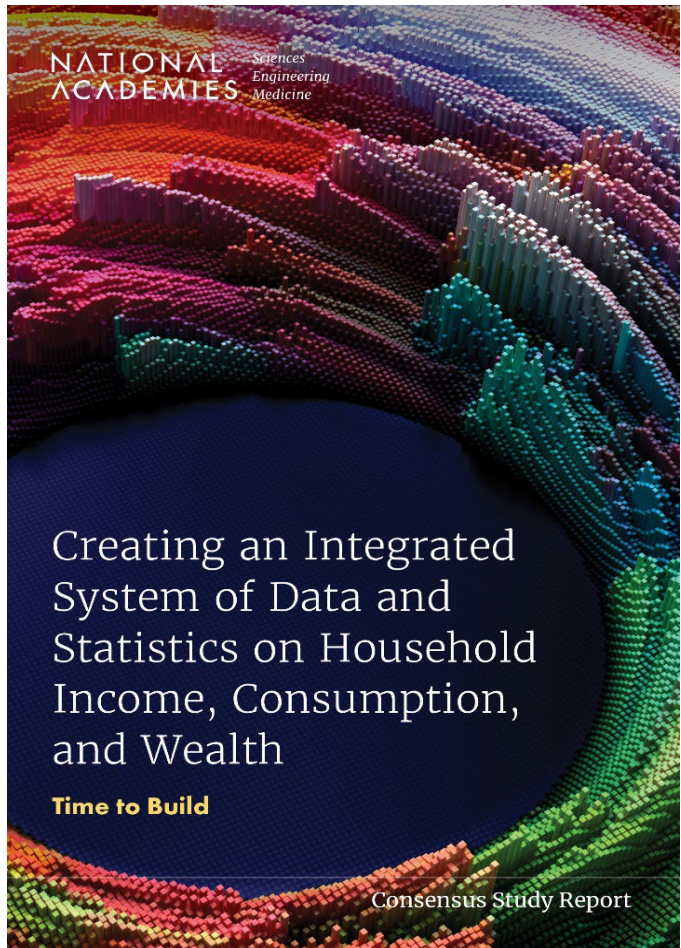# Creating an Integrated System of Data and Statistics on Household Income, Consumption, and Wealth

Time to Build

Tim Smeeding and David Johnson

FCSM October 23, 2024

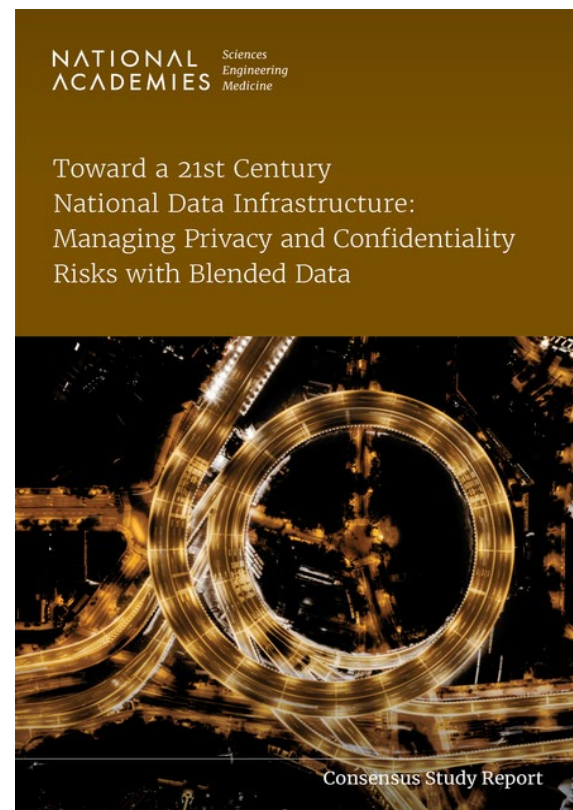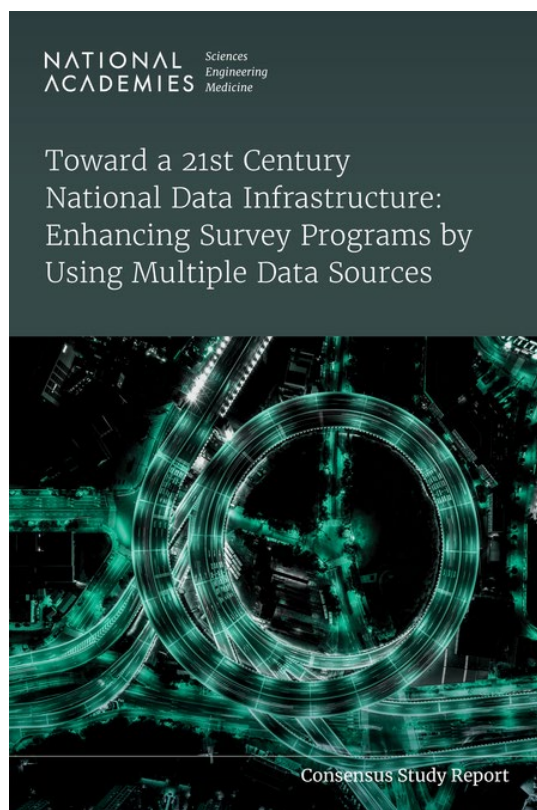**TIMOTHY M. SMEEDING** (*Chair*)**,** University of Wisconsin-Madison
KATHARINE G. ABRAHAM, University of Maryland
WILLIAM A. DARITY, JR., Duke University
KAREN DYNAN, Harvard University
JACOB W. FABER, New York University
ROBERT E. HALL, Hoover Institution and Stanford University
STEPHEN P. JENKINS, London School of Economics
DAMON JONES, University of Chicago
AMY B. O'HARA, Georgetown University
LUIGI PISTAFERRI, Stanford University
JOHN SABELHAUS, Brookings Institution
YAJUAN SI, University of Michigan
AMIR SUFI, University of Chicago
GABRIEL ZUCMAN, Paris School of Economics
JORRIT ZWIJNENBURG, OECD

DAVID JOHNSON, Study Director
CONSTANCE F. CITRO, Senior Scholar
CHRISTOPHER MACKIE, Senior Program Officer
ALEX HENDERSON, Senior Program Assistant

Builds upon recent 21ˢᵗ Century reports stressing the importance of blended data

NATIONAL ACADEMIES Sciences Engineering Medicine

Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good

Consensus Study Report

NATIONAL ACADEMIES Sciences Engineering Medicine

Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources

Consensus Study Report

NATIONAL ACADEMIES Sciences Engineering Medicine

Toward a 21st Century National Data Infrastructure: Managing Privacy and Confidentiality Risks with Blended Data

Consensus Study Report

NATIONAL ACADEMIES Sciences Engineering Medicine

21ˢᵗ Century report conclusions

Conclusion 2-1 from 1ˢᵗ 21ˢᵗ Century Report

- The United States needs a new 21st century data infrastructure that **blends data from multiple sources to improve the quality, timeliness, granularity, and usefulness of national statistics…**

Conclusion 9-3 from Multi-source report.

- **Use of multiple data sources is expected to play a major role in the future production of statistical information in the United States, …**

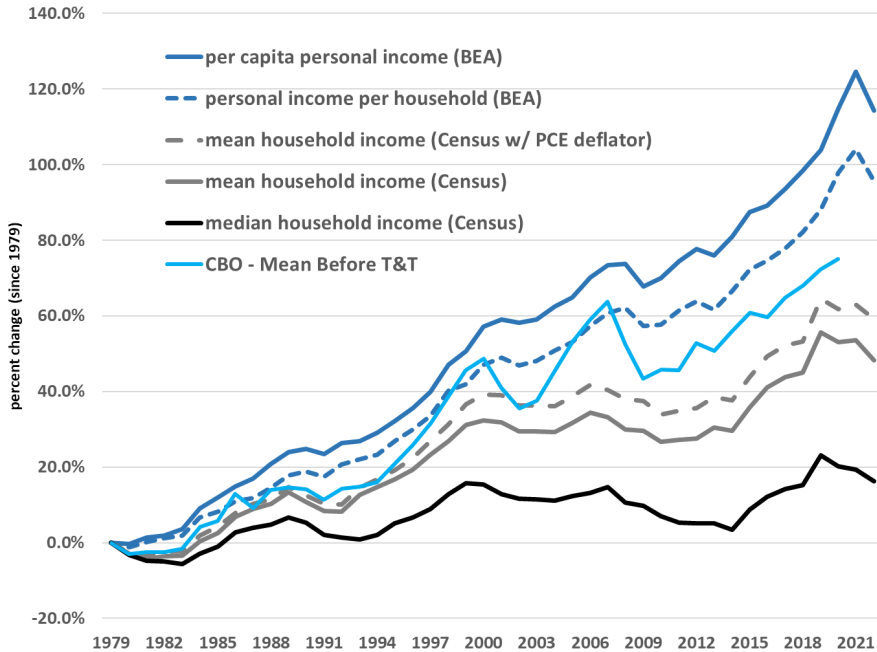**Conclusion 3-2 from Privacy report**

- **Tiered access for data users and agencies is a key component of a dynamic disclosure risk/usefulness framework**, to reflect differences in acceptable risks given policy priorities.

NATIONAL ACADEMIES *Sciences Engineering Medicine*

The Need: Integrated data and statistics on income (I), consumption (C) & wealth (W)

- For decades household and individual economic wellbeing has been measured annually for income (I) only—*and different measures of income give different estimates of means, medians, trends and levels of inequality and poverty*

- There is a need for I, C, W and changes in W (ΔW) to all be well measured from the same high-quality dataset or data construction in a transparent way

- There is a need for high-quality, consistent "official" annual statistics on I, C, W and their joint distributions, as well as access to the full range of data for research purposes

- But this is not possible with current data sources – we need to build better ones

Current estimates are confusing and inconsistent

How much has income increased?

Legend:
- per capita personal income (BEA)
- personal income per household (BEA)
- mean household income (Census w/ PCE deflator)
- mean household income (Census)
- median household income (Census)
- CBO - Mean Before T&T

y-axis: percent change (since 1979), from -20.0% to 140.0%
x-axis: 1979 to 2021

How much has income inequality increased?

Legend:
- Piketty, Saez & Zucman National Income
- Auten & Splinter Pre-tax National Income
- CBO Before Tax and Transfer Income
- Census Money Income
- BEA Personal Income
- LIS Disposable Household Income

y-axis: Gini Index, from 0.300 to 0.600
x-axis: 1979 to 2021

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Current estimates are confusing and inconsistent

How much has income increased?

Legend:
- per capita personal income (BEA)
- personal income per household (BEA)
- mean household income (Census w/ PCE deflator)
- mean household income (Census)
- median household income (Census)
- CBO - Mean Before T&T

Y-axis: percent change (since 1979), from -20.0% to 140.0%
X-axis: 1979 to 2021

with Wealth & Consumption – span the space

Legend:
- Piketty, Saez & Zucman National Income
- Auten & Splinter Pre-tax National Income
- CBO Before Tax and Transfer Income
- Census Money Income
- BEA Personal Income
- LIS Disposable Household Income
- Wealth (SCF)
- Consumption (BLS)

Y-axis: Gini Index, from 0.200 to 0.900
X-axis: 1979 to 2021

NATIONAL ACADEMIES *Sciences Engineering Medicine*
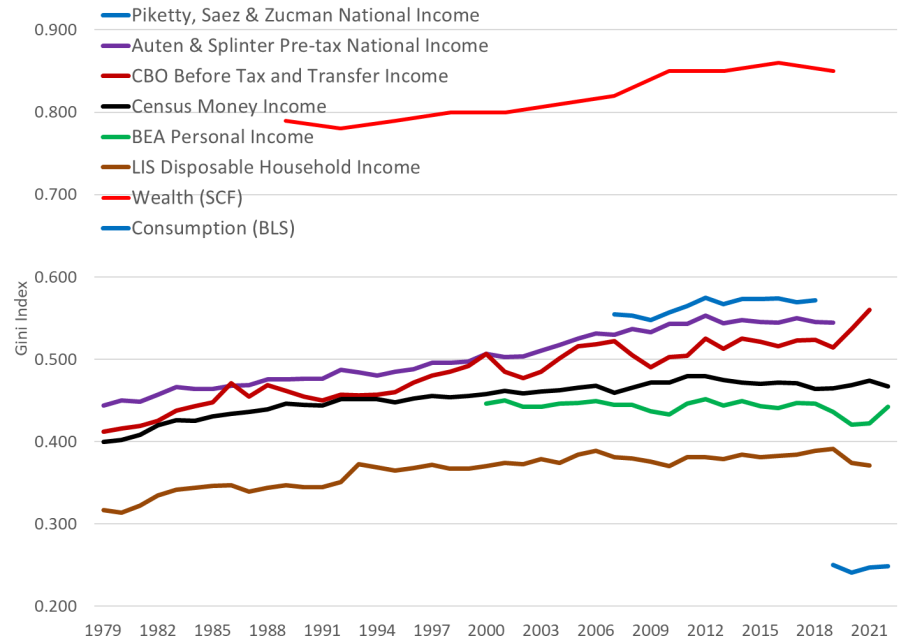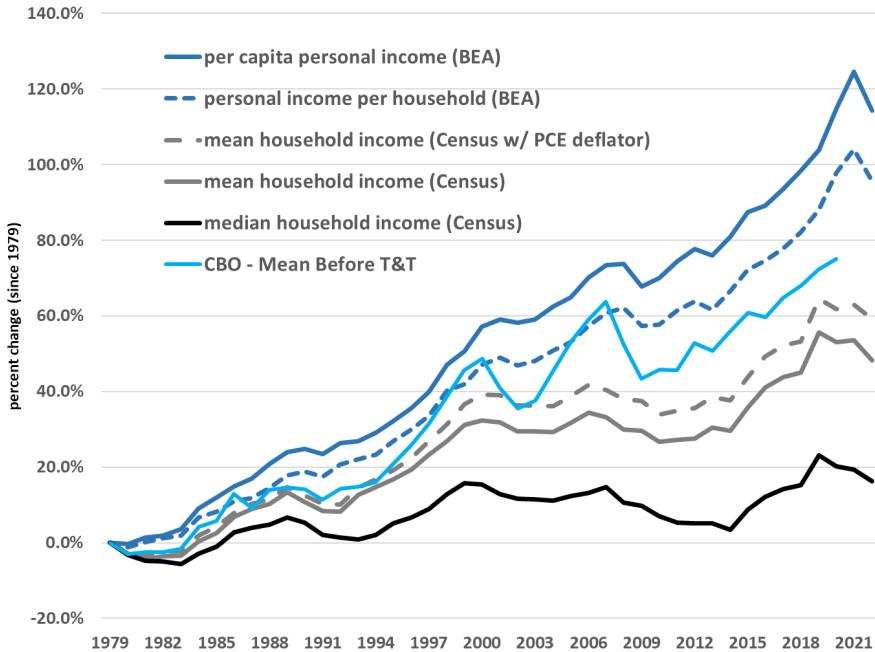
The Need: Fundamental research questions that could be informed by an integrated ICW system

- Trends and levels of **inequality, poverty, and economic mobility**

- Assessing **financial well-being**, **hardship**, and **exposure to risk**—at the business cycle frequency and over the longer-run

- **Implications of shifts** in demographics, household structure, and the economic environment for the dynamics of the economy

- **Assessing the role of accumulated wealth** is an important source of income and consumption insurance and an increasingly important source of intergenerational transfers

- Better understanding of consumer behavior—**propensities to consume, borrow, save, take on risk**

  - heterogeneity by income/wealth/access to financial opportunities/other characteristics

  - heterogeneity by type of resource shocks (permanent/transitory)

NATIONAL ACADEMIES *Sciences Engineering Medicine*

Need consistent ICW on same dataset to precisely estimate MPC by wealth
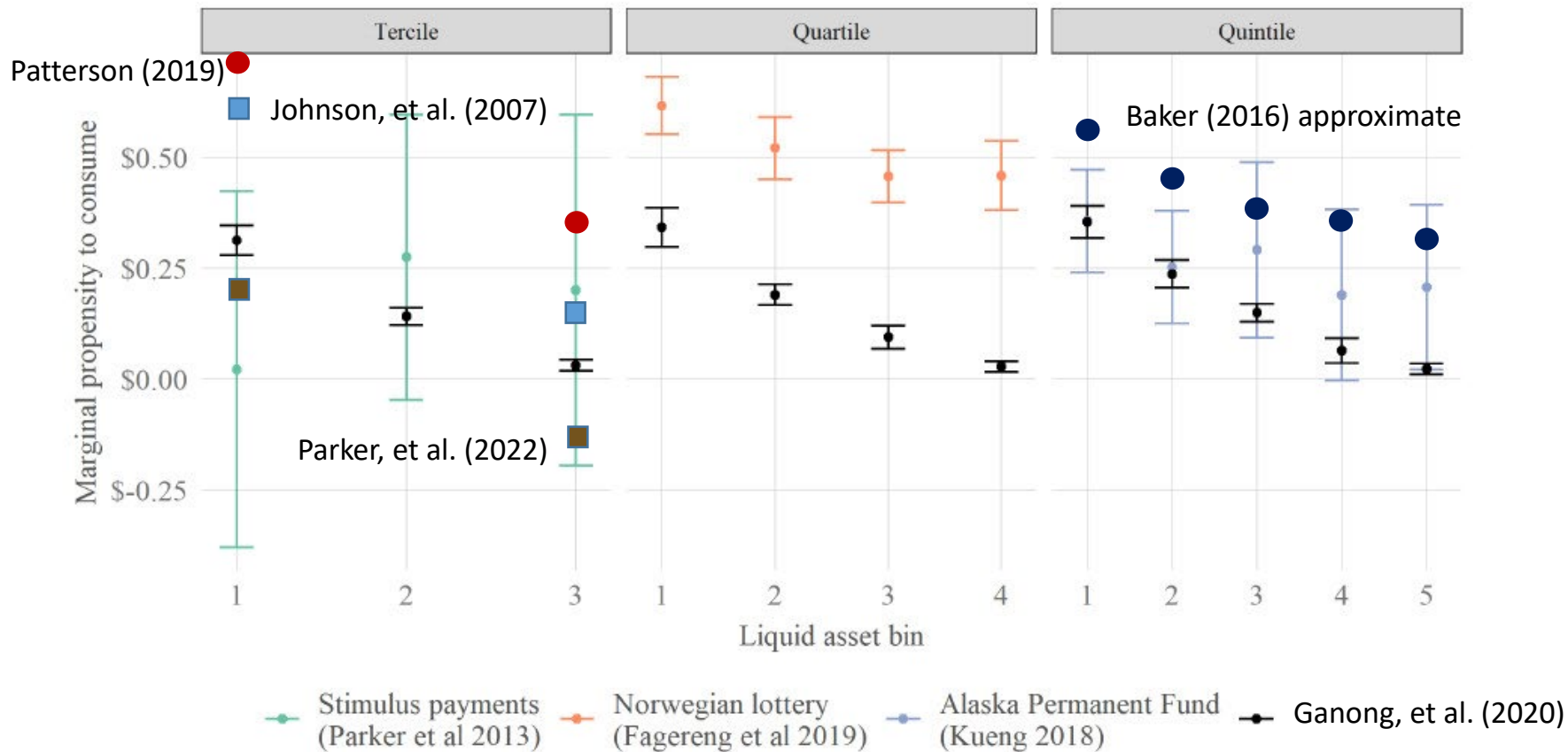
Key is for ICW measures to be consistent with budget identity and national accounts

- Conclusion 2-1: Conceptual definitions of household income, consumption, and wealth **are most useful when they are constructed to satisfy the budget identity, Consumption = Income – Saving**… A fundamental principle in specifying an integrated data system is that decisions about what to count as income have direct implications for what to count as consumption and wealth.

- Conclusion 2-3: **Multiple definitions** of household income, consumption, and wealth are needed **to examine different policy and research**...

- Conclusion 2-4: Definitions **embedded in systems of national accounts** provide a well-established starting point for the variety of ICW concepts envisioned for the dataset recommended in this report.

Research does not use consistent measures of Income, Consumption and Wealth

TABLE 2-1A Established Definitions of Income

| Source	Dispos-able personal income (BEA)	After tax & transfer national income (PSZ)	OECD (SNA)	Can-berra dispos-able HH income	Money income (Census Bureau)	After tax & trans-fer (CBO)	AGI	SCF family income (FRB)	CE after-tax income (BLS)
Employment income	+	+	+	+	+	+	+	+	+
Employer contribution to Social Security	a		a						
Employee contribution to Social Security	a		a		+		+	+	+
Employer-provided pensions	+	+	+				+ (some)		
Employer-provided health insurance	+	+	+			+			
Employer-provided other benefits	+	+	+						
Business income (including self-employment)	+	+	+	+	+	+	+	+	+
Investment income[b]	+	+	+	+	+	+	+	+	+

TABLE 2-1B Established Definitions of Consumption

	Personal consumption expenditures	CE expenditures (published)	BLS/CE consumption (proposed)	Fisher et al. consumption	Meyer/Sullivan consumption	OECD consumption expenditures
Durable goods						
Motor vehicles and parts						+
New motor vehicles	+	+				+
Net purchases of used motor vehicles	+	+				+
Motor vehicle parts and accessories	+	+	+	+	+	+

The Recommendations: Improved statistics and a report from federal statistical agencies

- Conclusion 3-1: The initiatives by [the federal agencies] are laudable. These initiatives deserve support to become established series as soon as practicable.

- Recommendation 3-1: *…***statistical agencies should build on their current initiatives** to publish improved statistics of household income, consumption, and wealth.
  - **BEA** should make its distributional household personal income (PI) estimates more useful by adding estimates for: (1) disposable (after-tax) PI that excludes nonprofit institutions serving households; and (2) disposable PI (excluding NPISH)…
  - The **Census Bureau's** household income estimates should include pre- and post-tax-and-transfer series…
  - The **BLS**…should expand and accelerate its program to publish annual estimates of household consumption...

- Recommendation 3-2: Relevant statistical agencies should **collaboratively publish a major report every three years** that compares levels and trends among the household income, consumption, expenditure, and wealth statistical series…

Possible new Report: Current estimates on ICW inequality use blended data

Top 10% shares

- Wealth (DFA)
- Income (PI)
- Consumption (PCE)
- PCE for top DPI decile
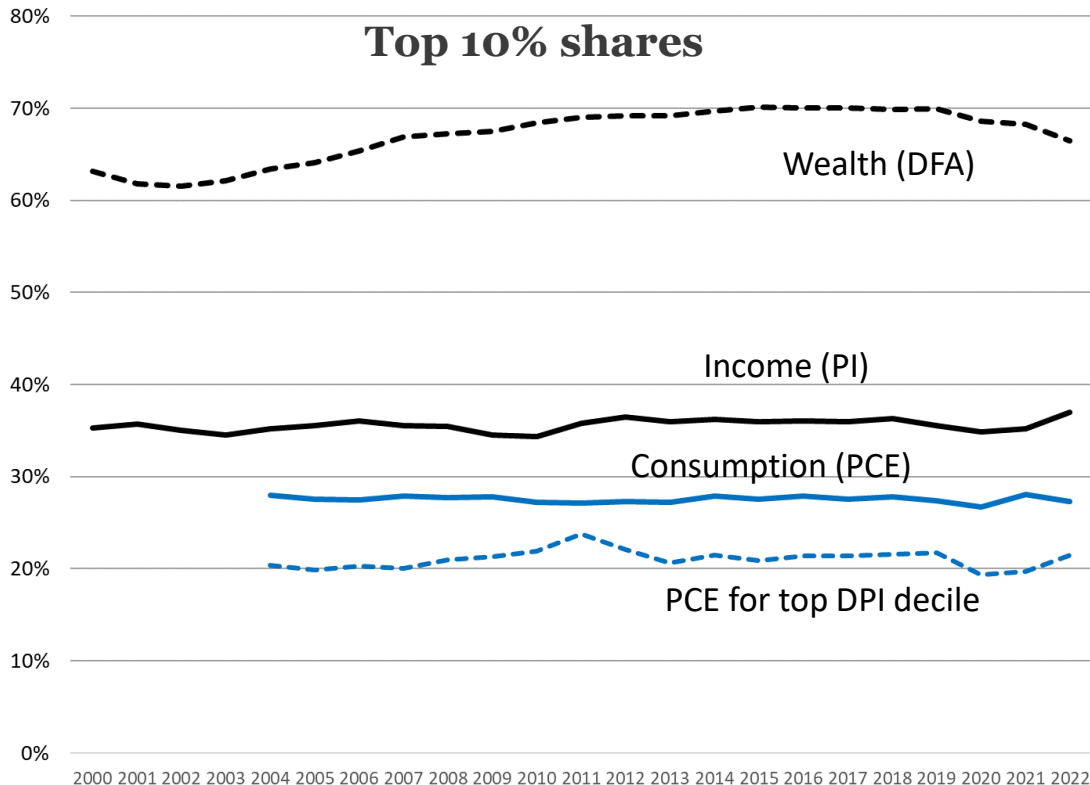
- Federal Reserve Board estimates quarterly distributional financial accounts (DFA) for wealth (SCF uses tax data)

- BEA estimates annual distributions of personal income (PI) and disposable personal income (DPI) (uses SOI tabulations)

- BLS estimates annual distributions of Personal Consumption Expenditures (PCE)

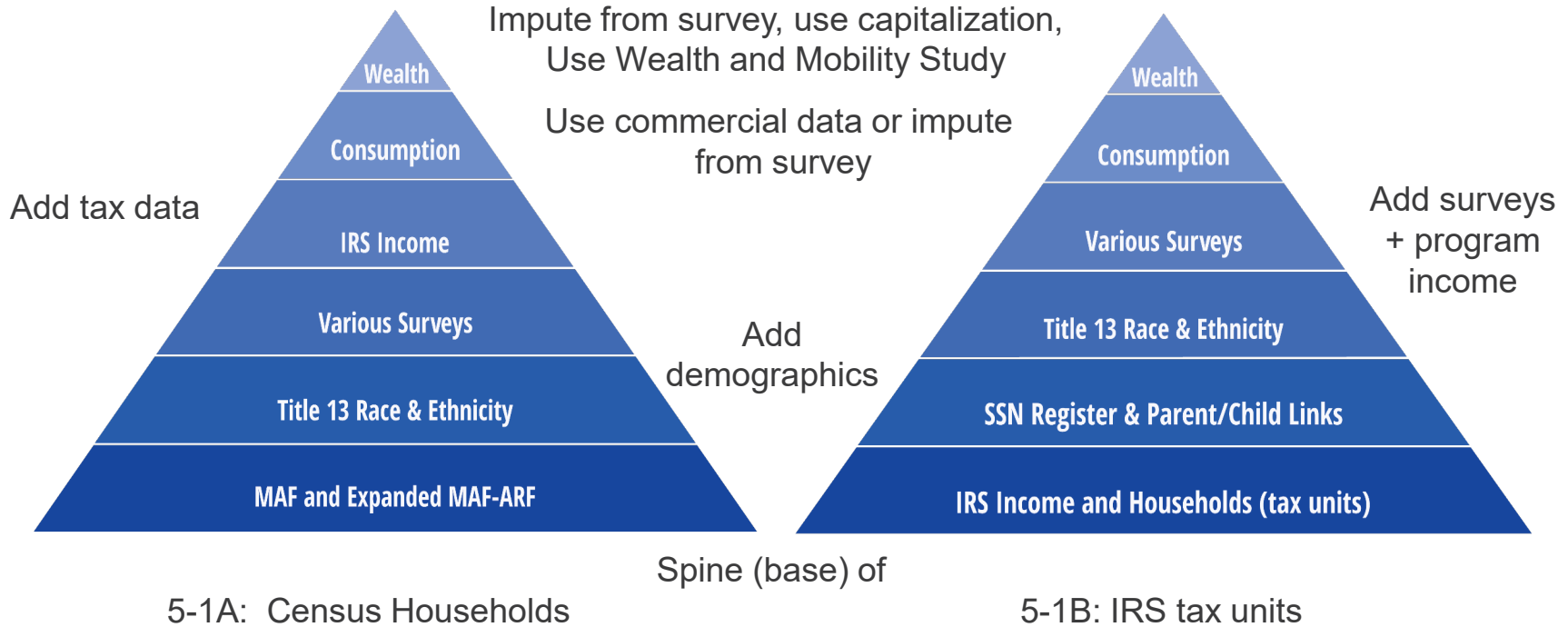- Joint distribution of DPI and PCE (from BEA and BLS)

Need to consider a wide variety of blended (linked) data sources operating under the oversight of a coordinating entity

- Recommendation 5-1: The Chief Statistician and the National Secure Data Service should work together to create a **coordinating entity to solve administrative, legal, and technical challenges to integrate data from multiple federal entities**…as well as private businesses that collect microdata on individual and household income, consumption, and wealth.

- Recommendation 5-2: …the coordinating entity should **expand on current efforts and coordinate new pilot studies to blend multiple datasets** that have the key components of income, consumption, and wealth;

- Recommendation 5-4: …the coordinating entity…should explore the **fusion of relevant publicly accessible datasets**

NATIONAL ACADEMIES *Sciences Engineering Medicine*

The Ideal Dataset

- The ideal data infrastructure for statistics and research on household ICW is a micro dataset with records for households with some data on the individuals living in the household. Such a dataset would allow users to precisely estimate the joint distribution of ICW for individuals aggregated to households and other units of analysis with as much geographic, race/ethnicity, and demographic detail as possible at various breakdown points in the distribution, including the very top (e.g., top 1% and top 0.1%). The ideal dataset is easily accessible, regularly updated, and comparable both over time and with other countries.

- The ideal dataset captures the desired units of observation, coverage, characteristics, consistency, frequency, timeliness, updateability, and flexibility. However, such a dataset does not exist in the United States.

Blended data begins with a spine

Impute from survey, use capitalization, Use Wealth and Mobility Study

Use commercial data or impute from survey

Add tax data

Add surveys + program income

Add demographics

Spine (base) of

5-1A: Census Households

5-1B: IRS tax units

5-1A pyramid (top to bottom):
- Wealth
- Consumption
- IRS Income
- Various Surveys
- Title 13 Race & Ethnicity
- MAF and Expanded MAF-ARF

5-1B pyramid (top to bottom):
- Wealth
- Consumption
- Various Surveys
- Title 13 Race & Ethnicity
- SSN Register & Parent/Child Links
- IRS Income and Households (tax units)

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Census National Experimental Wellbeing Statistics blends multiple data sources to measure income

Figure 4: Linkage Diagram for Address File
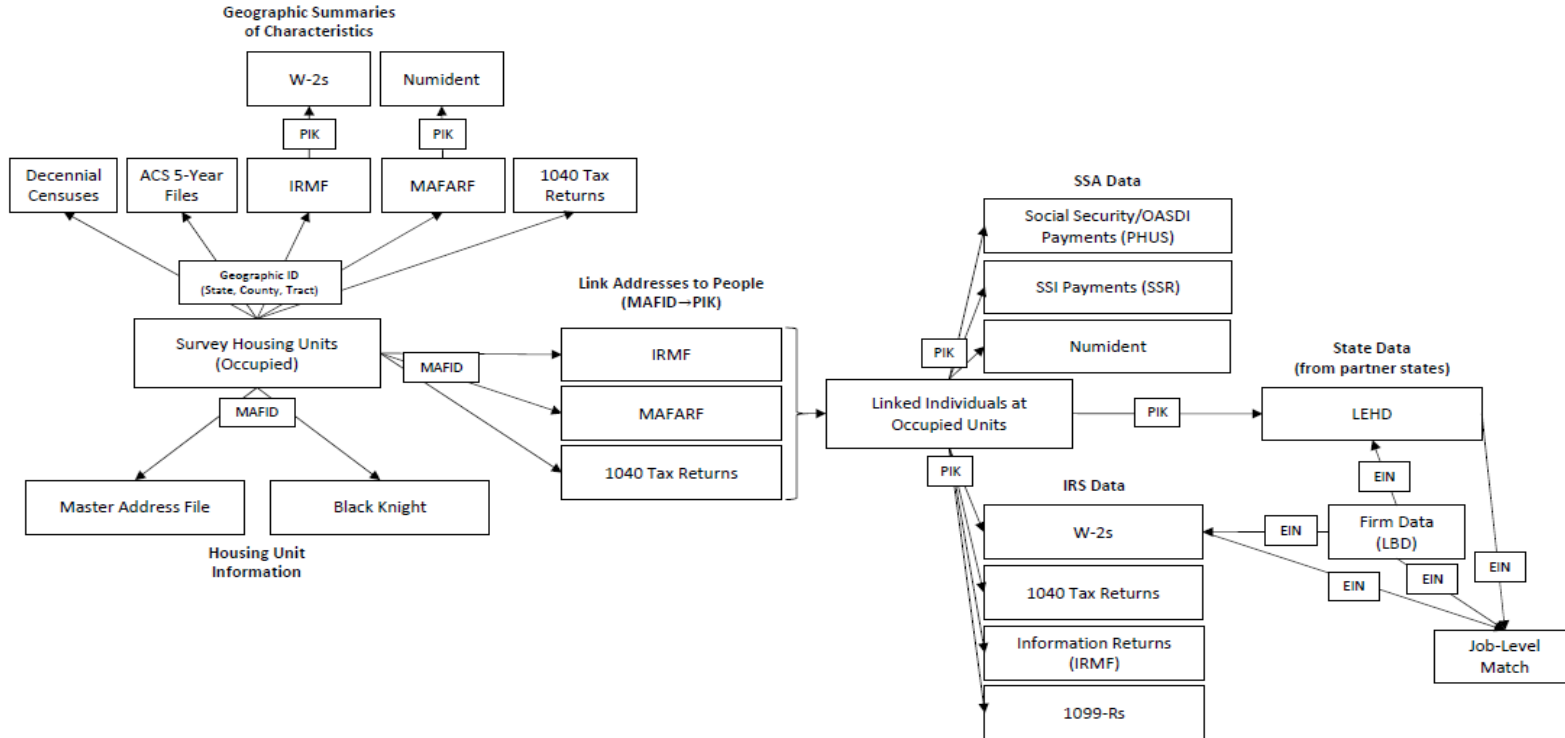
Source: Bee, A., Mitchell, J., Mittag, N., Rothbaum, J., Sanders, C., Schmidt, L., & Unrath, M. (2023). *National experimental wellbeing statistics* (SEHSD Working Paper No. 2023-02). U.S. Census Bureau

Need to ensure privacy while maintaining accessibility and usefulness

- Recommendation 6-1: …the coordinating entity, in collaboration with relevant statistical agencies, **should jointly develop a risk-utility framework** and a combination of traditional disclosure limitation strategies and **privacy enhancing technologies**…

- Recommendation 6-2: The coordinating entity…should **propose options for a National Secure Data Service pilot** to be created through which **approved researchers would be allowed to access** linked microdata…

The Roadmap: How to make progress and a steering committee to lead change

Conclusion 6-2: The success in using multi-source data and data integration in other countries that operate with a single centralized statistical agency can provide guidance to improve the collaboration within and integration of the current U.S. statistical system.

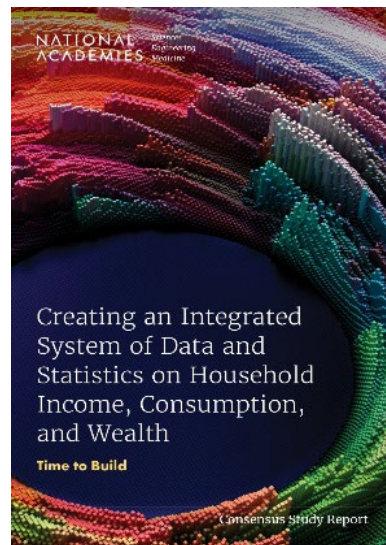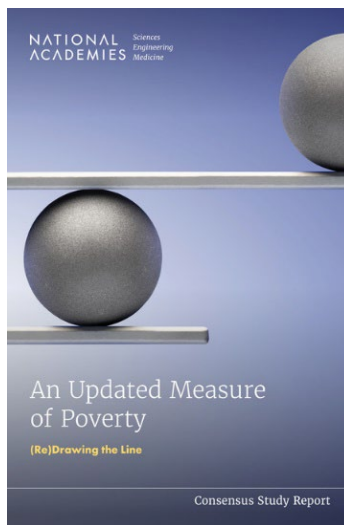Recommendation 6-3: The OMB Chief Statistician, in collaboration with the Interagency Council on Statistical Policy, should **establish an ICW technical steering committee…**
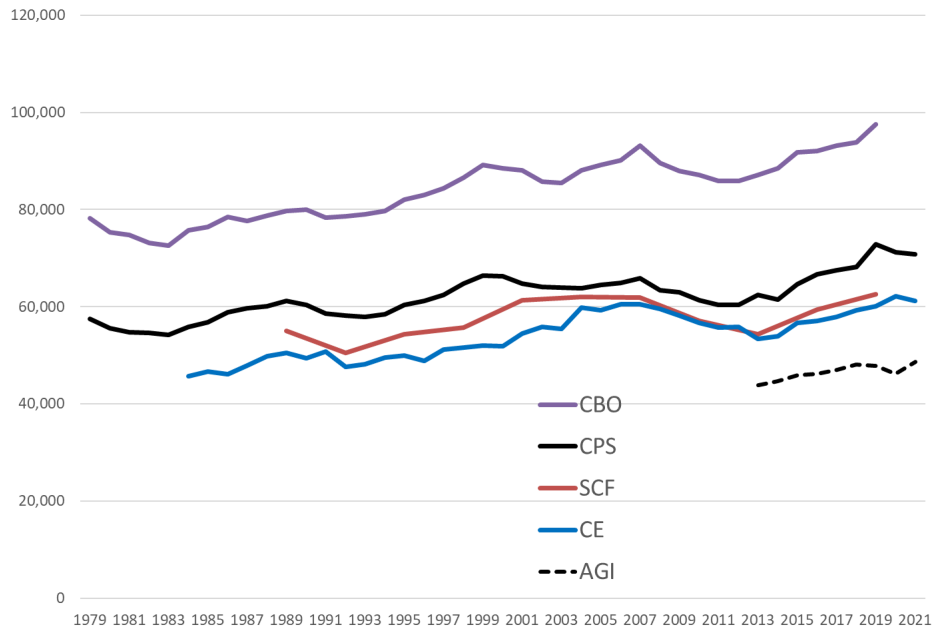

ICW Data and Products

Pre-Launch
- Create a coordinating entity to oversee the data acquisition, linkage, & access (Rec. 5.1)
- Establish a steering committee (Rec. 6.3) to guide coordinating entity
- Support current agency efforts to improve income measurement, linkages & privacy methods

Phase 1
1−3 years
- Identify & evaluate the administrative, legal, & technical challenges to integrating ICW from multiple federal entities & private businesses (Rec. 5.1, 5.2, 5.4)
- Expand on current efforts to blend data sets that have key components of ICW
- Evaluate modeling & statistical matching methods to improve ICW estimates & joint distributions (Rec 5.4)
- Resolve definitional concepts of ICW and produce consistent estimates of ICW distributions

Phase 2
2−5 years
- Begin publishing regular estimates & producing regular reports (Rec 3.2, 5.2)
- Blend survey & administrative data, using both IRS & Census data
- Expand NEWS, CID, and IRS privacy projects (Rec. 5.1, 5.2, 5.3)
- Impute wealth via capitalization methods & impute consumption on a blended dataset
- Initiate quality measures (Rec. 4.2, 4.3)
- Develop an NSDS pilot project to make data accessible for researchers (Rec. 6.2)
- Investigate a new survey or adding questions to a current commercial database (Rec. 5.5)

Phase 3
4−6 years
- Improve the data & methods that lie behind the regular report
- Establish secure methods for researchers to access the linked microdata
- Expand sources in the linked data (e.g., commercial datasets)
- Incorporate current projects, link NEWS & CID, into the ICW data infrastructure
- Improve ICW statistics (Rec. 3.1, 3.2) & create pilot research projects for data access & improving estimates
- Update models, improve geographic data (Rec 3.4), produce more timely data (Rec. 5.6), & include additional demographic groups
- Create a panel of ICW data (Rec. 5.2), linking income data across generations & integrating wealth & consumption

COMMITTEE ON NATIONAL STATISTICS REPORTS ON NEW DATA INFRASTRUCTURE

20

- EXTRA SLIDES

Possible new report: The joint report with consistent statistics

Comparing Medians across surveys/agencies

- Conclusion 2-3: Multiple definitions of household income, consumption, and wealth are needed…**each definition requires specification of accompanying purpose**(s) and a transparent guide to its construction.

- Recommendation 3-2: …annual reports should compare their estimates to those from other agencies and, to the extent possible, **identify the factors, including differences in definitions, that contribute to differences**.