

# National Experimental Wellbeing Statistics (NEWS)

Combining Survey and Administrative Data to Improve Income and Poverty Statistics

## Estadísticas nacionales experimentales de bienestar:

combinación de datos administrativos y de encuestas para mejorar las estadísticas de ingresos y pobreza

CEPAL seminar on reconciliation of information sources for the  
measurement of income distribution

October 12, 2023

Adam Bee, Joshua Mitchell, Nikolas Mittag, Jonathan Rothbaum,  
Carl Sanders, Lawrence Schmidt, and Matthew Unrath

*Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7524052, Disclosure Review Board (DRB) approval number: CDRB-FY23-SEHSD003-025).*

# What is NEWS?

- Rethink how we produce income and resource statistics
  - What is the best possible estimate given all the data currently available at Census for a given income/resource statistic?
- Address multiple sources of bias simultaneously
  - Apply research on addressing each

# How Does NEWS Do This?

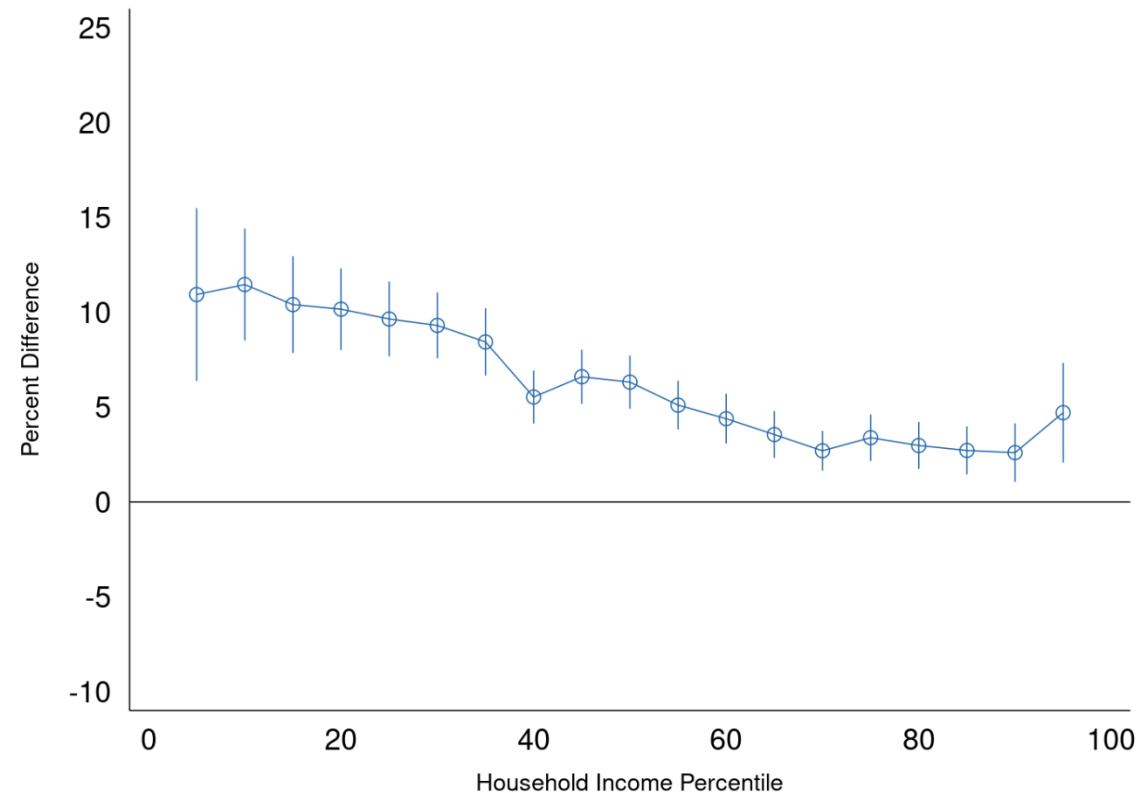
- Pull together all available data: survey, census, administrative records, commercial (third-party) data
  - Often need *linked* data to address bias correctly
- Do everything in a transparent, replicable, evidence-based manner
- Engage research community
  - Will create linked microdata and code database for access in FSRDCs
  - Code will be shared publicly (subject to disclosure constraints)

# What Have We Done?

- Version 1 Release – February 14
  - Proof of concept
    - 1 year
    - Mirror income and poverty releases – money income (no taxes, credits, in-kind benefits)
    - Present methods and approach for feedback
  - Paper and estimates available at
    - <https://www.census.gov/data/experimental-data-products/national-experimental-wellbeing-statistics.html>

# Household Income in 2018

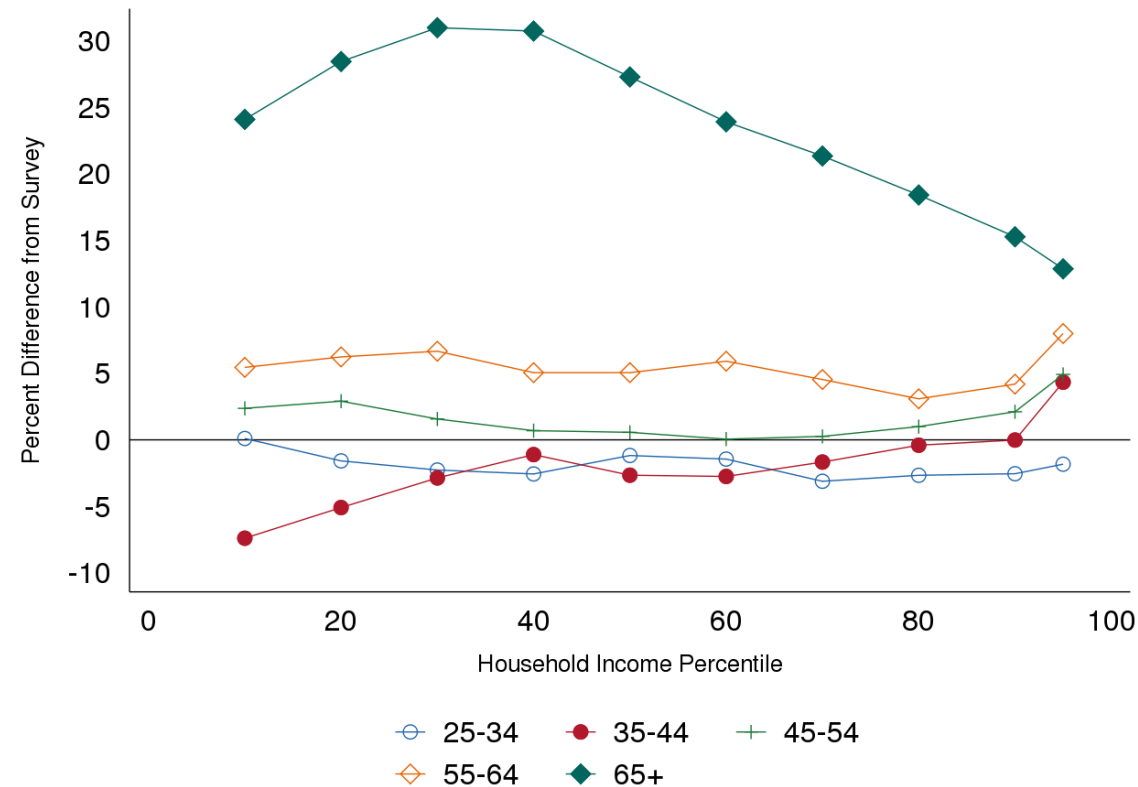
## NEWS Estimate Relative to Survey



Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data.

# Household Income in 2018

## NEWS Estimate Relative to Survey by Age



Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data.

# Measurement Challenges Survey Data

1. Unit Nonresponse Bias
  - Not answering the survey
  - Poverty biased **down** by 0.3-0.5 percentage points during the pandemic (Bee and Rothbaum, 2022)
2. Item Nonresponse Bias
  - Not answering income questions (~45 percent of income in the CPS ASEC is imputed!)
  - Poverty biased **down** by 0.5-1 percentage points (Bollinger et al., 2019; Hokayem et al., 2022)
3. Mis- and underreporting
  - Not answering accurately
  - Poverty biased **up** by 2.5 percentage points for individuals 65+ (Bee and Mitchell, 2017)

Biases can have different signs and magnitudes which can vary by group

# Measurement Challenges

## Administrative Data

### 1. Selection into administrative data

- Not everyone has to file taxes or gets a W-2 or other information return
- Larrimore, Mortenson, and Splinter (2020) estimate poverty from administrative data, but must impute the existence and poverty status of 4-6 million people

### 2. Administrative data “nonresponse”

- Some information not reported that should have been
- Under-the-table jobs without a W-2, for example – 5% of adults in CPS ASEC report wage and salary earnings on the survey with no W-2

### 3. Administrative mis- and underreporting

- Not always 100% accurate
- Unreported tips, underreported self-employment earnings (refer to IRS tax gap analyses)



# Measurement Challenges

## Administrative Data

### 4. Conceptual misalignment

- Administrative not always measuring what we want
- W-2s historically do not have earnings used to pay for health insurance premiums – understate true earnings (Census also doesn't get this information when it's available)

### 5. Incomplete data coverage

- Data not available for individuals or places

### 6. Selection into linkage

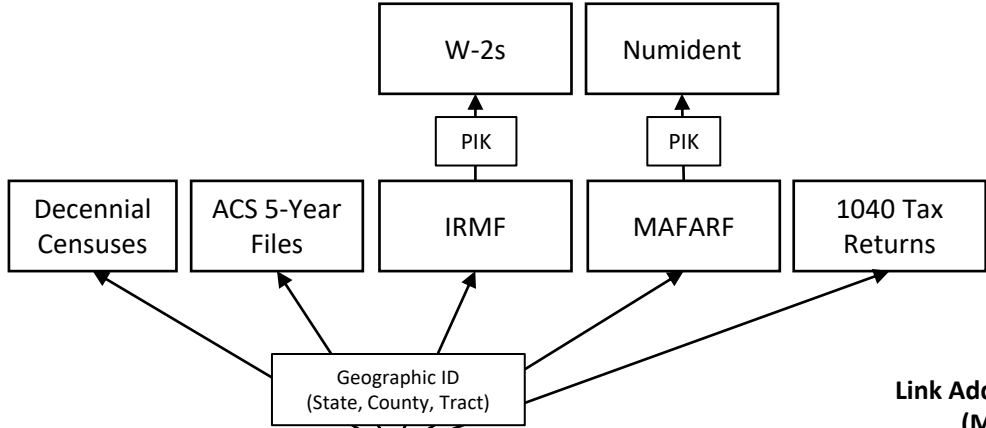
- Not all individuals can be linked across data sources (refer to Bond et al., 2014)

# Addressing the Measurement Challenges

Step	Description	Measurement Challenge	Related Work
Weighting	Use address-level data for all occupied housing units to weight respondent, linked sample to be representative of the target universe of households	Survey unit nonresponse Selection into administrative data Administrative data “nonresponse” Selection into linkage	Rothbaum et al. (2021) Rothbaum and Bee (2022)
Imputation			
Survey earnings	Impute survey earnings conditional on survey and administrative information	Survey item nonresponse	Hokayem et al. (2022)
Admin gross earnings	Impute gross earnings when missing in administrative data	Administrative data “nonresponse” Conceptual misalignment Incomplete data coverage	
Means-tested program data	Impute means-tested program data for states for which administrative data is not available	Incomplete data coverage	Fox et al. (2022)
Nonfiler income	Impute unemployment insurance compensation, interest, and dividends for nonfilers	Selection into administrative data Incomplete data coverage	Rothbaum (2023)
Estimation			
Combine survey and admin earnings	Combine survey and administrative wage and salary earnings according to the NEWS earnings measurement error model	Survey mis- and underreporting Administrative mis- and underreporting	Bee et al. (2023)
Income replacement	Use survey and administrative data, imputed income, and earnings from the measurement error model to construct household and family income	Survey mis- and underreporting Administrative mis- and underreporting	Bee and Mitchell (2017)

# Address-Linked Data (Weighting)

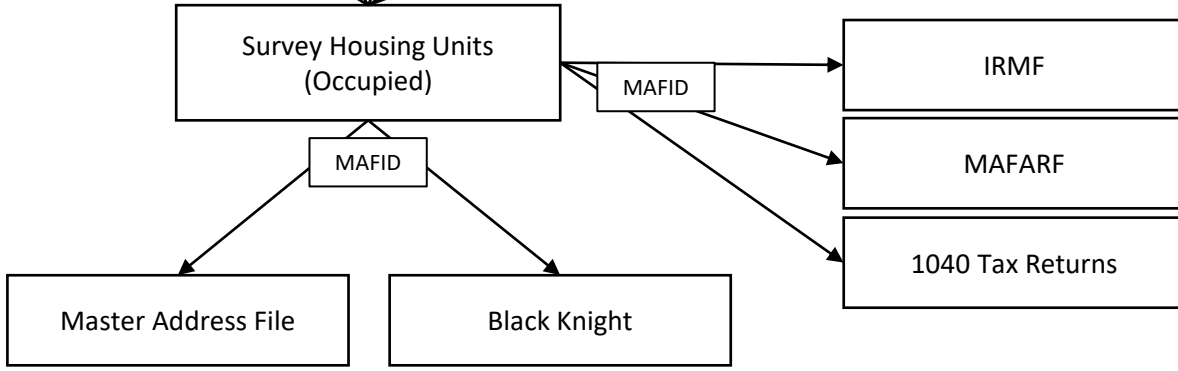
## Geographic Summaries of Characteristics



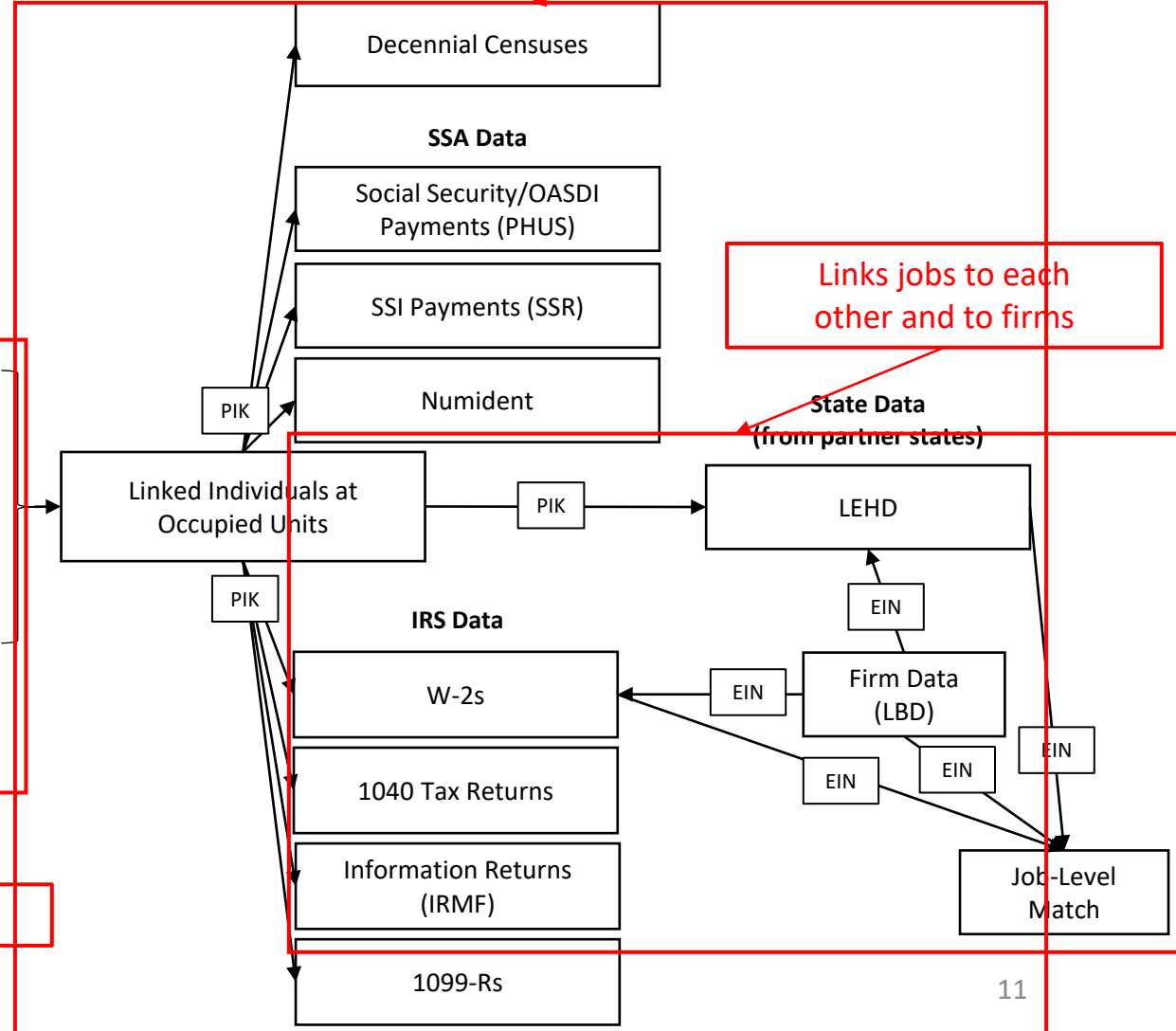
Links by Geography

Links to People in Adreacs at the Addresses

## Link Addresses to People (MAFID → PIK)



## Housing Unit Information



Links jobs to each other and to firms

State Data (from partner states)

Links by Address

# Estimation

## Combining Survey and Admin Earnings

- Five sources of wage and salary earnings information
  1. Survey
  2. W-2s
  3. Detailed Earnings Records
  4. LEHD
  5. 1040 wage and salary

# Different Earnings Sources

## W-2 vs. LEHD

	All Jobs	EIN and Indirect Matches	
		Unmatched Jobs	Share of Implied Total
Total Jobs			
W-2	256,800,000	25,680,000	0.097
LEHD	237,900,000	6,744,000	0.026
EIN Matches	216,100,000		0.820
Indirect Matches	15,040,000		0.057
Implied Total Jobs		263,600,000	

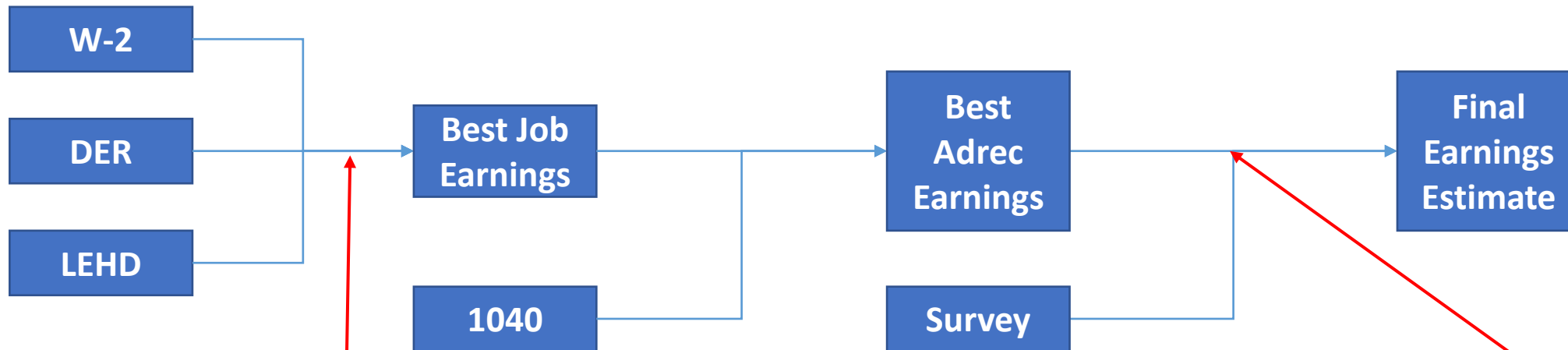
# The Full Picture – Wage and Salary Earnings

1. Use job-level Information to get “best possible” administrative job-level earnings

2. Compare to 1040 to check for missing earnings (at tax-unit level)

3. Compare to survey and decide for which individuals to use adrec or survey earnings

4. Final “best” estimate of earnings for each individual/household



If LEHD is missing (or has apparent data quality issues), impute gross earnings conditional on administrative and survey information for each job (up to 2)

Improve survey imputes

How to combine survey and administrative earnings?

# Different Earnings Sources W-2 vs. Survey Responses



Source: O'Hara et al. (2017) using the 2011 ACS linked to 2010 W-2 records.

# Earnings Measurement Error Model

- Define a function to select whether to use the survey or administrative report ( $y_s$  or  $y_a$ ) conditional on  $x$
- We need assumptions – true earnings unknown
  - Survey responses are unbiased conditional on  $x$ 
    - Prefer the survey if  $E(y_s|x) > E(y_a|x)$
  - Administrative reports capture true variance conditional on  $x$ 
    - Prefer the adrec if survey reports are noisier conditional on  $x$

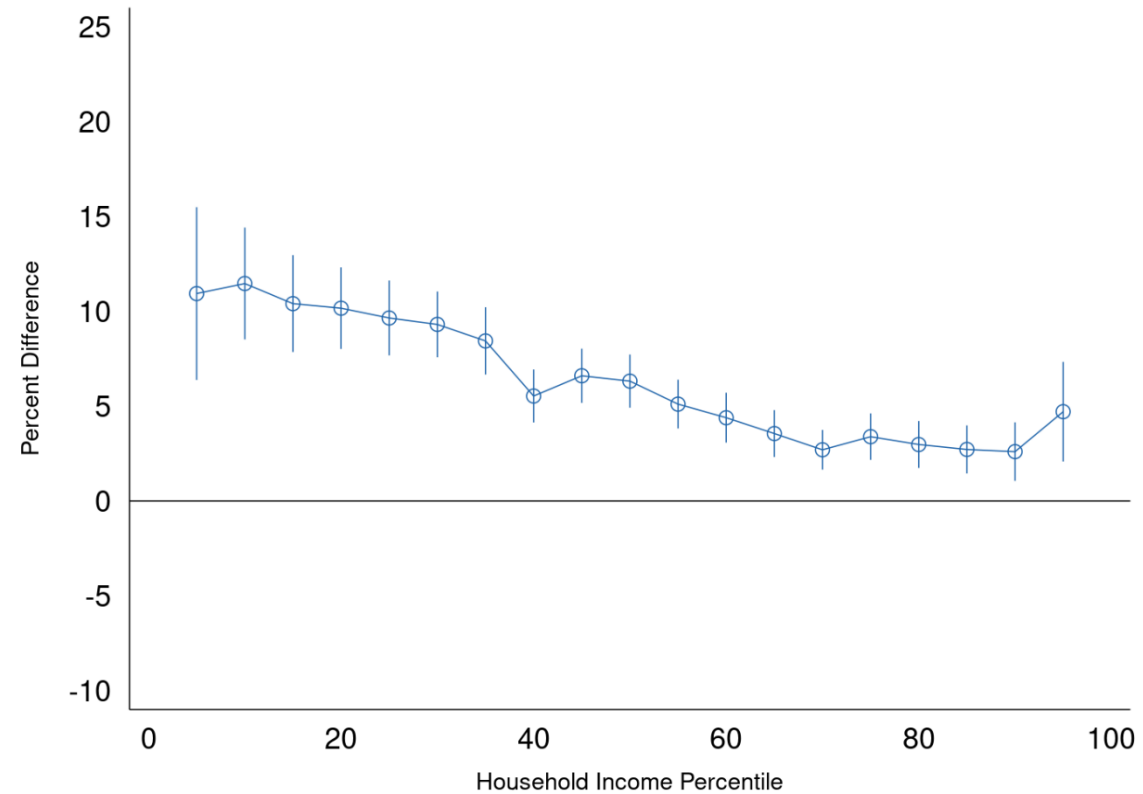


# Survey Earnings Use

- 21 percent of individuals
- More often for
  - Workers in real estate and construction
  - Younger workers (25-44 year-olds)
- Less often for
  - Workers in retail, education, management, and health care
  - Older workers (65+)
  - Black workers

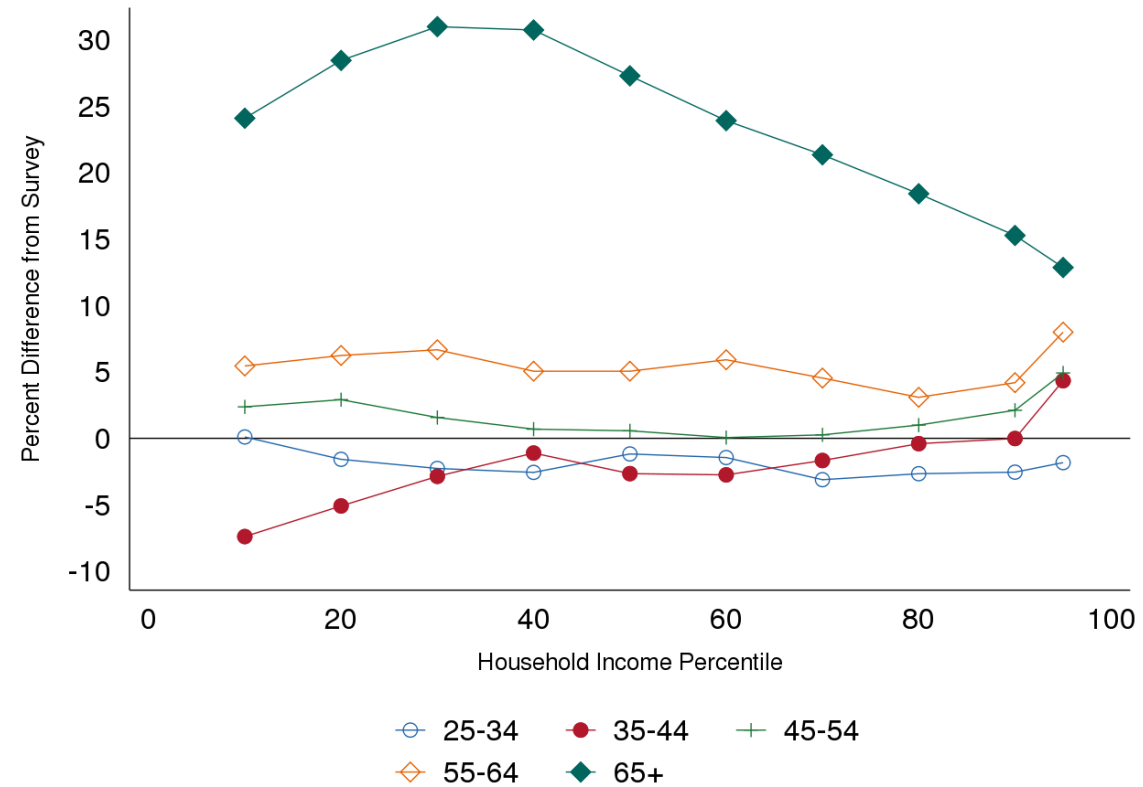
# Results

## Household Income Relative to 2019 CPS ASEC



Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data.

# Household Income Relative to Survey By Age



Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data.

# Future Plans

- More years
  - Not all adreCs are available in all years
  - Not all survey variables are available in all years
- More geographies
  - Use ACS – less detailed information makes combining surveys and adreCs more difficult
- More income/resource concepts
  - Include taxes, credits, and in-kind transfers
  - Supplemental Poverty Measure
- Address more sources of measurement error
  - Self-employment earnings
  - Income at the very top of the distribution (top 0.1%, 0.01%,...)
- Further investigate assumptions, issues for other subgroups of interest
  - Non-citizens, homeless/unhoused (or those with unstable living arrangements), group quarters
- Feedback into surveys to improve questions and processing

# Feedback

Paper and estimates available at:

<https://www.census.gov/data/experimental-data-products/national-experimental-wellbeing-statistics.html>

Please e-mail any comments, concerns, suggestions, and feedback to:

[census.newsproject@census.gov](mailto:census.newsproject@census.gov)