Penn Libraries
UNIVERSITY of PENNSYLVANIA

University of Pennsylvania
ScholarlyCommons

2018 ADRF Network Research Conference Presentations

ADRF Network Research Conference Presentations

11-2018

Linking Survey and Administrative Data to Measure Income, Inequality, and Mobility

Carla Medalia
US Census Bureau

Bruce D. Meyer
University of Chicago

Amy O'Hara
Georgetown University

Derek Wu
University of Chicago

Follow this and additional works at: https://repository.upenn.edu/admindata_conferences_presentations_2018

Linking Survey and Administrative Data to Measure Income, Inequality, and Mobility

Abstract

Income is one of the most important measures of well-being, but it is notoriously difficult to measure accurately. Income data are available from surveys, tax records, and government programs, but each of these sources has important strengths and major limitations when used alone. We are linking multiple data sources to develop the Comprehensive Income Dataset (CID), a restricted micro-level dataset that combines the demographic detail of survey data with the accuracy of administrative measures. By incorporating information on nearly all taxable income, tax credits, and cash and in-kind government transfers, the CID surpasses previous efforts to provide an accurate and comprehensive measure of income for the population of U.S. individuals, families, and households. We use models to evaluate differences across the data sources and explore imputation methods and trends over time. The CID can enhance Census Bureau surveys and statistics through investigating measurement error, improving imputation methods, and augmenting surveys with the best possible estimates of income. It can also be used to improve the administration of taxes by the Internal Revenue Service and forecast and simulate changes in programs and taxes. Finally, the CID has substantial advantages over other sources to analyze numerous research topics, including poverty, inequality, mobility, and the distributional consequences of government transfers and taxes.

Linking survey and administrative data to measure income, inequality, and mobility

Carla Medalia, U.S. Census Bureau

ADRF Network Research Conference

November 13, 2018

Joint with:

Bruce D. Meyer, University of Chicago, NBER, AEI, and U.S. Census Bureau

Amy O'Hara, Georgetown University

Derek Wu, University of Chicago

United States Census™ Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Motivation

Household surveys

Tax records

Program participation data

- Income is extremely important as both outcome and predictor
- Income is difficult to measure: sources have strengths and weaknesses
- Previous research has combined sources, but scope and generalizability are limited

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

A Comprehensive Income Dataset (CID)

Household surveys

Tax records

Program participation data
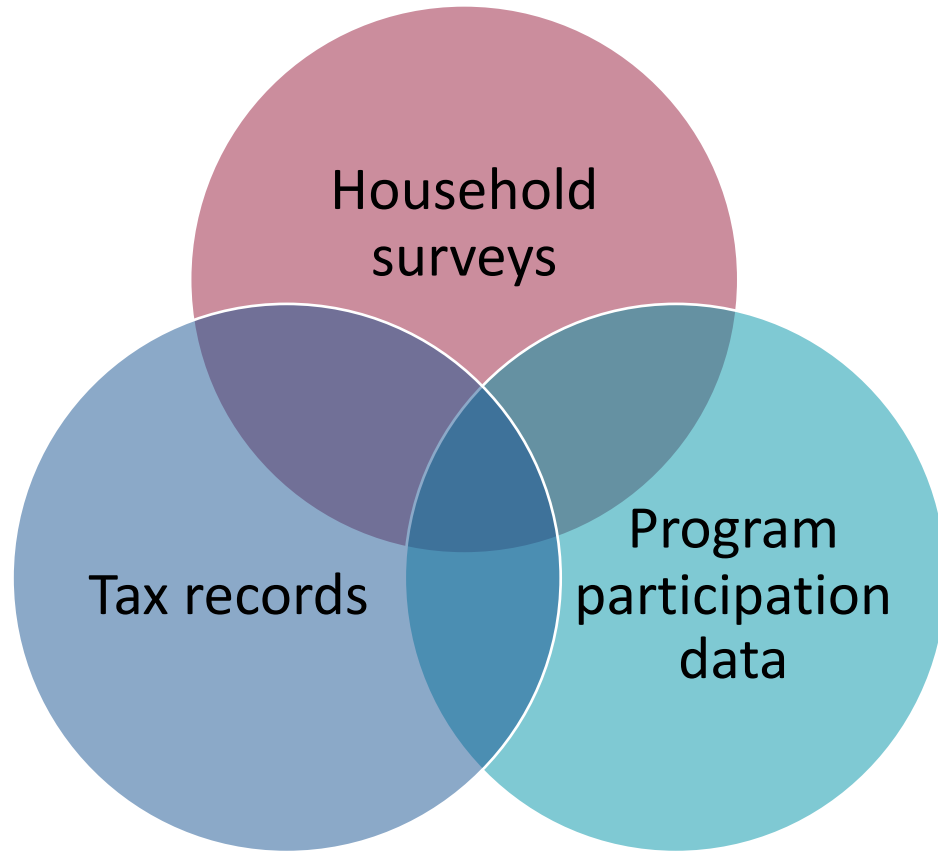
- Accurate and comprehensive measure of income for individuals, families and households

- Income, demographic characteristics, government programs and tax credits

- Uses: survey improvement, policy evaluation, and research

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Income is difficult to measure

Source	Strengths	Weaknesses
Household surveys	Rich demographics, representative of population, flexible platform	Measurement error, under-reporting, non-response[1]
Tax records	Accuracy, broad coverage (with information returns)	Lacks demographics, program information for in-kind benefits, and information on non-taxable income Tax units not necessarily economic unit Under-reporting
Program participation data	Accuracy, needed to evaluate programs, eligibility and other information not available on surveys	Limited history and demographics Misses non-recipients

[1]Bee and Mitchell 2017; Meyer, Mok, and Sullivan 2015; U.S. Census Bureau 1993

Income is difficult to measure

Source	Strengths	Weaknesses
Household surveys	Rich demographics, representative of population, flexible platform	Measurement error, under-reporting, non-response[1]

[1]Bee and Mitchell 2017; Meyer, Mok, and Sullivan 2015; U.S. Census Bureau 1993

Surveys underestimate income from government programs

Source: Meyer, Mok, and Sullivan (2015), by program and survey, 2000-2012

6

Income is difficult to measure

Source	Strengths	Weaknesses
Tax records	Accuracy, broad coverage (with information returns), time series	Lacks demographics, program information for in-kind benefits, and information on non-taxable income
		Tax units not necessarily economic unit
		Under-reporting

Income is difficult to measure

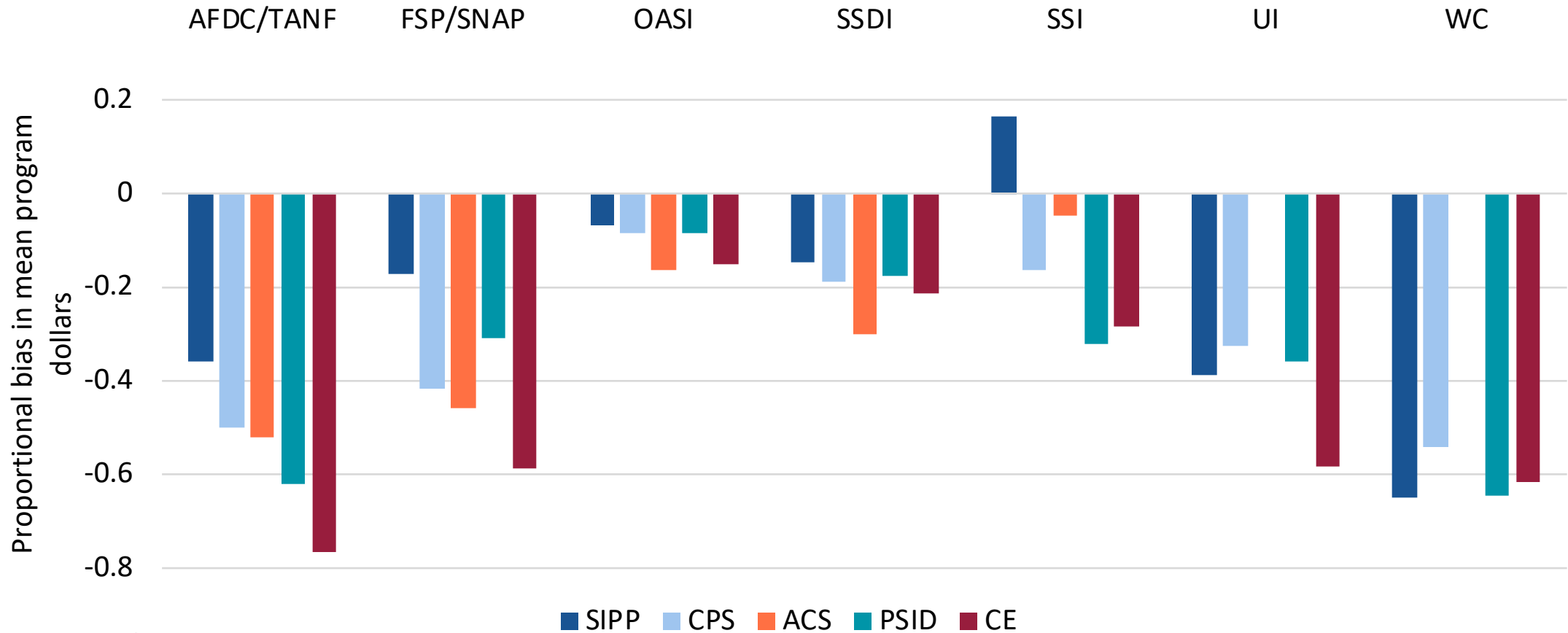| Source	Strengths	Weaknesses
Program participation data	Accuracy, needed to evaluate programs, eligibility and other information not available on surveys	Limited history and demographics Misses non-recipients

Defining income

- Focus on material well-being

- Includes income from all sources, some non-cash benefits

- State and federal income and payroll taxes

Trends in Benefits for Selected Transfers, 1996-2011

Total Expenditures in 2011 dollars (billions)

Legend: AFDC/TANF, SNAP, EITC
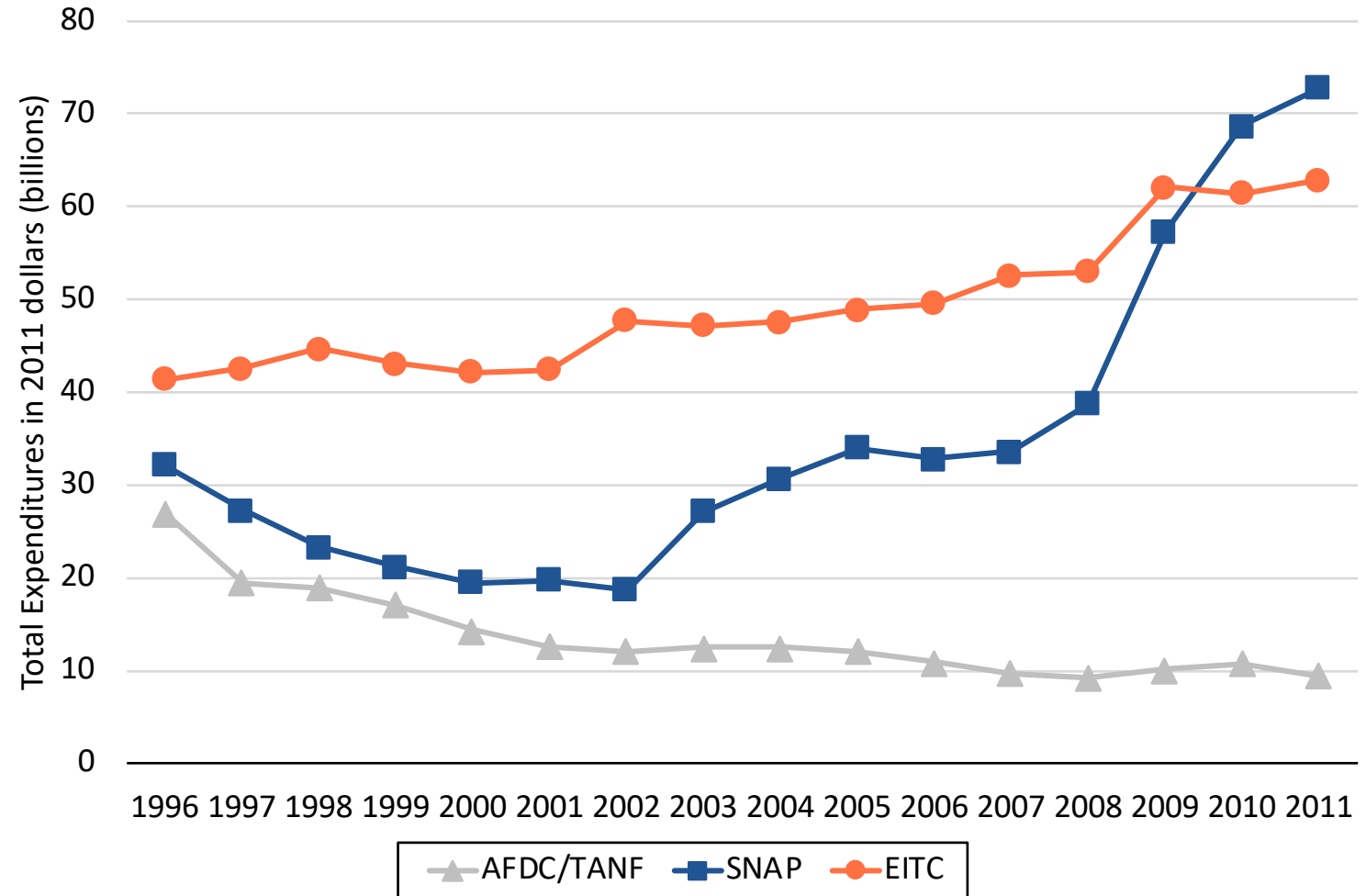
Data for CID

| Source type	Phase I	Phase II
Household surveys	Current Population Survey (CPS) Survey of Income and Program Participation (SIPP) American Community Survey (ACS)	Consumer Expenditure (CE) Survey
Tax data	Forms 1040, W-2, 1099-R	More detailed 1040 extracts, more extensive info returns Tax credits (e.g. EITC) Unemployment Insurance (UI)
Federal programs	SSA: Social Security and Supplemental Security Income HUD: Federal housing assistance HHS: Medicare and Medicaid enrollment, TANF	VA: Veterans Benefits
State programs	Public Assistance (e.g. TANF) SNAP, WIC LIHEAP	More Public Assistance, SNAP, WIC, LIHEAP Workers Compensation

Methods

- Records linked using Protected Identification Keys (PIKs)
  - Linkage rate: 99% of most admin records, 90-97% of survey data
  - Adjust sample weights to account for inability to link
- Unit of analysis reconciliation: tax unit, household rosters
- Link all dollars from admin records to survey households
- Assumptions regarding which income source is more accurate

Progress on CID

- Developing a prototype

- Interagency agreements

- Linking data and cleaning data

- Assessing quality of linkages, characteristics and accuracy of data

- Gathering documentation and metadata

- Three years of funding

- Initial projects

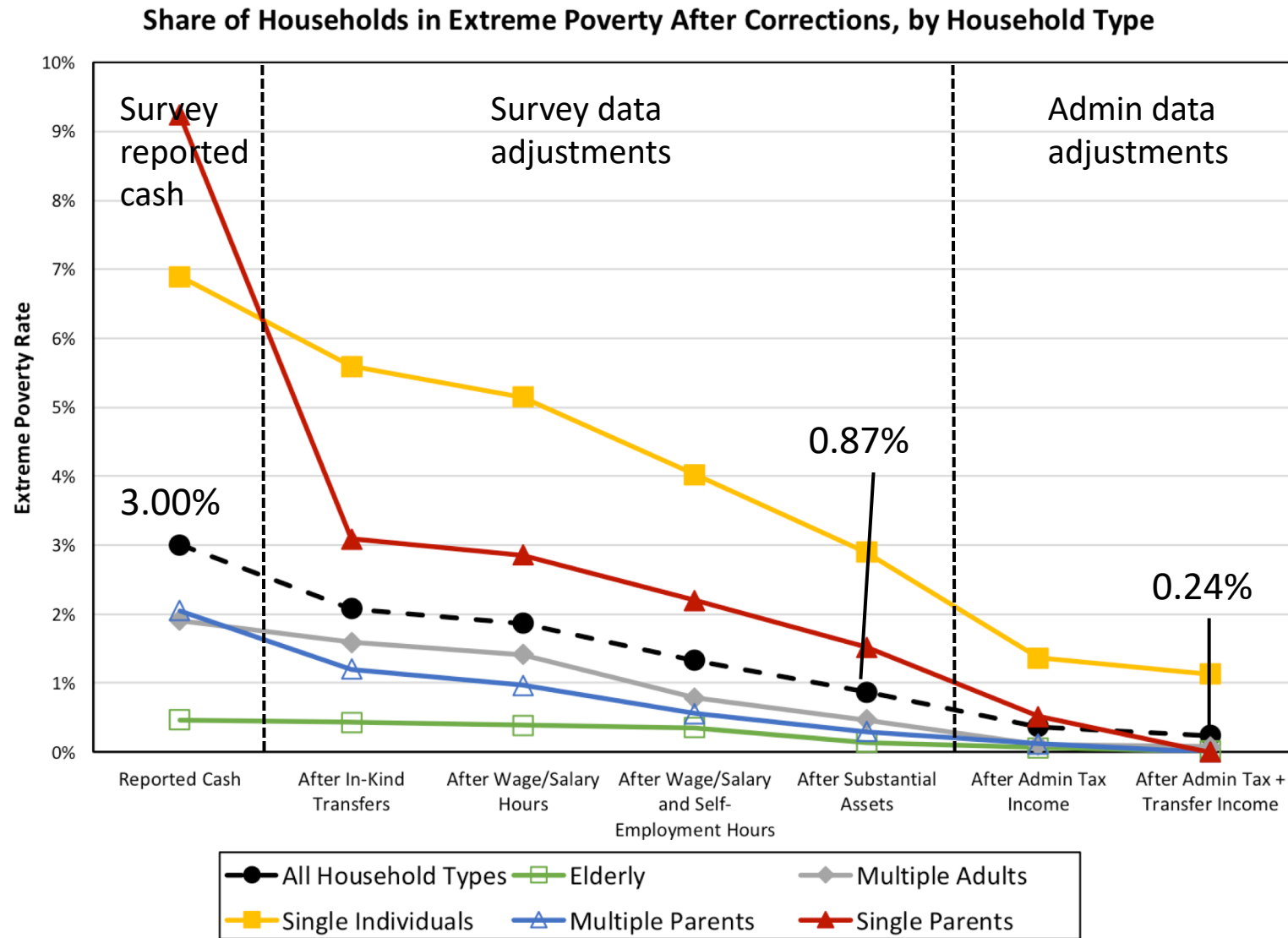# Estimated poverty rate reduction using administrative and survey data

The poverty reduction of Social Security and means-tested transfers (Meyer and Wu 2018)
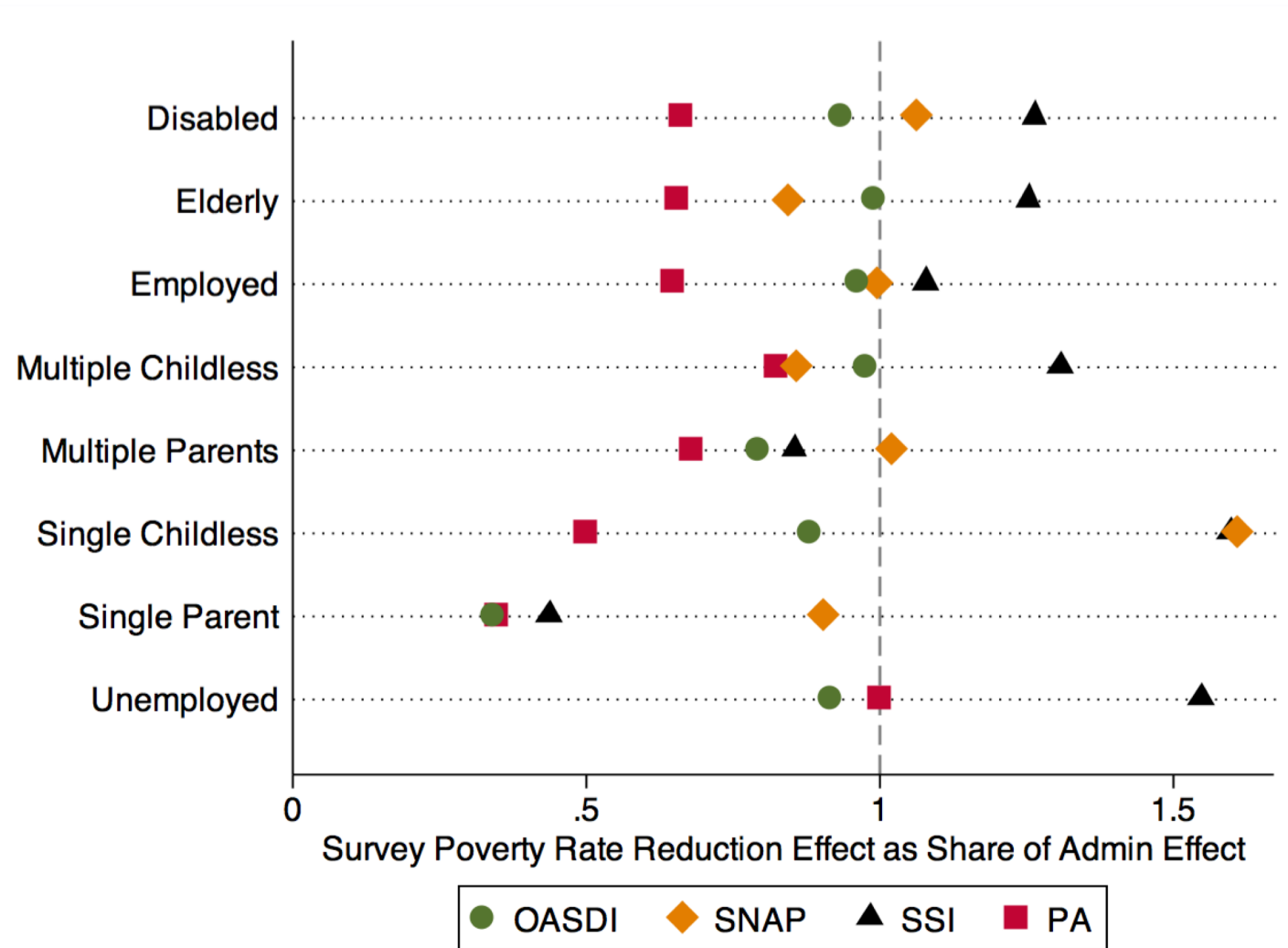
Source for survey data: 2008 SIPP Panel, Waves 1-14

Percent of households in extreme poverty – under $2/person/day

The use and misuse of income data and the rarity of extreme poverty in the United States (Meyer, Mooers, Wu, and Medalia 2018)

Share of Households in Extreme Poverty After Corrections, by Household Type

Survey reported cash

Survey data adjustments

Admin data adjustments

3.00%

0.87%

0.24%

Y-axis: Extreme Poverty Rate (0% to 10%)

X-axis: Reported Cash | After In-Kind Transfers | After Wage/Salary Hours | After Wage/Salary and Self-Employment Hours | After Substantial Assets | After Admin Tax Income | After Admin Tax + Transfer Income

Legend: All Household Types | Elderly | Multiple Adults | Single Individuals | Multiple Parents | Single Parents

Source for survey data: 2008 SIPP Panel, Wave 9

14

Future direction

Now

- Acquire new data; provide get-backs to agencies
- Better documentation of data and methods
- Gather feedback from potential users

Vision and challenges

- Make available in FSRDCs
- Develop path to get from prototype to production
- Challenges to implement CID in survey production environment
  - Mismatch of reference periods
  - Availability of data – not all states
  - Timeliness of data releases

Thank you!

Email addresses:

carla.medalia@census.gov

bdmeyer@uchicago.edu

amy.ohara@georgetown.edu

derekwu@uchicago.edu

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

References

Bee, Adam and Joshua Mitchell. 2017. "Do Older Americans Have More Income Than We Think?" SESHD Working Paper, U.S. Census Bureau.

Cajner, Tomaz, Leland Crane, Ryan Decker, Adrian Hamins-Puertolas, Christopher Kurz, and Tyler Radler. 2018. "Using Payroll Processor Microdata to Measure Aggregate Labor Market Activity," Finance and Economics Discussion Series 2018-005. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2018.005.

Congressional Budget Office. 2017. "Federal Spending for Means-Tested Programs, 2007 to 2027." https://www.cbo.gov/sites/default/files/115th-congress-2017-2018/reports/52405-means-tested-programs.pdf

Davies, Paul S. and T. Lynn Fisher. 2009. "Measurement Issues Associated with Using Survey Data Matched with Administrative Data from the Social Security Administration." Social Security Bulletin, 69(2): 1-12.

Ganong, Peter and Pascal Noel. 2017a. "Consumer Spending During Unemployment: Positive and Normative Implications." *Working Paper.*

Ganong, Peter and Pascal Noel. 2017b. "The Effect of Debt on Default and Consumption: Evidence from Housing Policy in the Great Recession." *Working Paper.*

Harron, Katie, Chris Dibben, James Boyd, Anders Hjern, Mahmoud Azimaee, Mauricio Barreto, and Harvey Goldstein. 2017. "Challenges in administrative data linkage for research." Big Data and Society, July-December 2017: 1-12. https://doi.org/10.1177/2053951717745678.

Johnson, Barry, and Kevin Moore. 2012. "Consider the Source: Differences in Estimates of Income and Wealth From Survey and Tax Data." *Compendium of Federal Estate Tax and Personal Wealth Studies, Volume II*. Chapter 9: Comparing Administrative and Survey Data." Pgs. 875-897. Internal Revenue Service, Statistics of Income. https://www.irs.gov/pub/irs-soi/11pwcompench9.pdf Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives*, 29(4): 199-226.

Meyer, Bruce D. and Nikolas Mittag. 2015. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net." NBER Working Paper 21676. Nicholas, Joyce and Michael Wiseman. 2010. "Elderly Poverty and Supplemental Security Income, 2002-2005." *Social Security Bulletin*, 70(2).

Meyer, Bruce D. and Derek Wu. 2018. "The Poverty Reduction of Social Security and Means-tested Transfers." *ILR Review*, 71(5): 1106-1153. https://doi.org/10.1177/0019793918790220

Meyer, Bruce D., Derek Wu, Victoria Mooers and Carla Medalia. 2018. "The use and misuse of income data and the rarity of extreme poverty in the United States." Unpublished working paper.

Oberski, DL, A Krichner, S Eckman, and F Kreuter. 2017. "Evaluating the Quality of Survey and Administrative Data with Generalized Multitrait-Multimethod Models." Journal of the American Statistical Association, 112(520): 1477-1489, DOI: 10.1080/01621459.2017.1302338.

Schnetzer, Matthias, Franz Astleithner, Predrag Cetovic, Stefan Humer, Manuela Lenk, and Matthias Moser. 2015. "Quality Assessment of Imputations in Administrative Data." *Journal of Official Statistics*, 31(2): 231-247, DOI http://dx.doi.org/10.1515/JOS-2015-0015.

U.S. Census Bureau. 1993. "Measuring the Effect of Benefits and Taxes on Income and Poverty: 1992". Current Population Reports, Series P60-186RD.

Walker, Ed. 1997. "The Census Bureau's Business Register: Basic Features and Quality Issues." Paper presented at the Joint Statistical Meetings, Anaheim, CA.

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov