

National Experimental Wellbeing Statistics (NEWS)

Combining Survey and Administrative Data to Improve Income and Poverty Statistics

2024 FCSM Research and Policy Conference
October 23, 2024

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 numbers: **CDRB-FY23-SEHSD003-025** and **CBDRB-FY25-SEHSD003-003**).*

National Experimental Wellbeing Statistics (NEWS) Project

- 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?

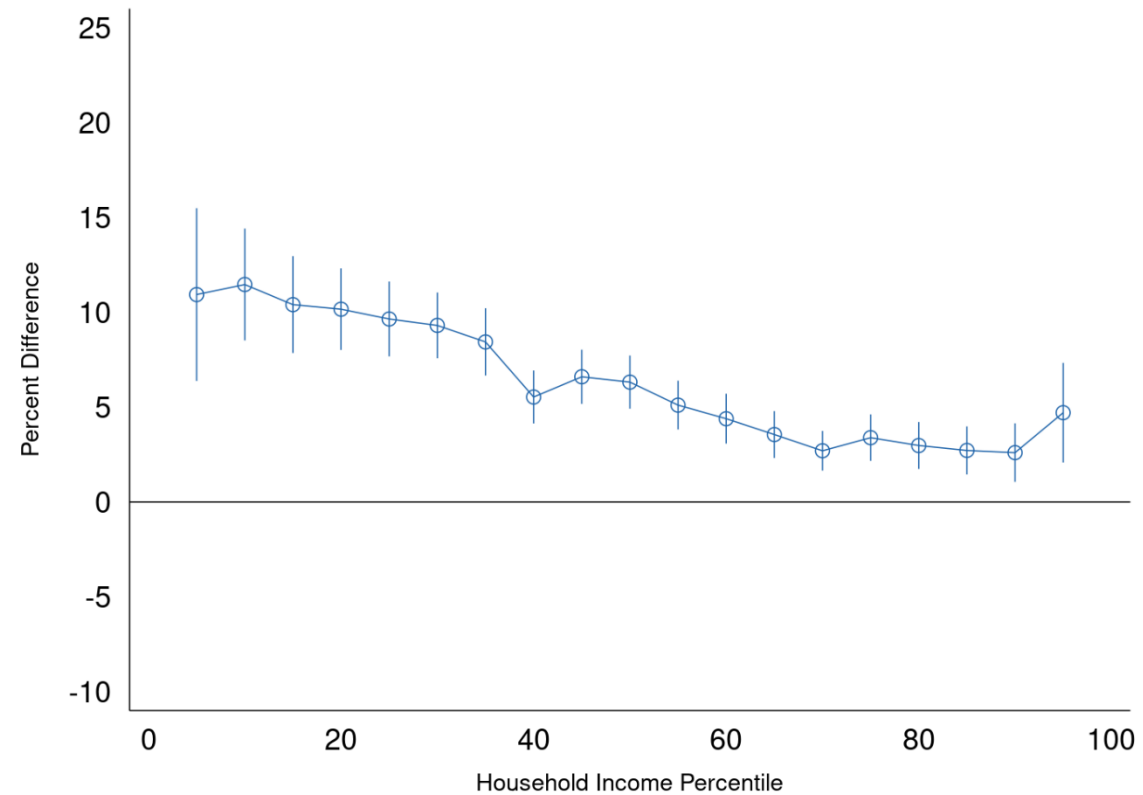
- Pull together all available data: survey, census, administrative records, commercial (third-party) data
 - Often need *linked* data to address bias correctly
- Do everything in a transparent, replicable, evidence-based manner
- Engage research community
 - Will create linked microdata and code database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

What Have We Done?

- Version 1 Release – February 14, 2023
 - 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

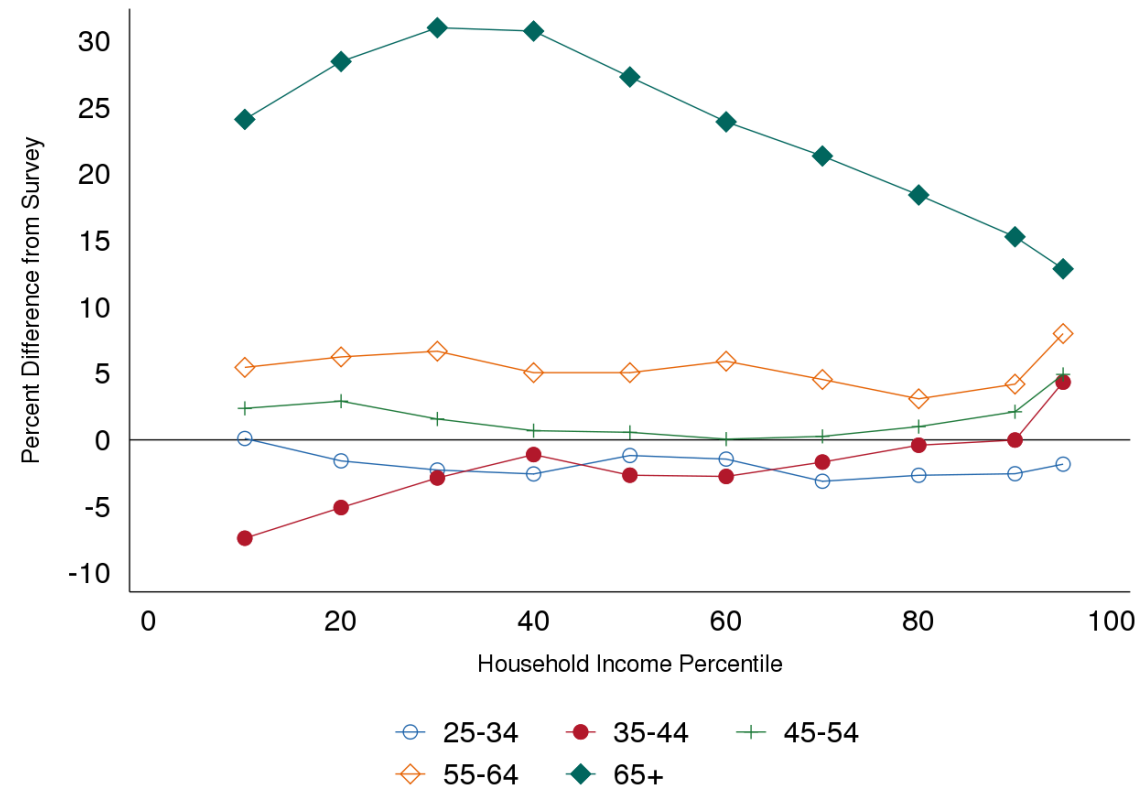
NEWS Estimate Relative to Survey



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
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

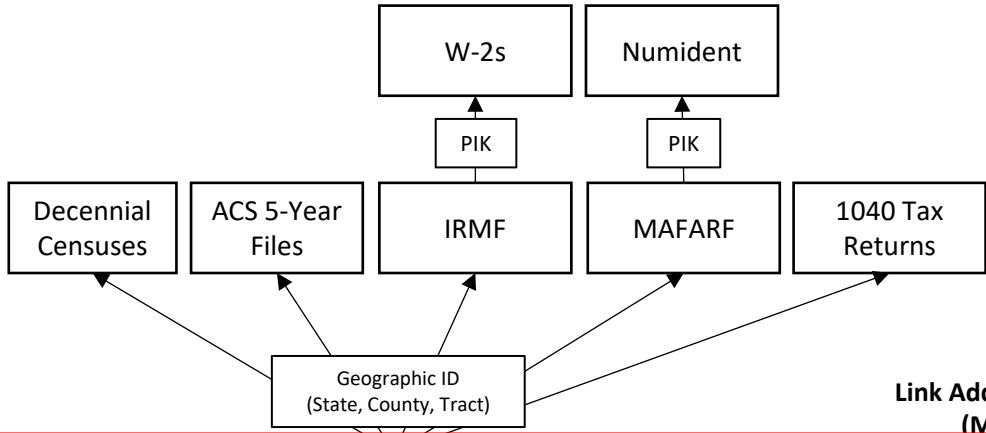
- Not all individuals can be linked across data sources (refer to Bond et al., 2014)

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)

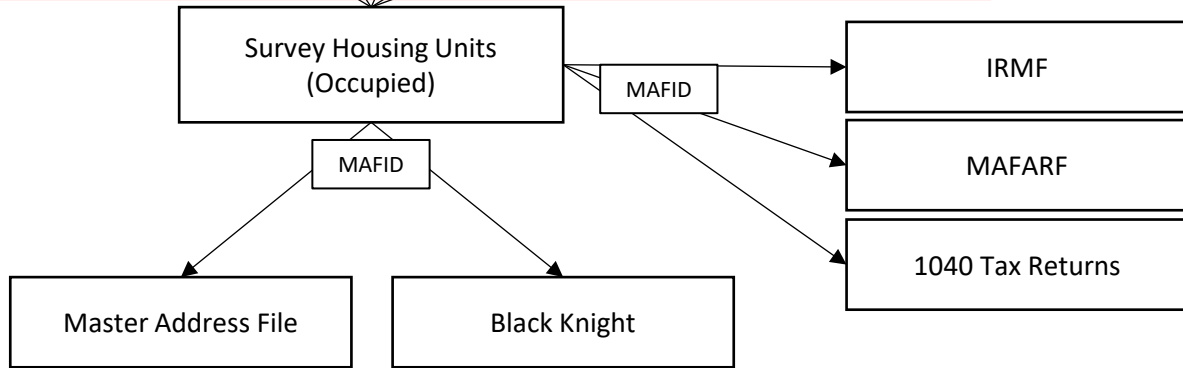
Geographic Summaries of Characteristics



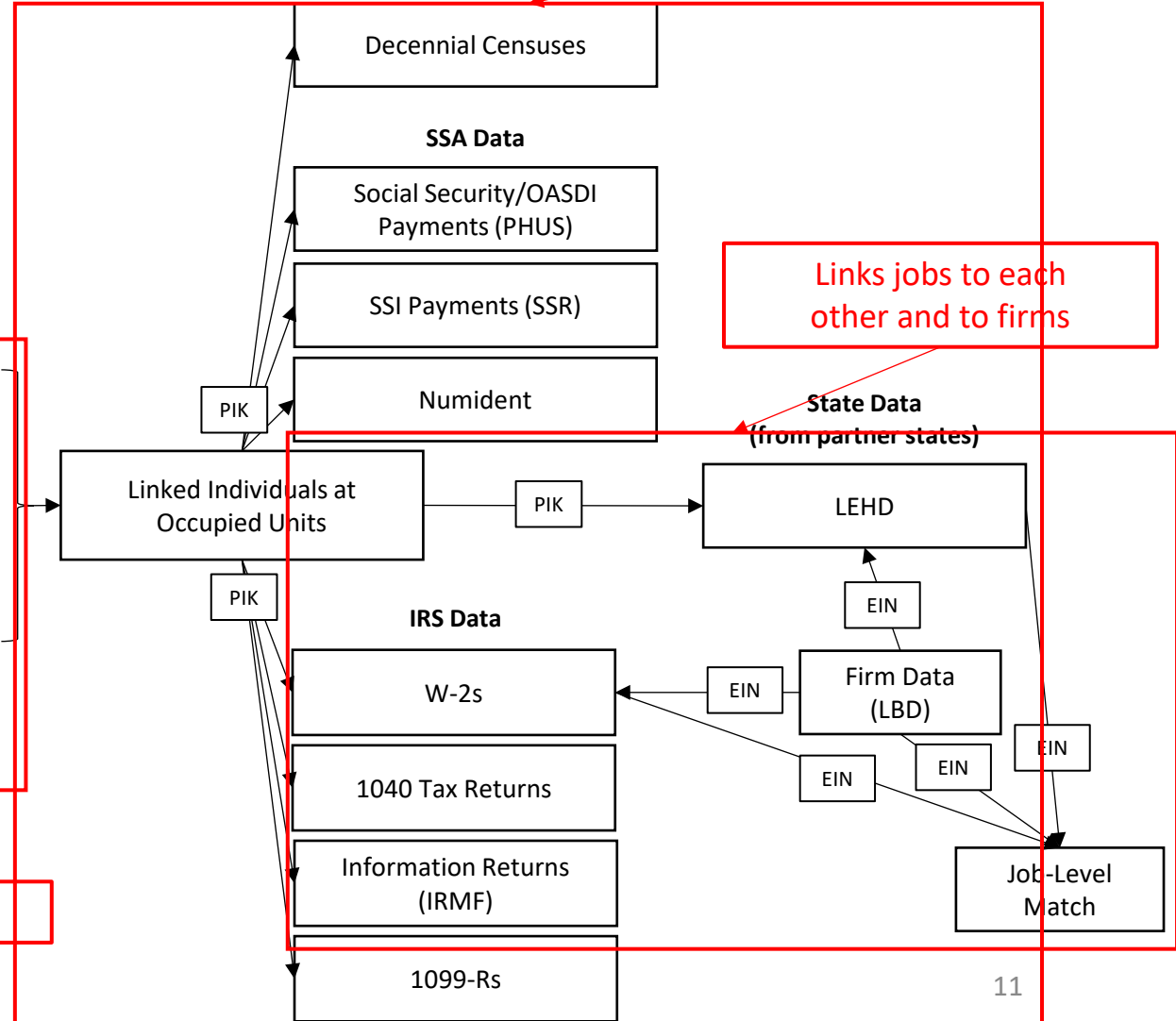
Links by Geography

Links to People in Adrecs 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

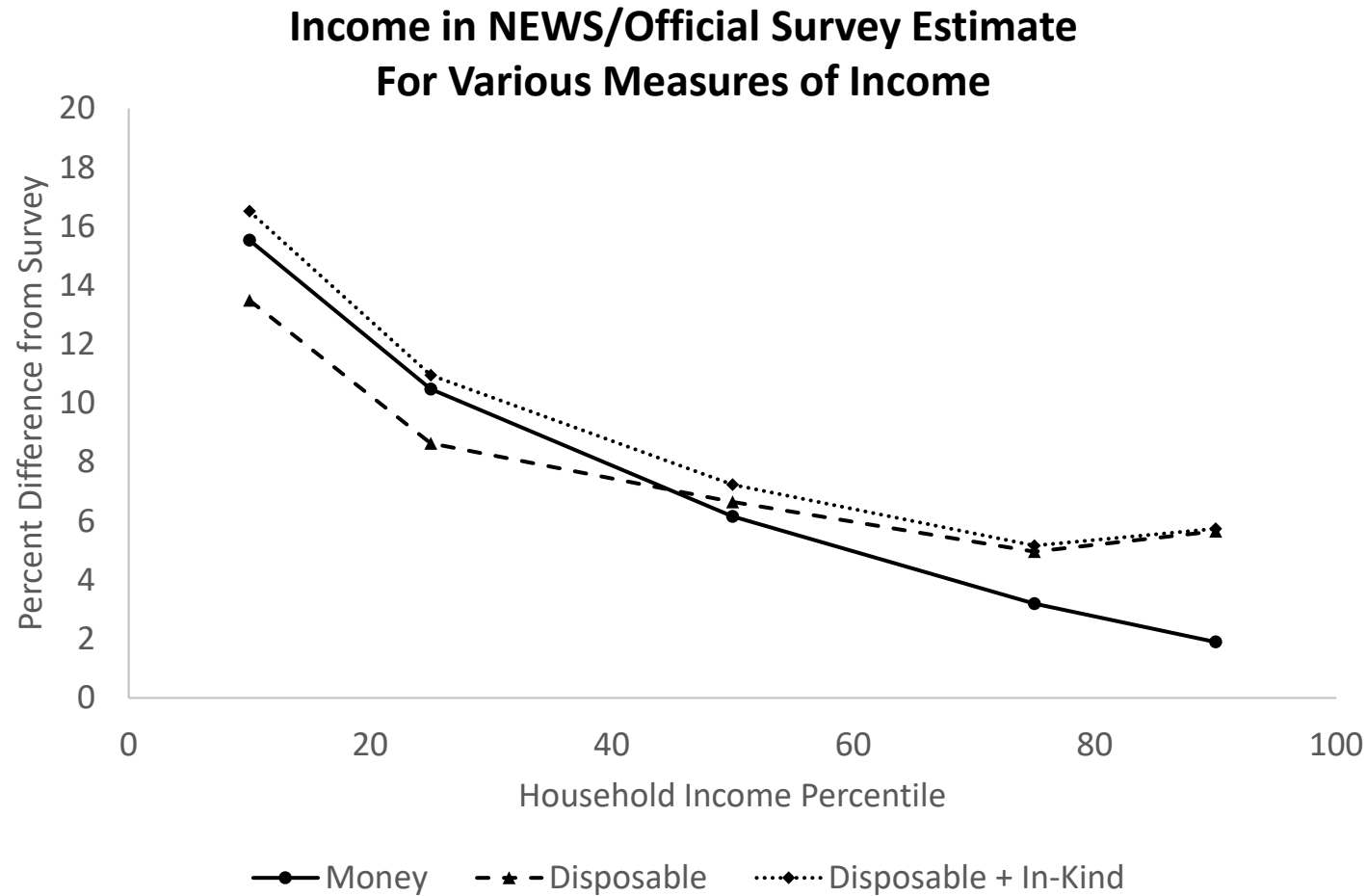
Combining Survey and Admin Earnings

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

What Are We Planning?

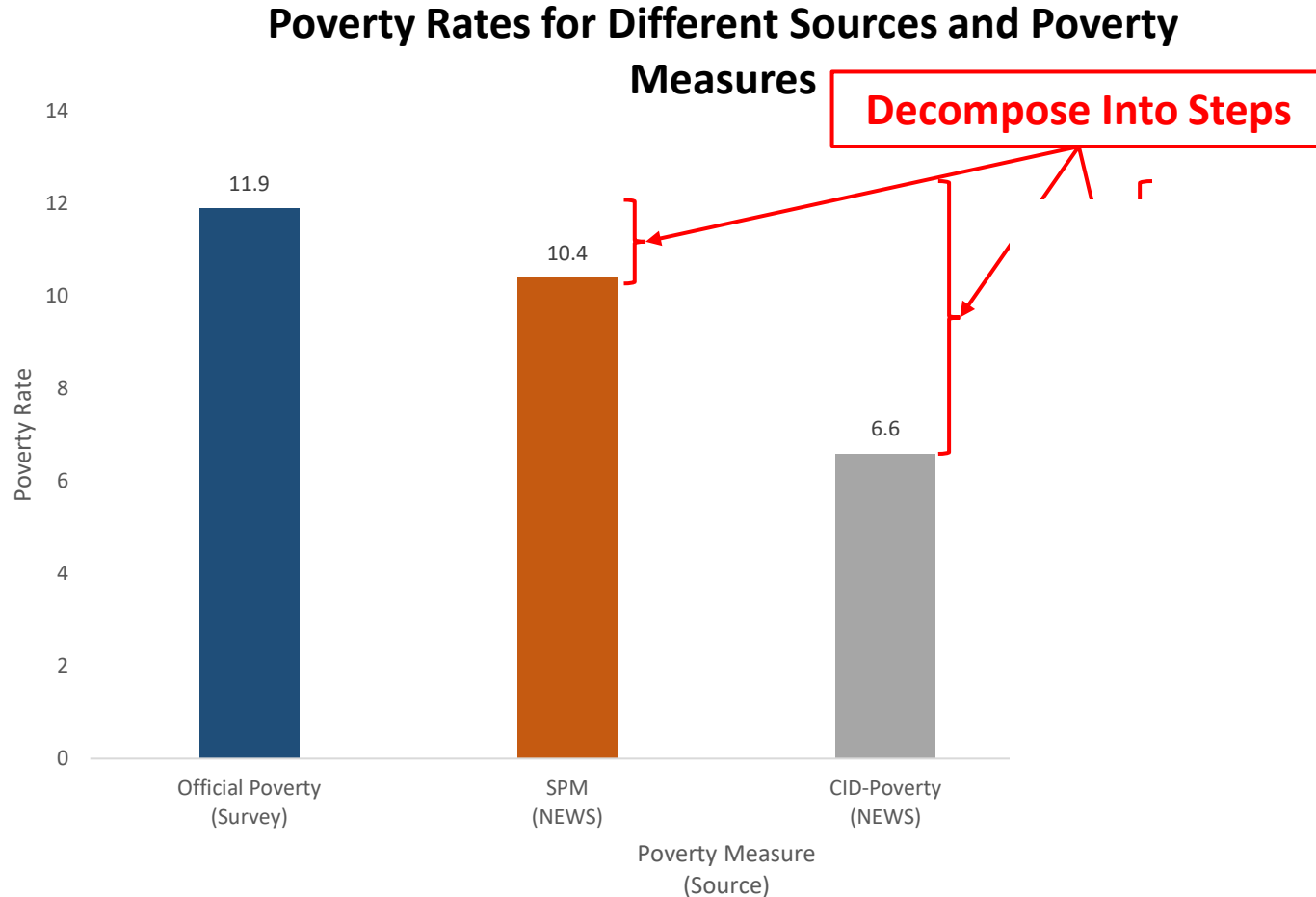
- Release by the end of this calendar year
 - More income/resource concepts - taxes, credits, and in-kind transfers, Supplemental Poverty Measure
 - More years
 - Improved methods
 - Better – integrate machine learning models into imputation, more flexible weighting algorithm, better model to combine survey and administrative earnings
 - Faster – to move toward a goal of rapid releases with incomplete data

Preliminary Results from Upcoming Release



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

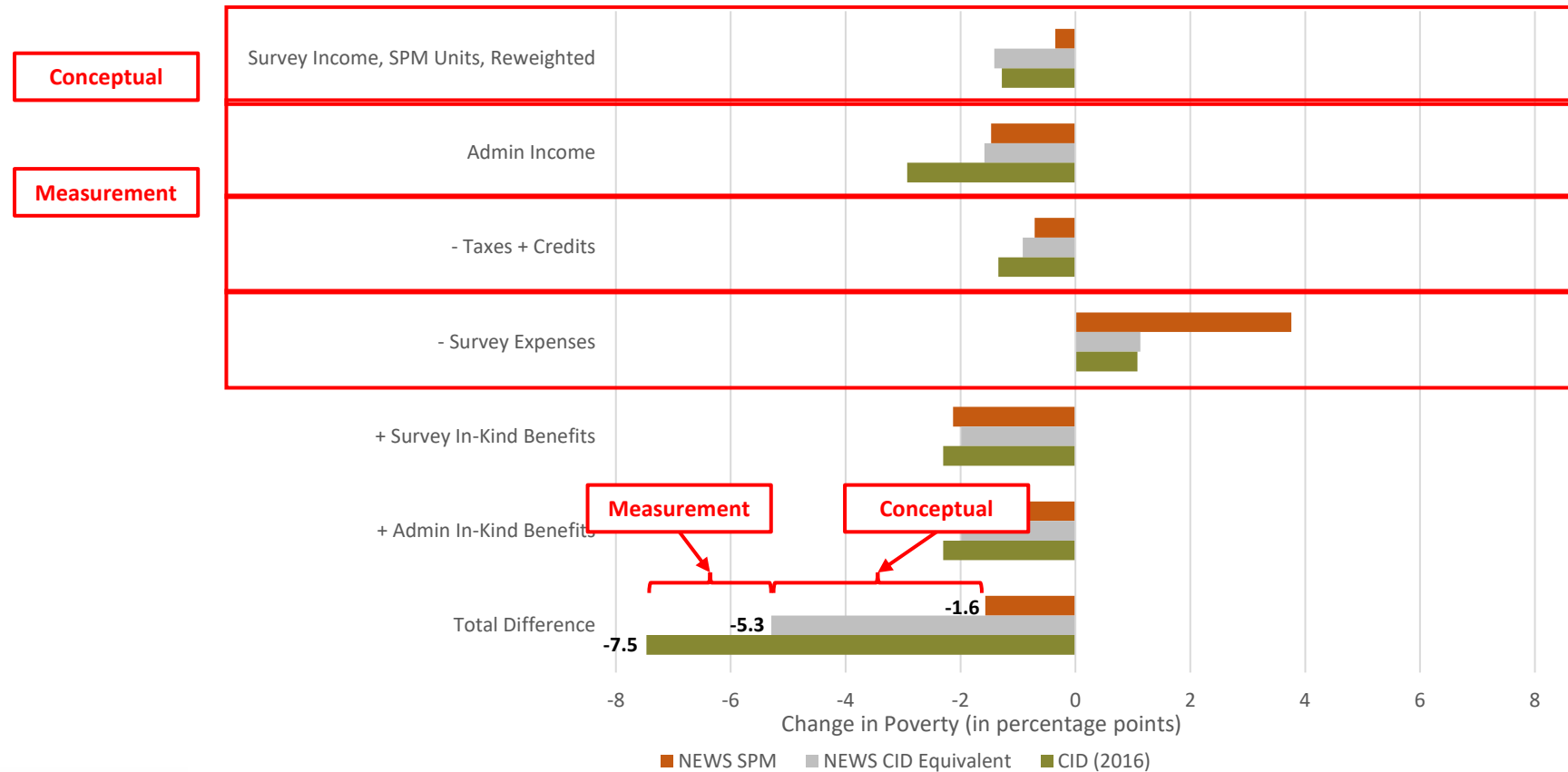
Preliminary Results from Upcoming Release



Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data and Meyer and Wu (2024) results from the Comprehensive Income Dataset (CID).

Preliminary Results from Upcoming Release

Changes in Poverty Moving From Official Poverty to Improved NEWS/CID Poverty Measures



Future Plans

- Prepare for more years
 - Handle incomplete administrative data
 - Either because not available in the past or not yet available for timely estimates
- Finer geographies
 - Integrate American Community Survey data (~33x as large a sample as CPS ASEC)
 - Better use full universe administrative data

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

Extra Slides

Comparing NEWS to CID

Conceptual Differences

Changes in Poverty Moving From Official Poverty to Improved NEWS/CID Poverty Measures



Comparing NEWS to CID

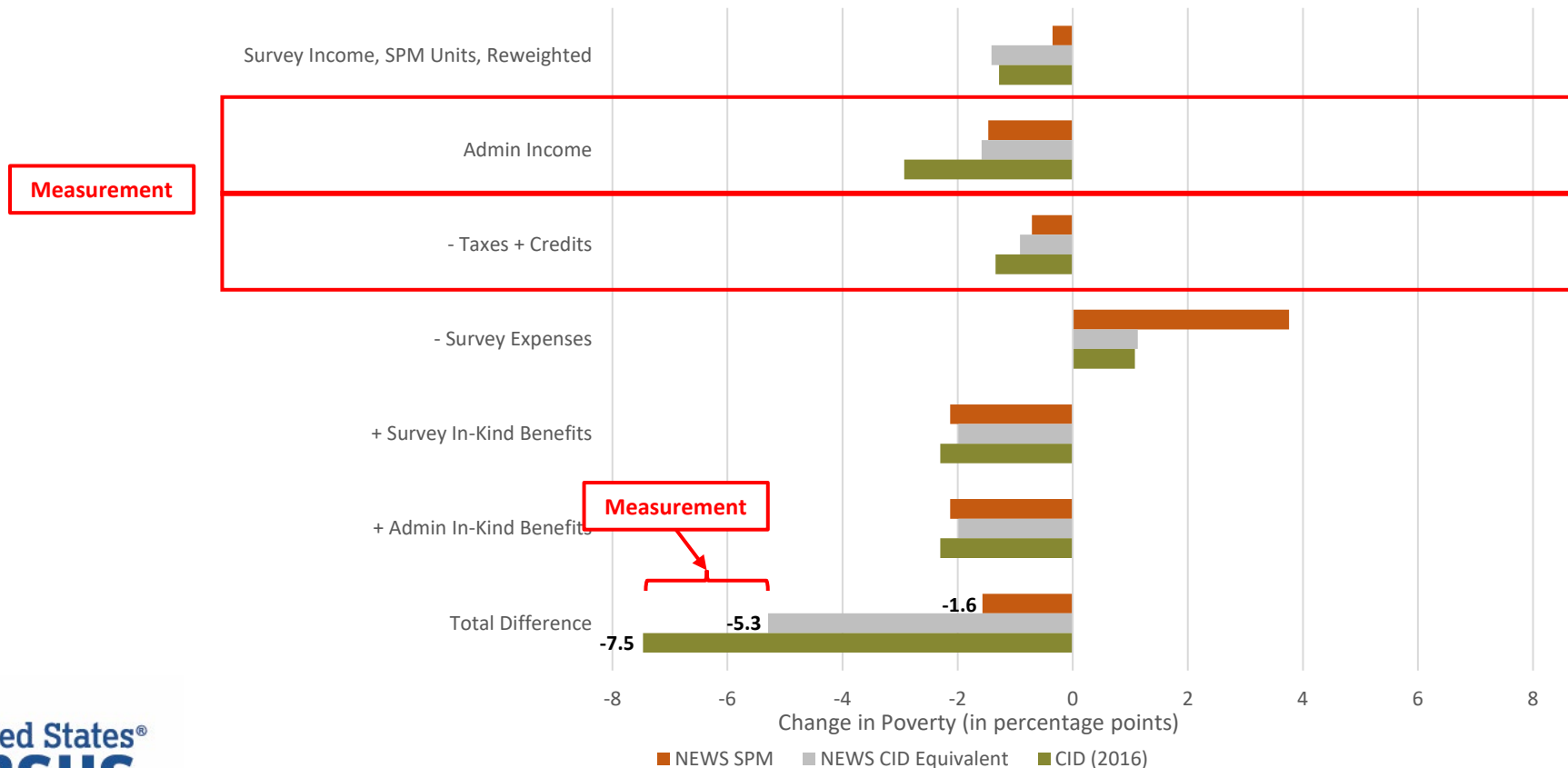
Conceptual Differences

- Thresholds (Survey Income, SPM Units, Reweighted): **1 pp**
 - SPM – thresholds vary by geography, housing tenure (renters and owners with w/out mortgage), and are based on consumption
 - CID – thresholds use OPM poverty line for 2 parent, 2 child households and use SPM equivalence scale to adjust for other units
- Survey Expenses: **2.3 pp**
 - CID – does not deduct child care expenses and medical out-of-pocket expenditures from resources, which the SPM does
- Impact – $(\text{NEWS CID} - \text{NEWS SPM}) / (\text{CID} - \text{Survey OPM})$ across all steps
 - 3.8 pp
 - 50 percent – share of difference between OPM and CID estimated CID poverty

Comparing NEWS to CID

Conceptual Differences

Changes in Poverty Moving From Official Poverty to Improved NEWS/CID Poverty Measures



Source: 2019 Current Population Survey Annual Social and Economic Supplement linked to administrative, decennial census, and third-party data and Meyer and Wu (2024) results from the Comprehensive Income Dataset (CID).

Comparing NEWS to CID Measurement Differences

- Admin Income: **1.3 pp**
 - NEWS – estimates a model of measurement error for earnings and draws earnings from a predicted distribution conditional on observables for individuals where we estimate the administrative data is not accurate
 - CID – Takes the maximum of administrative and survey earnings reports in specific cases, which comprise 37 percent of all individuals classified as poor with the Survey OPM
 - Both – use survey earnings when administrative earnings = 0 and survey earnings > 0
- Taxes and Credits: **0.4 pp**
 - Using the *actual* tax filing information, there is too little EITC for tax units in the survey
 - NEWS – we are exploring how to address the shortcoming, which seems to come from too few weighted units at AGI levels that receive high EITC amounts (compared to SOI aggregates). Fix pending in future releases.
 - CID – estimate taxes for units that *did not file taxes*. The additional EITC (and child tax credits) partially address the EITC shortfall relative to SOI aggregates and also decreases poverty.
- Impact – $(CID - NEWS\ CID)/(CID - Survey\ OPM)$ across all steps
 - 2.2 pp
 - 29 percent – share of difference between OPM and CID estimated CID poverty
 $(NEWS\ CID - NEWS\ SPM)/(CID - Survey\ OPM)$