
A Comprehensive Income Dataset

Bipartisan Policy Center

September 6, 2018

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

Joint with

Carla Medalia, U.S. Census Bureau

Amy O'Hara, Stanford University

Derek Wu, University of Chicago

Disclaimer: Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

A Comprehensive Income Dataset (CID)

- Combine the advantages of three sources of data
 - Surveys: rich demographics, random sample of population, flexible platform
 - Tax Data: accuracy, broad coverage (with information returns)
 - Program Data: accuracy, needed to evaluate programs, eligibility and other information not available on surveys
-

Goals of the CID

- Improve Census Bureau Surveys
 - Enhance tax administration
 - Provide the information to better examine public policies and support research
-

Survey Data Alone Inadequate

- Severe under-reporting of many types of income
 - Government transfers
 - Private pensions
 - Most of bias in government program dollars in surveys due to measurement error: Meyer, Mok, Sullivan (JEP Fall 2015)
 - One can compare weighted survey reports to administrative totals
 - Gives net reporting, net bias (false positives can cancel out false negatives)
-

Table 1

Proportional Bias in Survey Estimates of Mean Program Dollars and Months Received, by Program and Survey, 2000–2012

	<i>AFDC/ TANF</i>	<i>FSP/ SNAP</i>	<i>Social Security</i>		<i>SSI</i>	<i>UI</i>	<i>WC</i>	<i>NSLP</i>	<i>WIC</i>
			<i>OASI</i>	<i>SSDI</i>					
Dollars									
ACS	-0.519	-0.458	-0.165	-0.299	-0.046				
CE	-0.767	-0.587	-0.149	-0.214	-0.283	-0.583	-0.618		
CPS	-0.500	-0.417	-0.086	-0.187	-0.162	-0.325	-0.541		
PSID	-0.619	-0.308	-0.086	-0.176	-0.322	-0.360	-0.646		
SIPP	-0.357	-0.170	-0.070	-0.146	0.164	-0.388	-0.651		
Months									
ACS			-0.154	-0.261	-0.372				
CPS	-0.453	-0.422	-0.147	-0.154	-0.397			-0.503	-0.341
PSID	-0.574	-0.297	-0.114	-0.121	-0.502			-0.470	-0.180
SIPP	-0.232	-0.165	-0.008	0.041	0.023			0.141	-0.197

Source: Meyer, Mok and Sullivan JEP (2015)

Tax or Program Data Alone Inadequate

■ Tax Data

- ❑ Lack demographics
- ❑ Lack program information for in-kind benefits
- ❑ Lack information on non-taxable income
- ❑ Not for economic unit
- ❑ Miss nonfilers

■ Program Data

- ❑ Limited history and demographics
 - ❑ Miss non-recipients
-

Data

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	Better 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, GA) SNAP, WIC LIHEAP	More Public Assistance, SNAP, WIC, LIHEAP Workers Compensation

Linking

- Records are linked using an anonymized version of the SSN called a Protected Identification Key (PIK)
 - Usually can attach a PIK to 99 percent of admin data
 - Records usually have SSN that has been validated
 - Usually can attach a PIK to 90-97 percent of households in survey data
-

Uses of CID: Survey Improvement

- Assess extent and nature of errors
 - Improve imputation of missing values
 - Substitute administrative values for survey ones
 - Augment surveys, avoid asking some questions
-

Uses of CID: Tax Administration, Policymaking

- Tax administration (IRS)
 - EITC takeup
 - Identifying areas of noncompliance
- Forecasting and simulation of changes in programs and taxes

Uses of CID: Research

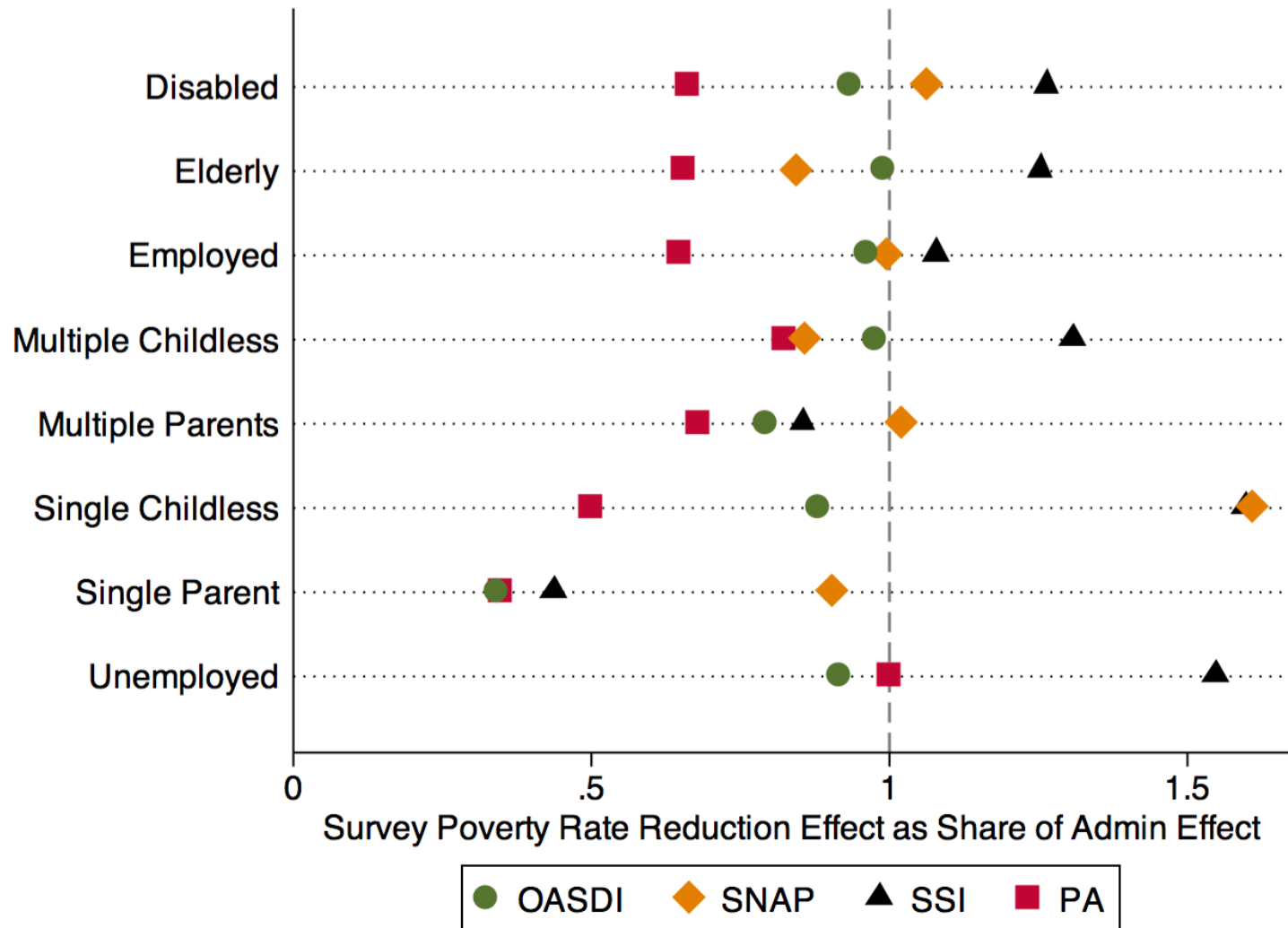
- Measure income distribution, poverty, inequality
 - Estimate effects of social insurance and means-tested transfers
 - Simulate changes in programs and taxes
-

The Poverty Reduction Due to Social Security and Means-Tested Transfers

Meyer and Wu (forthcoming)

- Worst case scenario for value of the CID because uses better survey data than any new data
-

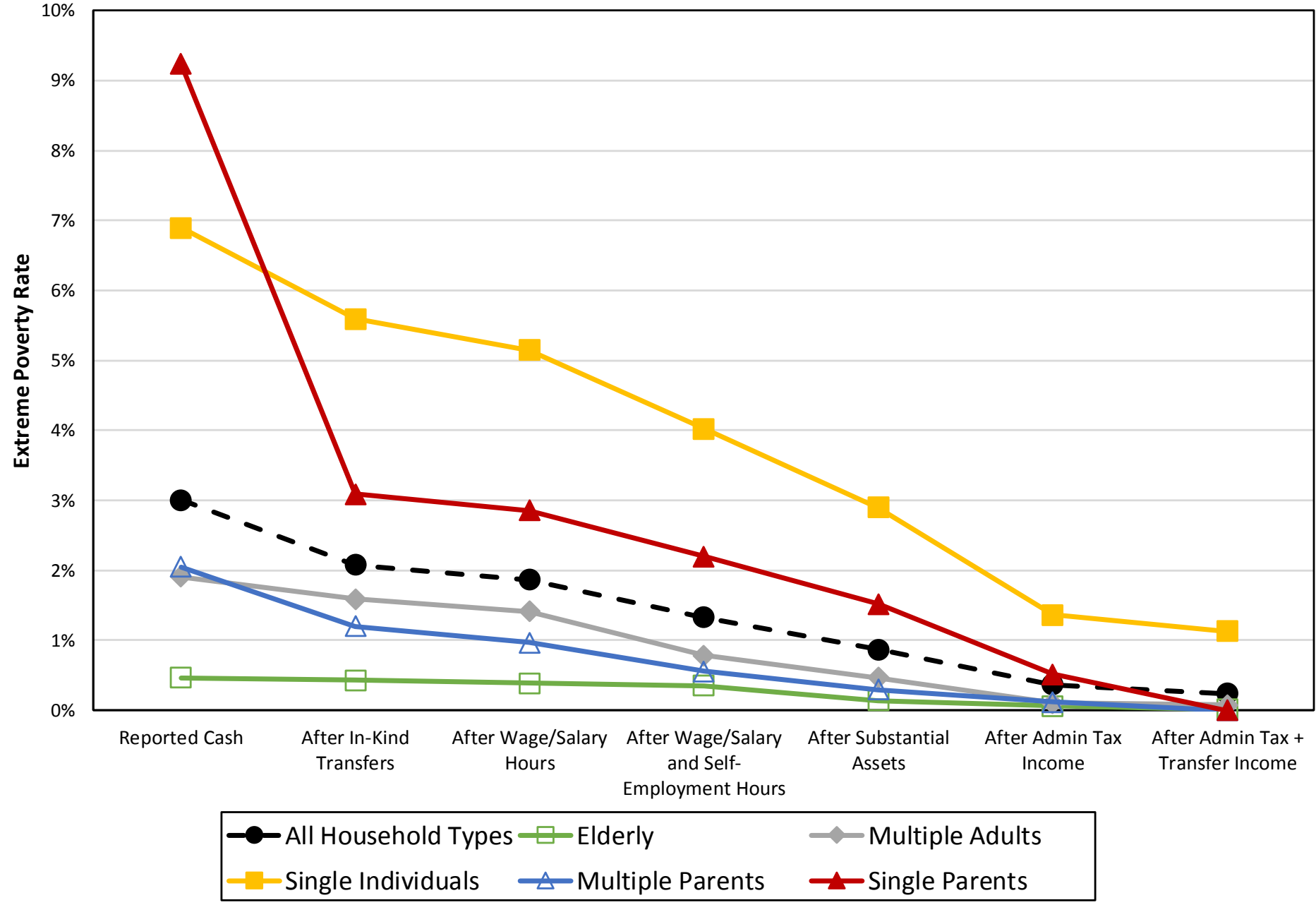
Poverty Rate Reduction from Combined vs. Survey Data: OASDI, SSI, SNAP, PA



The Use and Misuse of Income Data and the Rarity of Extreme Poverty in the United States

Meyer, Mooers and Wu (2018)

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



Future Plans for the CID

- Provide products to federal and state agencies
 - Work with constituencies to design them
 - Better documentation of data and methods
 - Advisory board
 - Possible Census adoption of CID as regular program
-