

The Distribution of Income and Consumption Using Linked Data

Bruce D. Meyer (University of Chicago)

Derek Wu (University of Virginia)

joint with

Aaron Hong, Connor Murphy (University of Chicago)

and James X. Sullivan (University of Notre Dame)

CNSTAT Public Seminar on ICW (NAS)

December 10, 2024

Disclaimer: Any conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results were approved for release by the Census Bureau's Disclosure Review Board, authorization numbers CBDRB-FY23-0184 and CBDRB-FY2023-CES005-016



Motivation

- Obtaining an accurate picture of the distribution of resources for the population is a critical issue for both researchers and policymakers
- Much of past work relies on survey income sources prone to income underreporting
 - For example, about half of private pension and SNAP receipt not recorded in the CPS and more than half of UI missed recently (Meyer, Mok and Sullivan 2015; Bee and Mitchell 2017; Larrimore et al. 2023; Meyer et al. 2024)
- Recent studies that link administrative data to major Census Surveys have found that incomes are often much higher than reported in the survey data alone, while the impact of transfer programs is often much larger as well (Bee and Mitchell 2017; Meyer and Wu 2018, 2024; Dushi and Trenkamp 2021)
- An alternative to survey or blended income is to rely on consumption data to assess well-being (Johnson et al.; Meyer and Sullivan, Garner et al., etc.), accounting for dis-saving, durable flows, income under-reporting
 - Past research has compared the distributions of income and consumption, but work comparing the measures directly at the individual or family level is more limited
 - Income in the CE Survey often criticized though it has its defenders (Attanasio; Johnson et al.).



Our Contributions

- We provide the first family level comparisons of survey income combined with a full range of administrative data to expenditure and consumption. This is part of the Comprehensive Income Dataset (CID) project
 - We link the Consumer Expenditure (CE) survey to administrative tax records and program data
- This work fills in a need highlighted in NASEM (2024) for data with both income and consumption to examine their joint distribution
 - We blend tax data and program records with the income from survey data that also includes consumption (reflecting the suggestions in Figure 5-1A of ICW report)
 - As consumption and income are difficult to measure, the report suggests pilot studies to improve the CE survey data.



Data: The CE Survey, Samples, Admin Data, and Linkages



The Consumer Expenditure (CE) Survey and our sample

- Use the Interview component of the Consumer Expenditure (CE) Survey
 - The reference period is the 12 months prior to the interview month for income and the 3 months prior to the interview month for expenditures
 - Also try combining multiple CE interviews into a calendar year
 - We use surveys conducted in January to April in the years 2015 to 2017 to best align with calendar year income
- Link administrative records using Protected Identification Keys (PIKs)



Administrative Data Sources

Income Source	Administrative Source	Income Unit	Income Frequency	States Covered
Earnings	W-2 (IRS), Form 1040 (IRS)	Individual & Tax Unit	Annual	All
AGI & Other Income	Form 1040 (IRS)	Tax Unit	Annual	All
Retirement Income	Form 1099-R (IRS)	Individual	Annual	All
Social Security	PHUS & MBR (SSA)	Individual	Monthly	All
SSI	SSR (SSA)	Individual	Monthly	All
Veterans' Benefits	US VETS Data	Individual	Annual	All
Taxes (simulated)	Form 1040 (IRS)	Tax Unit	Annual	All
SNAP	State Agencies	Household	Monthly	20+ States
Housing Assistance	PIC & TRACS (HUD)	Household	Monthly	All



Constructing Resource Measures by Combining Survey & Administrative Data



Income Concept

- Our income concept of interest is post-tax, post-transfer income to align with expenditures
- Components include:
 - Pre-tax money income (e.g., earnings, asset income, retirement income, taxable transfers like OASDI and UI, non-taxable transfers like SSI and TANF, other cash income)
 - Tax liabilities and credits
 - Selected in-kind transfers (specifically housing assistance and SNAP)
- We create a survey-only version (relying on survey responses only) and a blended version (combining survey and administrative data)



Constructing Expenditures and Consumption

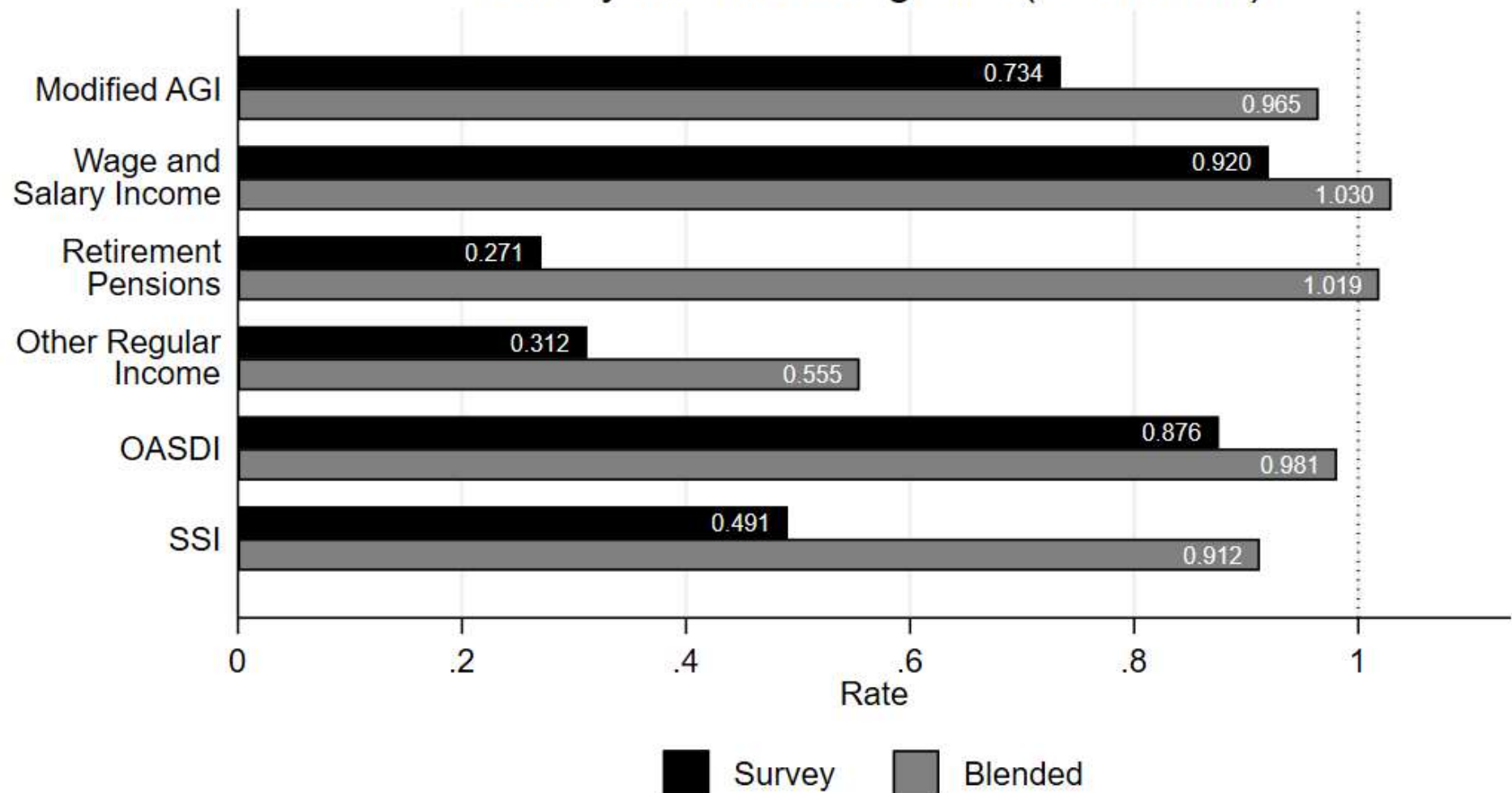
- To convert reported expenditures in the CE to a measure of consumption, we make a number of adjustments
 - We convert vehicle and housing spending to service flows
 - We impute a rental value for those living in government or subsidized housing
 - We exclude saving such as outlays for retirement including pensions and social security



Comparing Weighted Income Totals to Publicly Available Aggregates



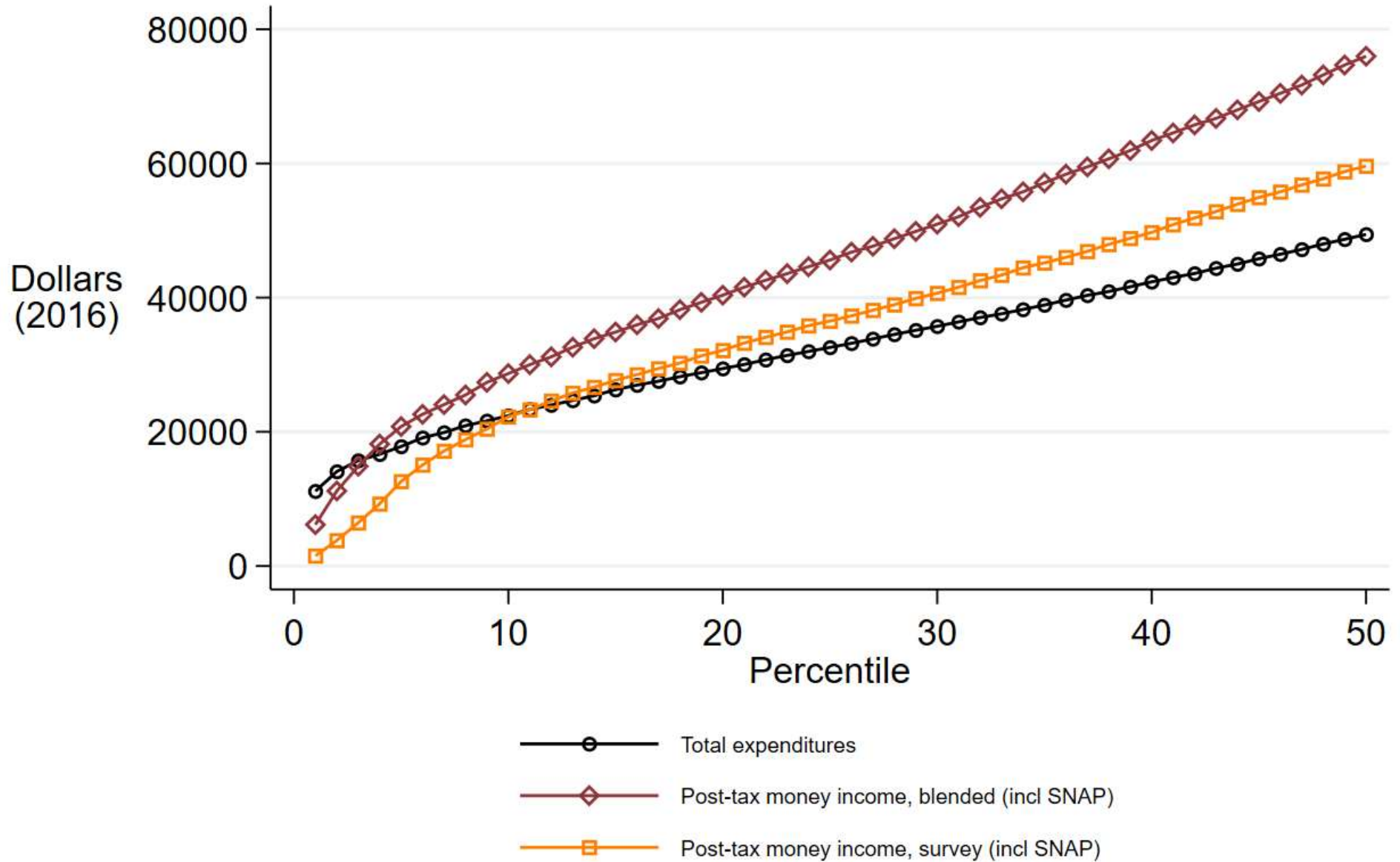
Average Reporting Rates of Blended versus Survey Income Categories (2014-2016)



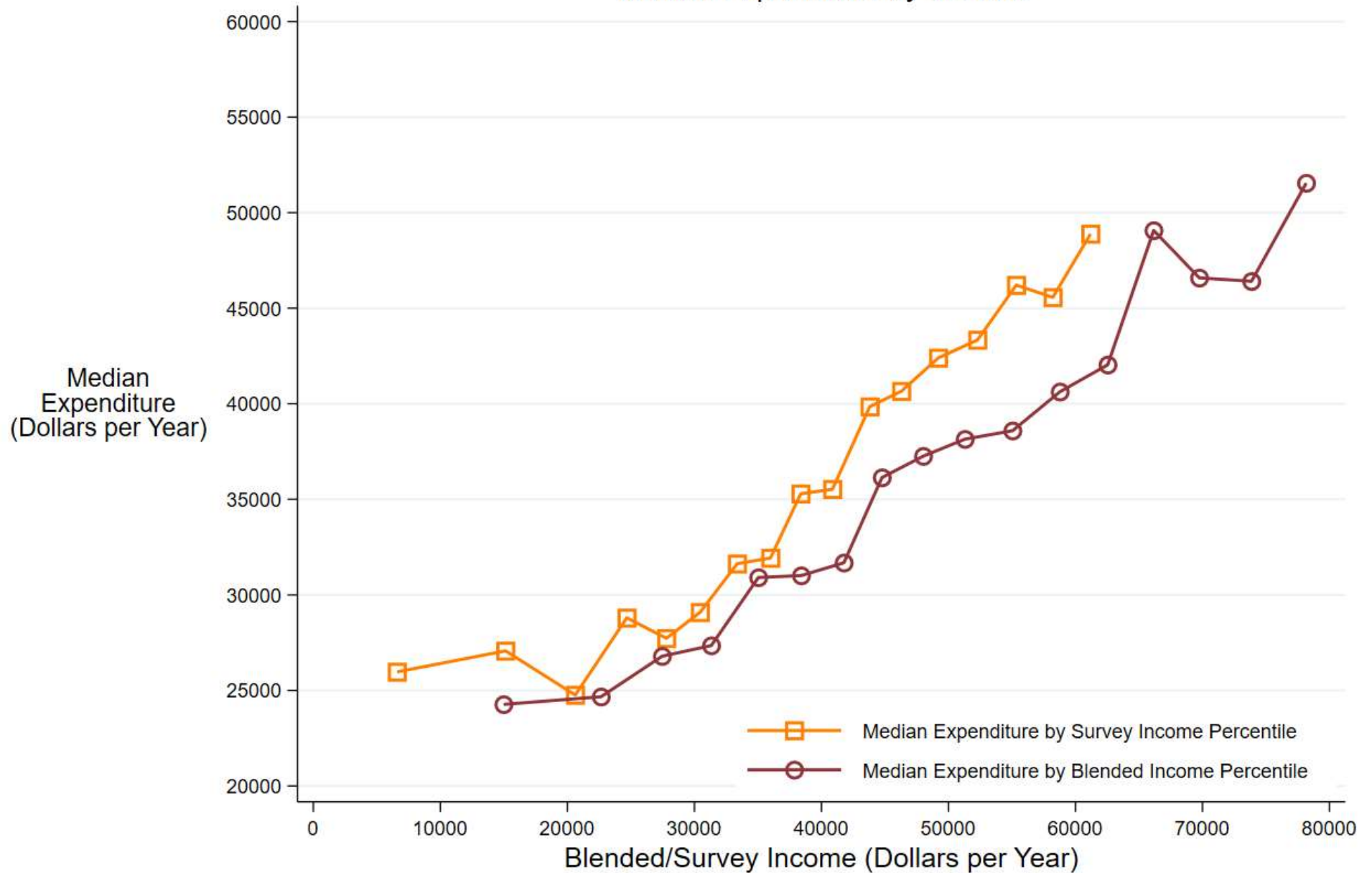
- (1) Modified AGI consists of wages and salary income + self employment income + retirement pensions + interest and dividends + rental income and royalties + other REGULAR income + other NONRENTAL income.
- (2) While the survey definition of Other Regular Income consists of income components such as VA benefits, Worker's Compensation, UI, child support, and alimony we only have the corresponding administrative sources for VA benefits. Hence, our blended measure is relatively understated.



Distribution of Resource Measures Bottom 50 Percentiles



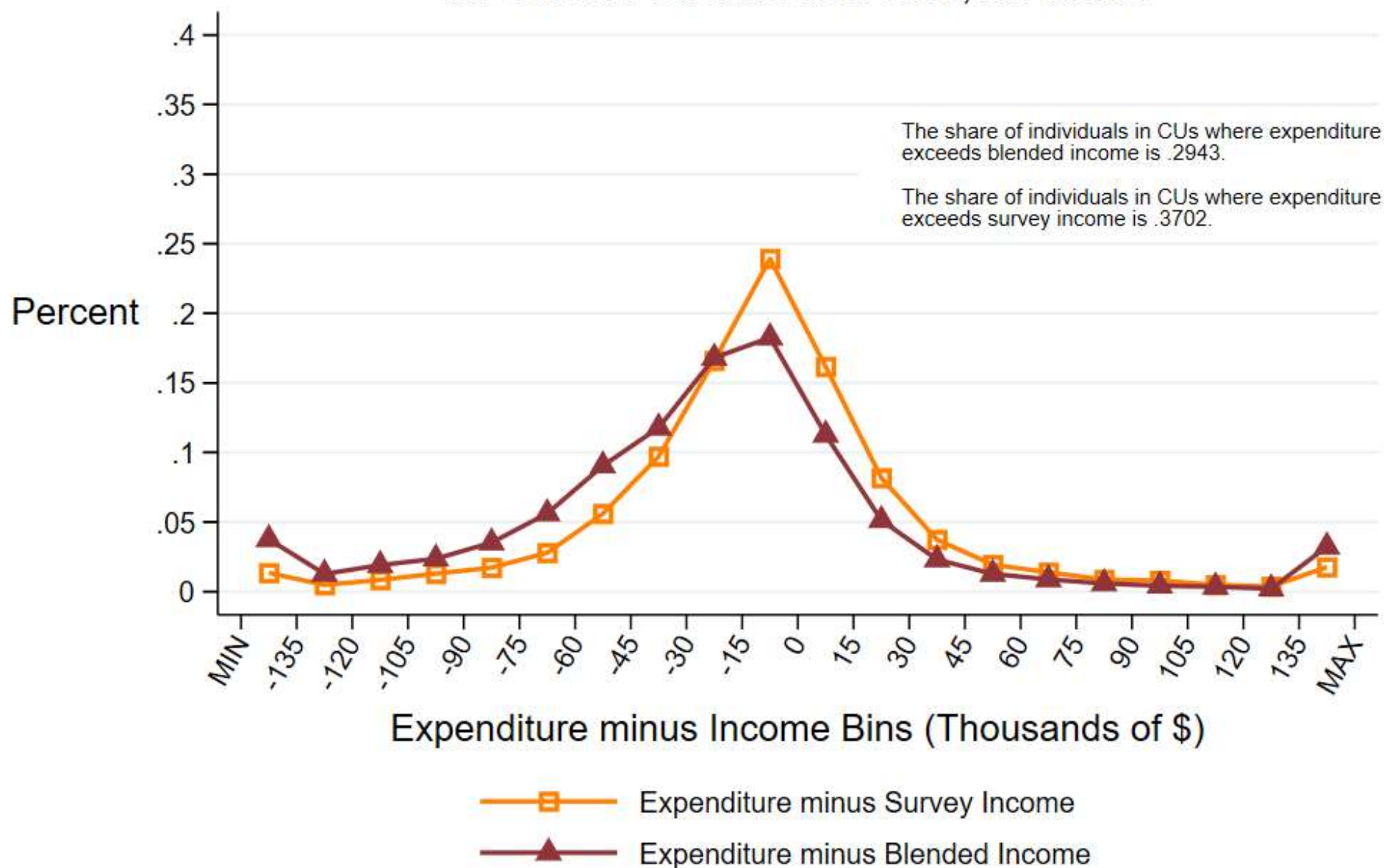
Median Expenditure by Income



(1) Each point represents groups of three percentiles.



Distribution of Expenditure minus Income for 1st and 4th Interview CUs, All States

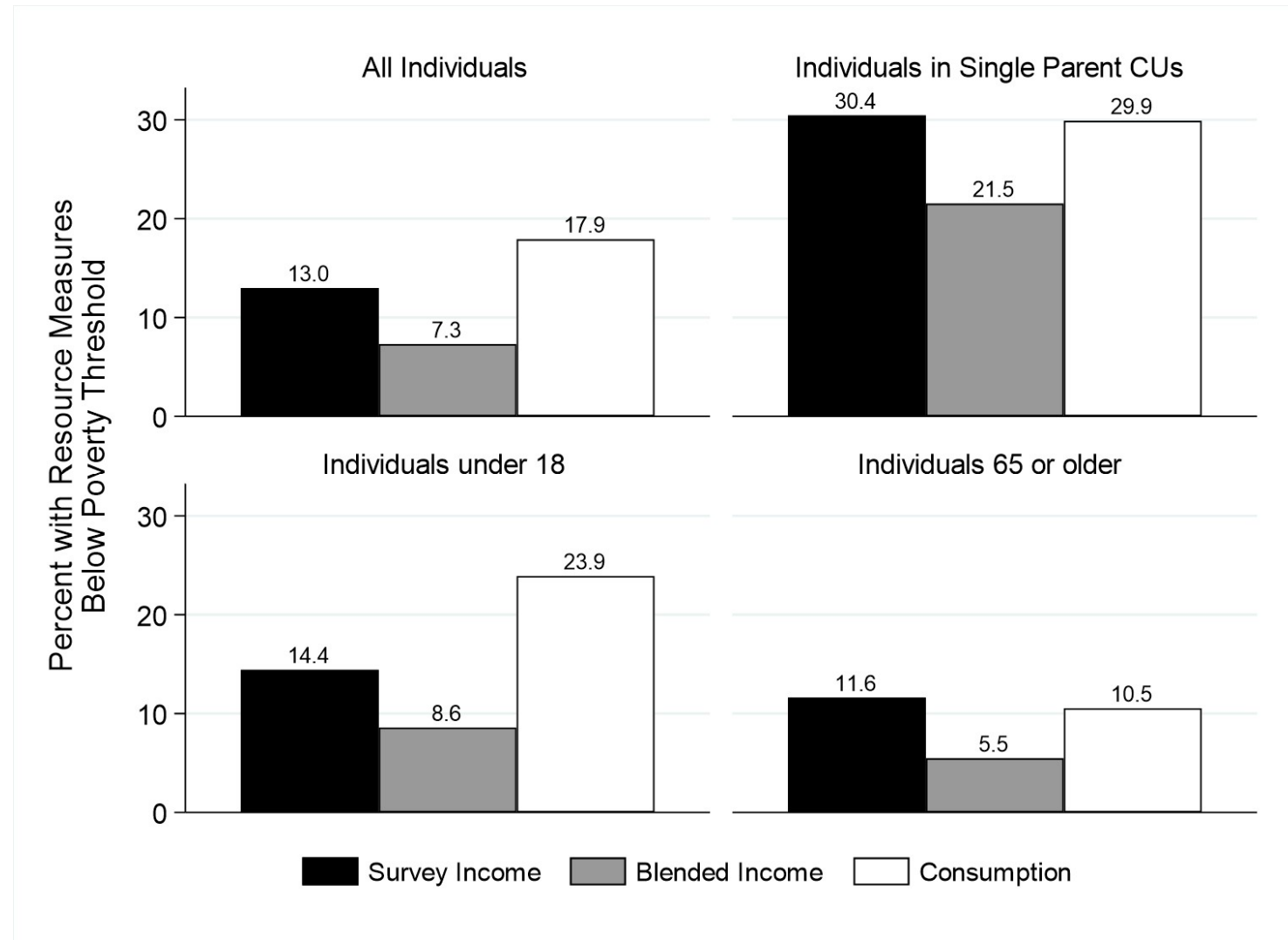


Poverty Measures in the CE

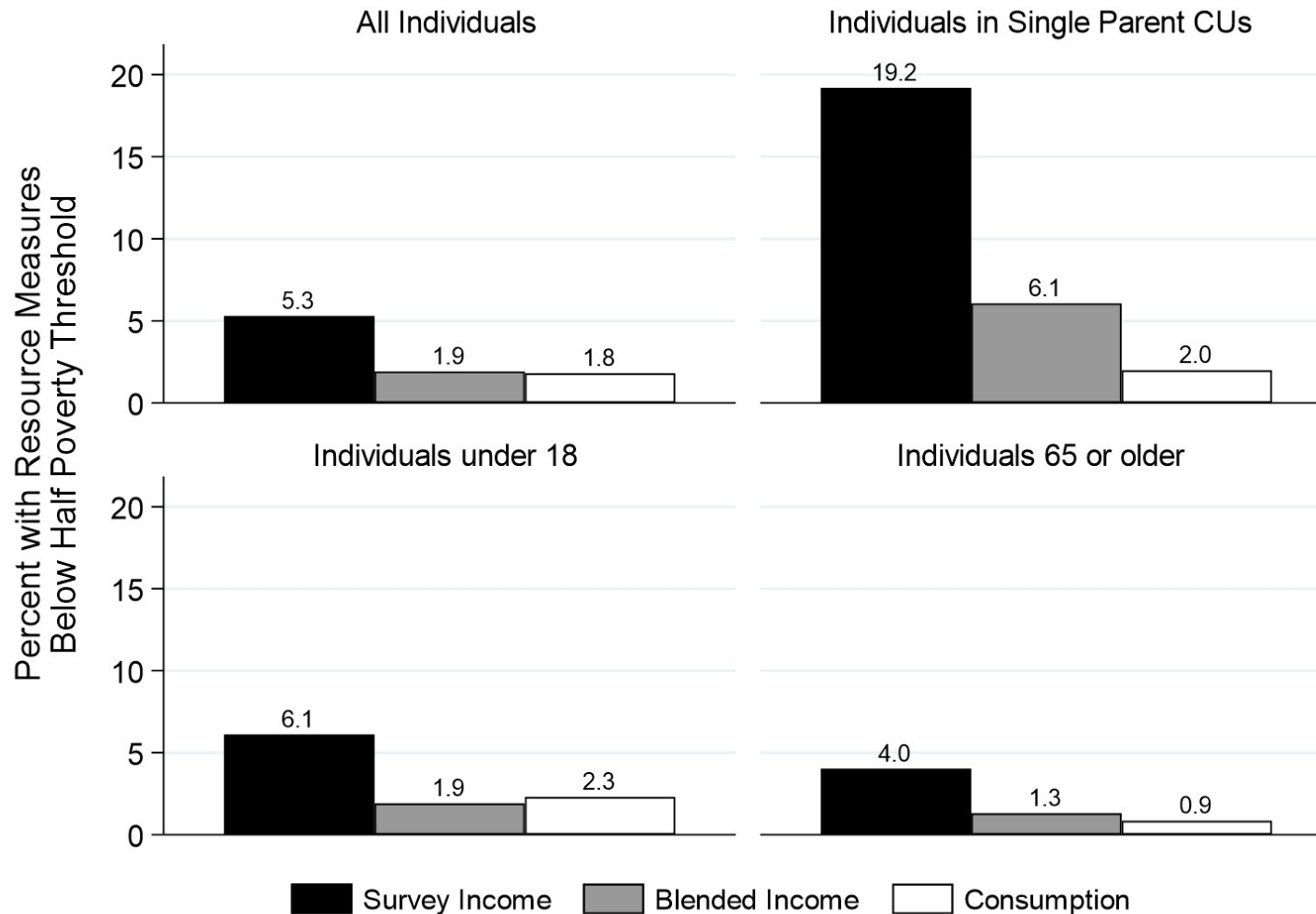
- To assess the economic security of the population and see the implications of improving our income measures or using consumption to measure well-being, we calculate the share of individuals in CUs below the SPM poverty line using three income or consumption series, These resource measures are:
 - Post-tax money income plus SNAP plus Housing benefits (entirely survey based)
 - Post-tax money income plus SNAP plus Housing benefits (blended)
 - Consumption (which includes housing subsidies, and the rental equivalent of owned homes and cars; excludes pension contributions)



Share of People in CUs Below Poverty Line



Share of People in CUs Below Half Poverty Line



Robustness Checks

- We tried an alternative sample, restricted to just those with 4th interviews early in the calendar year allowing us to calculate yearly expenditures



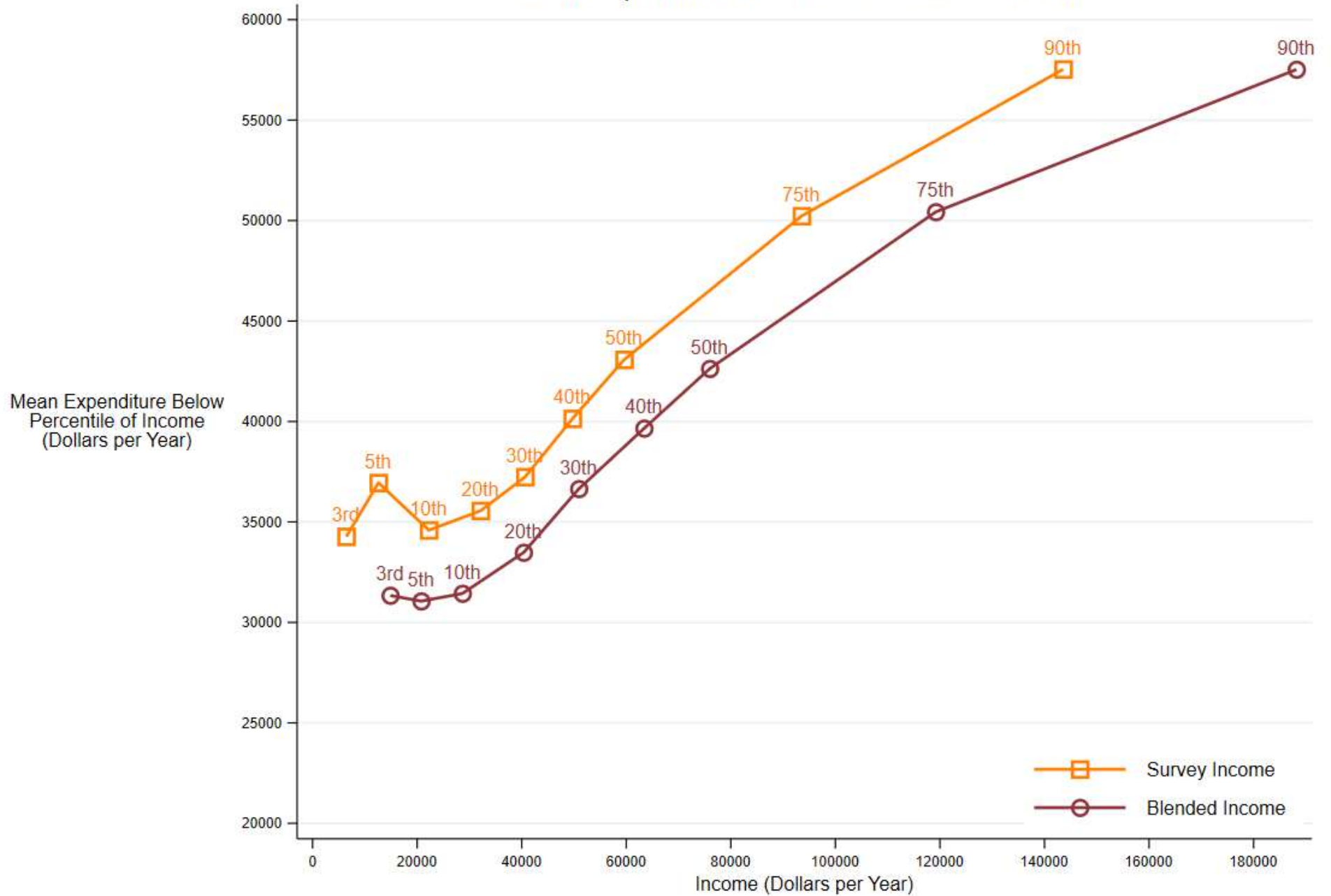
Conclusions and Next Steps

- Our results show the feasibility of improving CE Survey income measures using admin data
- Differences between income and expenditures narrowed at very bottom, but overall largely unresolved
- We break new ground in linking admin data to income and consumption microdata in the same survey; the process literally took years to just get approvals and gain access
- In future work:
 - Analyze top of distribution in great detail and focus also on particular subgroups for whom income and expenditures may differ greatly
 - Incorporate savings data from CE to help address discrepancies between income and consumption

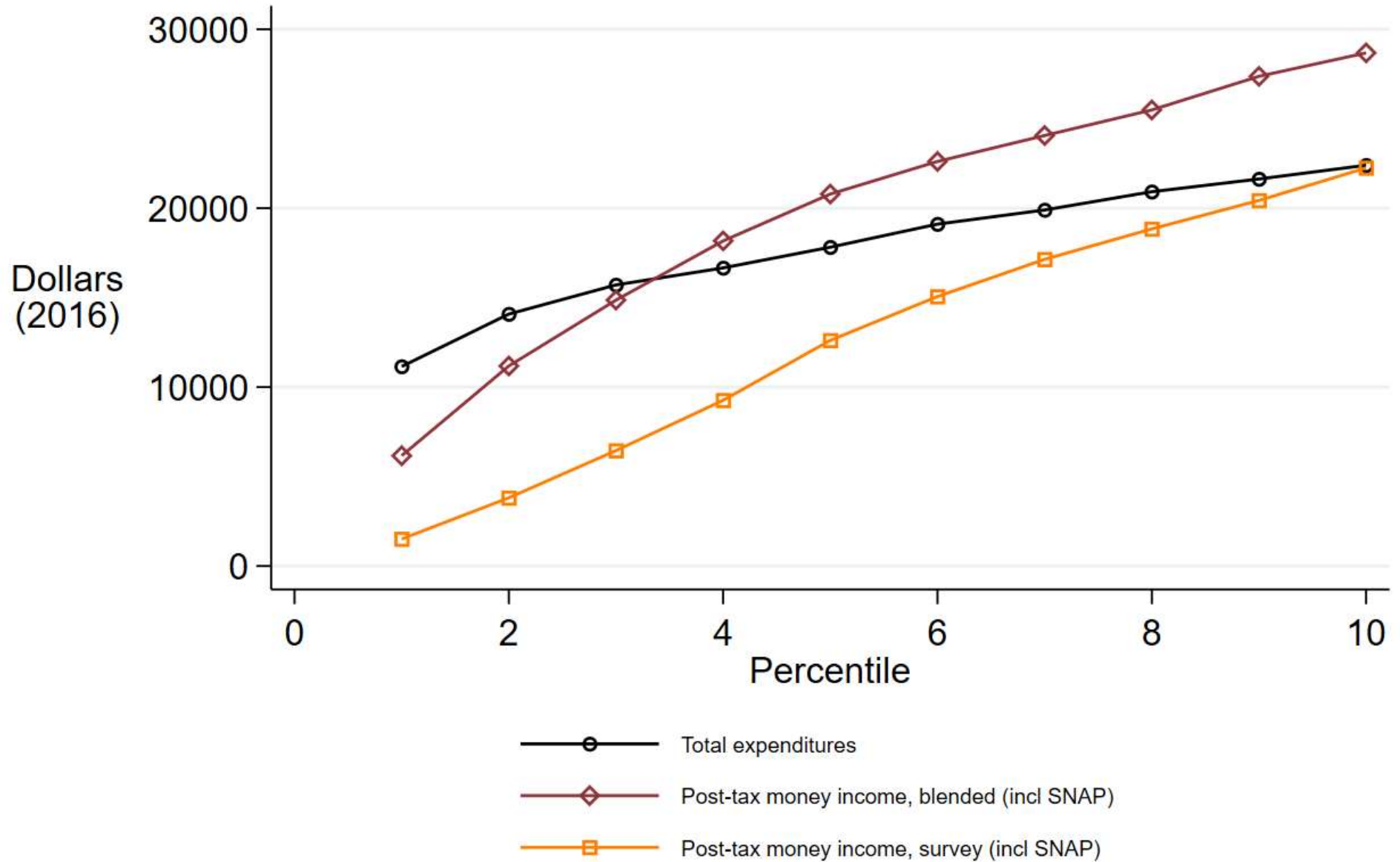
Appendix Slides



Mean Expenditure below Percentile of Income



Distribution of Resource Measures Bottom 10 Percentiles



Blending Pre-Tax Money Income

- The vast majority of Consumer Units (CUs) file tax returns. For these CUs, we use AGI reported on Form 1040 as a starting point
- But, admin AGI is net of deductions and may miss some jobs and informal income sources
 - Therefore, we continue to use survey analog of AGI when it is higher and reflects income plausibly missed in admin records
- We cannot perfectly align survey income components with the admin AGI concept so we modify admin AGI (AGI*) to match better
- We continue to use survey AGI* when higher only if survey earnings are non-imputed and at least one of the following holds:
 - Key administrative data (W2s or 1040s) are missing for an entire CU; or a CU member is primarily self-employed or works in a “high-tip” industry.



Blending Pre-Tax Money Income (cont.)

- For non-filers in CUs, we don't have 1040 AGI so begin with survey post-tax, post-transfer income as our baseline and substitute admin data whenever possible to create blended income
- We also account for additional cash income sources that aren't accounted for in AGI*



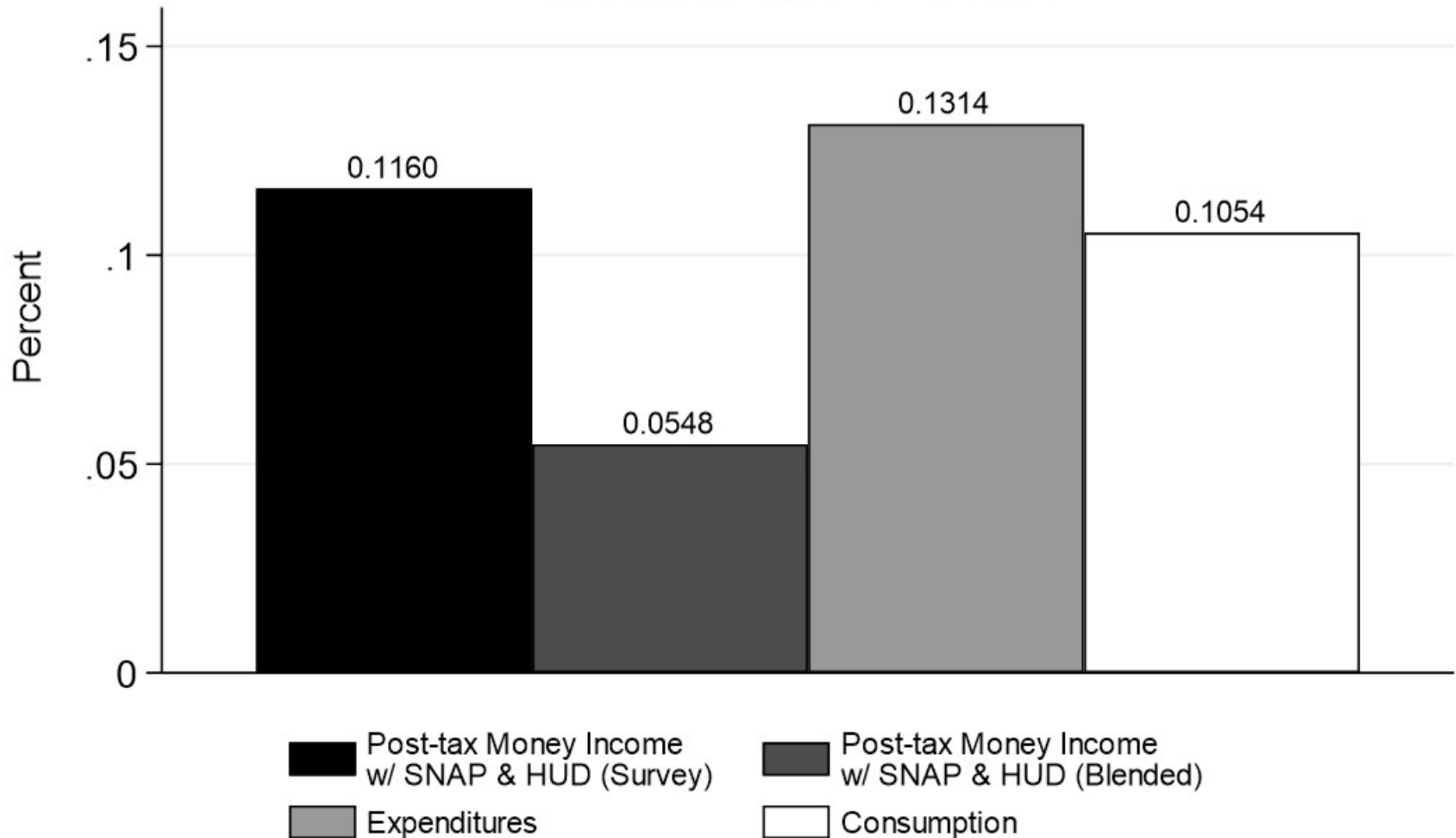
Incorporating Taxes and In-Kind Transfers

- Using TAXSIM we subtract tax liabilities and add credits based on 1040 variables (for non-filers use W2s and other forms)

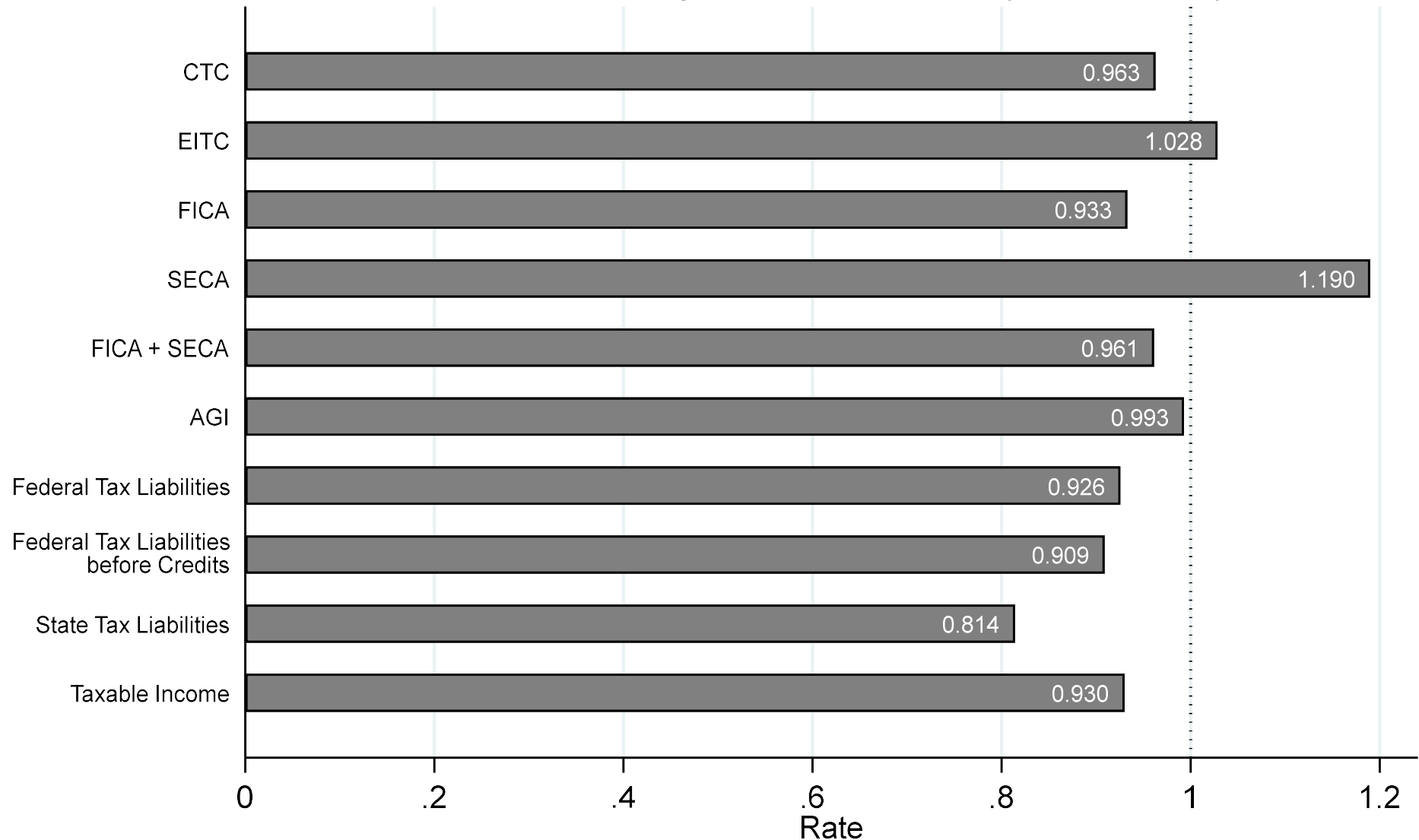


Share of People in CUs Below Poverty Line Key Resource Measures, All States

Individuals aged 65 or older



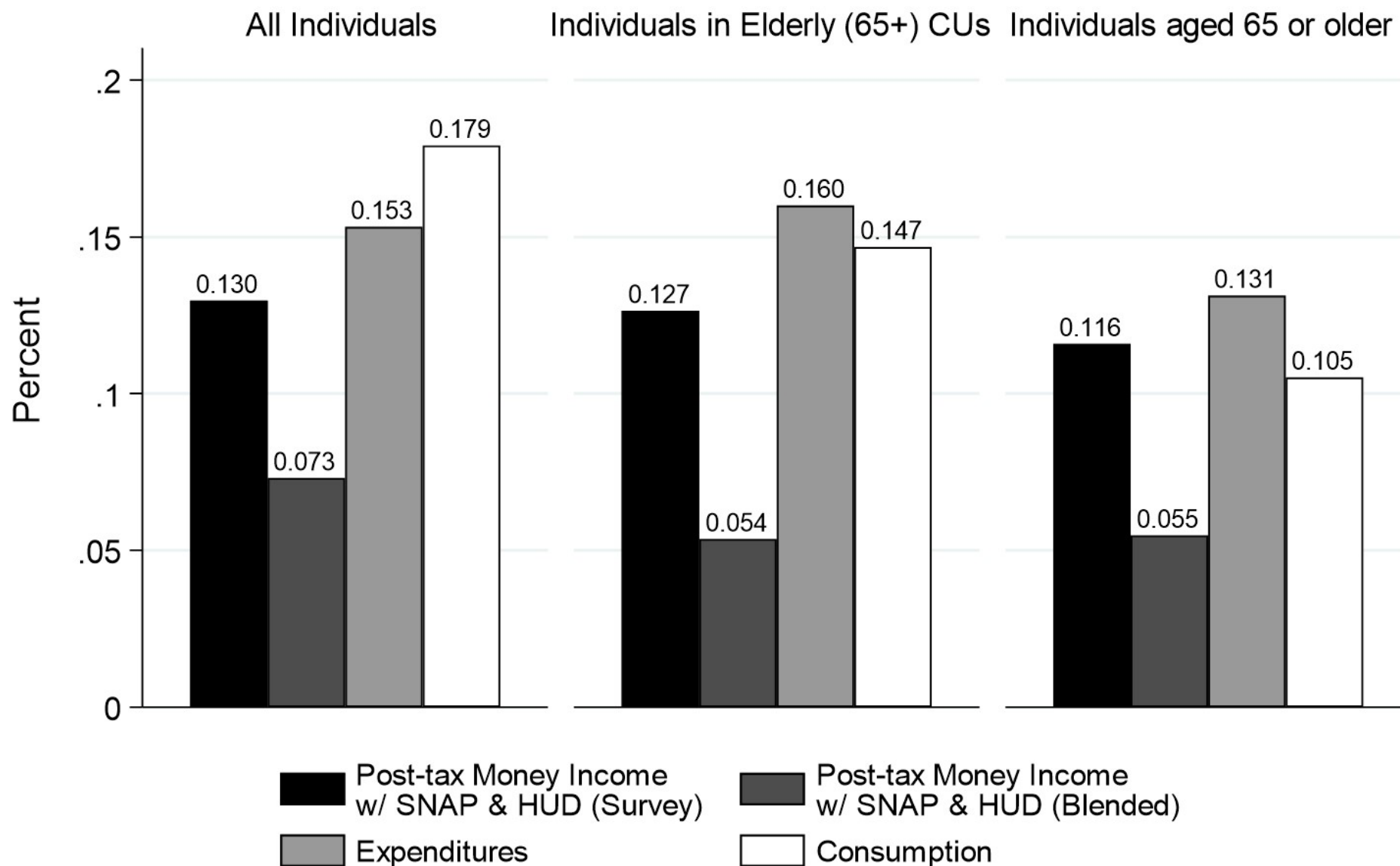
Average Reporting Rates of TAXSIM Outputs Created using Blended Inputs (2014-2016)



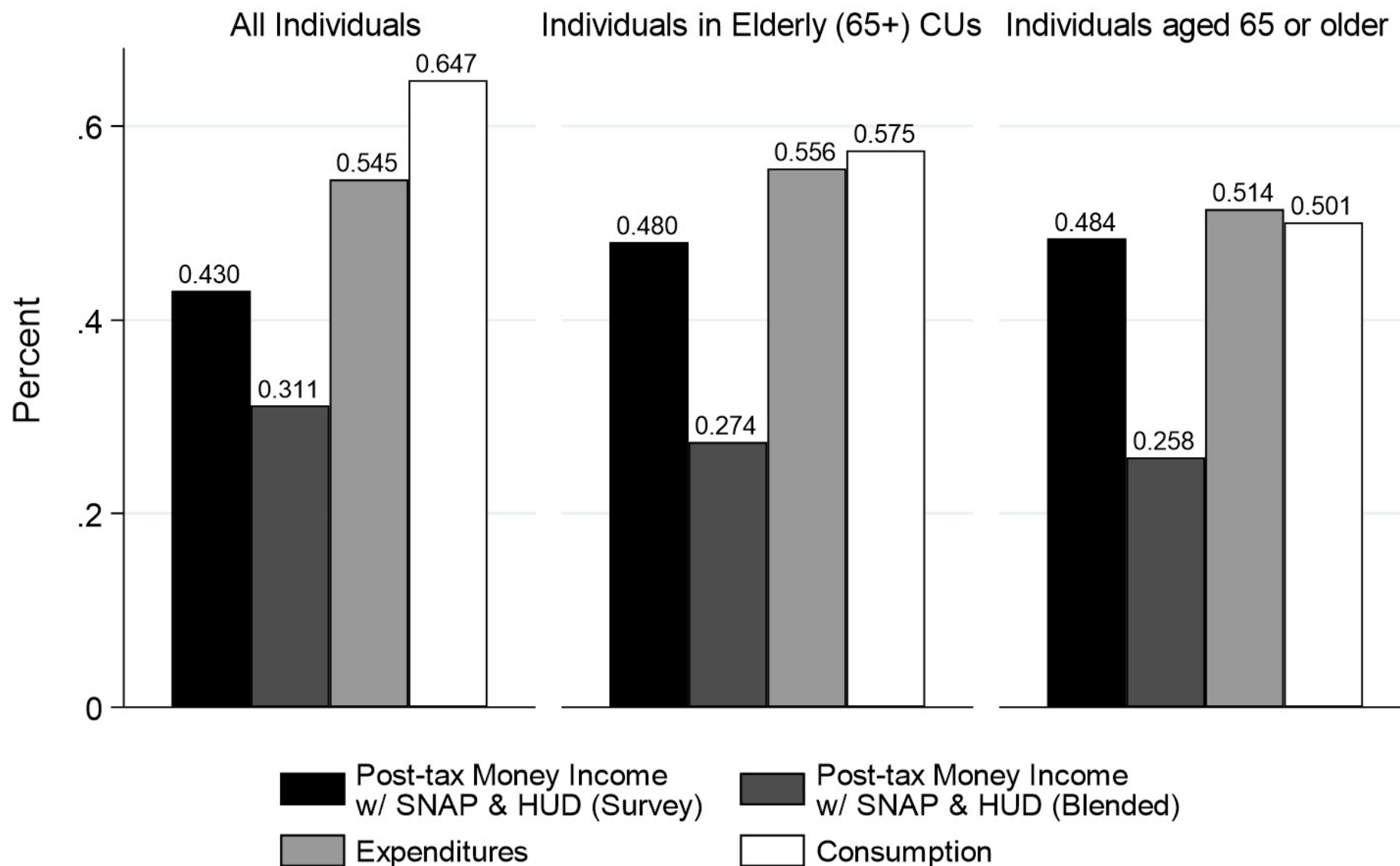
(1) We did not get the opportunity to disclose the reporting rates for TAXSIM outputs created using survey inputs, but the eventual goal would be to compare the two in a figure like this.



Share of People in CUs Below Poverty Line Key Resource Measures, All States

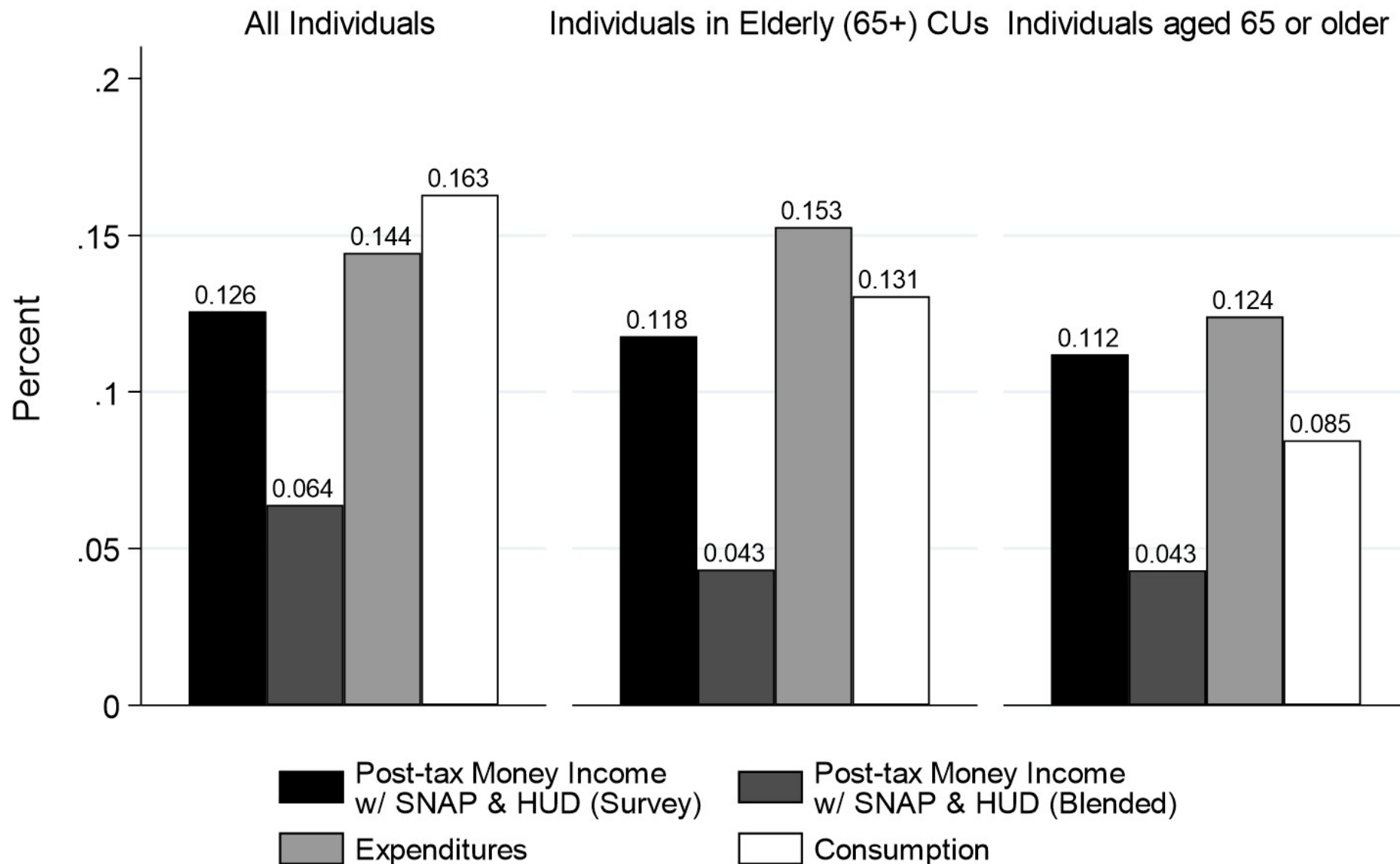


Share of People in CUs Below Twice Poverty Line Key Resource Measures, All States



Share of People in CUs Below Poverty Line

Key Resource Measures, States with Administrative SNAP available



Blending Pre-Tax Money Income

- Vast majority of CUs file tax returns. For these CUs, we use AGI reported on Form 1040 as a starting point
 - AGI includes important income sources (e.g. alimony, capital gains, and unemployment insurance) that we don't have in other administrative data
- But, admin AGI is net of deductions and may miss some jobs and informal income sources
 - Therefore, we continue to use survey analog of AGI when it is higher and reflects income plausibly missed in admin records
- Because we cannot perfectly align survey income components with admin AGI concept, we modify admin AGI to conceptually match the union of selected components of survey pre-tax money income
 - $AGI^* = AGI + \text{non-taxed wage/interest income} + VA\ DC - \text{taxable OASDI}$
- We continue to use survey AGI^* when higher only if survey earnings are non-imputed and at least one of the following holds:
 - Administrative Data (W2s or 1040s) is missing for the entire CU
 - Anyone in the CU indicated self-employment being the primary source of their income in the past 12 months
 - Anyone in the CU indicated that they work in a “high-tip” industry



Percentiles and Means of Income and Expenditures

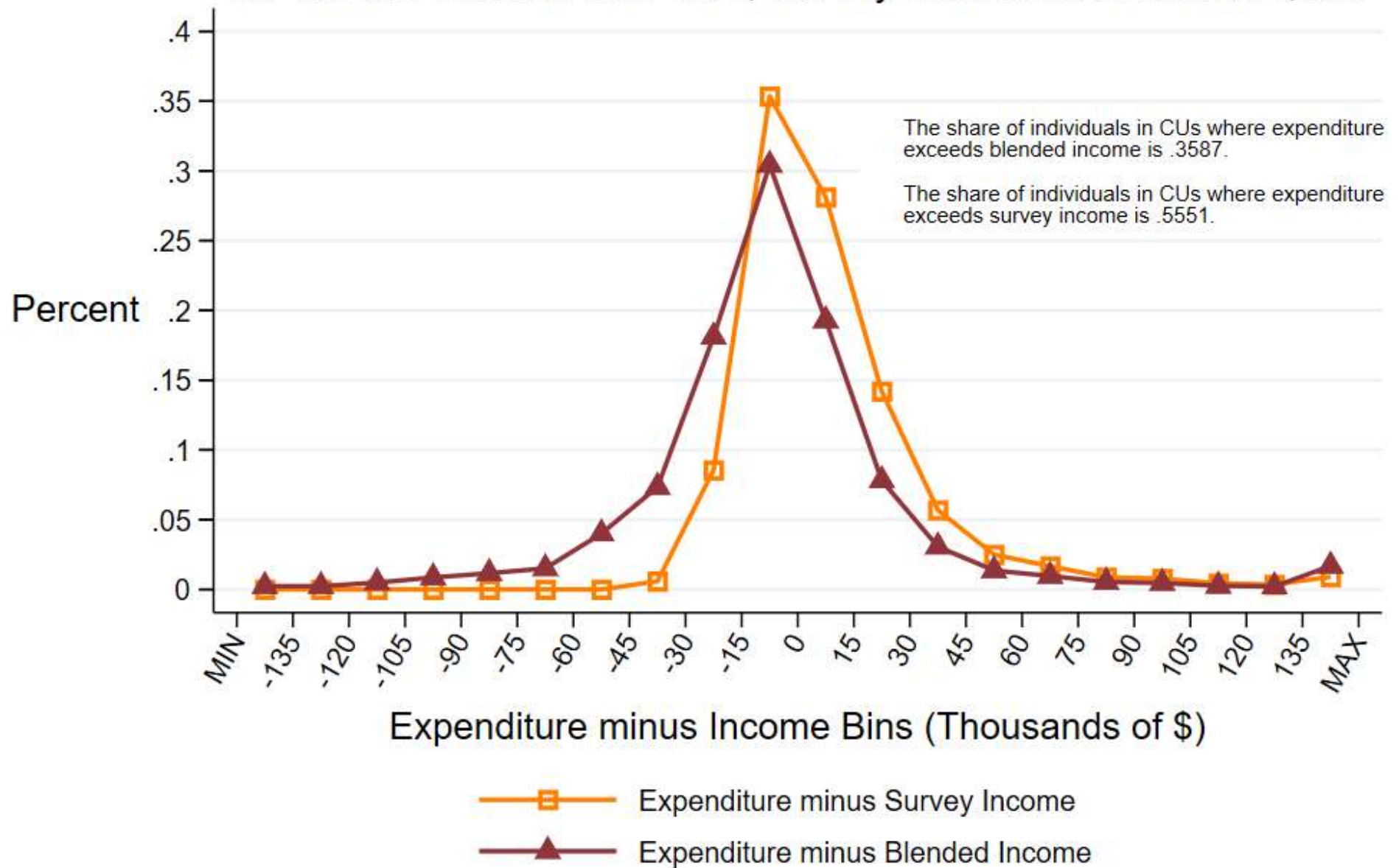
Quarterly Expenditures, Interview 1-4	Percentiles								
	3 rd	5 th	10 th	20 th	30 th	40 th	50 th	75 th	90 th
Total Family Survey Income (CE)									
(1) Percentile of Survey Income	6448	12610	22250	32150	40660	49720	59620	93550	143600
(2) Mean for those below given percentile of survey income	2501	5287	11610	19520	25260	30380	35270	48580	59800
(3) Mean for those below given percentile of expenditures	21940	24110	27200	33140	36870	41050	45200	57120	66080
Total Family Blended Income (CE)									
(4) Percentile of Blended Income	14860	20790	28680	40370	50940	63420	75980	119200	188200
(5) Mean for those below given percentile of blended income	8282	12390	18830	26710	33070	39120	45240	62000	76620
(6) Mean for those below given percentile of expenditures	33020	37170	39880	46600	50700	55580	60210	74540	86420
Total Family Expenditures (CE)									
(7) Percentile of expenditures	15710	17820	22400	29420	35730	42360	49430	75990	119800
(8) Mean for those below given percentile of expenditures	11840	13780	17100	21640	25300	28790	32240	42070	51050
(9) Mean for those below given percentile of survey income	34270	36950	34590	35560	37240	40140	43080	50230	57530
(10) Mean for those below given percentile of blended income	31340	31060	31440	33480	36650	39660	42630	50430	57520

*All values are in 2016 dollars

**All percentiles displayed are not true percentiles of the weighted income distribution. Rather, they are interpolated percentiles, obtained by taking the mean of 15 greatest observations within each percentile.

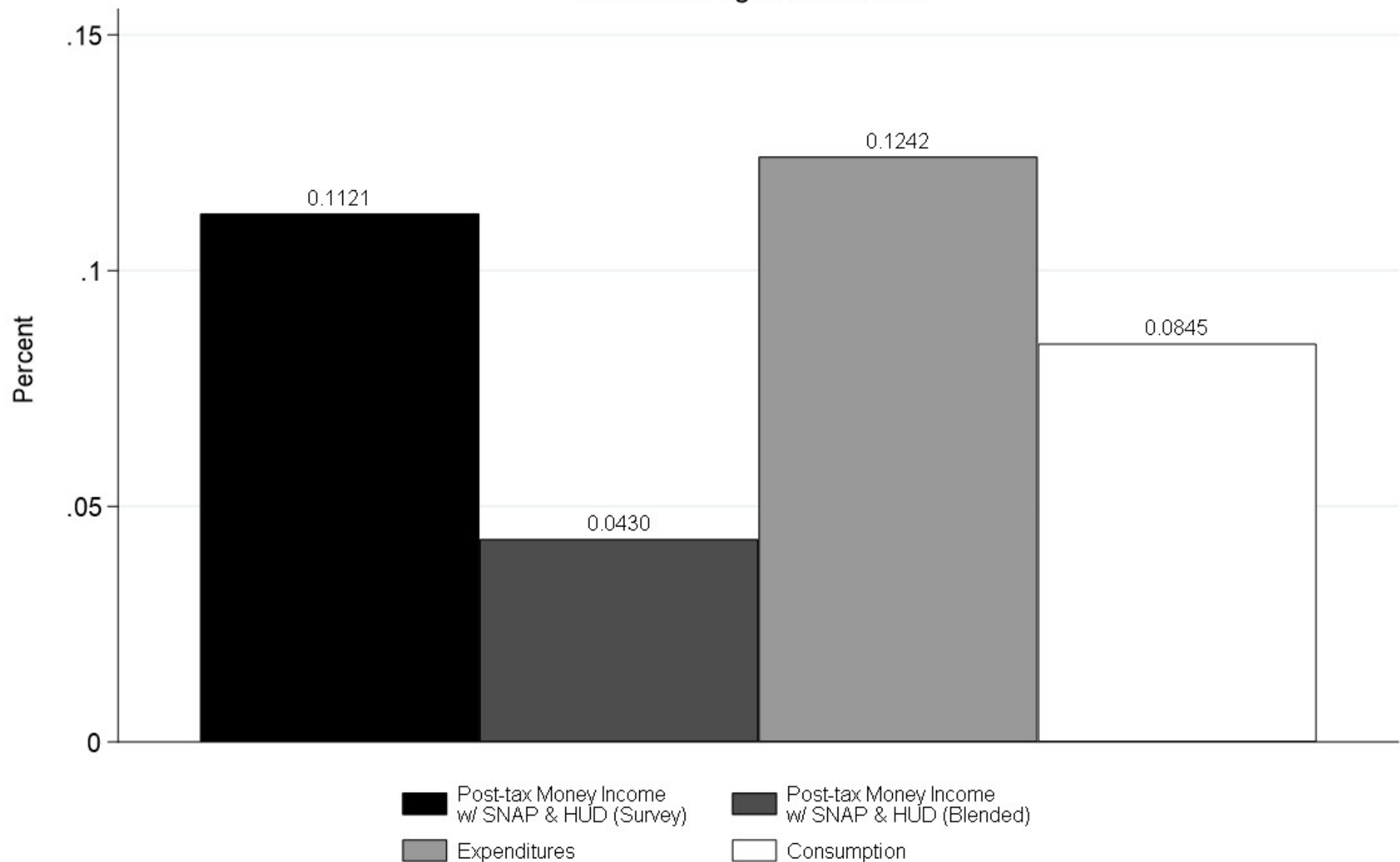


Distribution of Expenditure minus Income for 1st and 4th Interview CUs, Survey Income less than \$50,000



Share of People in CUs Below Poverty Line Key Resource Measures, States with Administrative SNAP available

Individuals aged 65 or older



Slide 34

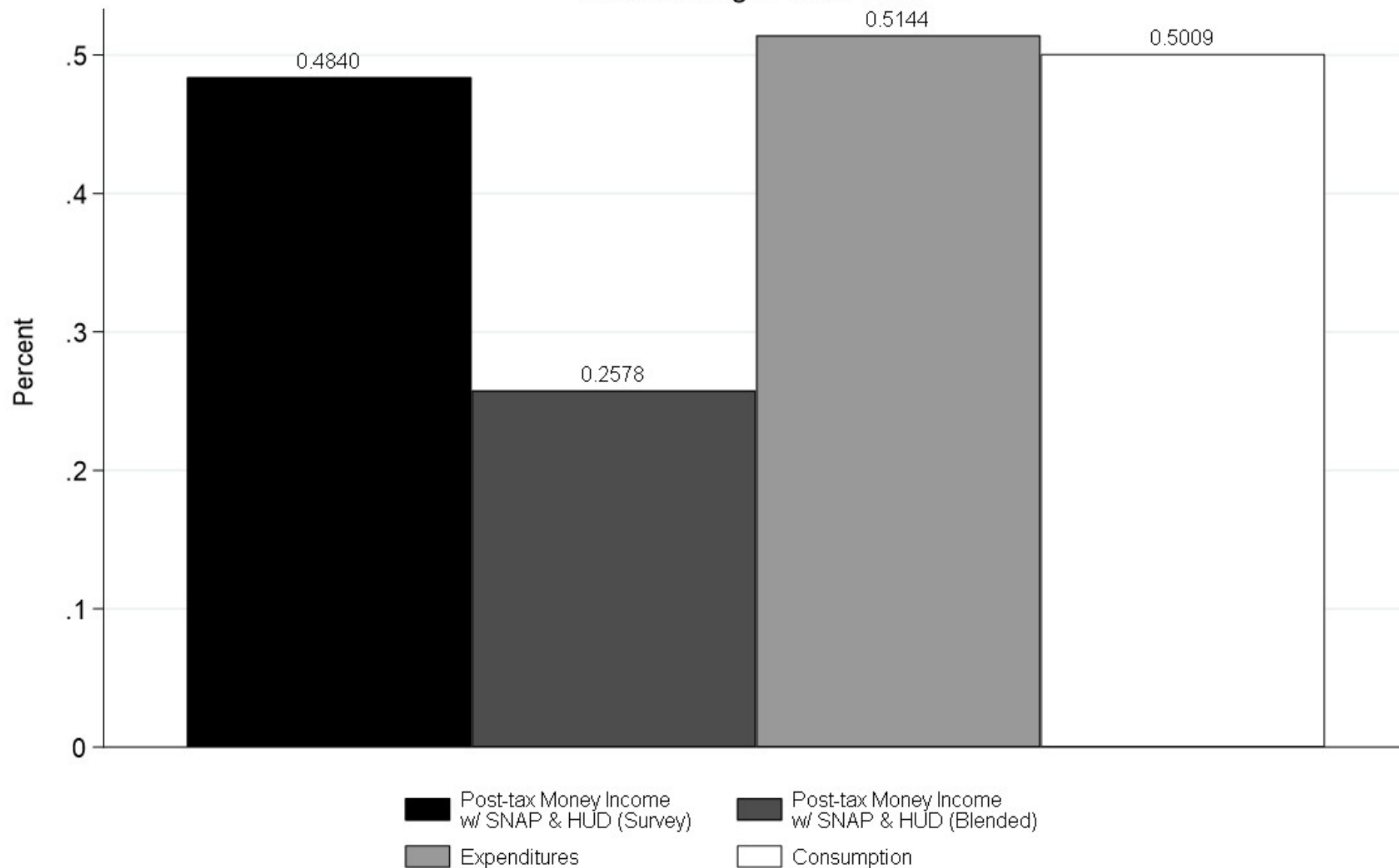
AH0

Slide 42 and 43 can probably be deleted since they contain the same information as the right most figure in slide 34 and 35.

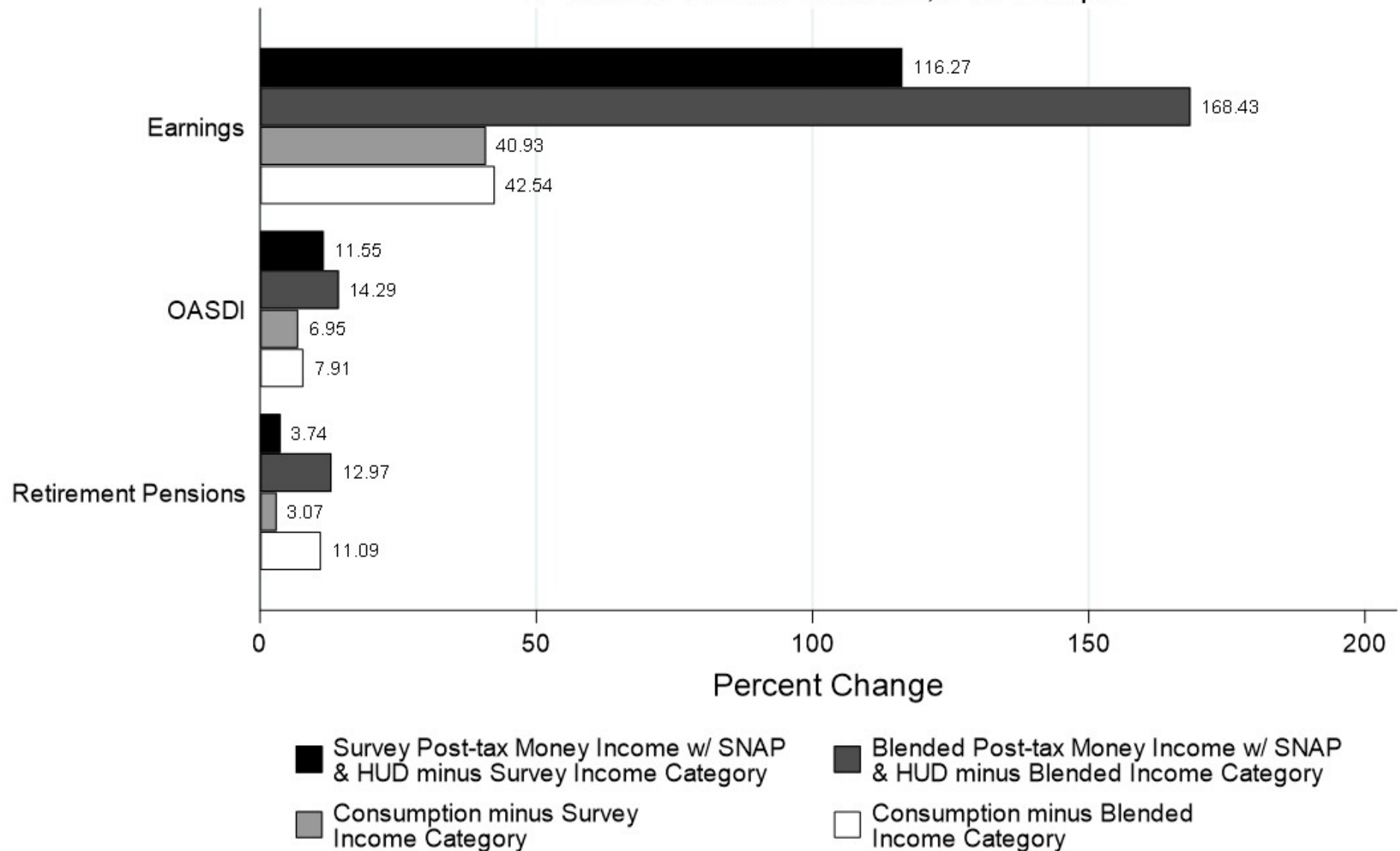
Aaron Hong, 2023-08-02T23:52:41.416

Share of People in CUs Below Twice Poverty Line Key Resource Measures, All States

Individuals aged 65 or older



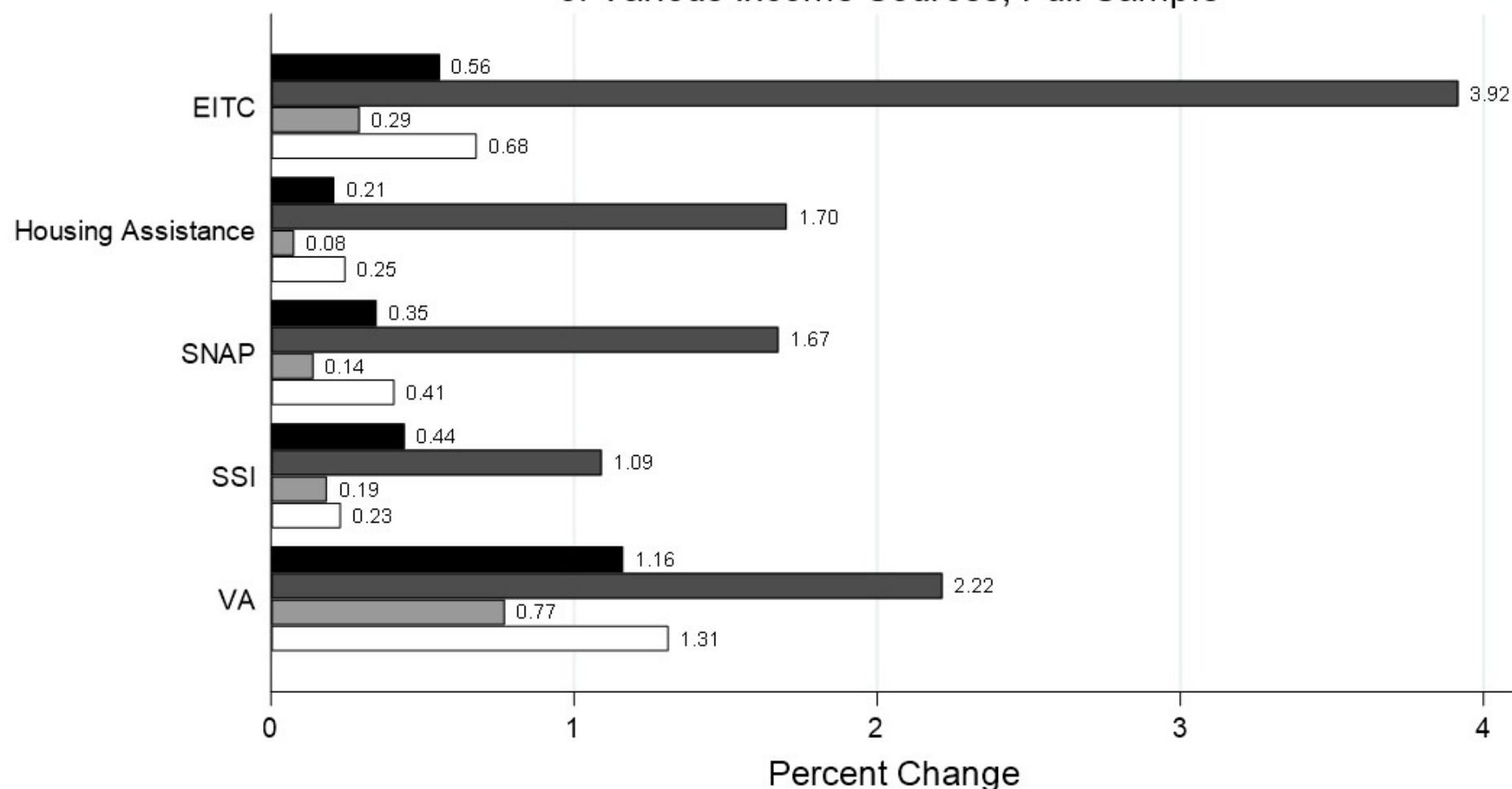
Percent Increase in Share Below Twice Poverty Line with Loss of Various Income Sources, Full Sample



(1) These income categories have so large an effect on the share of the population below multiples of the poverty line for many that we display their effects in a separate figure.

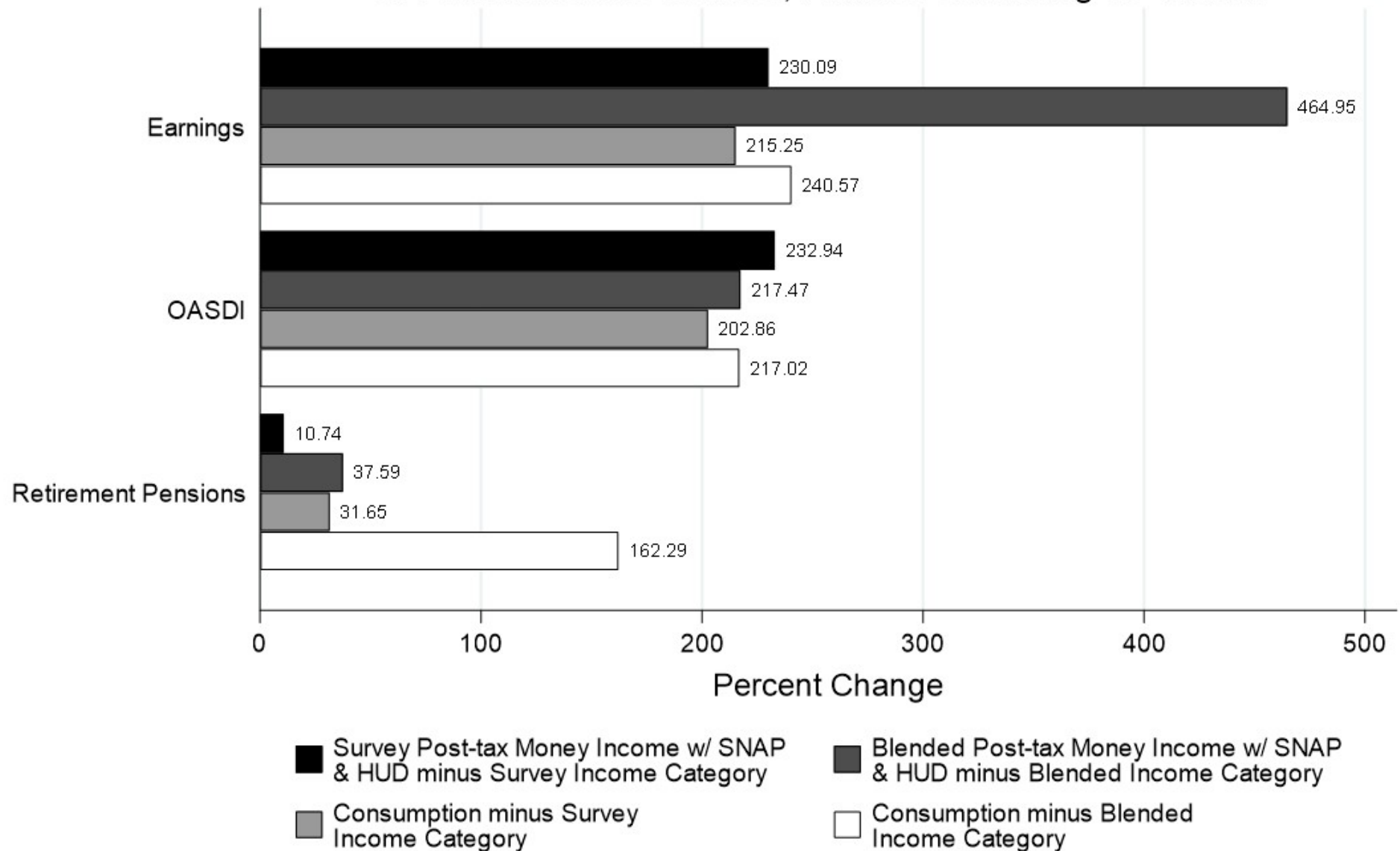


Percent Increase in Share Below Twice Poverty Line with Loss of Various Income Sources, Full Sample



(1) The survey category for VA contains all other sources of what the CE defines as regular non-rental income.

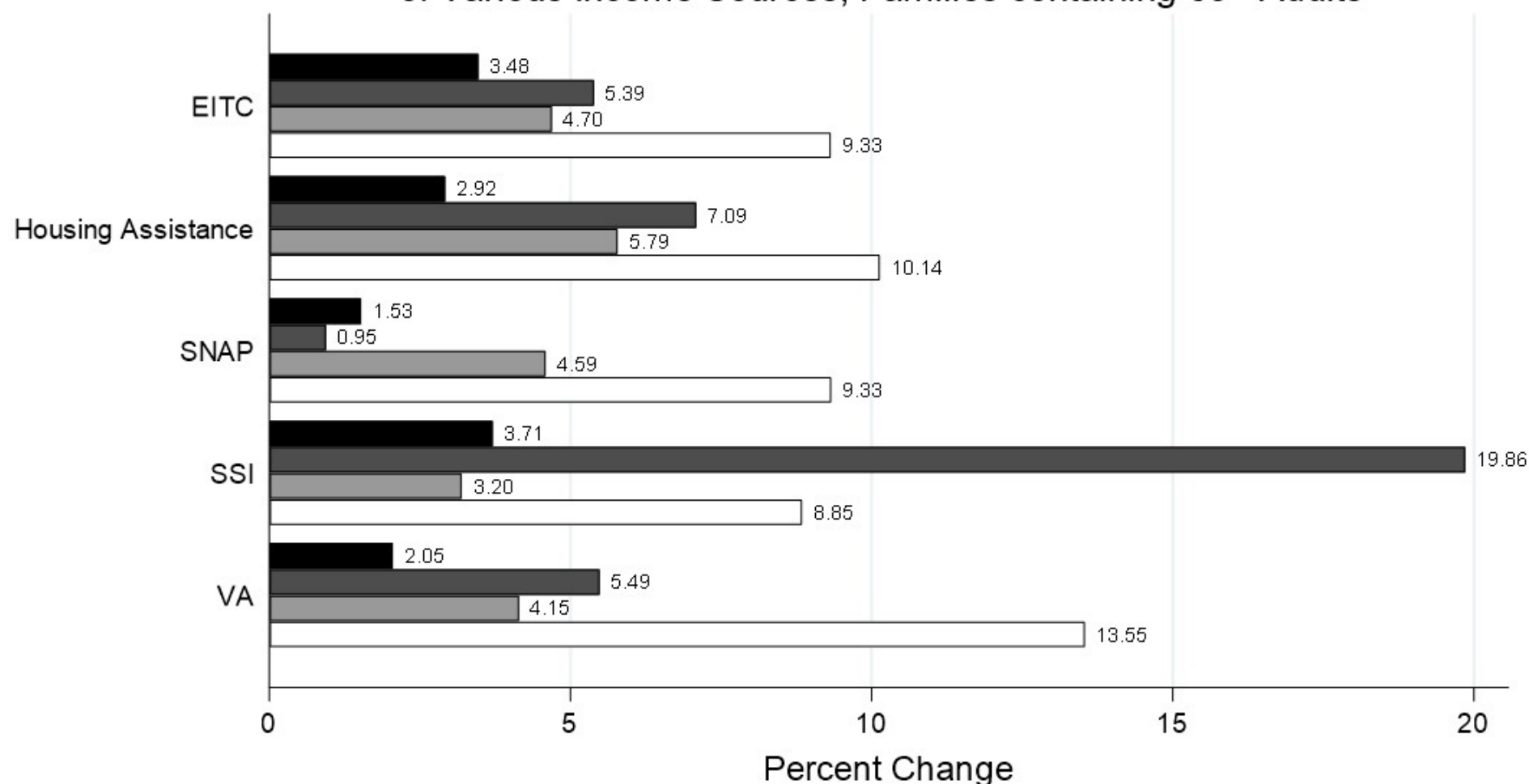
Percent Increase in Share Below Poverty Line with Loss of Various Income Sources, Families containing 65+ Adults



(1) These income categories have so large an effect on the share of the population below multiples of the poverty line for many that we display their effects in a separate figure.

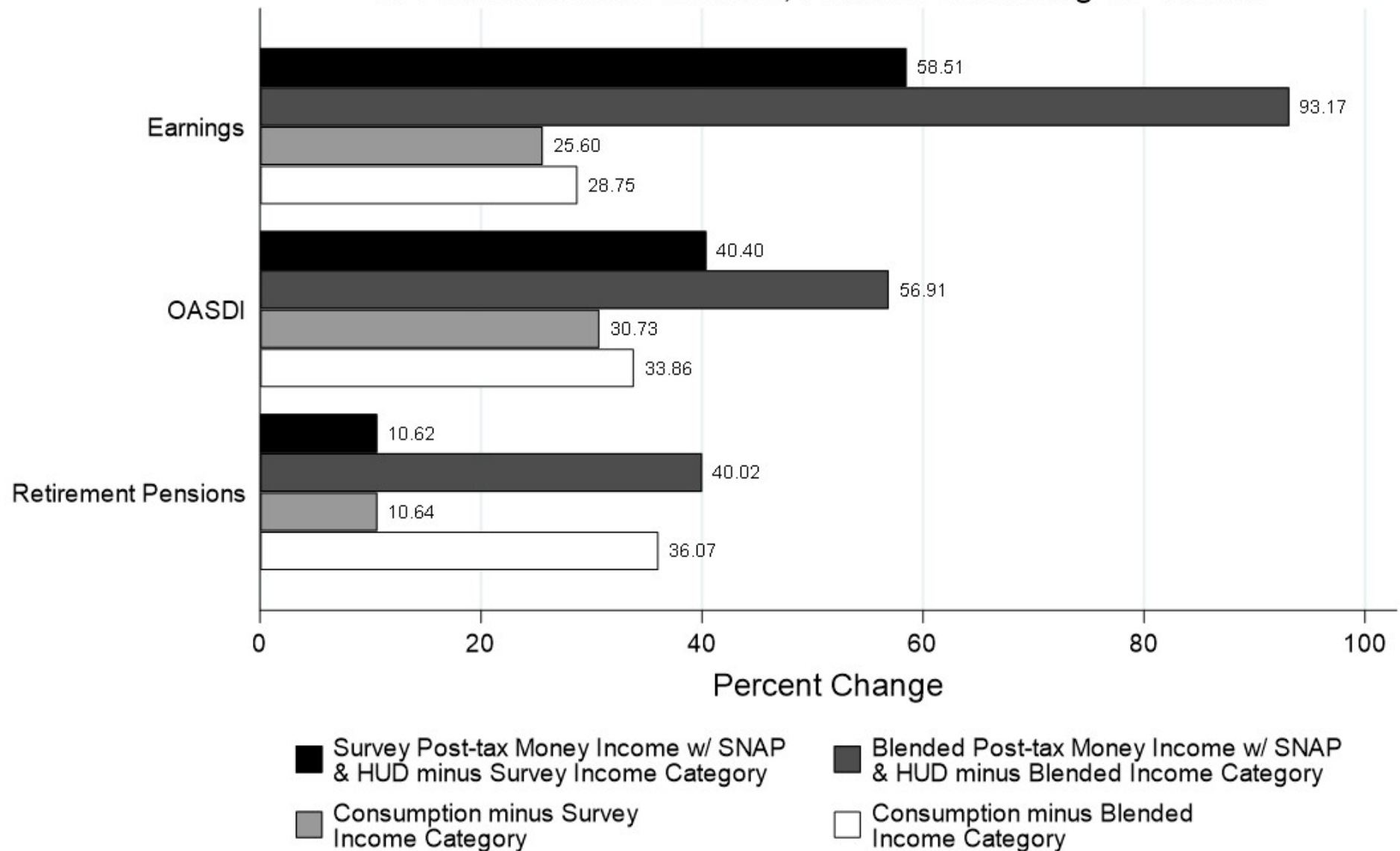


Percent Increase in Share Below Poverty Line with Loss of Various Income Sources, Families containing 65+ Adults



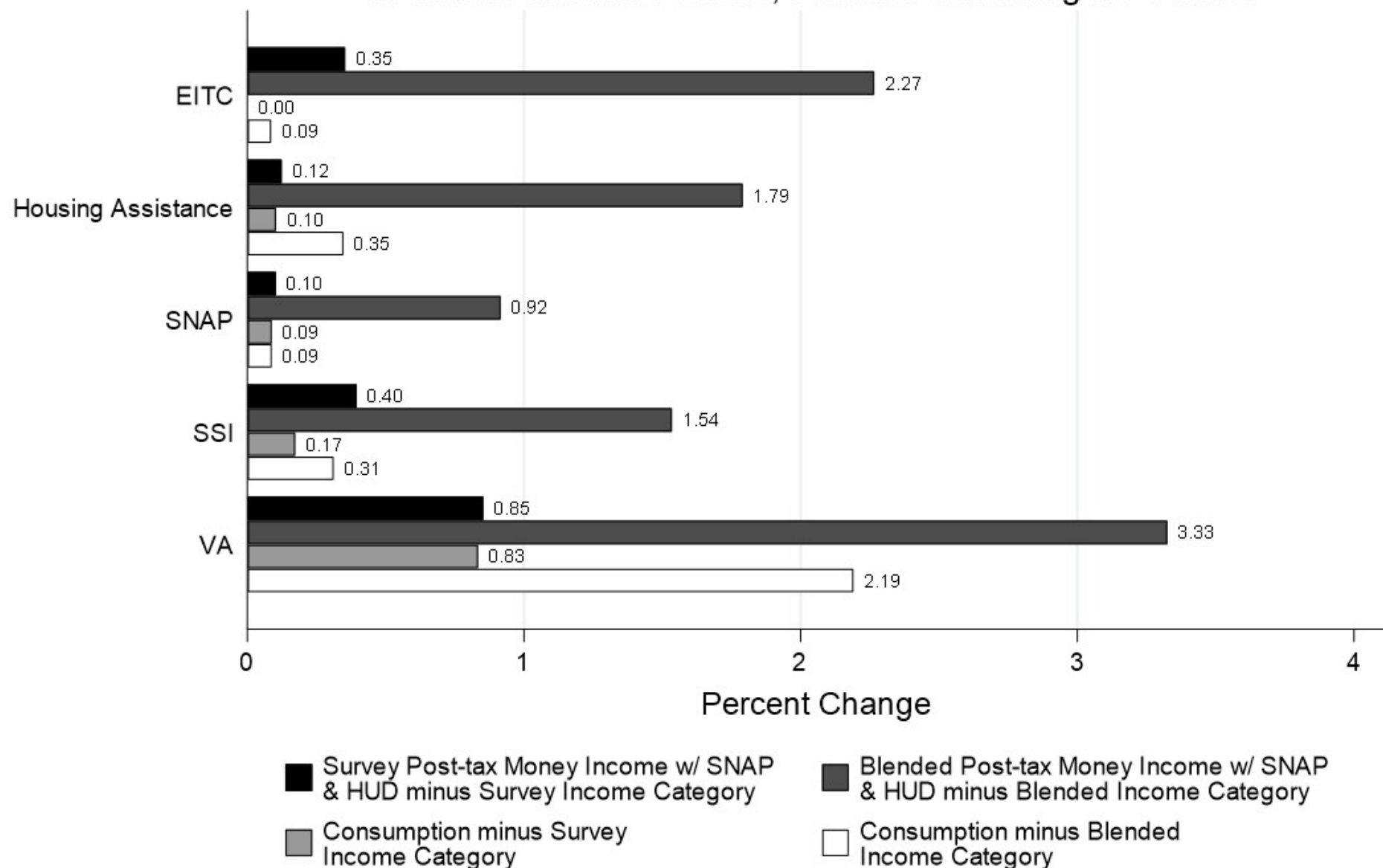
(1) The survey category for VA contains all other sources of what the CE defines as regular non-rental income.

Percent Increase in Share Below Twice Poverty Line with Loss of Various Income Sources, Families containing 65+ Adults



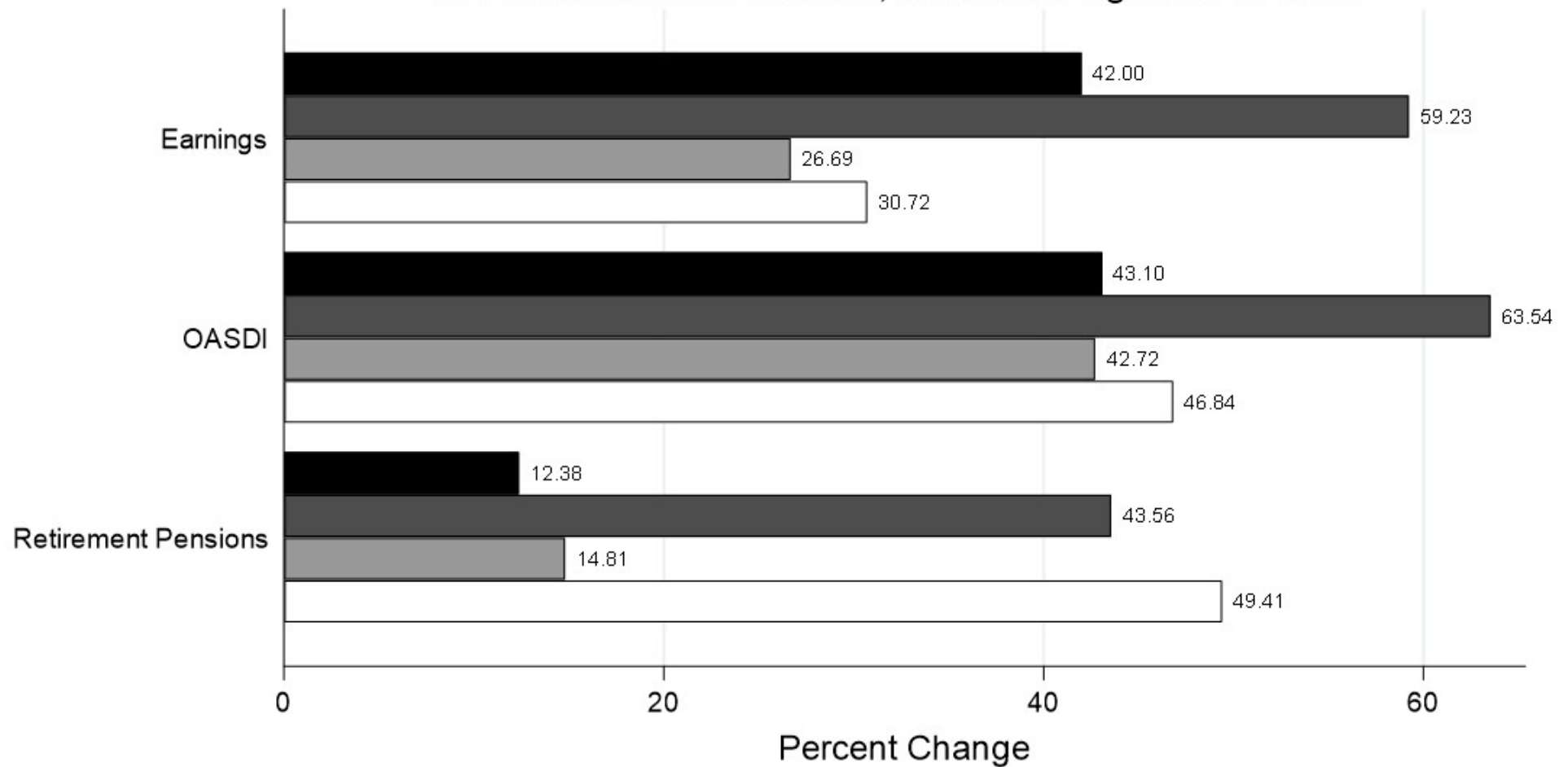
(1) These income categories have so large an effect on the share of the population below multiples of the poverty line for many that we display their effects in a separate figure.

Percent Increase in Share Below Twice Poverty Line with Loss of Various Income Sources, Families containing 65+ Adults



(1) The survey category for VA contains all other sources of what the CE defines as regular non-rental income.

Percent Increase in Share Below Twice Poverty Line with Loss of Various Income Sources, Individuals aged 65 or older

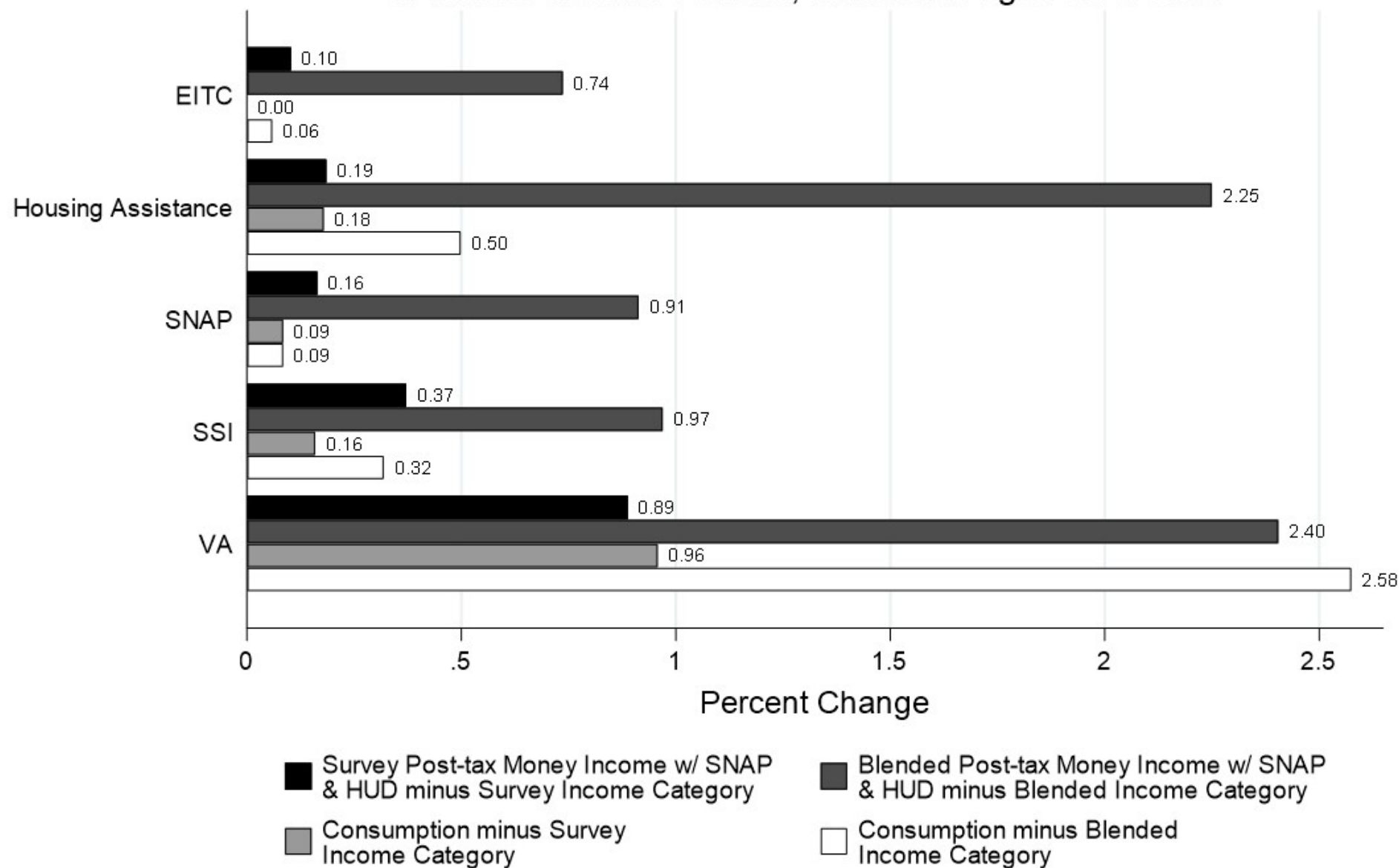


- Survey Post-tax Money Income w/ SNAP & HUD minus Survey Income Category
- Blended Post-tax Money Income w/ SNAP & HUD minus Blended Income Category
- Consumption minus Survey Income Category
- Consumption minus Blended Income Category

(1) These income categories have so large an effect on the share of the population below multiples of the poverty line for many that we display their effects in a separate figure.



Percent Increase in Share Below Twice Poverty Line with Loss of Various Income Sources, Individuals aged 65 or older



(1) The survey category for VA contains all other sources of what the CE defines as regular non-rental income.

Comparing Income and Expenditure Distributions



Univariate Distributions of Resource Measures

- As a part of our analyses comparing income to expenditures, we plot the univariate distributions of three different resource measures:
 - Post-tax money income plus SNAP benefits based on survey data only
 - Post-tax money income plus SNAP benefits obtained by blending administrative data with survey data
 - Yearly expenditures obtained by scaling up quarterly expenditures
- Values are equivalence scaled to a two-adult, two-child family unit and inflation adjusted to 2016 dollars



Joint Distributions of Resource Measures

- We create two figures that report the relationship between expenditures and blended/survey income
 - The first figure reports the relationship between median expenditure conditioned on survey and blended income for the lower half of the income distribution
 - Each marker represents groups of three percentiles

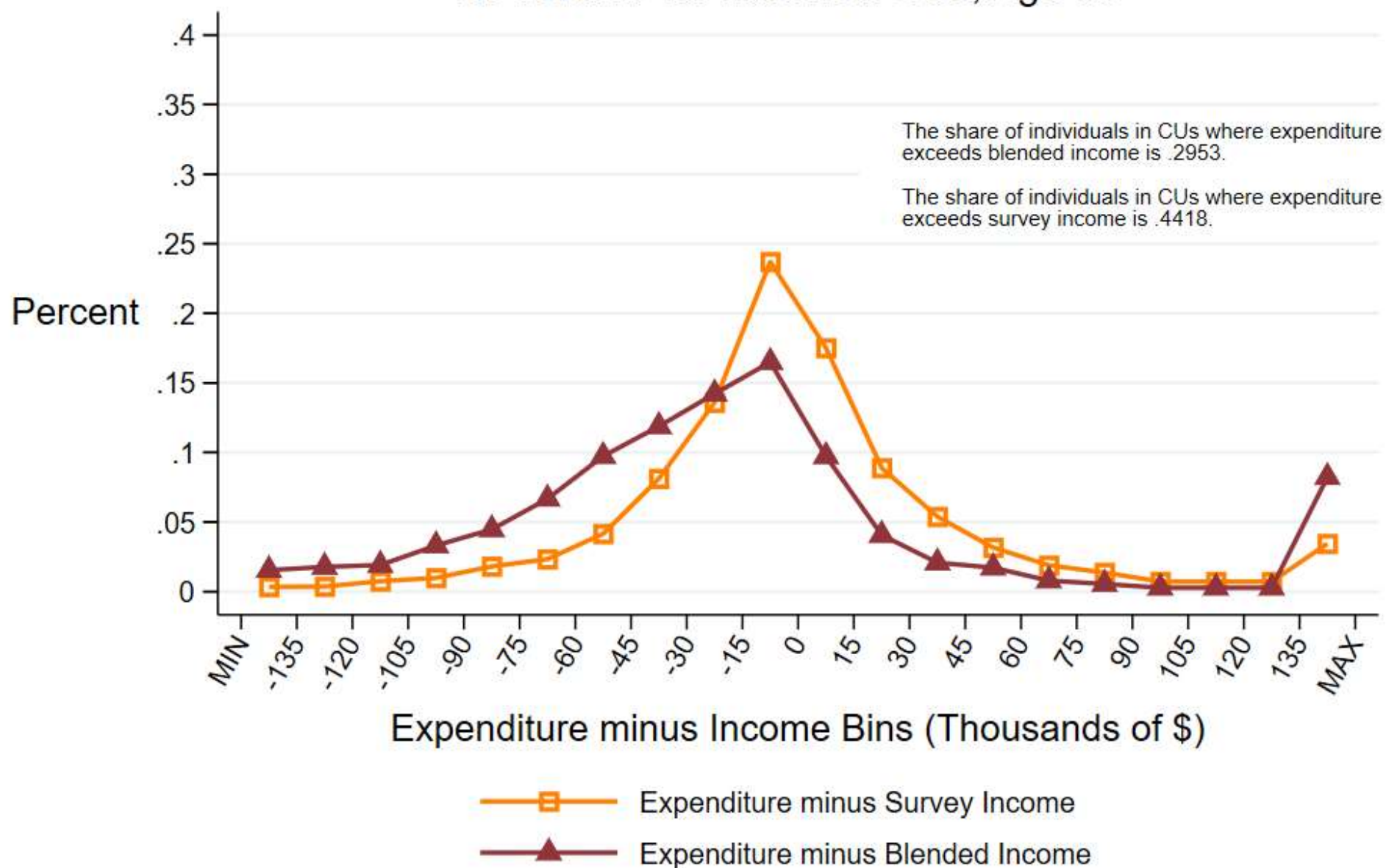


Comparisons of Expenditures minus Income

- As part of our analysis looking at individual differences between expenditure and income, we compare the distributions of expenditure minus blended and survey income across samples containing individuals interviewed in the 1st and 4th wave
- We plot the densities of expenditures minus blended and survey income for those aged 65 or older
 - The vertical axis is the percent of individuals is the Percent of CUs, Size Weighted
 - The horizontal axis is expenditure minus income bins in the thousands of dollars
 - The markers are at the midpoint of each bin



Distribution of Expenditure minus Income for 1st and 4th Interview CUs, Age 65+



Outline of Paper

1. Describe our data
2. Summarize how we combine survey and administrative data to construct resource measures
3. Evaluate a blended income measure by comparing weighted totals of select income categories to publicly available aggregate data
4. Compare income and expenditure distributions
5. Examine the difference in poverty across key resource measures
6. Examine the effect of key income sources on income and consumption poverty rates



Income and Consumption Poverty Measures



Poverty Reduction of Key Income Sources



THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL
OF PUBLIC POLICY

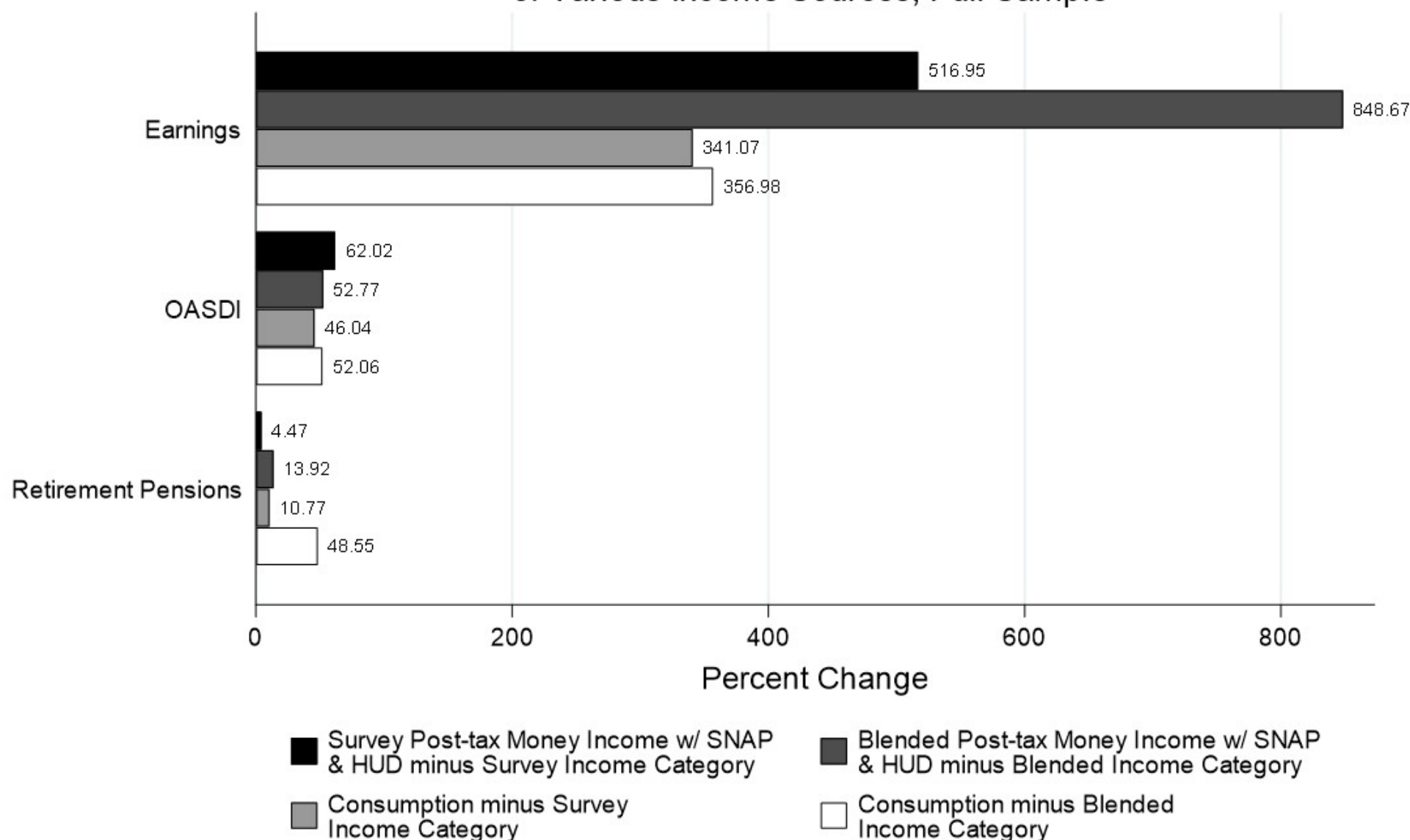
Comprehensive Income Dataset Project

What is the poverty reduction of different programs measured using income and consumption?

- To assess the impact of various income sources on poverty reduction, we recalculate the share of people below the poverty line after excluding the value of various income sources and programs in four ways
 - Survey post-tax money income + SNAP + HUD benefits minus survey income source
 - Blended post-tax money income + SNAP + HUD benefits minus blended income source
 - Consumption minus survey income source
 - Consumption minus blended income source
- We do this for OASDI, retirement pensions, earnings (sum of wage and salary and self-employment income), SNAP, EITC, CTC, Housing Benefits, Veteran's Disability, SSI, and welfare
- Caveat: Assuming no behavioral responses to the transfers and $MPC=1$; Probably best thought of as medium to long run calculation, which makes MPC close to 1 more plausible

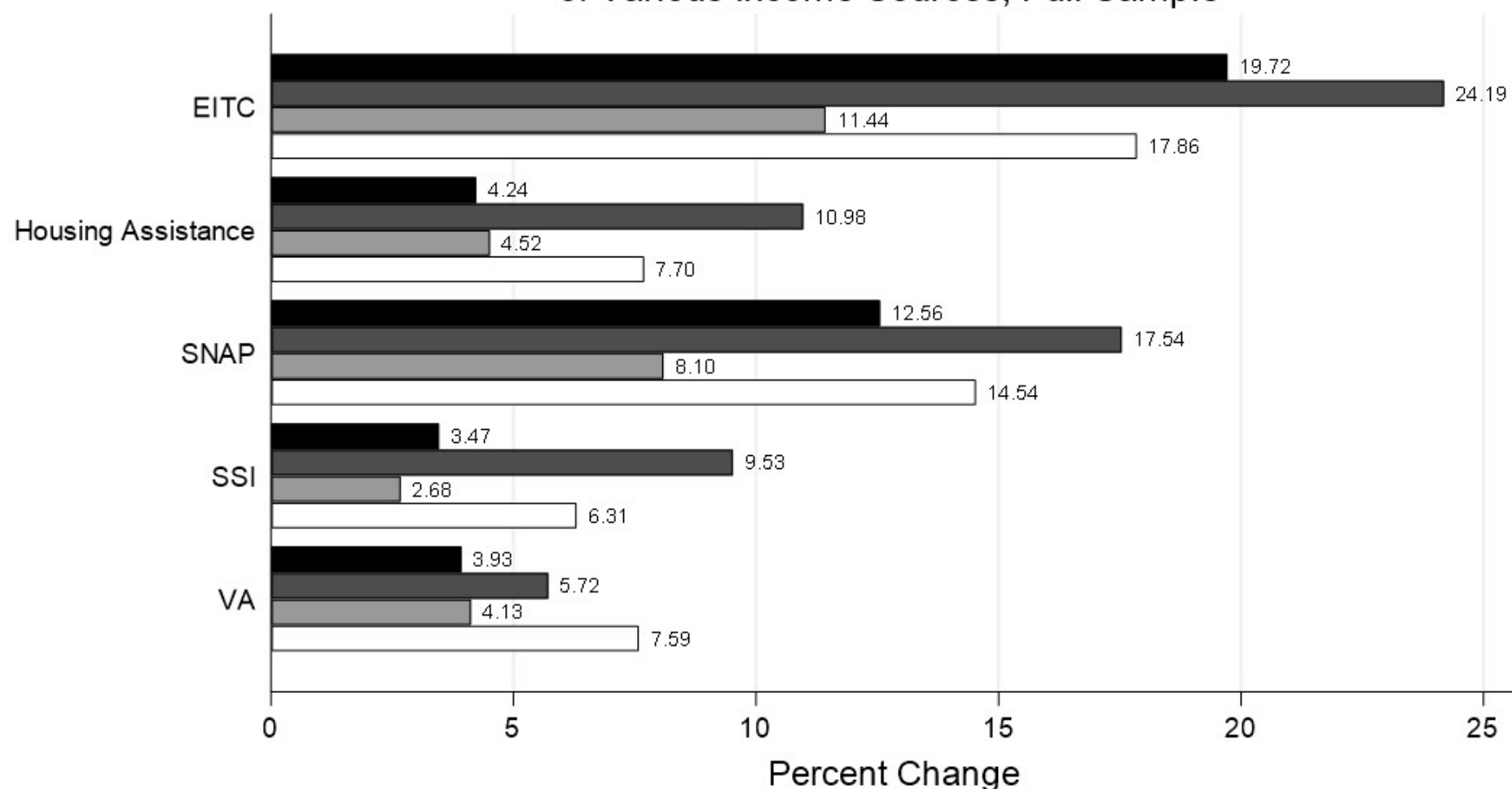


Percent Increase in Share Below Poverty Line with Loss of Various Income Sources, Full Sample



(1) These income categories have so large an effect on the share of the population below multiples of the poverty line for many that we display their effects in a separate figure.

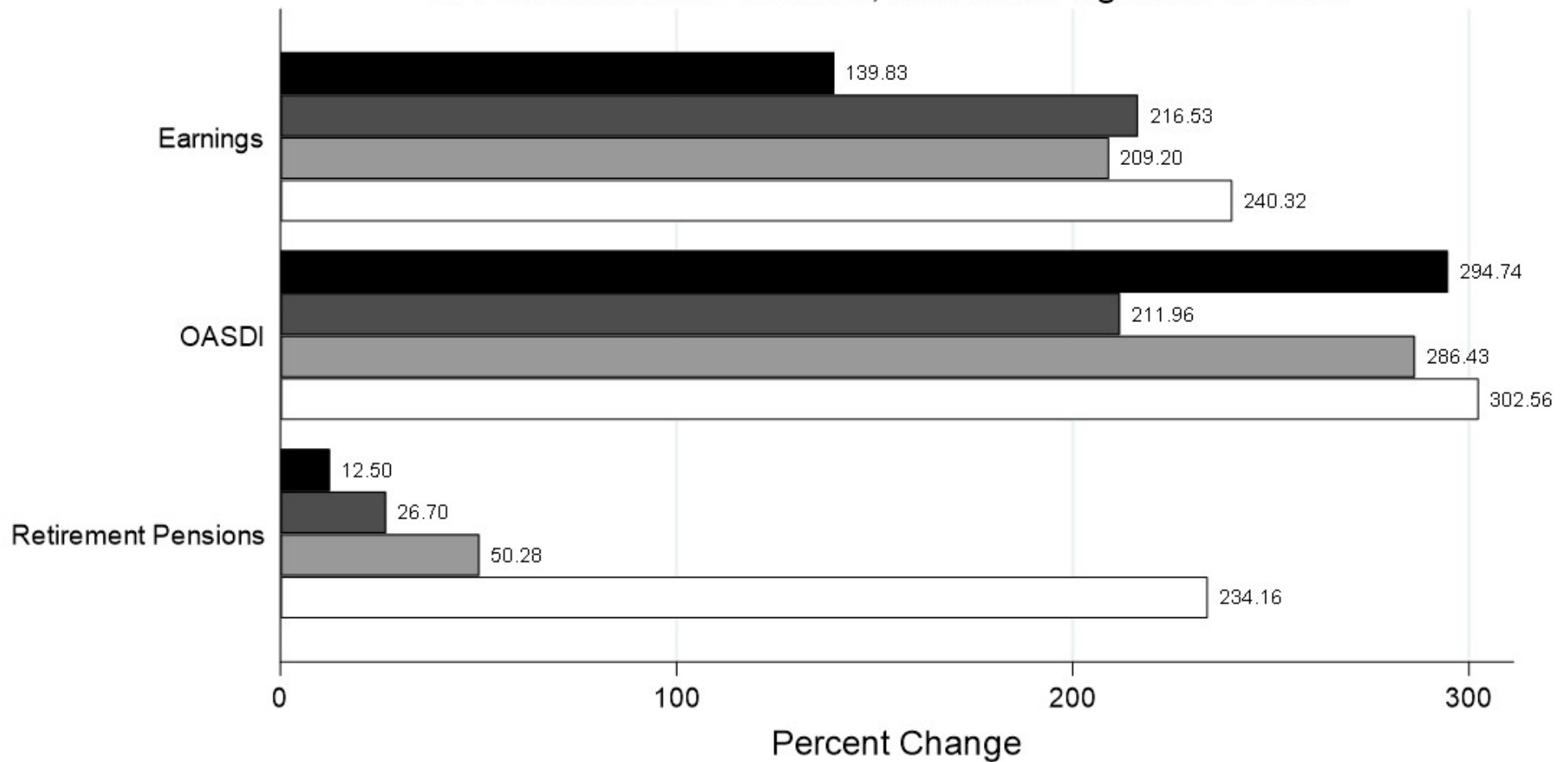
Percent Increase in Share Below Poverty Line with Loss of Various Income Sources, Full Sample



- Survey Post-tax Money Income w/ SNAP & HUD minus Survey Income Category
- Blended Post-tax Money Income w/ SNAP & HUD minus Blended Income Category
- Consumption minus Survey Income Category
- Consumption minus Blended Income Category

(1) The survey category for VA contains all other sources of what the CE defines as regular non-rental income.

Percent Increase in Share Below Poverty Line with Loss of Various Income Sources, Individuals aged 65 or older



■ Survey Post-tax Money Income w/ SNAP & HUD minus Survey Income Category

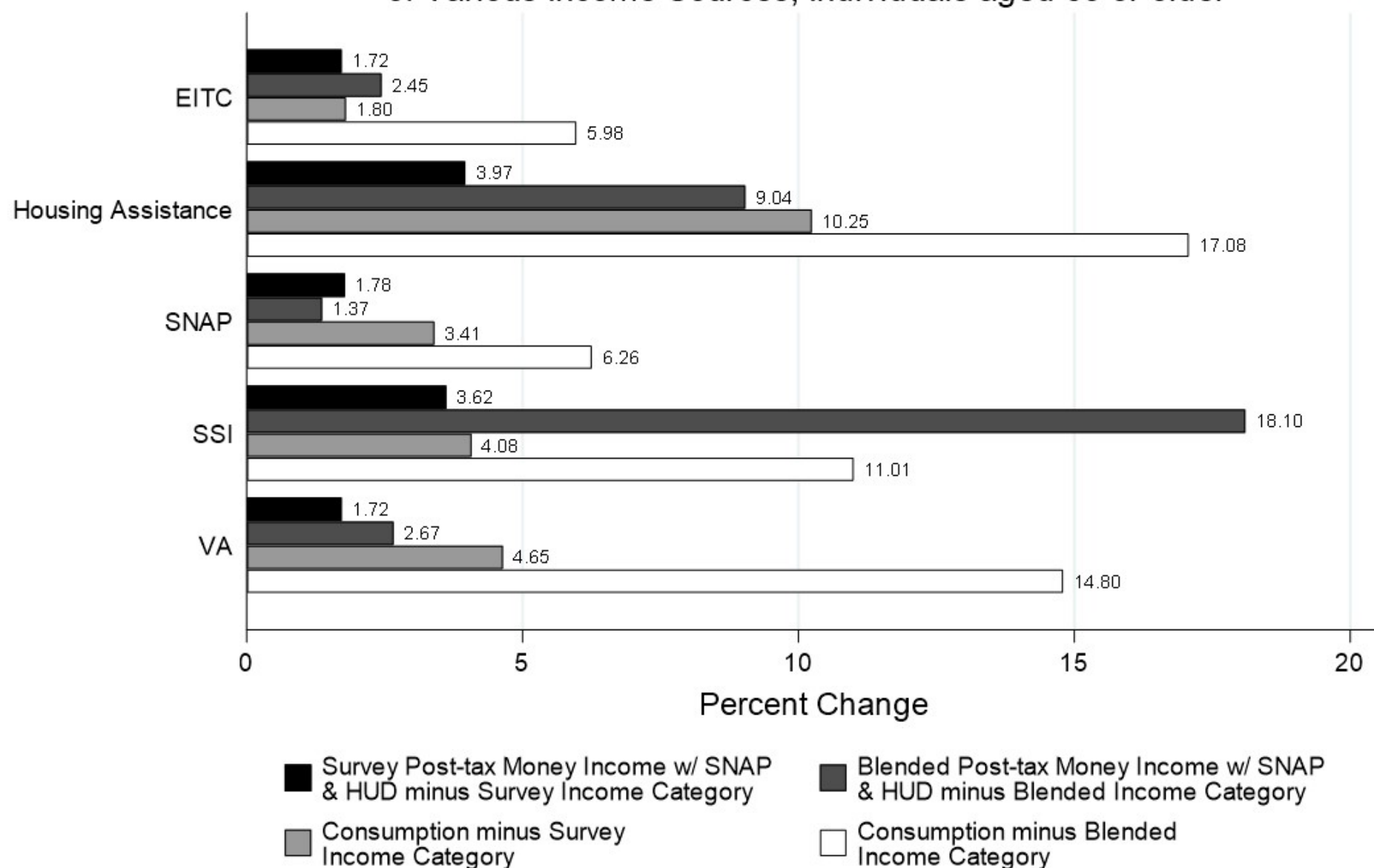
■ Blended Post-tax Money Income w/ SNAP & HUD minus Blended Income Category

■ Consumption minus Survey Income Category

■ Consumption minus Blended Income Category

(1) These income categories have so large an effect on the share of the population below multiples of the poverty line for many that we display their effects in a separate figure.

Percent Increase in Share Below Poverty Line with Loss of Various Income Sources, Individuals aged 65 or older



(1) The survey category for VA contains all other sources of what the CE defines as regular non-rental income.