.0%		100.0%	100.0%	50.0%		
FRAL8	6.5%	0.0%	1.9%	6.7%	1.6%	10.2%	0.0%	1.5%	40.9%	3.1%	10.8%	5.1%	3.0%	2.1%	4.6%	8.7%	17.4%	35.5%	32.7%	23.6%	9.5%	0.0%
FRALB	40.0%		0.0%		0.0%	0.0%	0.0%	50.0%	100.0%		33.3%	0.0%	50.0%	50.0%	50.0%	0.0%	0.0%	0.0%	0.0%		100.0%	100.0%
FRALE	0.0%	0.0%	0.0%	0.0%	3.2%	0.0%	5.0%	4.8%	18.2%	10.0%	45.5%	45.0%	0.0%	0.0%	0.0%	0.0%	14.3%	0.0%	0.0%	6.7%	0.0%	
FRALF	0.0%		0.0%	0.0%	1.5%	12.1%	0.0%	0.0%	9.1%	0.0%	35.3%	11.8%	6.3%	0.0%	8.0%	33.3%	0.0%	30.0%	40.0%	20.9%	25.0%	0.0%
FRALM	0.0%		0.0%		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	46.7%	0.0%	0.0%	0.0%	50.0%	0.0%	0.0%	0.0%	0.0%		
FRAM1	9.1%	25.0%	6.1%	0.0%	6.1%	29.4%	0.0%	20.0%	44.4%	6.3%	29.2%	4.2%	0.0%	0.0%	14.3%	0.0%	0.0%	0.0%	0.0%	36.4%	25.0%	50.0%
FRAM2	3.7%	0.0%	0.0%	20.0%	0.6%	4.0%	0.0%	4.8%	12.8%	1.5%	37.5%	37.9%	3.2%	9.8%	10.9%	10.6%	0.0%	28.9%	37.8%	10.7%	6.3%	6.7%
FRAM3	10.6%	55.6%	2.4%	4.5%	1.6%	15.9%	3.3%	7.3%	14.3%	1.9%	47.0%	20.8%	17.2%	20.1%	21.4%	10.8%	10.9%	30.8%	40.0%	27.3%	42.1%	40.0%

Reinterview Pilot Programs

- Began leveraging FQM metrics for reinterview (quality control) selection
  - Time in Instrument
  - Item Nonresponse
- FQM metrics found to be more successful at identifying data irregularities than other reinterview types

The Impact of FQM on Data Quality

- Investigations to date
  - Thousands FQM investigations have been closed since 2021
  - About 70 percent of total FQM investigations have ended in actionable resolution
  - Most prominent resolution type is interviewer counseled/retrained

Looking Ahead: The Future of FQM

- Utilizing response data and case notes to create metrics
- FY25 Data Quality Score
  - Scoring interviewer data quality using FQM metrics
  - Interviewers will receive monthly cumulative reports and then a final quality rating

Contact Information
Elizabeth.Mahoney@census.gov