Examining the Causes of Roster Error: A Comparison of Human Coding and Automated Coding Approaches

D.C. AAPOR Preview Conference
Washington, D.C.
September, 16, 2025

Kathleen Kephart, Kristen Kohm, Eric Stone, Alexandra Piccirillo
(US Census Bureau)

DRB Delegated Authority Approval #s CBDRB-FY25-ACSO003-B0021
DRB Delegated Authority Date # 9/9/2025
CBDRB-FY23-ACSO003-B0066CBDRB-FY24-ACSO003-0010
DRB/DAO Approval Date: April 25, 2024

United States® Census Bureau

Background

- The ACS household roster questions have changed very little since the 1990s, however, research has found issues with the rostering questions and instructions.
- Rostering error is more likely in households with:
  - Young children (age 0-4) who are not the biological or adopted children of the householder (Jensen & Hogan 2017; Jensen, 2019).
  - Complex households, or those that contain extended family, multigenerational, blended families, or non-relatives.
  - The last few decades have seen a significant increase in complex households (Jensen, Schwede, Griffin & Konicki 2018).
- There is evidence of respondent confusion and burden during the ACS roster questions
  - In particular, many people who are added during the follow-up questions are ultimately deleted (Ashenfelter et al., 2012; Clark, 2017)

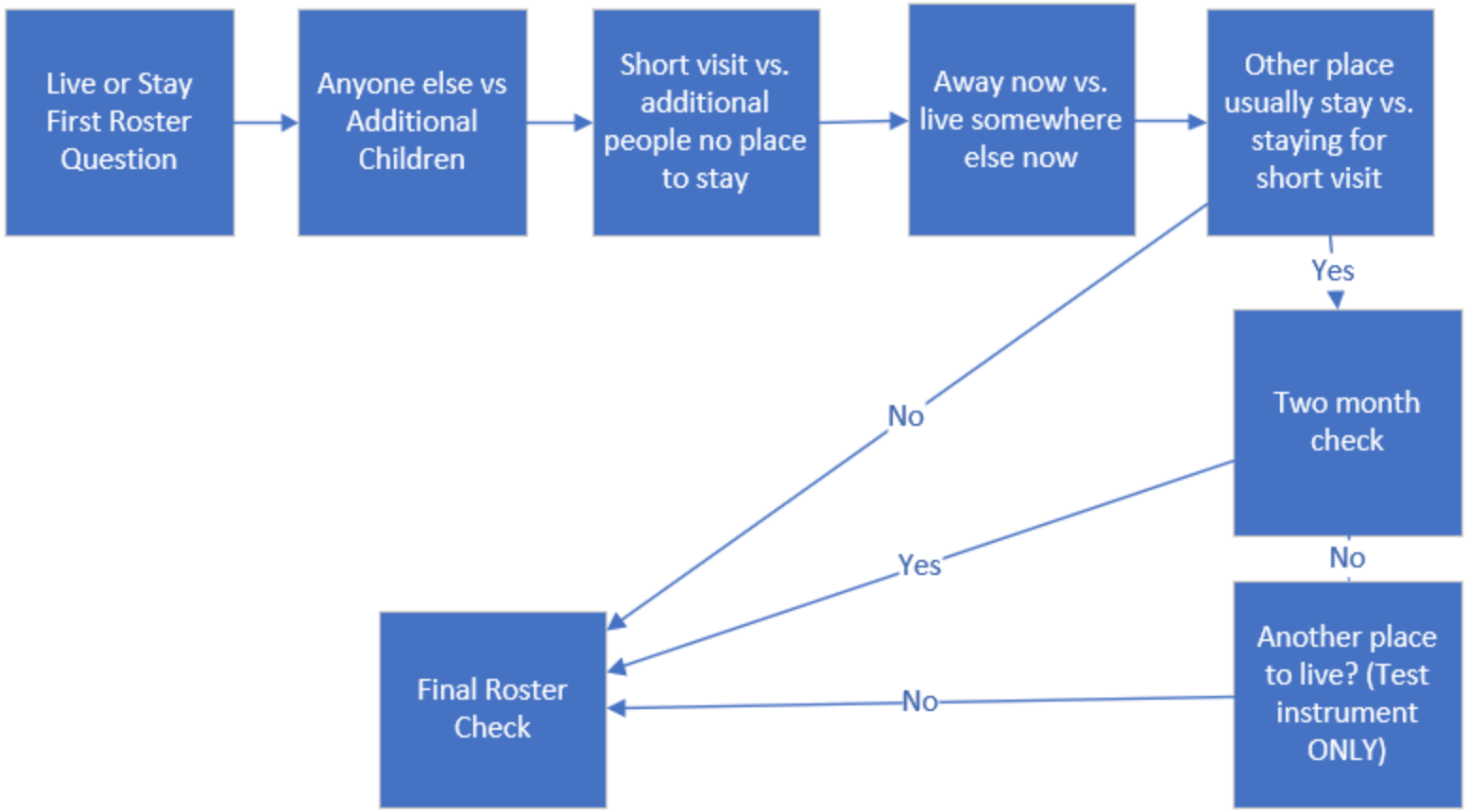
United States® Census Bureau

Roster Research

- Most of it has focused on either large split panel experiments (with quantitative measures) or small qualitative studies of 15-40 participants
- In the 2022 ACS Content Test we employed a mixed method research study on a nationally representative sample
- Attempt to get at the “why” of rostering error in a nationally representative sample
- If people added or removed anyone after the first roster question they received an open ended probe

United States® Census Bureau

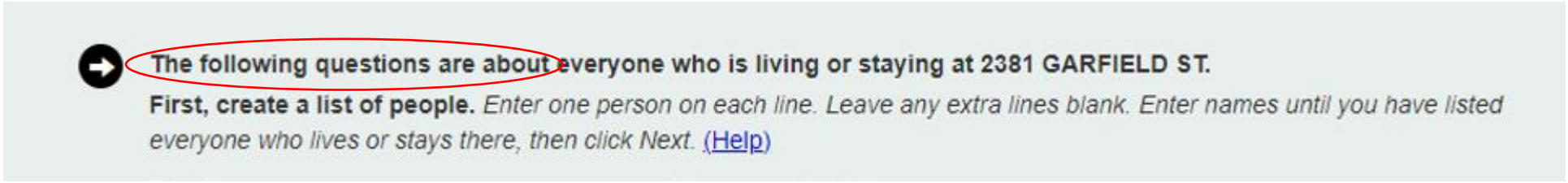
Roster Question Flow (Control vs Test)



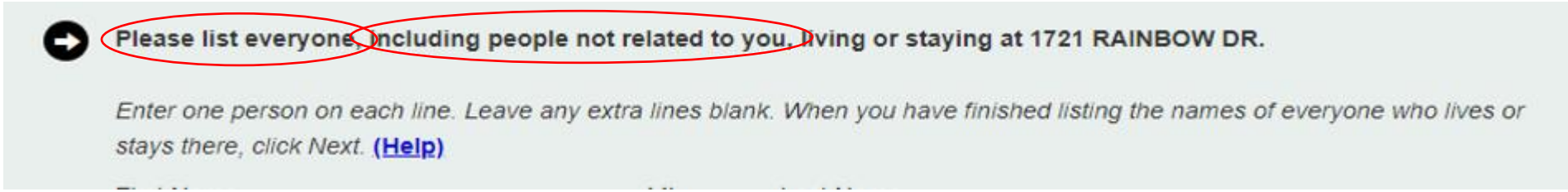
United States® Census Bureau

Questions Tested (Internet)
Live or Stay (First Roster Question)

Control Version



Test Version



United States® Census Bureau

Results

A significantly higher percentage of people were rostered on the first screen in the Test version than in the Control version

Table 1. People Originally Rostered on the First Roster Question

Mode	Control	Test	Difference	P-value
Across Mode	98.0	98.6	0.7	<0.01*
Internet	97.7	98.5	0.8	<0.01*
CAPI	98.8	99.1	0.2	0.52

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066 Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Results

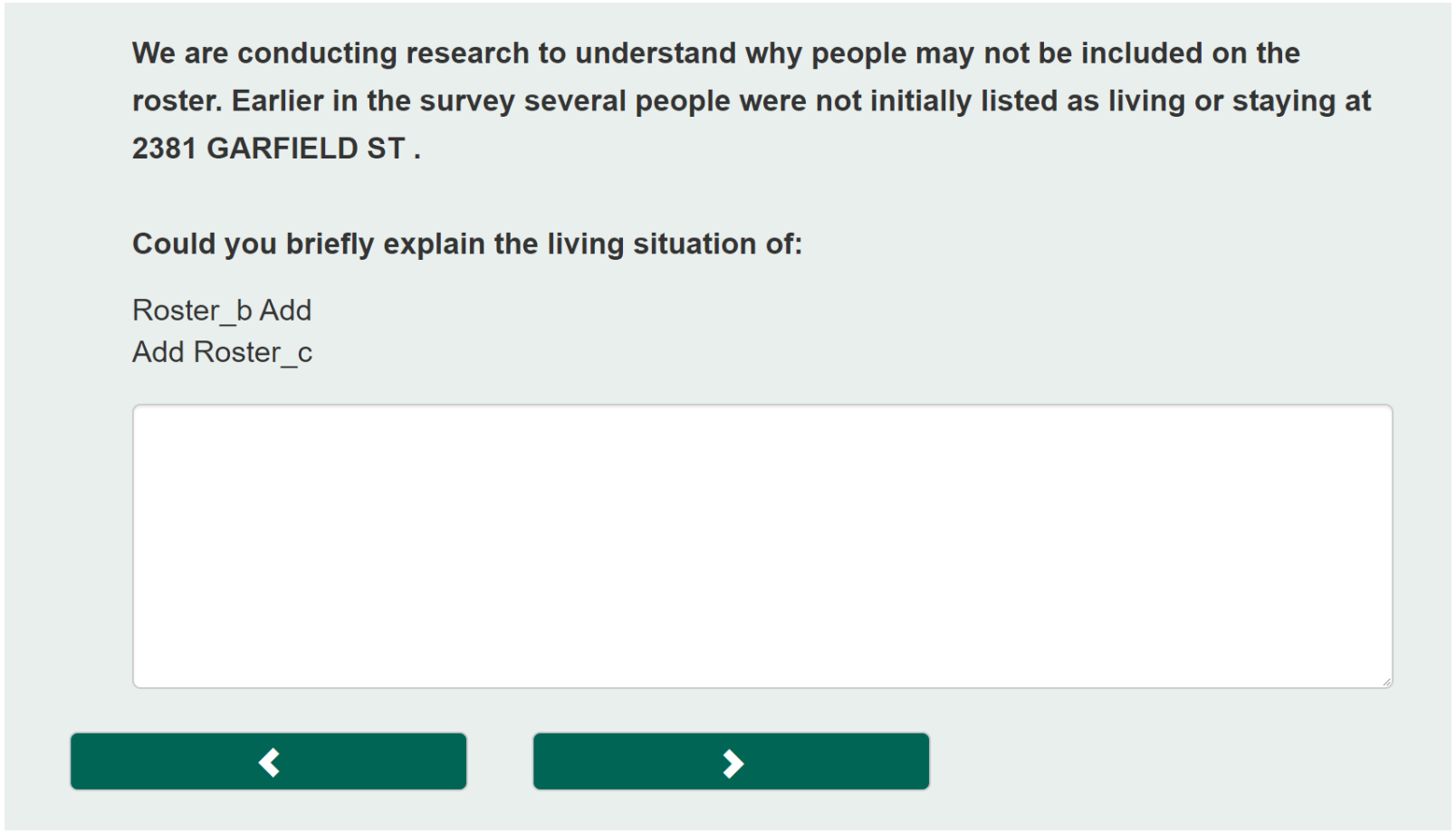
A significantly lower percentage of people were ultimately deleted from the roster, across all screens including the first screen, in the Roster Test version

Table 2. Percentage of People Deleted from Roster

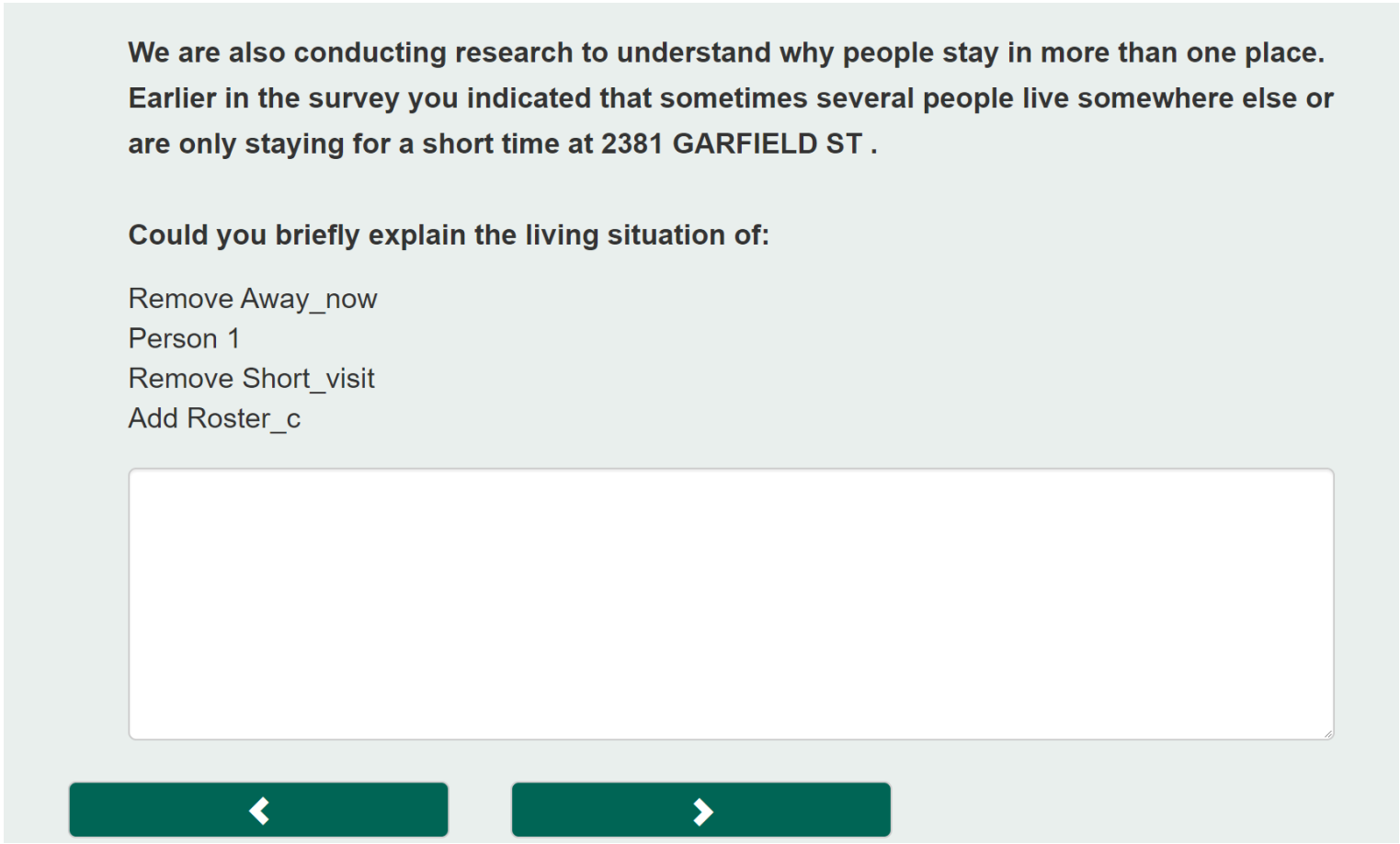
Mode	Control	Test	Difference	P-value
Across Mode	3.0	2.3	-0.6	<0.01*
Internet	3.2	2.5	-0.8	<0.01*
CAPI	2.1	1.9	-0.2	0.52

Source: U.S. Census Bureau, 2022 American Community Survey Content Test. DRB Approval Number: CBDRB-FY23-ACSO003-B0066 Note: Minor additive discrepancies are due to rounding. Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two tailed t-test at the α=0.1 level of significance.

Roster Open Web Probe
(If someone was added after the first roster question)



Roster Open Web Probe
(If someone was removed after the first roster question)



United States® Census Bureau

DRB Clearance Number: DRB/DAO Approval Number: CBDRB-FY24-ACSO003-0010
DRB/DAO Approval Date: April 25, 2024

Number of Weighted Cases Answering Probe

- In the Treatment version there were significantly fewer people who modified their roster after the first question, however after weighting we had about ~3.2 million (SE=182k) people who did.
- In the control version we saw significantly more adds and removals of people after the first question, so we had significantly more people ~4.5 million (SE=285k)

Qualitative Human Coding Scheme

- Four human coders developed a series of codes for the open web probes
- Some sample codes are here:
  - Away for college
  - Adult who cycles through several places on a routine basis (for work, care taking, or another reason)
  - Visiting here or temporarily visiting somewhere else
  - They are a child who splits time multiple places (formal custody or informally)
  - Statement of they are a baby or a child with no other information
  - Statement of family tie or a nonfamily relationship no other information
  - Statement they live or reside here

DRB Clearance Number: DRB/DAO Approval Number: CBDRB-FY24-ACSO003-0010
DRB/DAO Approval Date: April 25, 2024

Qualitative Auto Coding Scheme

- Limited in existing applications that can be used on Title 13 Data
- Used an internal tool created by Curtiss Chapman TopNTool
- Provided word frequencies, and generates dichotomous variables for common words being present
- Then used top 50 words to quickly create categories
- Leverages an automated feature to help one human coder develop codes/themes

Human Code Reason Ranking (weighted)

	Top 5 Code Reasons for Treatment Cases	Percentage (SE)
1	Attending College	61.2 (3.2)
2	Child in split custody or who lives multiple places informally	8.0 (1.9)
3	Baby/ young child (under 4)	5.4 (1.6)
4	Statement of live or reside	3.7 (1.1)
5	Away for a job/ work, situationally here for job/work	2.8 (1.1)

	Top 5 Code Reasons for Control Cases	Percentage (SE)
1	Attending College	46.1 (2.8)
2	Adult who splits time multiple places (not for work/job or medical reasons)	7.1 (1.2)
3	Visiting but seems to live somewhere else, temporarily away/ traveling/ visiting somewhere else	7.1 (1.8)
4	Statement of family tie or lack thereof	4.7 (0.9)
5	Recent mover	4.5 (1.2)

Auto Code Reason Ranking (Weighted)

	Top 5 Code Reasons for Treatment Cases	Percentage (SE)
1	Attending college	61.8 (3.1)
2	Statement of live or reside	15.3 (2.1)
3	Unable to code*	10.0 (2.2)
4	Statement of family tie	4.8 (1.2)
5	Temporal words (part, year, summer, months)	4.8 (1.2)

	Top 5 Code Reasons for Control Cases	Percentage (SE)
1	Attending college	46.5 (2.6)
2	Statement of live or reside	17.9 (2.0)
3	Unable to code*	16.0 (1.8)
4	Visiting	7.3 (1.3)
5	Temporal words (part, year, summer, months)	6.2 (1.4)

United States® Census Bureau

*Unable to code with current methodology, can be coded by humans

Conclusions

- The new roster wording was adapted into the ACS production instrument in 2025
  - It demonstrated evidence of less respondent confusion
  - It had similar overall demographic coverage to the control
- In this instance the human coders resulted in much more nuanced and useful conclusions
  - The auto coder is very rudimentary and some AI applications would likely result in more useful codes
  - When human resources are unavailable the auto coder can give a good place to start with codes

United States® Census Bureau

Recommendations

- Human coders with the level of experience and skills are unlikely to be available to most projects with large scale open ended text analysis
  - AI is potentially a great resource in cases where it can be used (with human supervision and training)
  - A rudimentary count with human coding can still have value and one researcher would be able to more carefully refine codes on a large number of cases
    - Removing the most common code, such as college, may allow for more patterns to be discovered, more training and code refinement might also help
    - Future studies we are planning will employ more sophisticated LLMs to code similar open ended text

United States® Census Bureau

Acknowledgements

- We would like to thank all the people who helped to design the new roster procedure as well as Curtiss Chapman for creating the TopNTool

Questions or Comments?

Kathleen.m.kephart@census.gov