Nonresponse Bias Monitoring and Interventions in the American Community Survey

Jonathan Eggleston (Presenter), Stephanie Coffey, Kendall Houghton, Carl Lieberman
U.S. Census Bureau

FCSM, Session B-2
Thursday, October 26, 2023

Any opinions and conclusions expressed herein are those of the author and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7531477: CBDRB-FY23-0260, CBDRB-FY23-0518, CBDRB-FY2023-CES005-029).

United States® Census Bureau

Introduction

- Over past few years, Census has ramped up its effort to use administrative data to improve the representativeness of survey samples
- However, much of this work has entailed
  - Research on surveys years after they are released
  - Nonresponse and weighting work during data processing (after all the data is collected)
- Not as much effort has been applied to use administrative data to monitor or improve representativeness *during* data collection

United States® Census Bureau

ACS CAPI Workload Optimization

- Effort to innovate how the American Community Survey conducts its nonresponse follow-up operations
- Many projects designed to balance ACS field costs and data quality
- Part of this effort has entailed setting up a server space that has both
  - Administrative data
  - ACS daily response outcomes
- Benefit: New improved framework for using administrative data to **monitor** and **improve** data quality *during* data collection
- Rest of talk: Some highlights from this work

United States® Census Bureau

Administrative Data

- Match the following administrative data to addresses sampled for the ACS
  - IRS 1040 and 1099
  - SSA program benefit data
  - Demographic data from 2010 Census, 2020 Census, and SSA
  - Third-party home value data
- Have these data not only for many respondents, but also ACS nonrespondents and not-yet respondents.

United States® Census Bureau

1. Monitoring Data Quality

We use these administrative data to monitor quality during data collection. Our research questions are:

1. How has nonresponse bias changed recently in the ACS?
2. How does nonresponse bias vary across the country?

Our analysis looks at the representativeness of the ACS sample *before* the weighting adjustments done during data processing

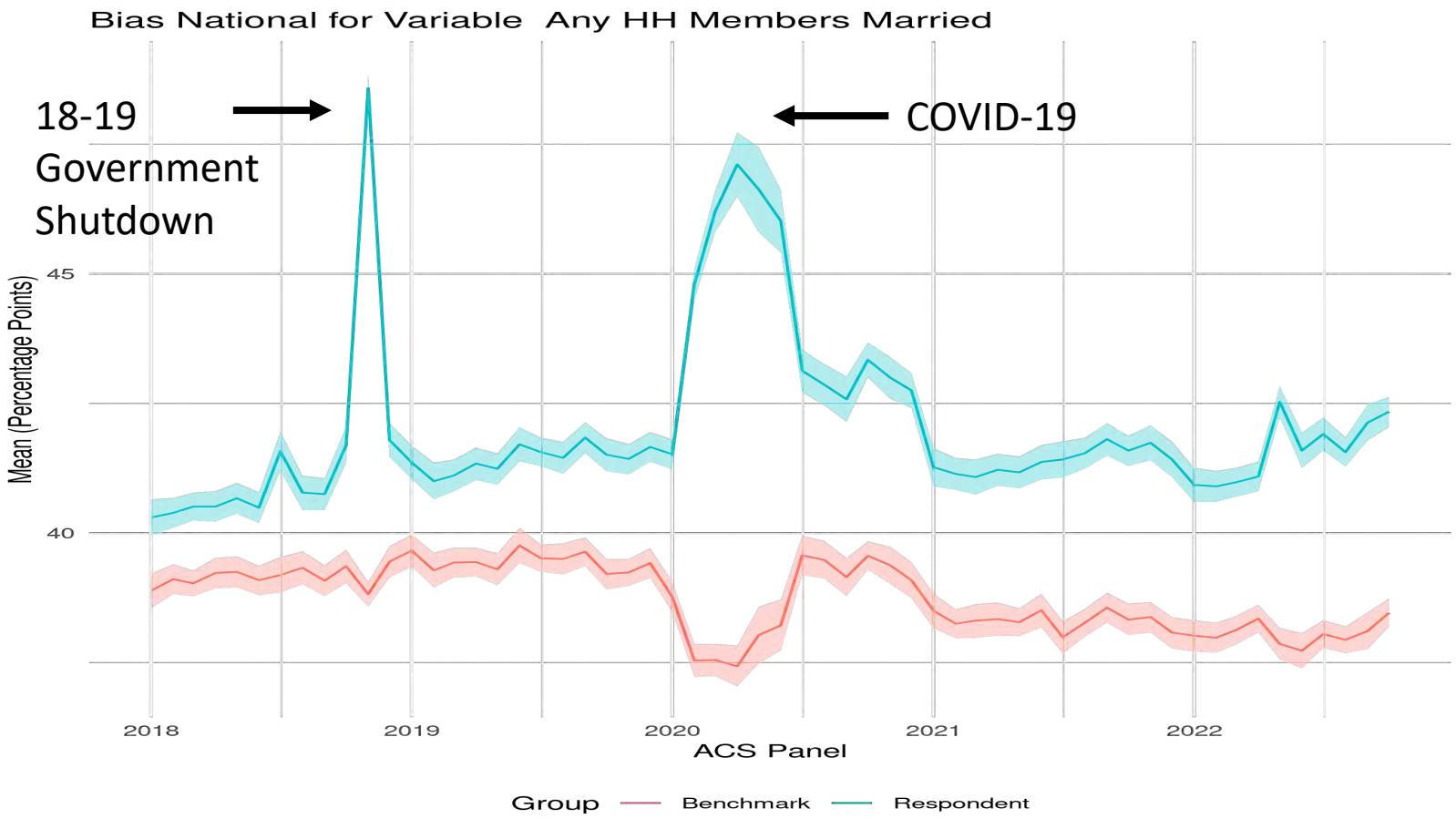
United States® Census Bureau

Nonresponse Bias – National Level

- To show how representative the respondent sample is, we used our administrative and sample frame data to compare statistics between:
    1. All occupied housing units-including nonrespondents. This is our measure of the benchmark
    2. Just respondents
- Bigger gap between #1 and #2 (above)→ Bigger nonresponse bias, before any weighting adjustments

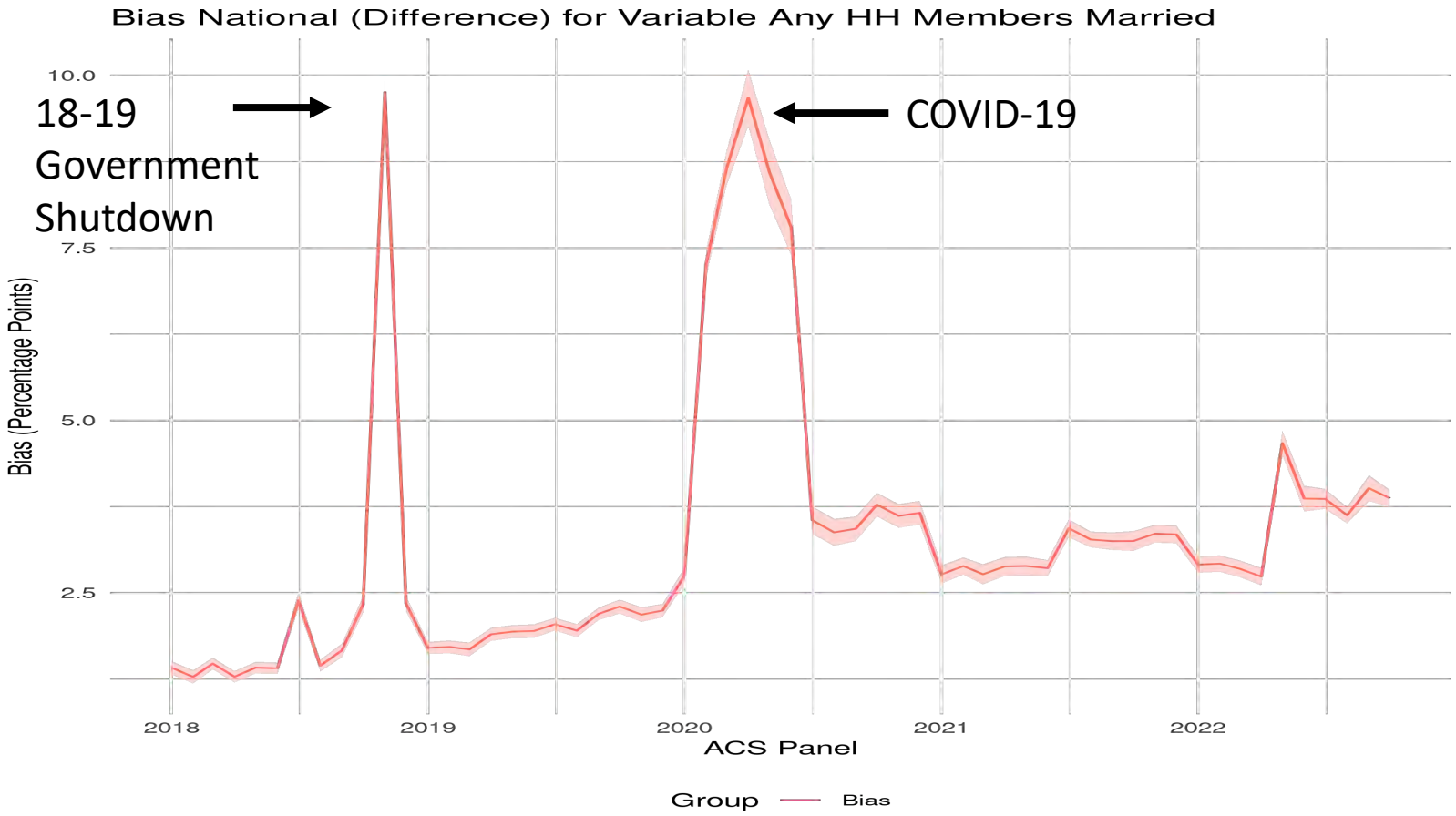
United States® Census Bureau

Married Proxy- Filing Status from IRS 1040

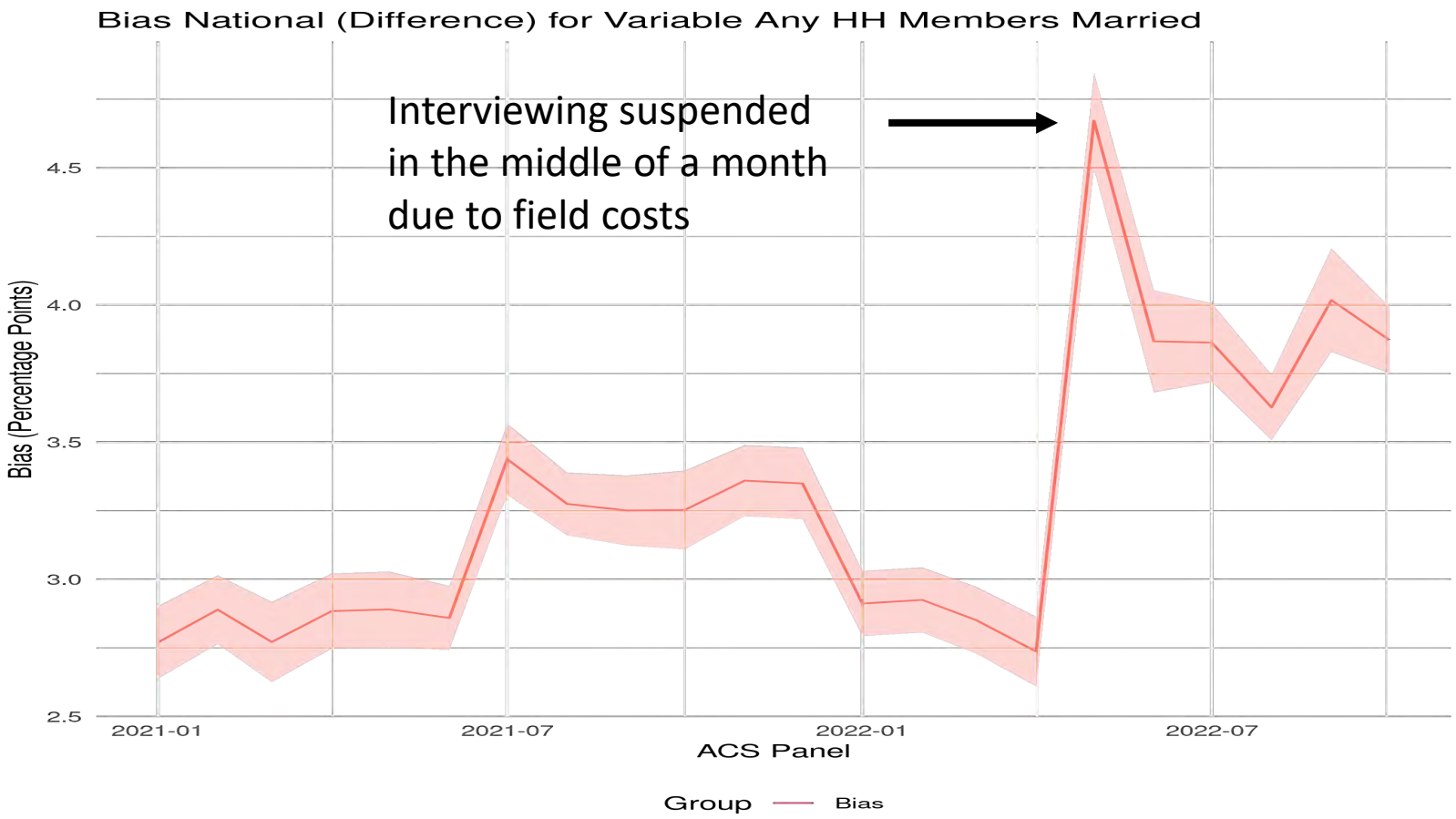


United States® Census Bureau

Married Proxy- Filing Status from IRS 1040



Married Proxy- Filing Status from IRS 1040

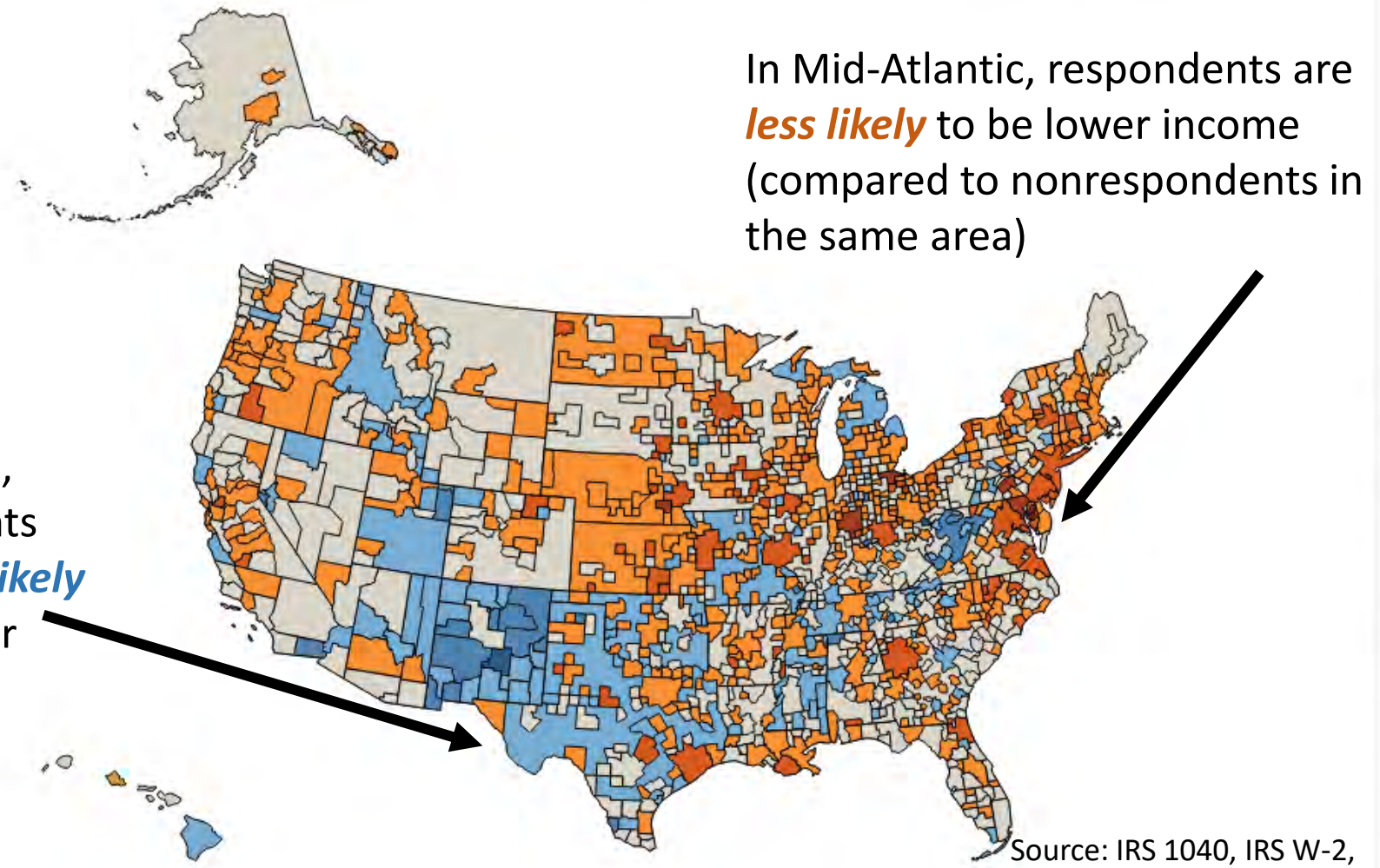


Bias Maps

- To show how representative the respondent sample is, we used our administrative and sample frame data to compare statistics between:
  1. All occupied housing units-including nonrespondents. This is our measure of the benchmark
  2. Just respondents
- Bigger gap between #1 and #2 (above)→ Bigger nonresponse bias, before any weighting adjustments
- For each CBSA (or parts of the state not in any CBSA) we plot $\bar{Y}^R - \bar{Y}$ for various variables

United States® Census Bureau

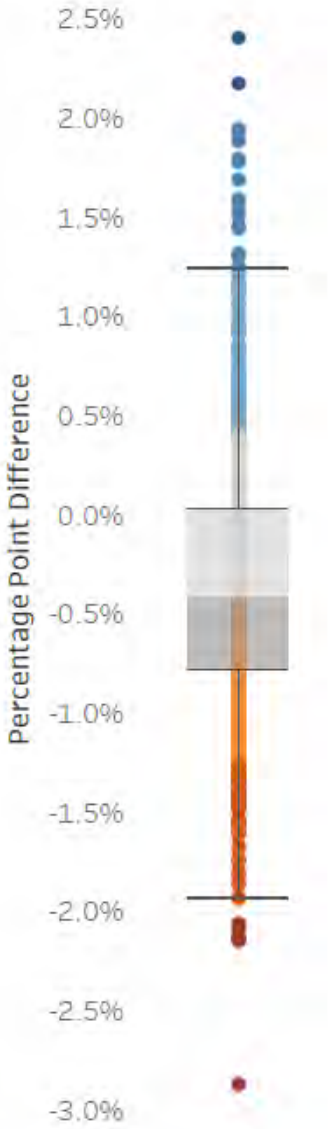
Nonresponse Analysis of Household Administrative Data Income of Less than $40,000



In Mid-Atlantic, respondents are ***less likely*** to be lower income (compared to nonrespondents in the same area)

In rural Southwest, respondents are ***more likely*** to be lower income

Source: IRS 1040, IRS W-2, IRS 1099-R, SSA SSR, SSA MBR



Percentage Point Difference

-2.9% 2.9%

2. Improving Data Quality

- CAPI Optimization: Prioritizing Contacts in Geographic Areas
  - Bias statistics used along with other criteria to decide which areas of the country should have shorter windows of nonresponse follow-up
  - Effective January – June 2023
- CAPI Optimization: Adapting Contact Efforts
  - Applying adaptive survey design to the ACS: Discussed in more detail in session C-4
  - Effective starting July 2023 (full implementation) to present
  - Administrative data used as inputs for creating “balance propensities” that summarize the representativeness of the sample into one continuous variable

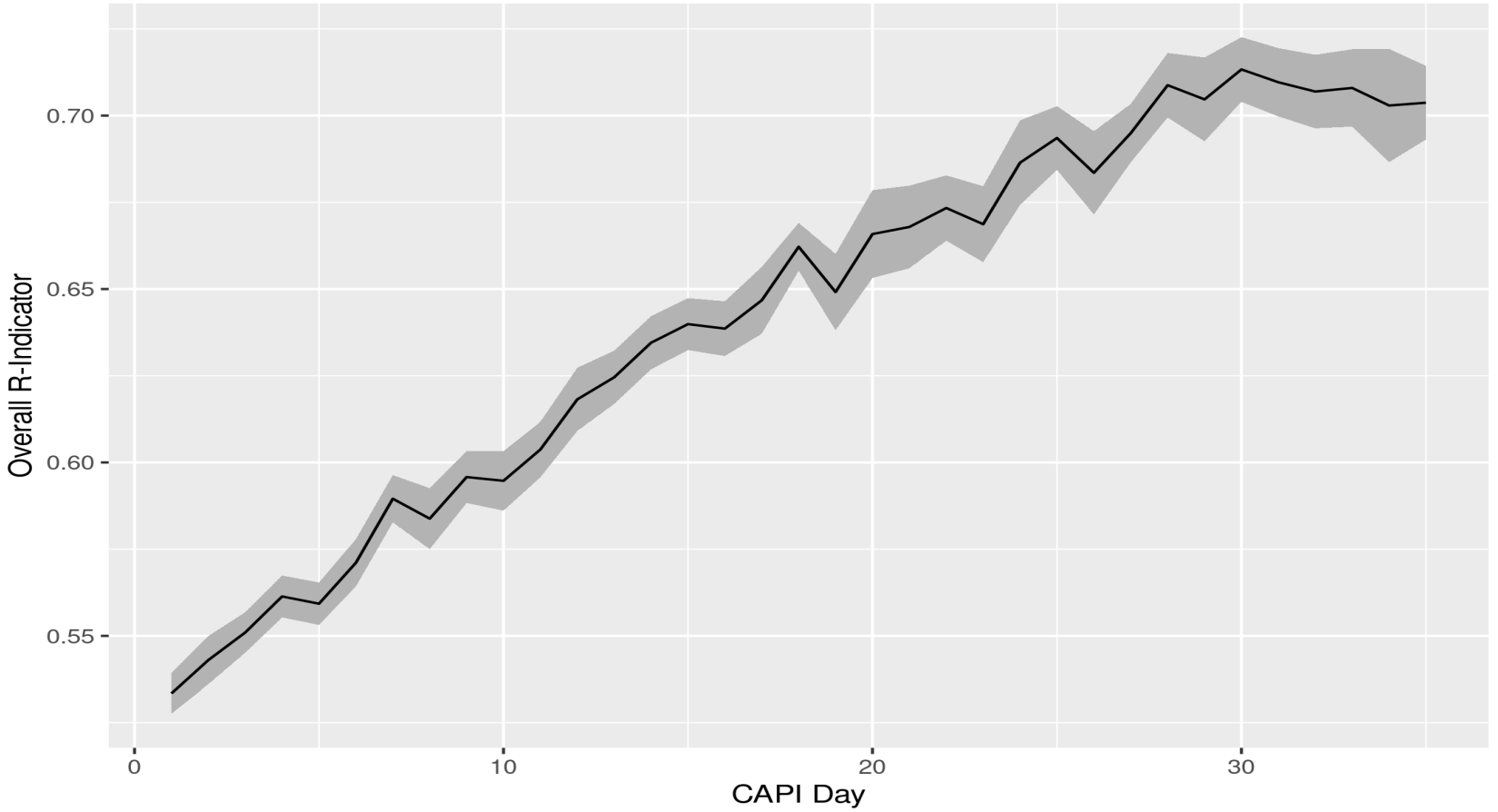
United States® Census Bureau

What has this information been used for?

- Quality-monitoring
  - Use admin data to estimate balance propensities: P(response|observables)
- Compute R-indicators for quality of data/effect of intervention
  - Groups consistently underrepresented? Quality affected by intervention? Etc.
  - Idea: want low variation in balance propensities
  - Overall data representativeness
  - Which subgroups are over/under-represented
  - Which variables are most imbalanced

United States® Census Bureau

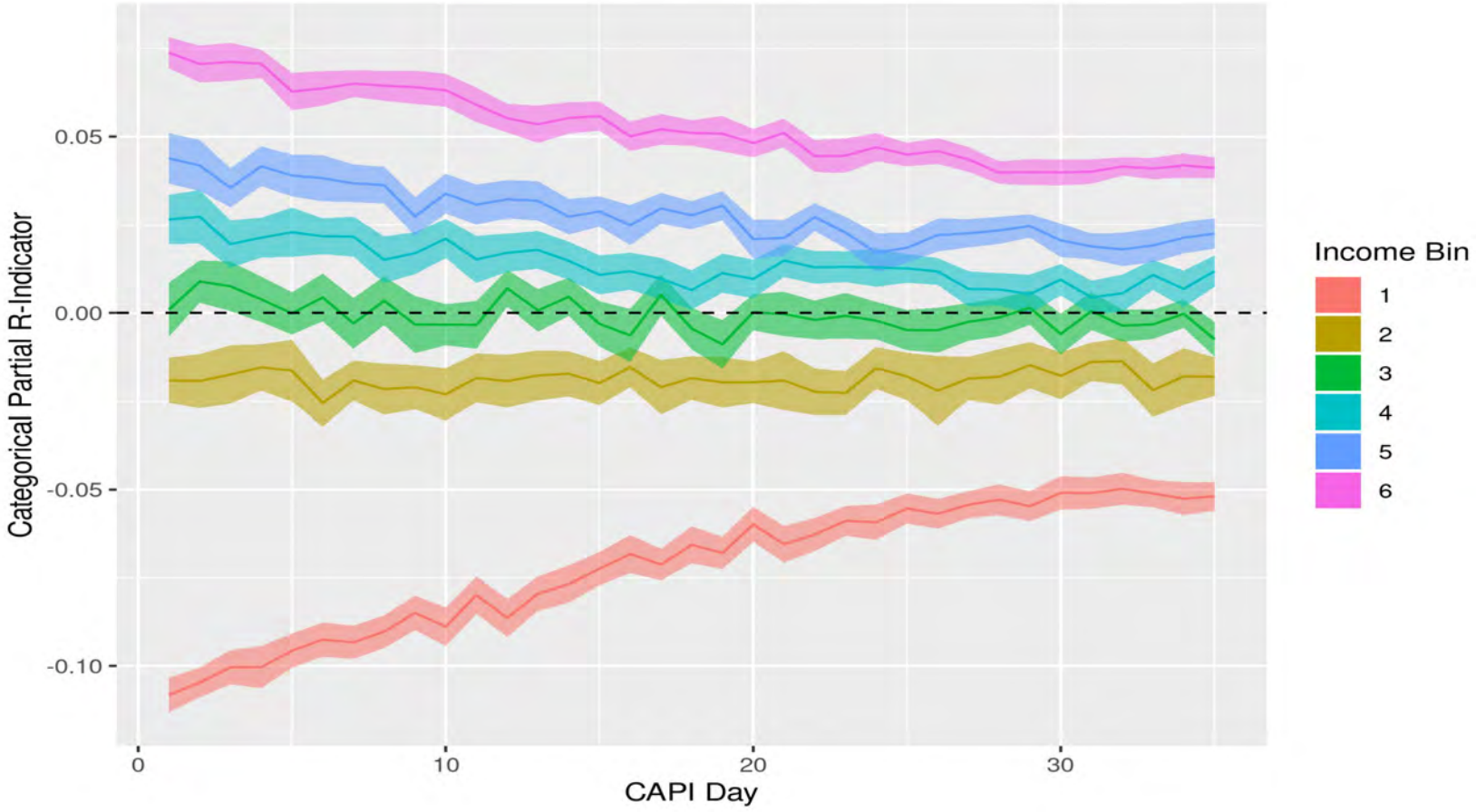
Overall R-indicator



Source: September 2022 ACS Panel + Administrative Data

United States® Census Bureau

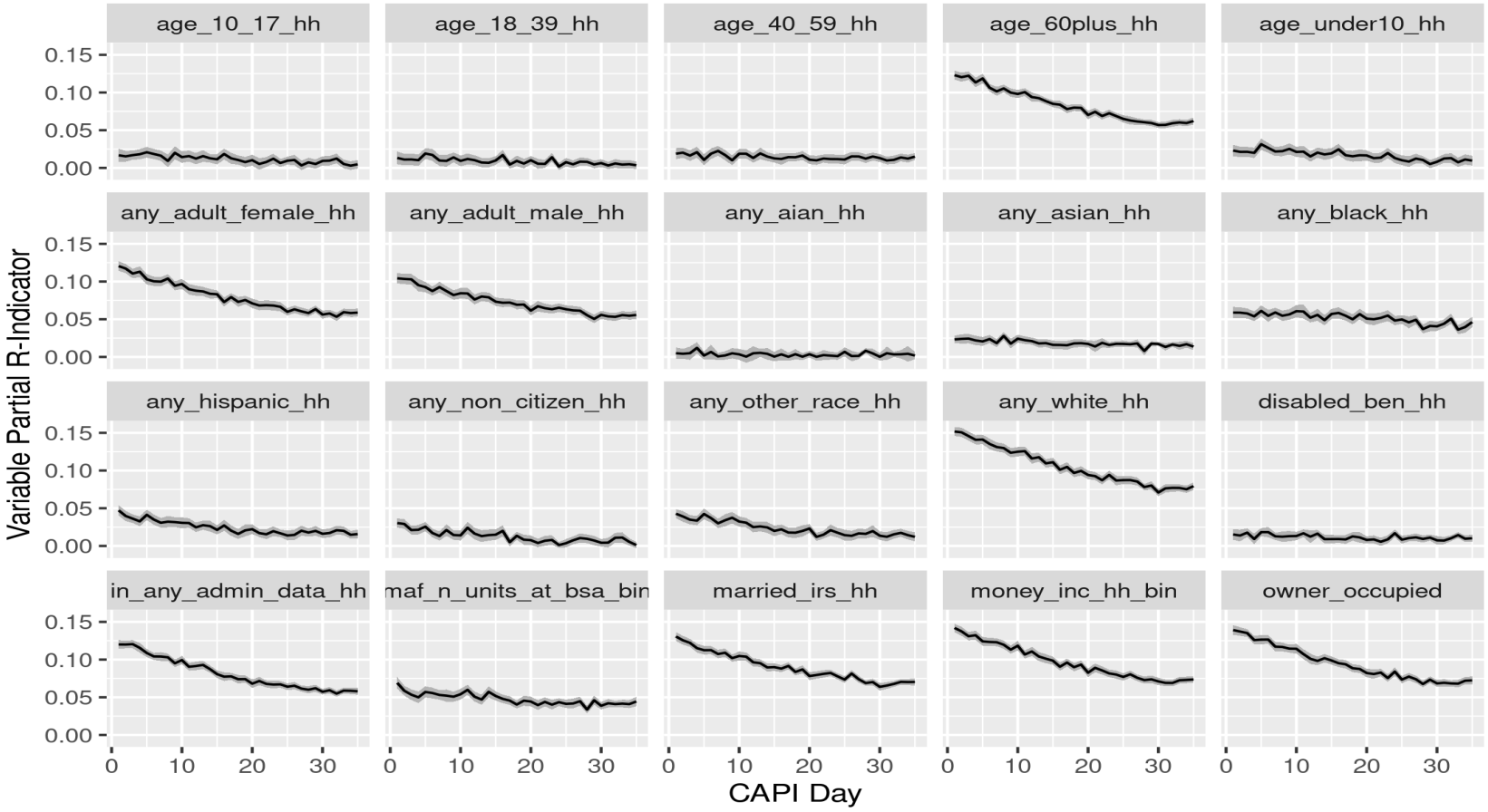
Partial R-indicator for Income Categories



Source: September 2022 ACS Panel + Administrative Data

United States® Census Bureau

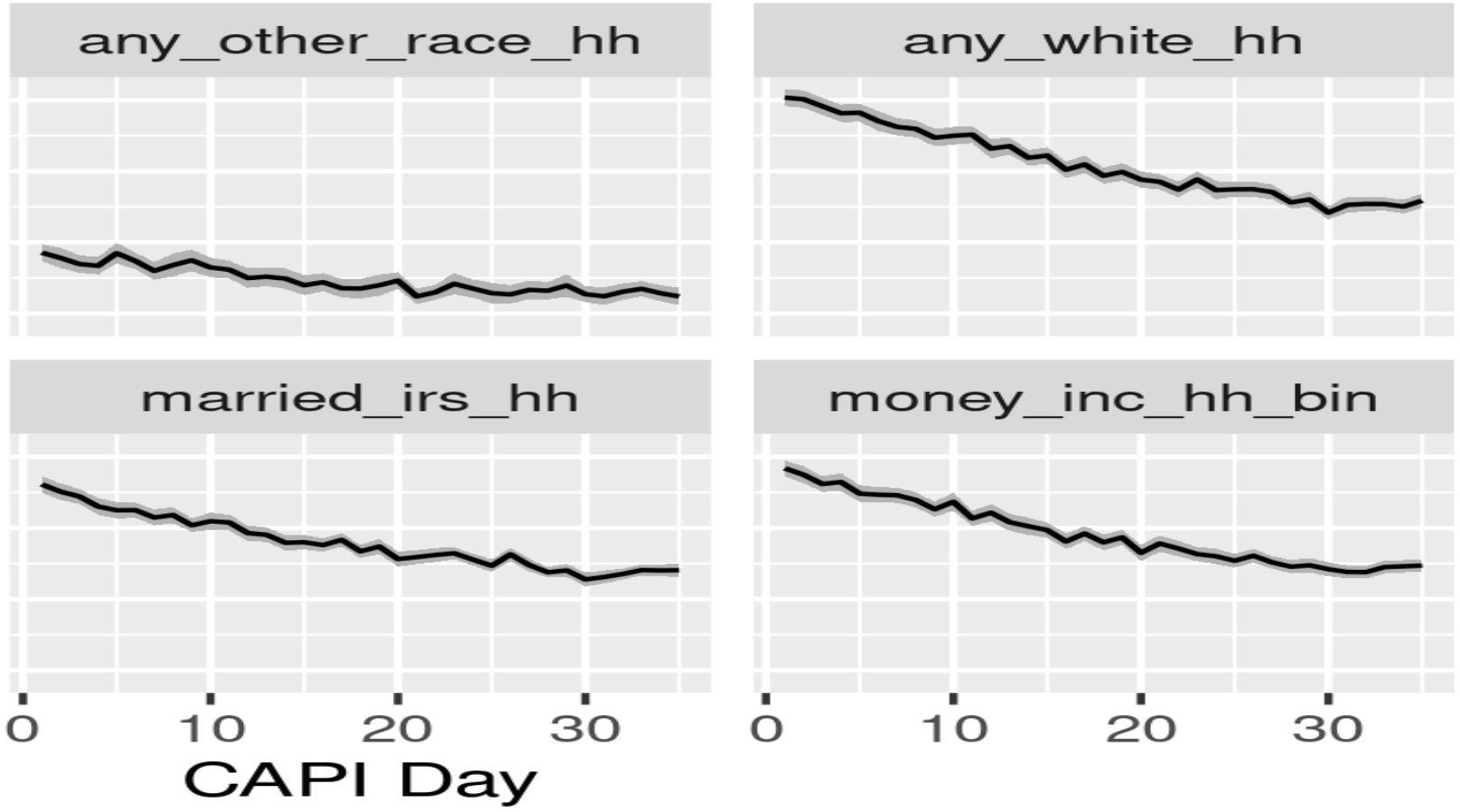
Partial R-indicator for Specific Variables



Source: September 2022 ACS Panel + Administrative Data

United States® Census Bureau

Partial R-indicator for Specific Variables



United States® Census Bureau

My Contact Information

Jonathan Eggleston
Senior Economist
Survey Improvement Technical Lead

Survey and Economic Research Group
Center for Economic Studies
U.S. Census Bureau

Office: 301.763.2357
jonathan.s.eggleston@census.gov

United States® Census Bureau