# Modernizing the U.S. Census Bureau's Demographic Programs

**Instructions:** Click on the link to access each author's presentation.

**Organiser:** Jennifer Ortman

**Chair:** Victoria Velkoff

**Participants:**

**Victoria Udalova:** Enhancing Health Data at the U.S. Census Bureau

**Jennifer Ortman:** Demographic Frame: Leveraging Linked, Person-Level Data

**Jonathan Rothbaum:** National Experimental Wellbeing Statistics (NEWS)

**Jason Fields:** Census Bureau Household Panel Survey

# Enhancing Health Data at the U.S. Census Bureau:

## Leveraging Administrative Records and Record Linkages to Develop a Physician-Business Linked Database

**Victoria Udalova**, PhD
Program Director, Enhancing Health Data

May 15-17, 2024
IAOS-ISI 2024 Conference

Any opinions and conclusions expressed herein are those of the presenter and do not reflect the views of the U.S. Census Bureau. No confidential data are used in this presentation.

United States® Census Bureau

# Why Enhance Health Data at Census?

- Direct data collection via surveys of patients and health care providers continues to experience declining response rates and increased costs
- Greater availability of health and other administrative records
- Greater understanding of the importance of non-health care factors (social determinants of health)
  - but this information is limited in health data
- Complex and rapidly evolving structure of the U.S. healthcare system

United States® Census Bureau

# Paradigm Shift

- Enhance health data using record linkages and existing survey and administrative records uniquely available at the Census Bureau
- Link multiple data sources to create new statistical products and research
- Conduct original research and collaborate with external partners for health records data and expertise
- Lower cost, more granularity, more timely, higher quality than direct data collection
- A way to complement direct data collection

United States® Census Bureau

# Health Frame - Physicians

- No single authoritative data source on the supply of health care
- Rapid reshaping of the U.S. health care system's structure
  - Consolidation of physician practices
  - Private equity investments
- Leverage Census unique data assets on workers, taxpayers, and employers to measure characteristics
  - physicians (race/ethnicity, employment status, earnings)
  - institutions where physicians work (geography, business structure, revenue, ownership)

United States® Census Bureau

# Connecting the Dots for Complex Organizations

## Simple Example

general practitioner with one office


## Complex Example

hospitalist in a 15-hospital health system

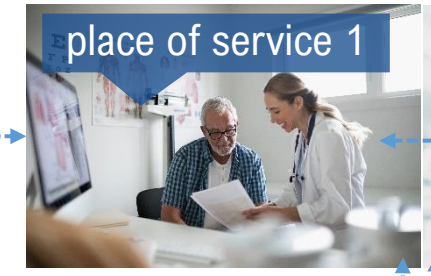

NASHVILLE

physician's regional hospital group recently merged with a national, private-equity backed firm

physician receives her salary from a physician staffing firm which provides MDs under a contract with her hospitals


HOUSTON

Splits time across multiple hospitals in a regional health system and also interacts with patients via telemedicine.




DEARBORN


DETROIT


billing largely handled within the admin offices of an affiliated hospital

TROY


patient data managed locally but scheduled to transition to new owner's central EHR system

JACKSONVILLE

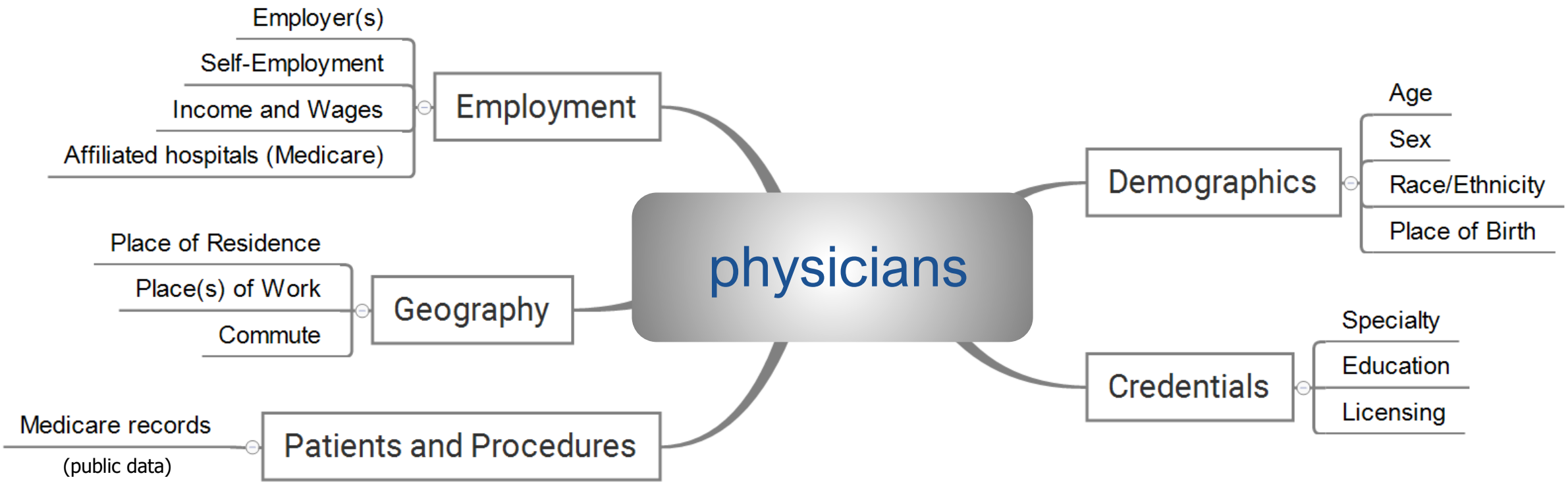
United States® Census Bureau

# Data Sources: Overview

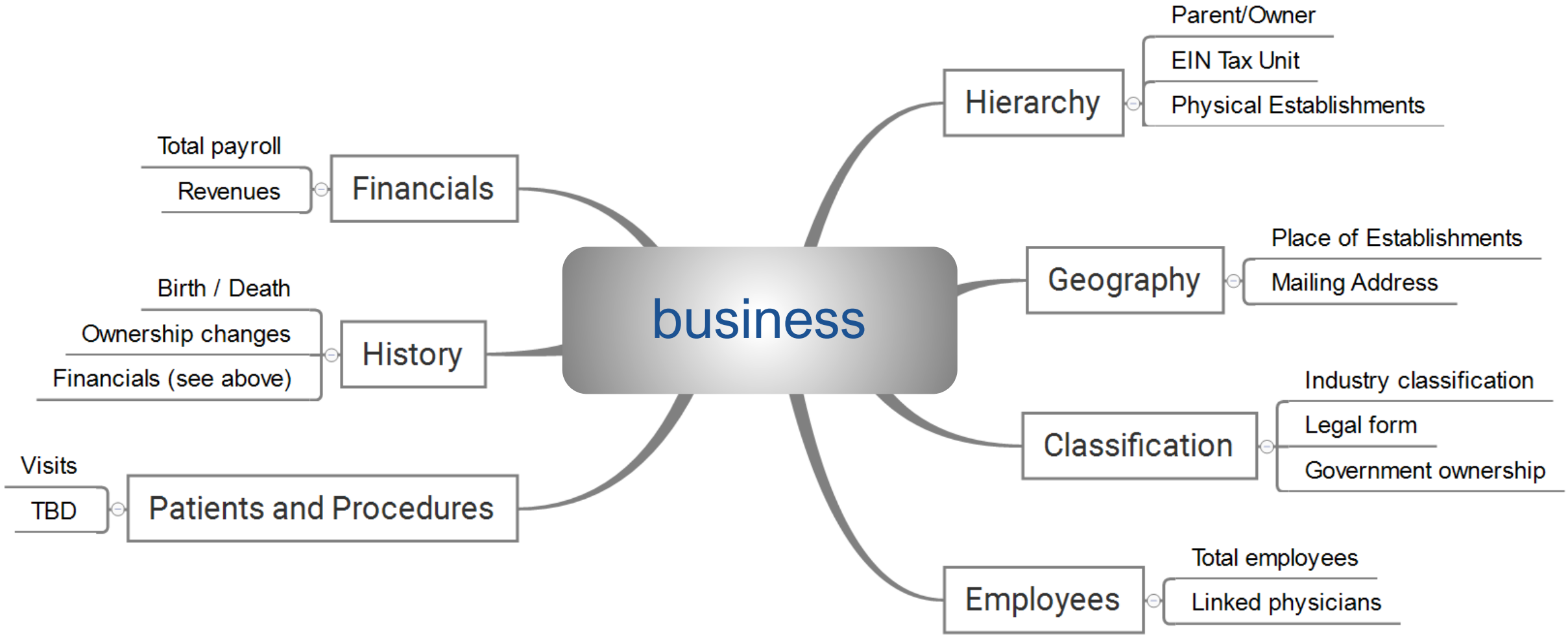
- Integrating 34+ datasets across 8+ years
- Physician-level data
  - American Community Survey (race/ethnicity, hours of work), Decennial Census (race/ethnicity)
  - Numident (date of birth/death, place of birth)
  - Master Address File (residence location)
  - IRS Forms W2/1040/1099/K1 (ownership, self-employment, income, retirement)
  - CMS MPUPD (Medicare charges, service provision), CMS Care Compare (places of work, credentials, affiliations), CMS NPPES/NPI Registry (specialty, places of work)
- Business-level data
  - Business Register (payroll, employment, business structure-establishments, location, industry)
  - Longitudinal Business Database (merger/acquisitions, entry/exit, revenue)
  - Service Annual Survey (care revenue, visits, expenses)
  - AHRQ Compendium of U.S. Health Systems (hospitals, health systems)

United States® Census Bureau

# Health Frame - Physicians



# Health Frame - Physicians



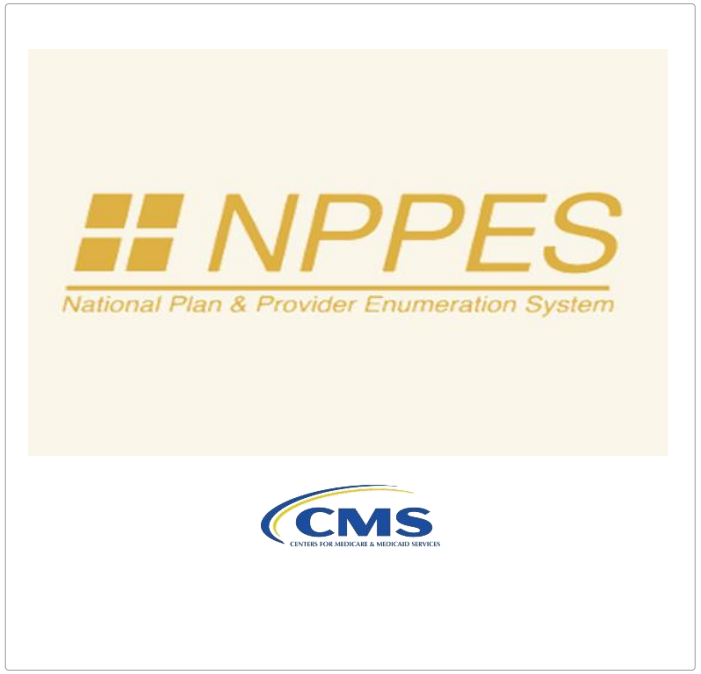
# Record Linkage at the Census Bureau

- Personally Identifiable Information (PII) is used to assign internal anonymized Census Bureau linkage keys, Protected Identification Keys (PIKs)
  - PII is matched to the internal reference file, a conglomeration of multiple survey and administrative records data
- Internal and approved external researchers do not have access to identified data
- Restricted microdata must remain within the Bureau's secure firewall
- Microdata can only be used for statistical purposes
- All aggregate statistical output is reviewed before it is released publicly

United States® Census Bureau

# Data Sources: Identifying Physicians

- NPPES/NPI Registry
  - National Provider Identifier (NPI) for each health care practitioner—mandated by HIPAA and launched in 2005
  - Over 5.7M active records for practitioners ranging from physicians (1.1M) to nurse practitioners (360K) to athletic trainers; 1.6M registered organizations
  - Restricted version to assign internal linkage keys – PIKs




United States® Census Bureau

# Data Sources: Identifying Physicians' Employers

- IRS Form W2
  - Employer Identification Number (EIN), wages
- IRS Form K1
  - Ownership data


IRS
Department of the Treasury
Internal Revenue Service

United States® Census Bureau

# Data Sources: Attaching Business Characteristics

- Business Register (BR)
  - Comprehensive database of business establishments
  - Updated continuously since 1972 using administrative records from IRS, SSA, BLS, state unemployment insurance records
  - Covers 5M single-establishment firms, 160K multi-establishment firms with 1.8M establishments, 21M non-employer businesses
  - Provides info on geography, industry, organization structure, payroll, employment


United States® Census Bureau

# Use Case: Earnings and Labor Supply of U.S. Physicians

- *"Who Values Human Capitalists' Human Capital? The Earnings and Labor Supply of U.S. Physicians"* (Joshua D. Gottlieb, Maria Polyakova, Kevin Rinz, Hugh Shiplett, Victoria Udalova)

  https://www.nber.org/papers/w31469

- Examined income and labor supply of U.S. physicians using rich administrative tax data linkages

**Who Values Human Capitalists' Human Capital? The Earnings and Labor Supply of U.S. Physicians**

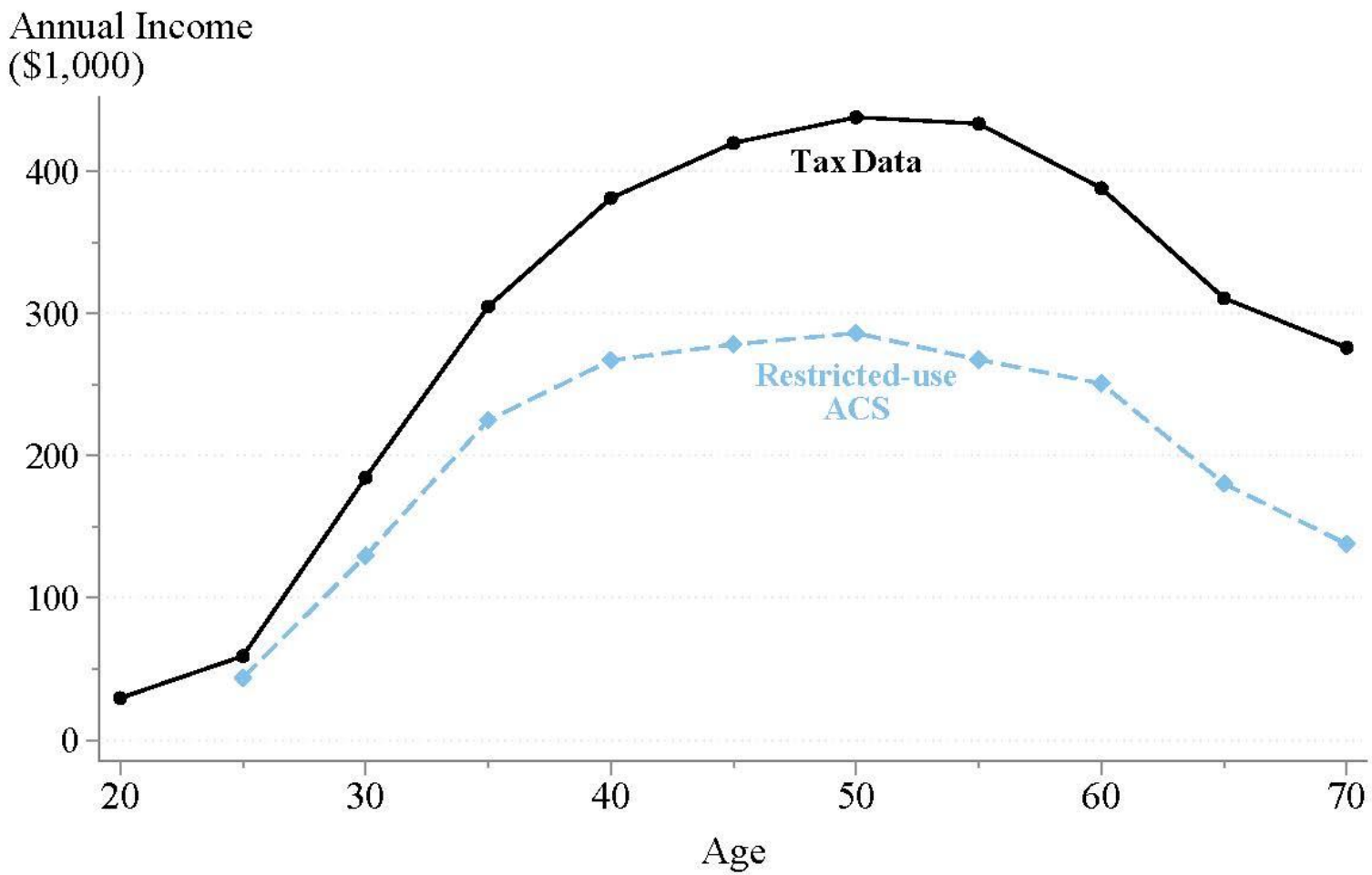
Joshua D. Gottlieb, Maria Polyakova, Kevin Rinz, Hugh Shiplett & Victoria Udalova

WORKING PAPER 31469 DOI 10.3386/w31469 ISSUE DATE July 2023

Is government guiding the invisible hand at the top of the labor market? We use new administrative data to measure physicians' earnings and estimate the influence of healthcare policies on these earnings, physicians' labor supply, and allocation of talent. Combining the administrative registry of U.S.–physicians with tax data, Medicare billing records, and survey responses, we find that physicians' annual earnings average $350,000 and comprise 8.6% of national healthcare spending. The age-earnings profile is steep; business income comprises one-quarter of earnings and is systematically underreported in survey data. There are major differences in earnings across specialties, regions, and firm sizes, with an unusual geographic pattern compared with other workers. We show that health policy has a major impact on the margin: 25% of physician fee revenue driven by Medicare reimbursements accrues to physicians personally. Physicians earn 6% of public money spent on insurance expansions. We find that these policies in turn affect the type and quantity of medical care physicians supply in the short run; retirement timing in the medium run; and earnings affect specialty choice in the long run.

United States® Census Bureau

# Underreported Physician Earnings in ACS




United States® Census Bureau

# Potential Use Cases of Health Frame - Physicians

- Prepare profiles of health care facilities (staffing, size, age, location)
  - Federally Qualified Health Centers (FQHCs)
  - Emergency departments
- Identify other health care providers (ex.: nurse practitioners)
- Allows to conduct research on the impact of major shifts in health care (ex.: vertical integration's impact on cost, access, and health outcomes)
- Create estimates for comparisons across countries (ex.: OECD data request on provider earnings)
- May aid policy implementation (ex.: COVID Provider Relief Fund distribution)

United States® Census Bureau

# Growing list of countries relying on blended data to report healthcare workforce statistics

- Approaches vary across countries
  - Healthcare system organization, administration, and funding vary
  - Data collection and processing vary
  - Data governance policies vary
- But increasingly countries are relying on blended data
  - Linking a registry of providers to administrative tax records
  - Linking survey data and administrative records
  - Linking administrative records across various government agencies
- Some examples
  - UK, Israel, the Netherlands, France, Ireland, Austria, Slovenia

United States® Census Bureau

# Conclusion

- Leveraging 30+ datasets, we are developing a physician-business linked database which is flexible and can be used for many different uses
  - Version 1.0 of physician frame is near completion, methodology paper and aggregate statistics forthcoming
- Navigating challenges (restricted data, disclosure avoidance, data use agreements)
- Strategically reusing existing administrative records and linking multiple data sources can be a great way to complement direct data collection

United States® Census Bureau

# Contact Information

census.gov/ehealth

Victoria Udalova, PhD

EHealth Program Director

U.S. Census Bureau

victoria.m.udalova@census.gov

United States® Census Bureau

# Demographic Frame: An Overview

Jennifer M. Ortman

May 15-17, 2024

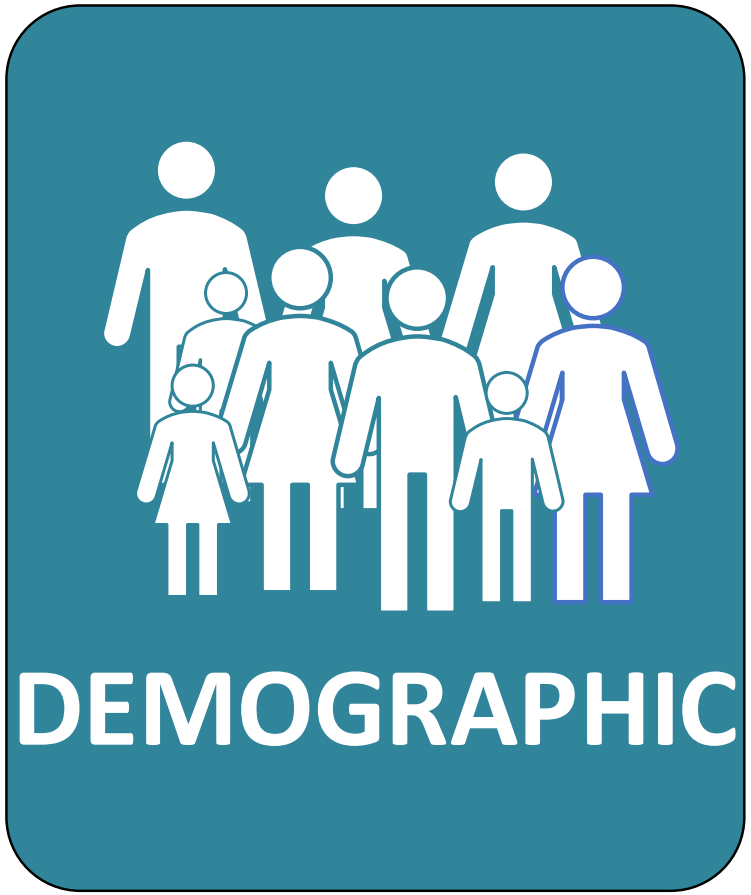
Presentation for the IAOS-ISI 2024 Conference



United States® Census Bureau

## The Demographic Frame is a comprehensive database of person-level data that:

- Contains demographic characteristics and addresses associated with each person
- Is derived from administrative, third-party, census, and survey data sources
- Includes unique person-level identifiers used to link individuals across datasets
- Can be linked to information in other enterprise frames
- Is available only to approved, internal users in a secured computing environment



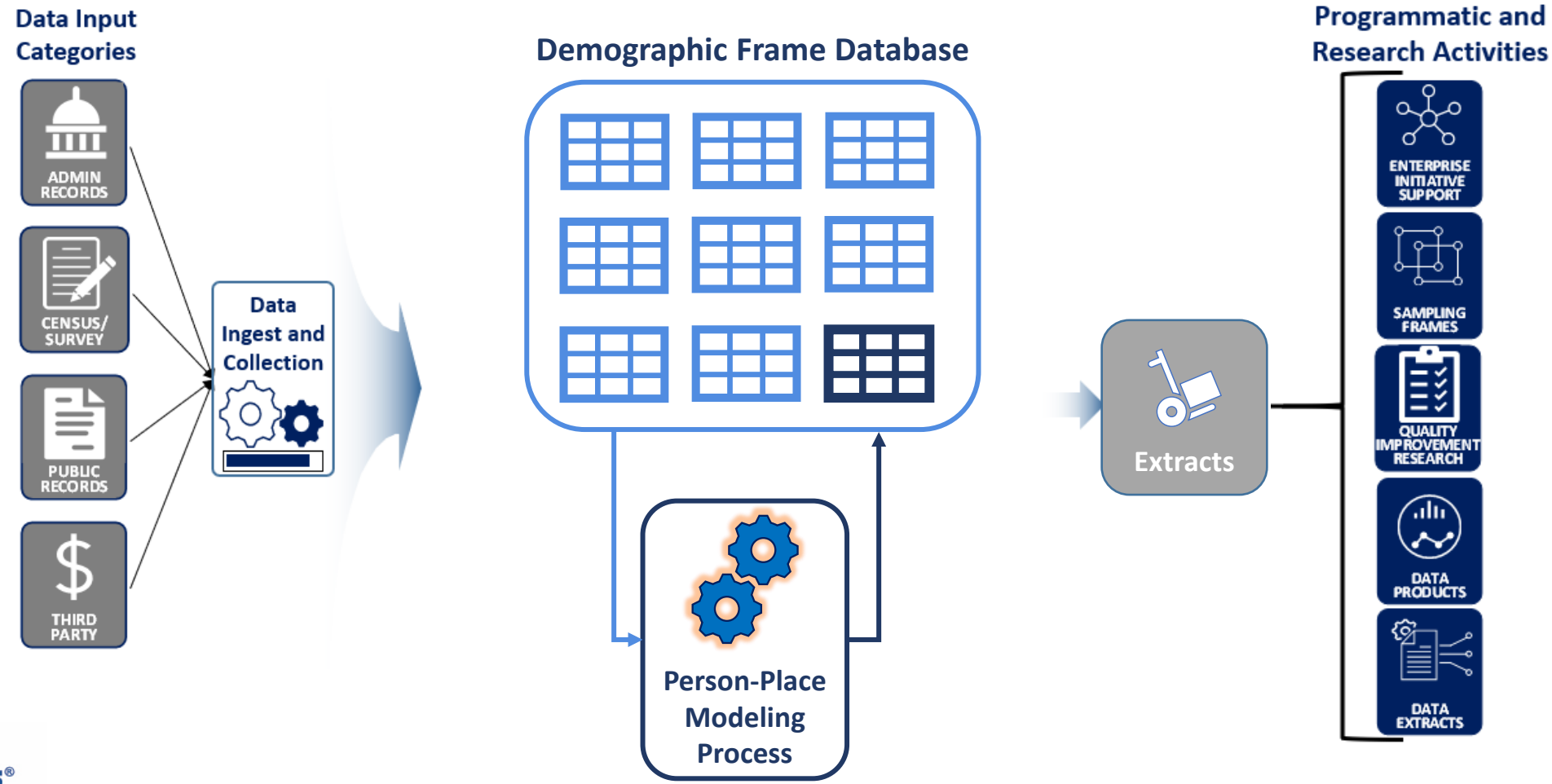

United States® Census Bureau

## Capabilities under development include:

- Working with others across the enterprise to **improve record linkage**
  - Review and enhance the Person Identification Validation System (PVS) process to reduce the number of individuals without a Protected Identification Key (PIK)
  - Evaluate quality and determine how best to handle remaining individuals without a PIK
- Aggregating to **household and family units**
- Adding more years of data to facilitate **linking information about individuals across time and space**
- **Establishing a production team** to maintain the Demographic Frame for enterprise use

United States® Census Bureau

# Building the Demographic Frame




United States® Census Bureau

# Demographic Frame Extracts

*Dates available: 2017, 2018, 2019, 2020, 2021, 2022*

| | | | | |
|---|---|---|---|---|
| Protected Identification Key (PIK) | Master Address File Identifier (MAFID) | Modeled Probabilities | First and Last Date Observed in a Source | Date of Birth |
| Age | Date of Death | Sex | Hispanic Origin | Race |
| Block | Block Group | Tract | County | State |

United States® Census Bureau

# Applications of the Demographic Frame

- **Serve as a sampling frame for population surveys** where a focus on ensuring a representative sample around the characteristics of people is a better foundation than an address-based approach
- **Support enterprise initiatives such as census and survey taking**, including reducing the burden on our respondents by using information already available to the federal government instead of asking questions.
- **Creating blended data products**, including merging Frame data on topics not currently asked about on a survey or the census, to provide even more detailed information about each of America's communities
- **Improving data quality by drawing upon the Frame for data editing and imputation**, rather than just statistical approaches to assigning values


United States® Census Bureau

# Current and Prospective Demographic Frame Customers

- 2030 Census Program
- Continuous Count Study (CCS)
- Self-Response Quality Assurance (SRQA) for Special Census
- Census Household Panel Survey
- American Community Survey (ACS)
- Community Resilience Estimates (CRE)
- Survey of Income and Program Participation (SIPP)
- Non-Employer Statistics – Demographics (NES-D)
- Longitudinal Employer-Household Dynamics (LEHD)

United States® Census Bureau

# Ongoing Evaluations

## Person- and Address-Level Comparisons

- Longitudinal Employer-Household Dynamics (LEHD) Residence Candidate File (RCF)
- 2020 Census
- Non-Employer Statistics by Demographics (NES-D)

## Benchmarking (Aggregate Counts)

- 2020 Census
- Population Estimates Program
- Administrative Records Population Estimates
- American Community Survey (ACS)

## Person-Place Model

- Evaluating 4 models
- Establishing model metrics for ongoing evaluation

United States® Census Bureau

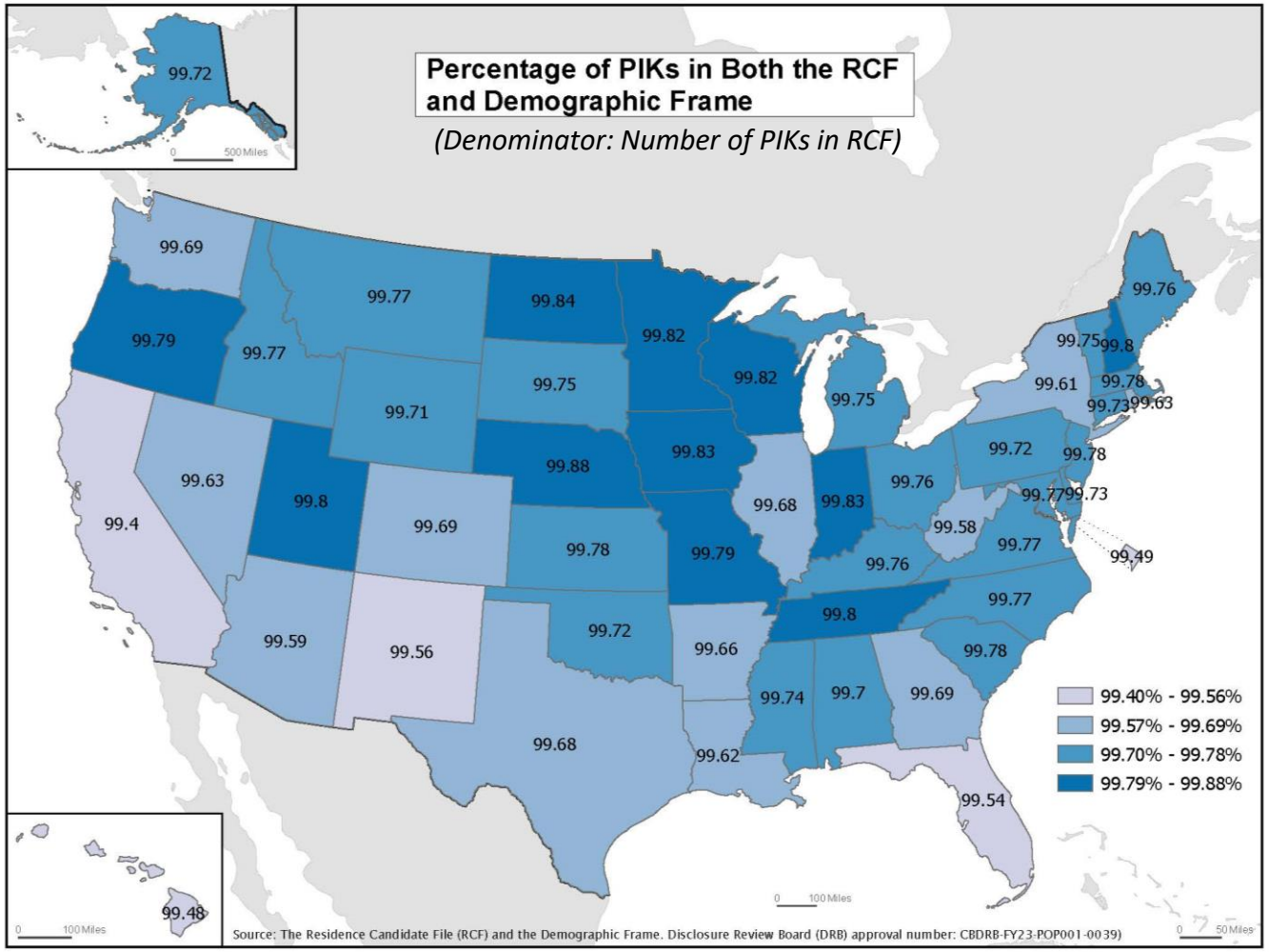
# Highlights from Our Evaluations

- Does the Demographic Frame **find the same people and put them at the same location** as the LEHD's Residence Candidate File?
- Does the Demographic Frame **provide similar counts** compared to the 2020 Census or Population Estimates?
- Do **household counts** match between the Demographic Frame and the ACS?

United States® Census Bureau

*Does the Demographic Frame **find the same people** as the Residence Candidate File?*

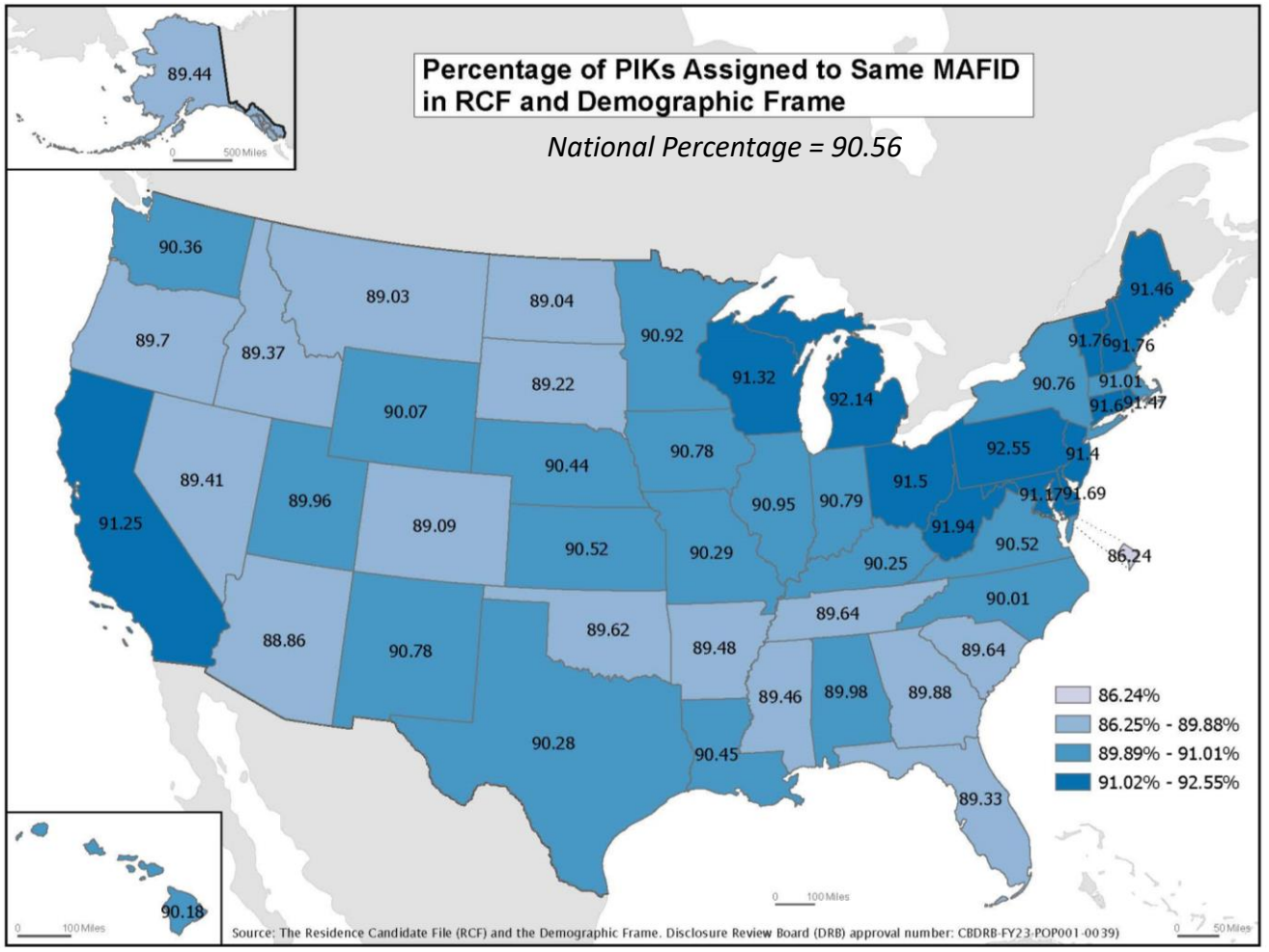
Over **99 percent** of PIKs observed in the RCF were also found in the Demographic Frame at the national level and when looking by state.



*Does the Demographic Frame* ***put people at the same address*** *as the Residence Candidate File?*

**90.56 percent** of PIKs were assigned to the same MAFID in the RCF and Demographic Frame.

Agreement ranges from **86 to 93 percent** by state.



# *Does the Demographic Frame provide similar counts compared to the 2020 Census or Population Estimates?*

The Demographic Frame is:

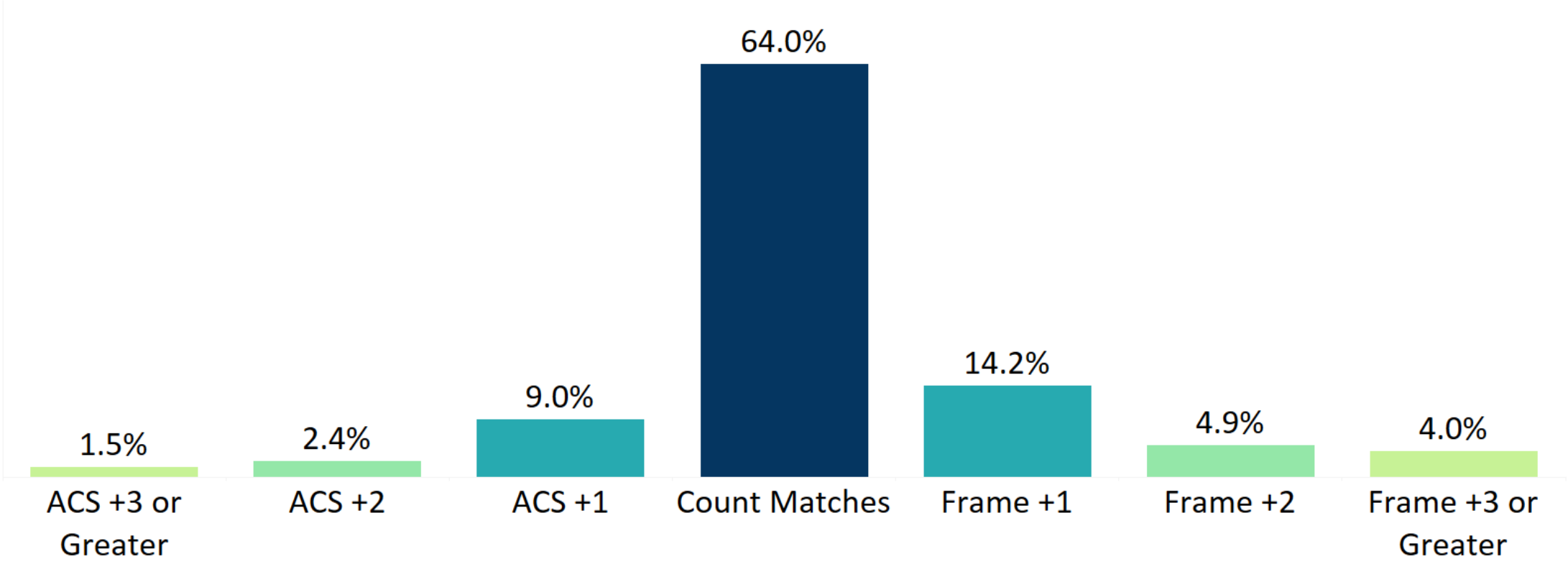
**2.45%** less than the 2020 Census
**1.87%** less than the 2020 Population Estimates

| | Demographic Frame | | 2020 Census | | 2020 Population Estimates | | Demo Frame Net Coverage Error (%) | |
|---|---|---|---|---|---|---|---|---|
| Geography | Number | Percent | Number | Percent | Number | Percent | Relative to Census | Relative to Pop Est. |
| United States | 323,300 | 100.0 | 331,500 | 100.0 | 329,500 | 100.0 | 2.45 | 1.87 |
| *Numbers in thousands and rounded to 4-significant digits* | | | | | | | | |

United States® Census Bureau

Source: U.S. Census Bureau, 2020 Census, 2020 Population Estimates, and 2020 Demographic Frame Prototype
Disclosure Review Board (DRB) approval number: CBDRB-FY23-POP001-0039

# *Do household counts match between the Demographic Frame and the ACS?*




Source: U.S. Census Bureau, 2019 American Community Survey and 2020 Demographic Frame Prototype
Disclosure Review Board (DRB) approval number: CBDRB-FY23-ACSO003-B0056

# For More Information

jennifer.m.ortman@census.gov

Census.gov/Frames

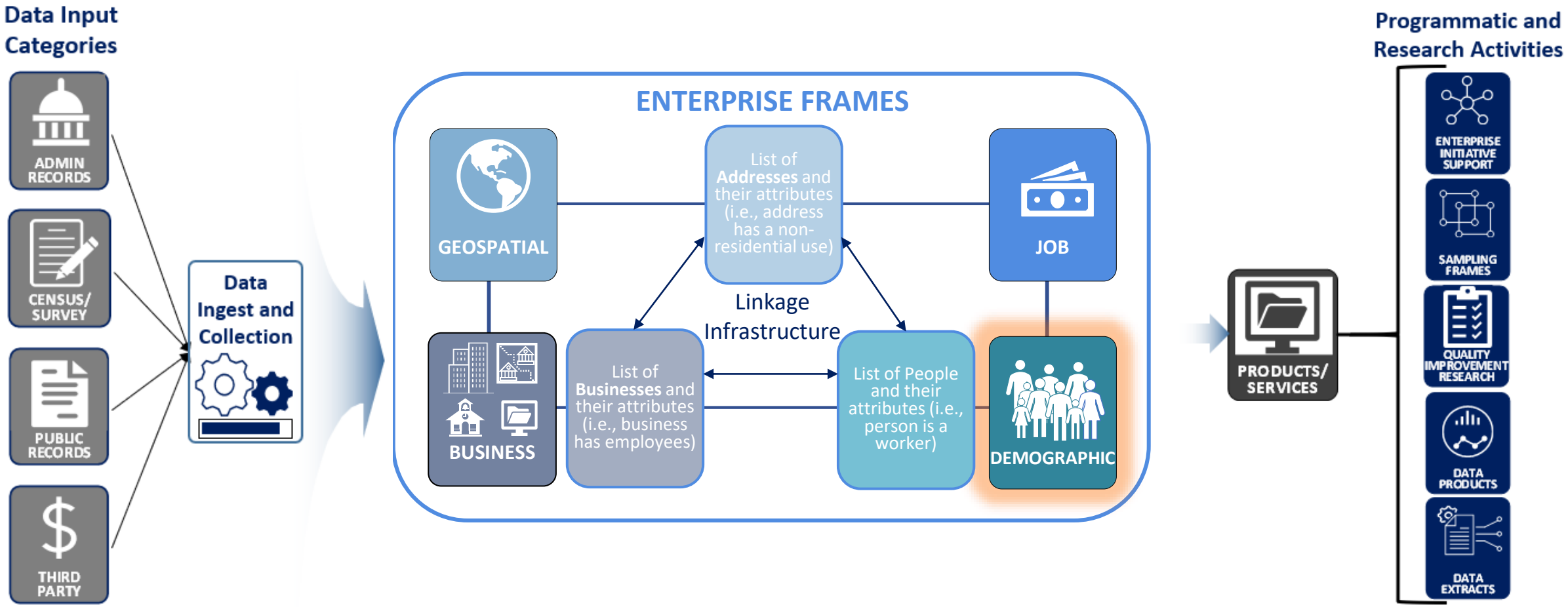
United States® Census Bureau

# Appendix Slides

United States® Census Bureau

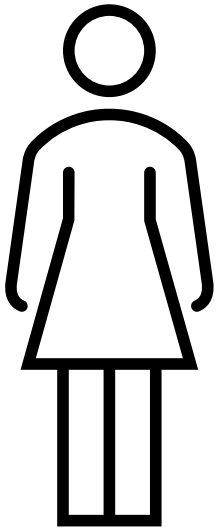
# Frames Program

*Creating an Infrastructure to Modernize the Census Bureau's Statistical Foundation*

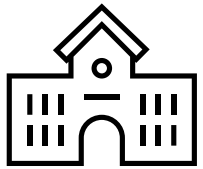



United States® Census Bureau

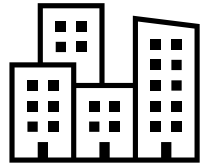
# Where does this person live?



THIS HOUSE?



THIS DORM?



THIS APARTMENT?

United States® Census Bureau

# The Person-Place Model (PPM)

- A process for assigning probabilities to PIK-MAFID (i.e., person-address) records to determine a person's residence on a particular reference date
- Learns from "true" PIK-MAFID pairs (ACS) to make guesses for PIK-MAFID pairs from administrative records
  - Over 100 model features including info about: source file, IRS/income, address, and aggregate demographics


United States® Census Bureau

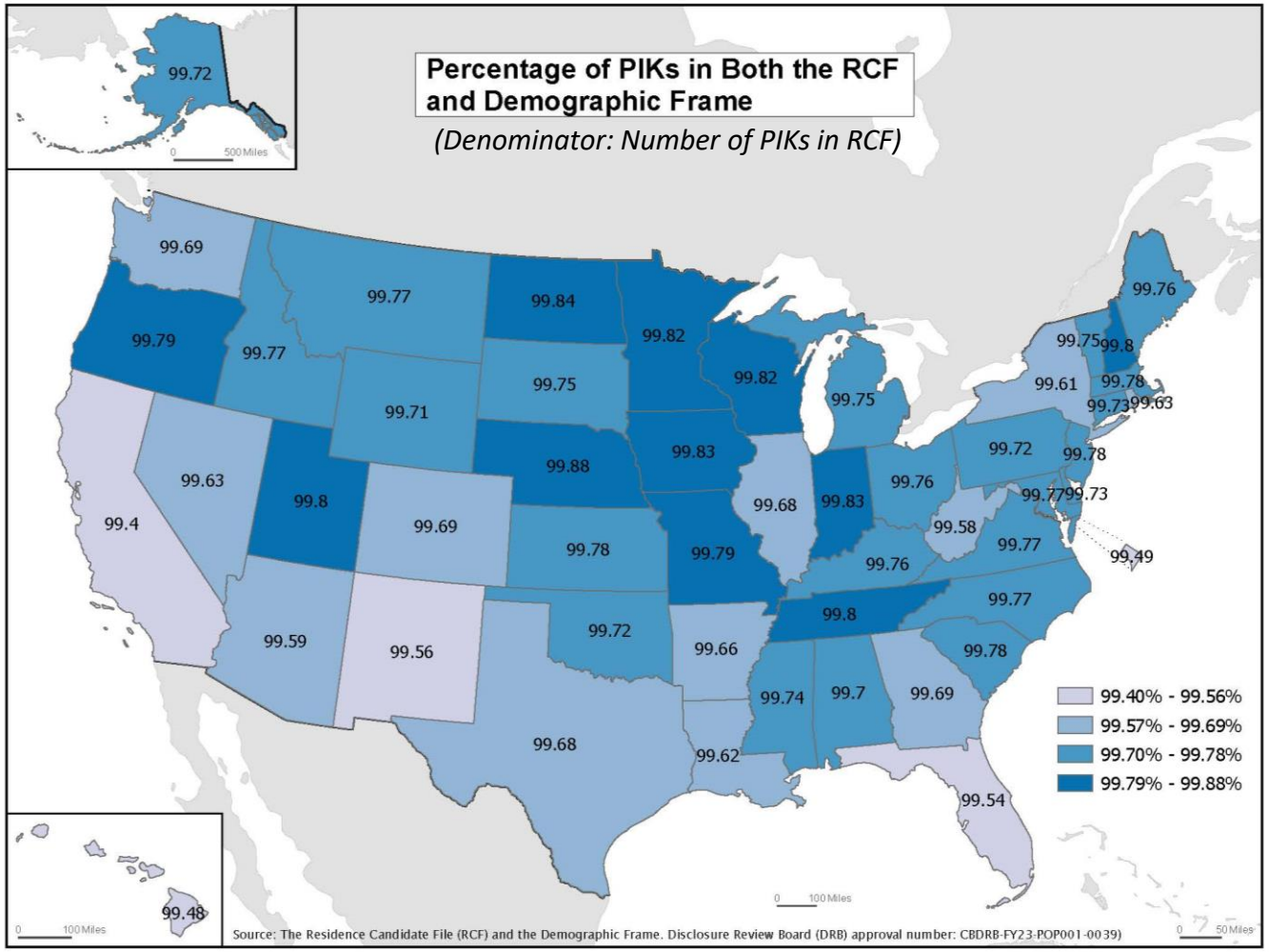
# Additional Evaluation Results

United States® Census Bureau

# RCF Analysis

*Does the Demographic Frame **find the same people** as the Residence Candidate File?*

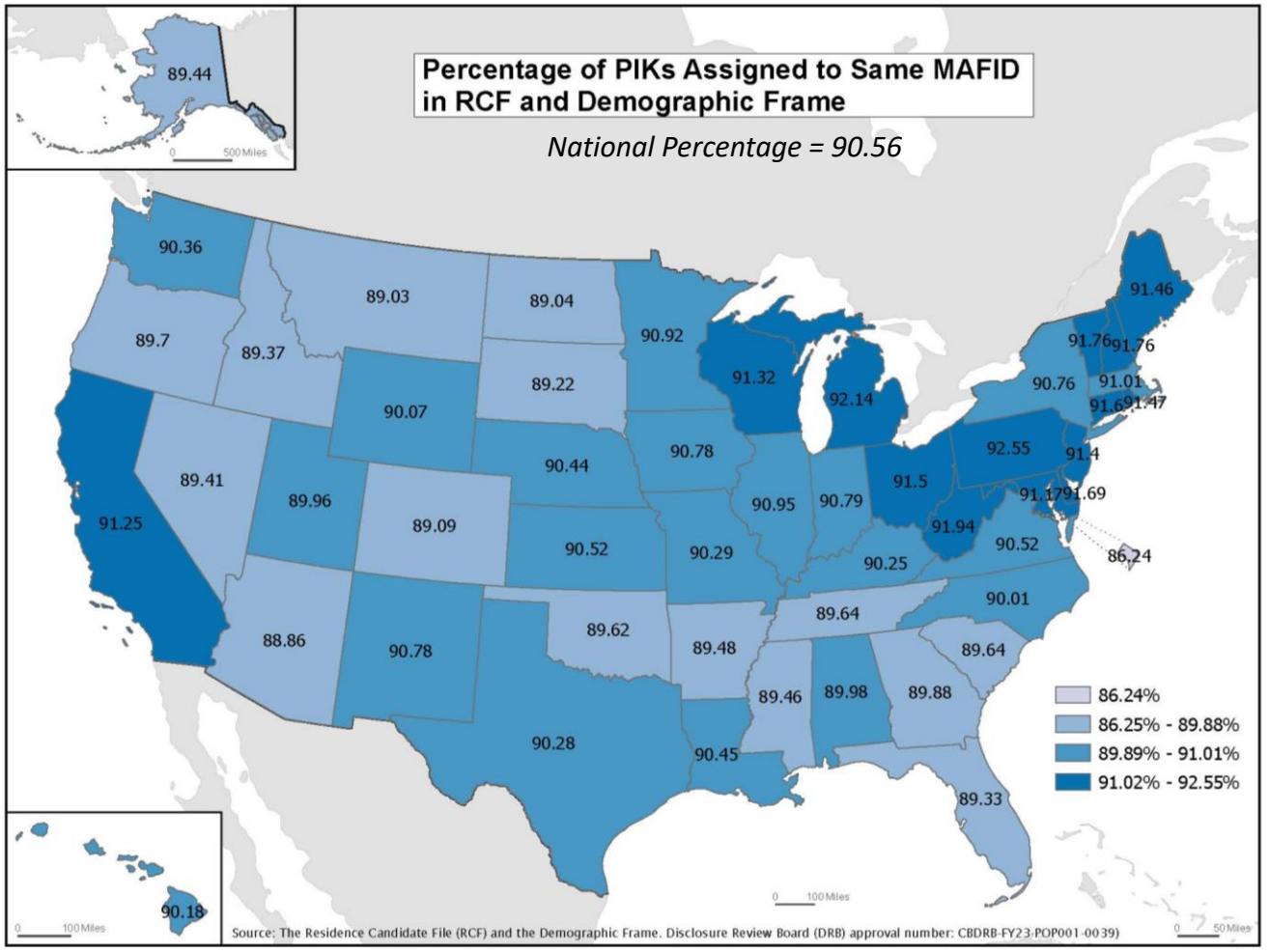
Over **99 percent** of PIKs observed in the RCF were also found in the Demographic Frame at the national level and when looking by state.



*Does the Demographic Frame **put people at the same address** as the Residence Candidate File?*

**90.56 percent** of PIKs were assigned to the same MAFID in the RCF and Demographic Frame.

Agreement ranges from **86 to 93 percent** by state.




United States® Census Bureau

# Nonemployer Statistics by Demographics (NES-D)

# Nonemployer Statistics by Demographics (NES-D)

- Annual series of nonemployer business counts, receipts, and business owner characteristics based on administrative records and Census Bureau data

  https://www.census.gov/programs-surveys/abs/data/nesd.html

- NES-D is investigating ways to use **demographic characteristics of business owners** from the Demographic Frame to reduce existing efforts to consolidate data sources

United States® Census Bureau

# Nonemployer Business Owners

Match Rates between 2018 NES-D and 2020 Demographic Frame by Type

| **Legal Form of Organization** | **Percent Matched** |
|---|---|
| Partnerships: *Two or more people who combine their resources to form a business* | 99% |
| Individual proprietorships: *Unincorporated business owned and run by one person or couple* | 97% |
| S-Corporations: *Smaller corporation with board of directors, corporate officers, bylaws, and management structure* | 97% |

*Results from Marquette (2023)*
*Disclosure Review Board (DRB) approval number: CBDRB-FY23-ESMD002-027*

United States® Census Bureau

# Demographic Consistency between 2018 NES-D and 2020 Demographic Frame

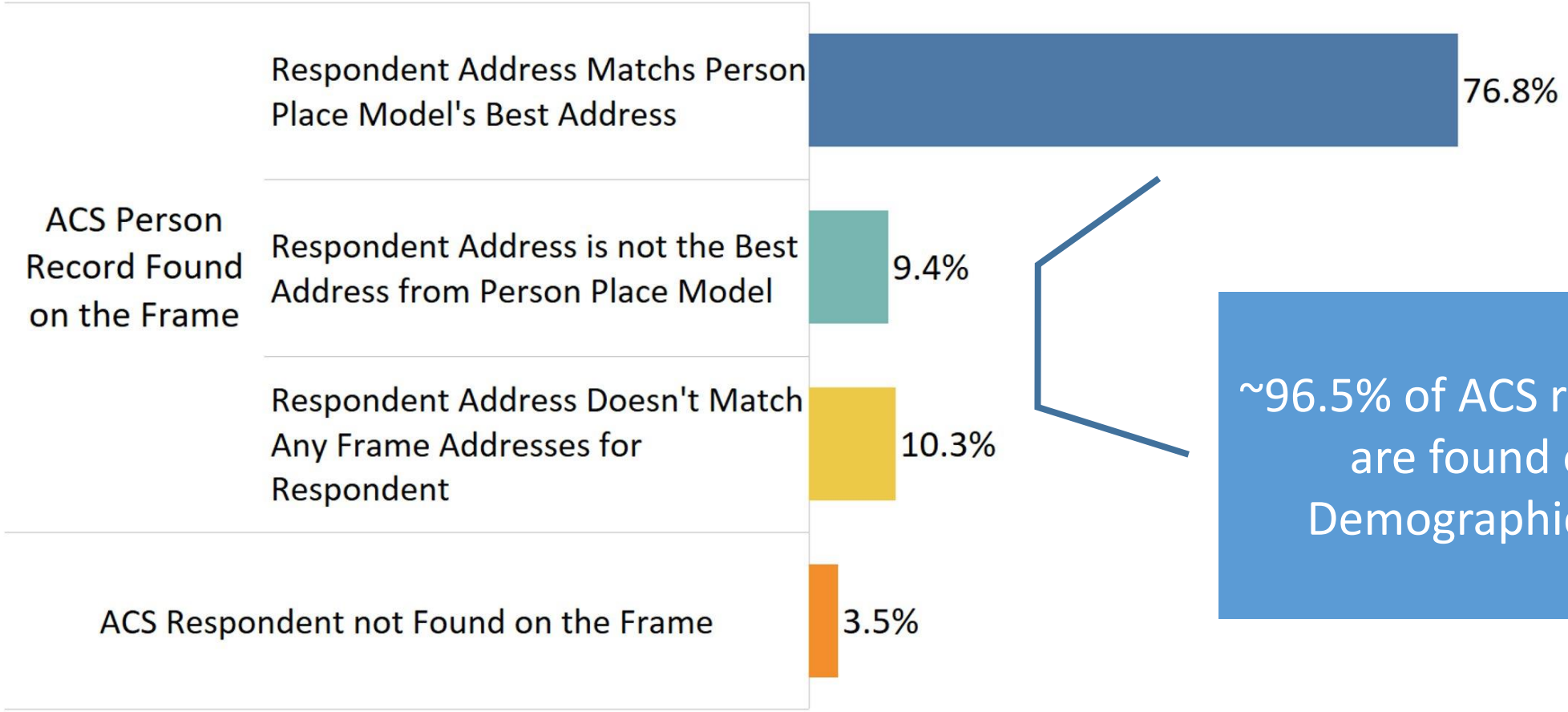
| Demo Frame Characteristics | Agreement Rate with NES-D |
|---|---|
| **Overall Race*** | **90%** |
| White | 96% |
| Black or African American | 92% |
| American Indian or Alaska Native | 83% |
| Asian | 73% |
| Native Hawaiian or Pacific Islander | 71% |
| Missing | 48% |
| **Overall Hispanic Origin** | **93%** |
| Hispanic | 92% |
| Non-Hispanic | 98% |
| Missing | 18% |
| **Overall Sex** | **99%** |
| Male | 99% |
| Female | 99% |
| Missing | 77% |
| **Overall Age** | **88%** |

United States® Census Bureau

**Following OMB's rules for business statistics and surveys, NES-D does not publish the race category "Some Other Race" or "Two or More Races."*
*Results from Marquette (2023). Disclosure Review Board (DRB) approval number: CBDRB-FY23-ESMD002-027*

# ACS Household Size Evaluation

# *Does the Demographic Frame put people at the same address as the ACS?*



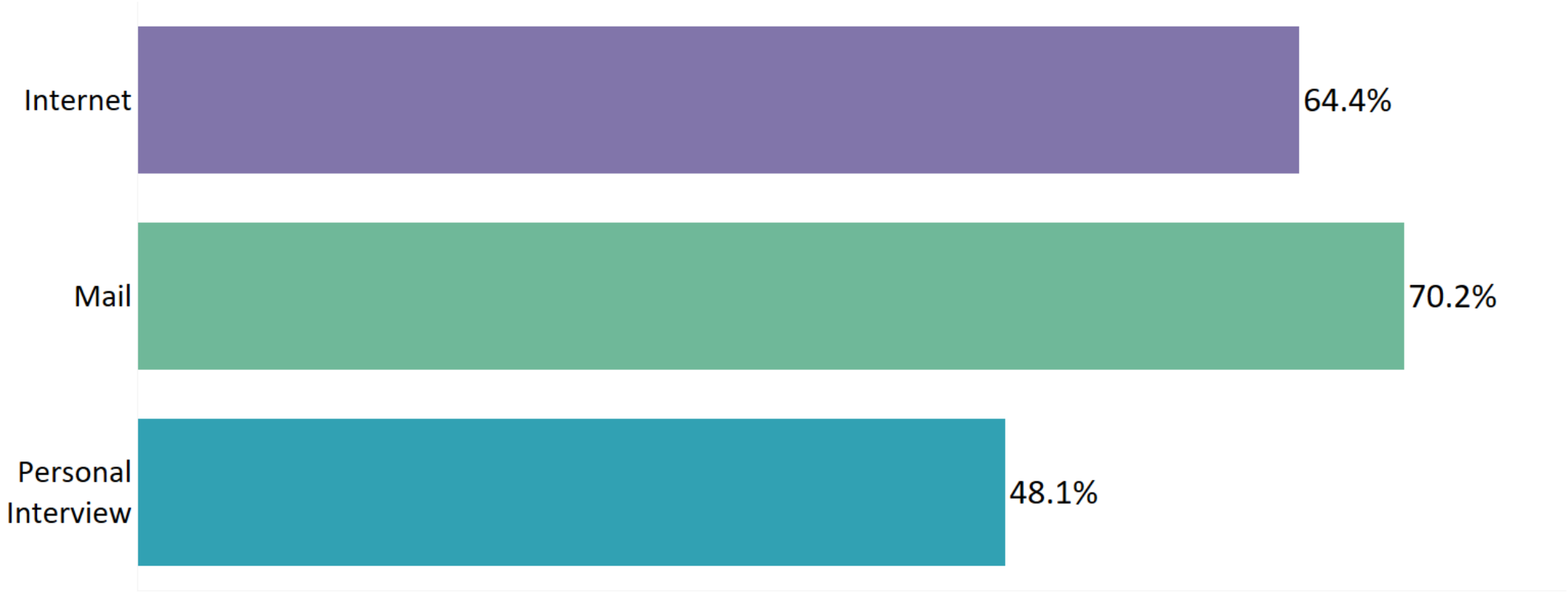

~96.5% of ACS respondents are found on the Demographic Frame.

United Census Bureau

Source: U.S. Census Bureau, 2019 American Community Survey and 2020 Demographic Frame Prototype
Disclosure Review Board (DRB) approval number: CBDRB-FY23-ACSO003-B0056

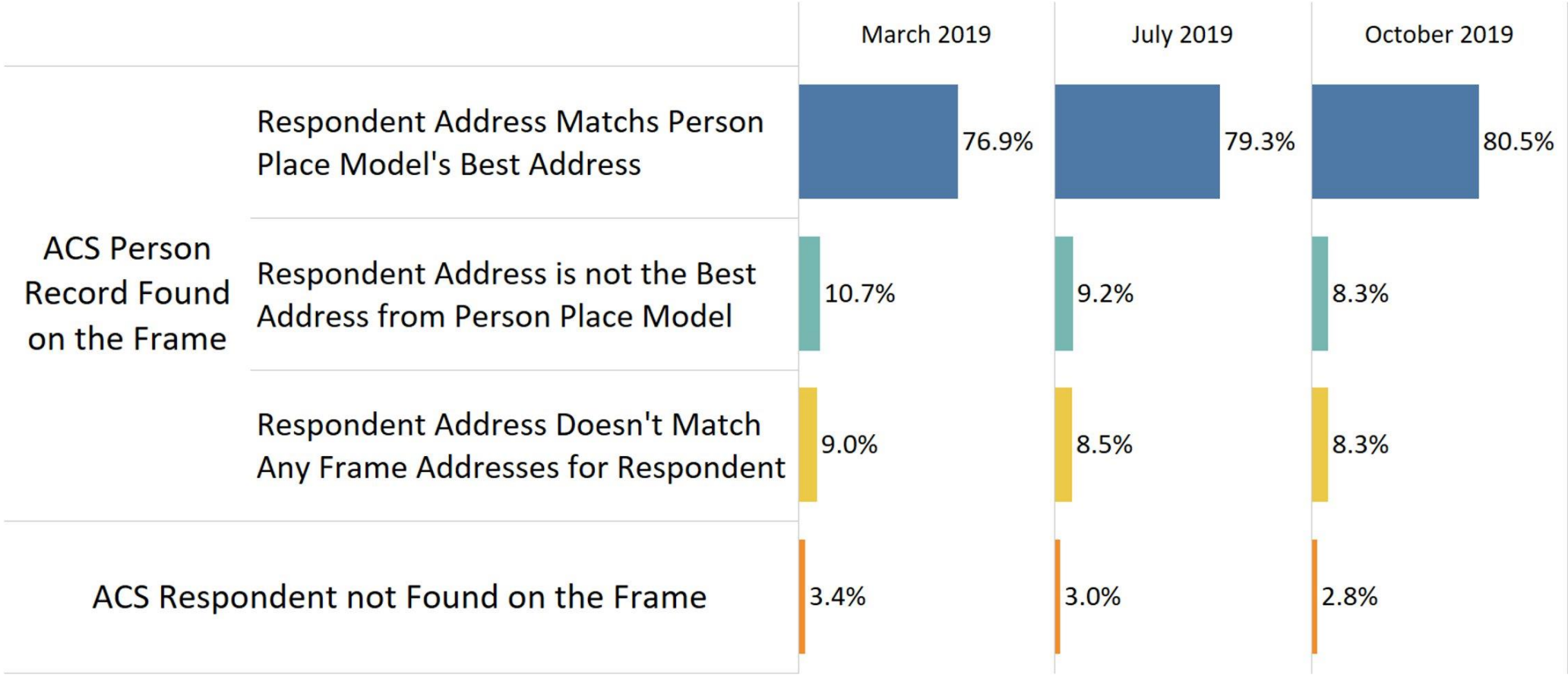
# Do Household Counts Match Between the ACS and Demographic Frame?

Percent of Households by Response where Count Matches in 2019 ACS and 2020 Demo Frame




*Results from Sawyer (2023). Disclosure Review Board (DRB) approval number: CBDRB-FY23-ACSO003-B0056*

# Does the Demographic Frame put people in the same address they are found in the ACS?




United Census Bureau

# National Experimental Wellbeing Statistics (NEWS)
## Combining Survey and Administrative Data to Improve Income and Poverty Statistics

IAOS-ISI Conference 2024
May 15, 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 number: **CDRB-FY23-SEHSD003-025**).*

United States® Census Bureau

# Income and Poverty Estimates

- Nonresponse to surveys is increasing
  - 11% in 2013 to 31% in 2023 in primary income and employment survey
- For those that respond to the survey, many do not answer income questions
  - ~45% of income in official U.S. poverty estimate imputed for nonresponse
- For those that answer income questions, many underreport
  - We estimate 9.4 percent fewer people in poverty (~3.5 million people) than official estimates

United States® Census Bureau

# What is NEWS?

- Rethink how we produce income and resource statistics
  - What is the best possible estimate given all the data currently available at the U.S. Census Bureau for a given income/resource statistic?
- Address multiple sources of bias simultaneously
  - Apply research on addressing each

United States® Census Bureau

# 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)

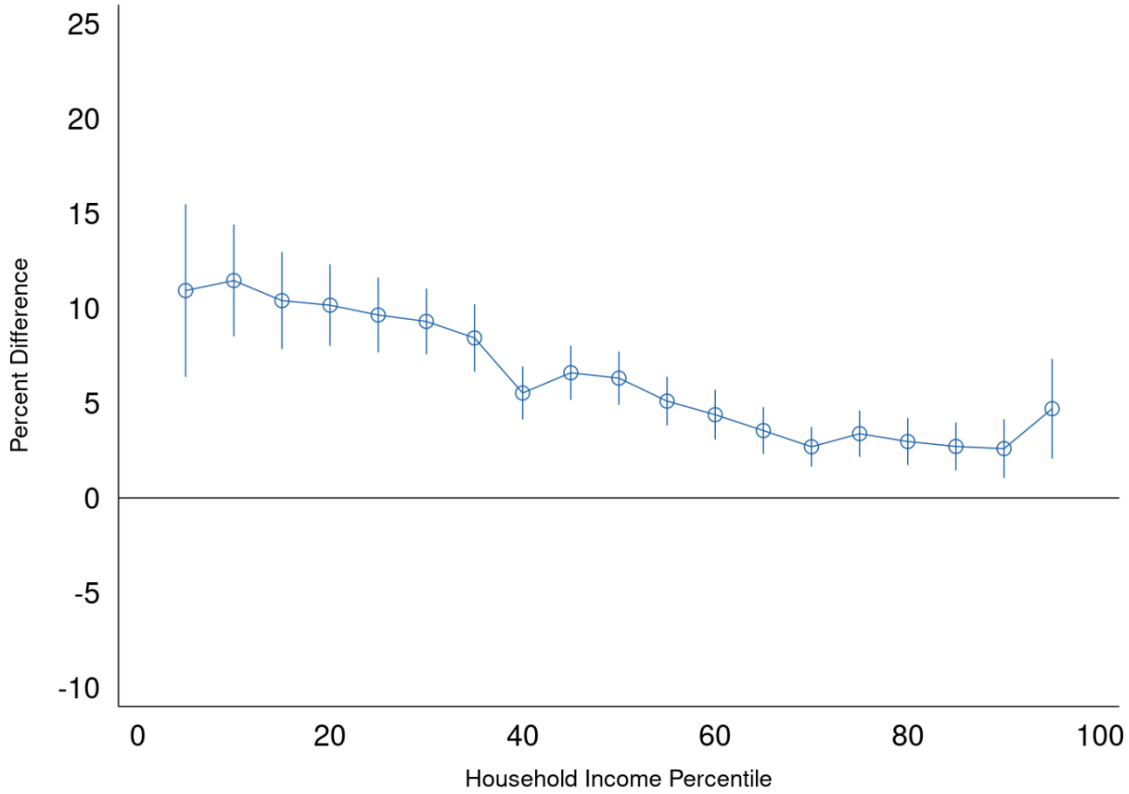
United States® Census Bureau

# What Have We Done?

- Version 1 Release – February 14
  - 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

United States® Census Bureau

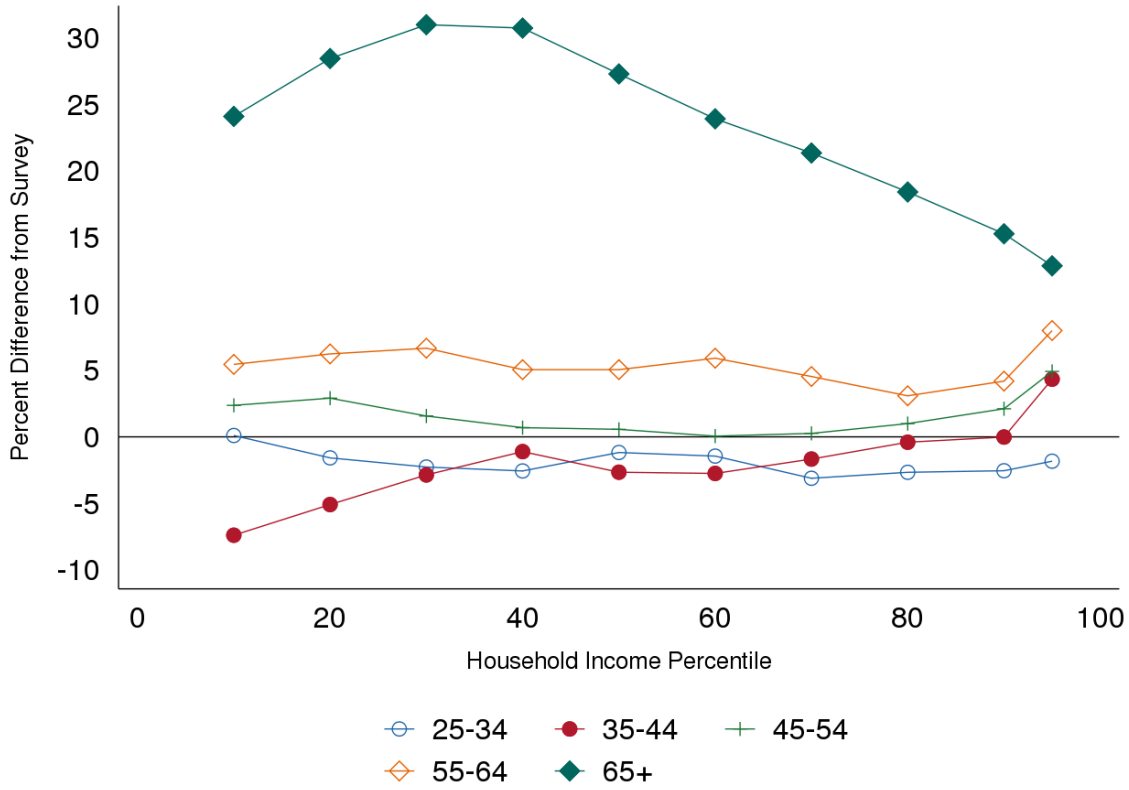
# Household Income in 2018
# NEWS Estimate Relative to Survey




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

United States® Census Bureau

# 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.

United States® Census Bureau

# 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

United States® Census Bureau

# Measurement Challenges
# Administrative Data

1. Selection into administrative data
   - Not everyone has to file taxes or gets a W-2 or other information return
   - Larrimore, Mortenson, and Splinter (2020) estimate poverty from administrative data, but must impute the existence and poverty status of 4-6 million people
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)

United States® Census Bureau

# 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)

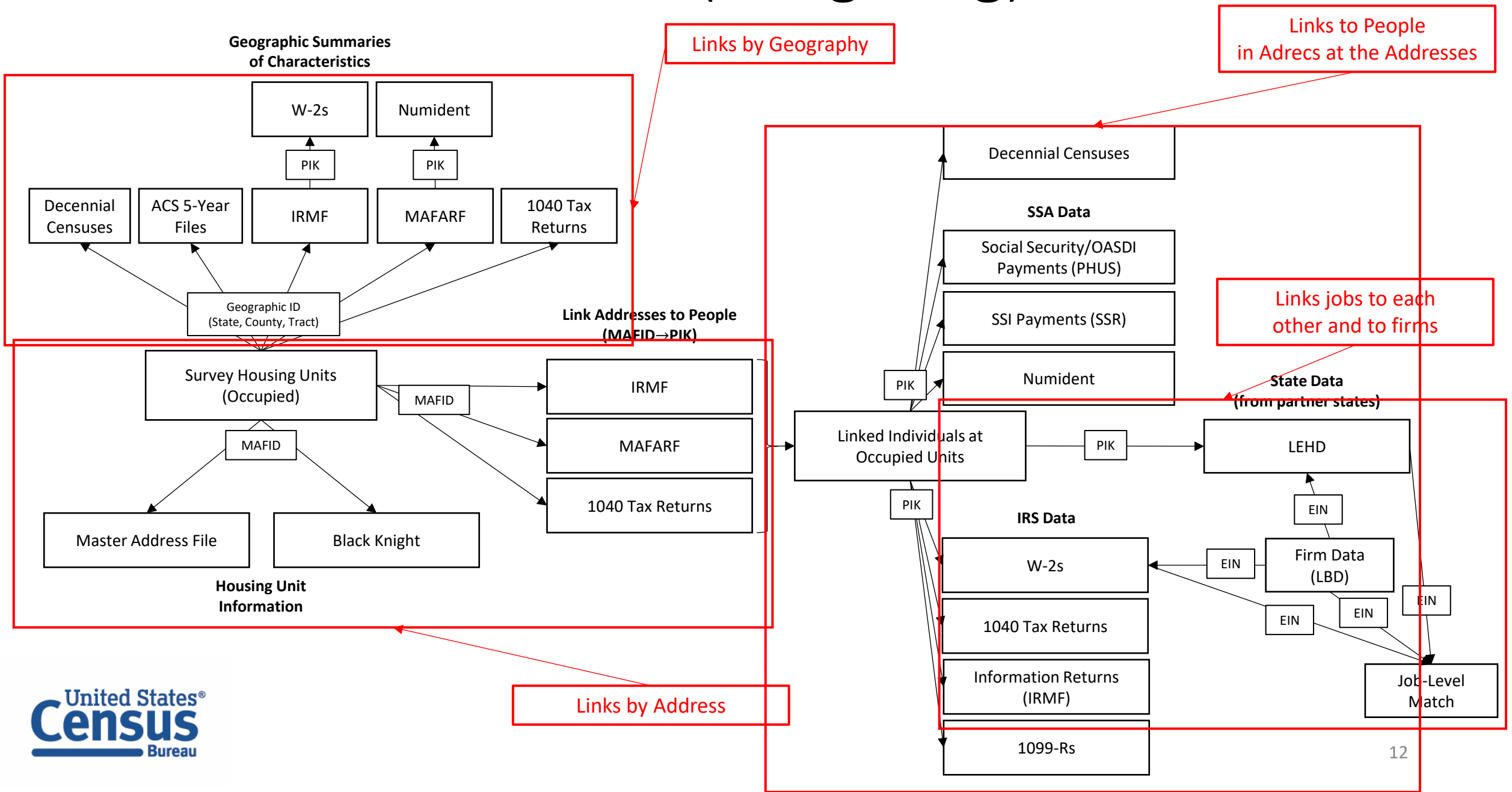
United States® Census Bureau

# 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<br>Selection into administrative data<br>Administrative data "nonresponse"<br>Selection into linkage | Rothbaum et al. (2021)<br>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"<br>Conceptual misalignment<br>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<br>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<br>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<br>Administrative mis- and underreporting | Bee and Mitchell (2017) |

United States® Census Bureau

# Address-Linked Data (Weighting)



# 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

United States® Census Bureau

# Different Earnings Sources
# W-2 vs. LEHD

| | | EIN and Indirect Matches | |
|---|---|---|---|
| | All Jobs | Unmatched Jobs | Share of Implied Total |
| Total Jobs | | | |
| W-2 | 256,800,000 | 25,680,000 | 0.097 |
| LEHD | 237,900,000 | 6,744,000 | 0.026 |
| EIN Matches | 216,100,000 | | 0.820 |
| Indirect Matches | 15,040,000 | | 0.057 |
| Implied Total Jobs | | 263,600,000 | |

United States® Census Bureau

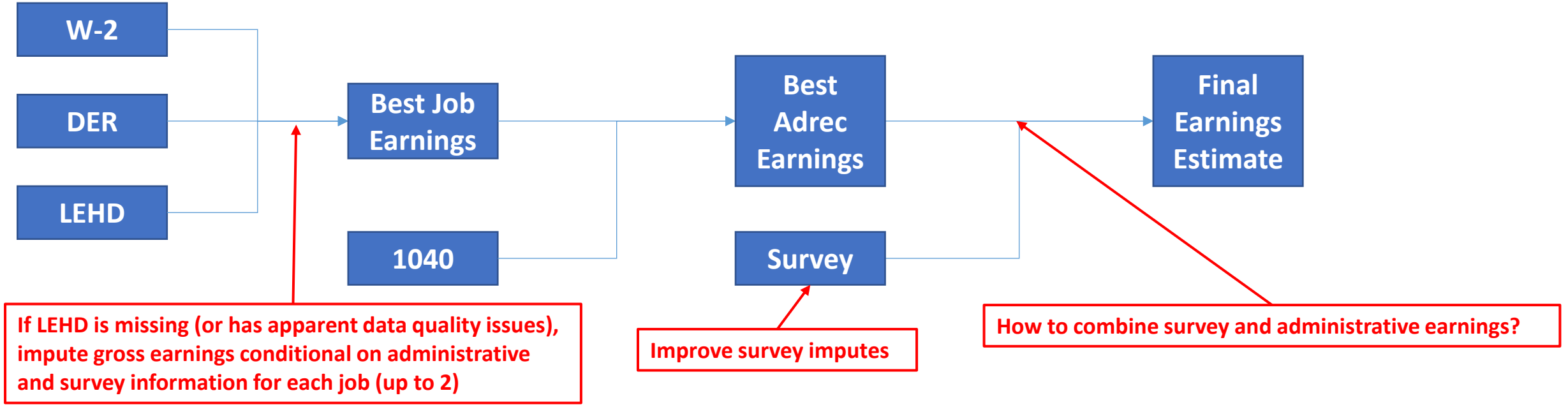
# The Full Picture – Wage and Salary Earnings

1. Use job-level Information to get "best possible" administrative job-level earnings

2. Compare to 1040 to check for missing earnings (at tax-unit level)

3. Compare to survey and decide for which individuals to use adrec or survey earnings

4. Final "best" estimate of earnings for each individual/household



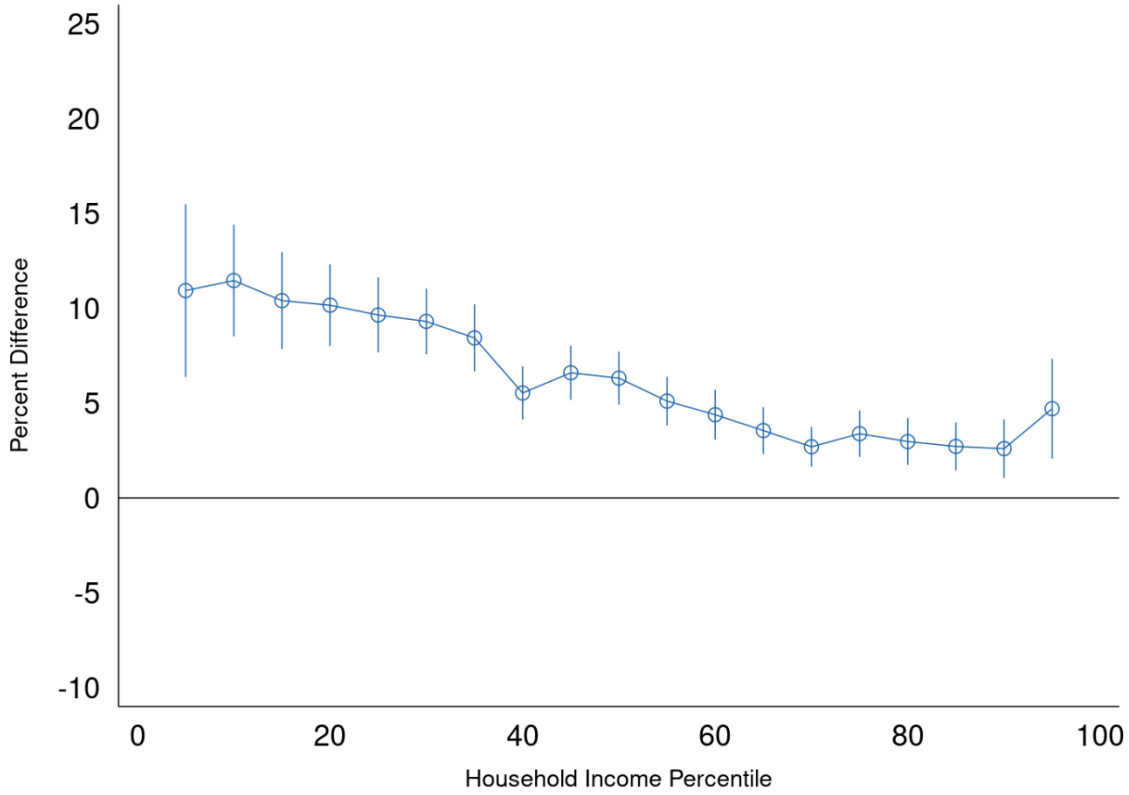

United States® Census Bureau

# Survey Earnings Use

- 21 percent of individuals
- More often for
  - Workers in real estate and construction
  - Younger workers (25-44 year-olds)
- Less often for
  - Workers in retail, education, management, and health care
  - Older workers (65+)
  - Black workers

United States® Census Bureau

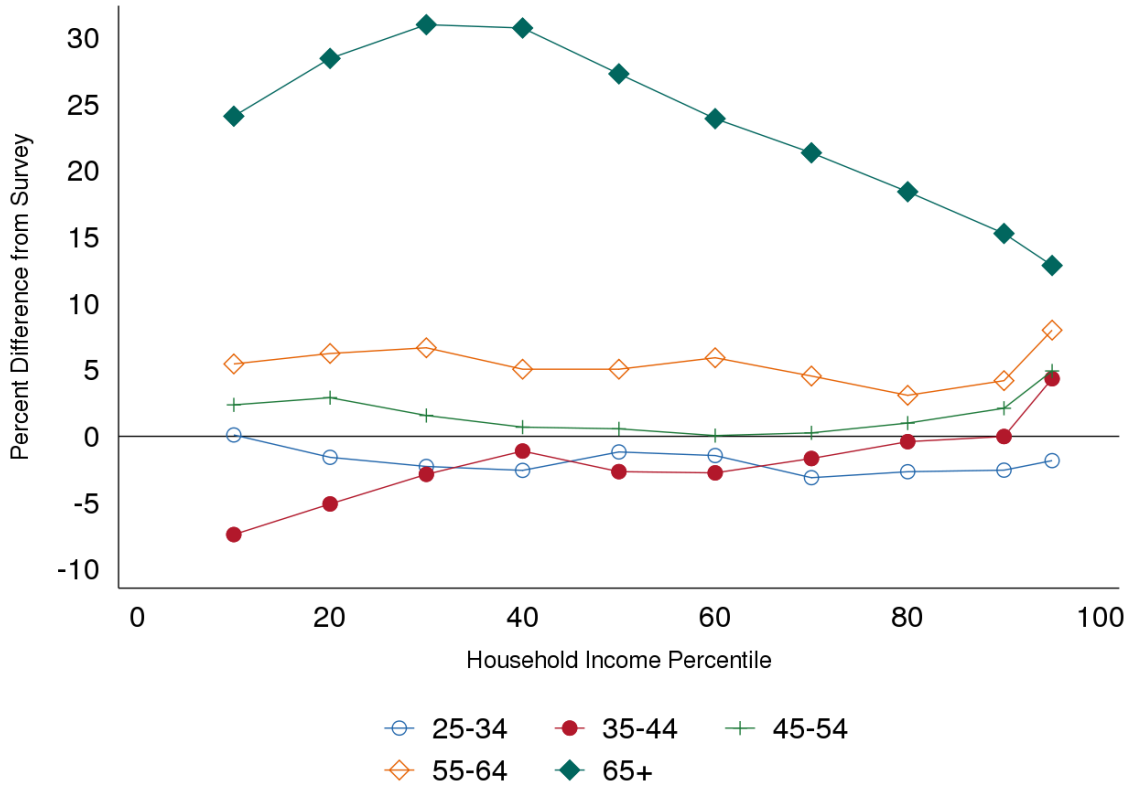
# Results
# Household Income Relative to 2019 CPS ASEC




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

United States® Census Bureau

# Household Income Relative to Survey By Age




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

United States® Census Bureau

# Next Release and Future Plans

- Next Release – planned for later this year
  - More income/resource concepts - Include taxes, credits, and in-kind transfers
- Prepare for more years
  - Handle incomplete administrative data
  - Either because not available in the past or not yet available for timely estimates
- Better address earnings measurement error

United States® Census Bureau

# 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

United States® Census Bureau

# The Census Household Panel Survey

Jason Fields, Jenny Hunter Childs, Aleia Fobia, Cassandra Logan, Stephanie Coffey, Jennifer Ortman and David Hornick

U.S. Census Bureau

Session: Modernizing The U.S. Census Bureau's Demographic Programs

IAOS-ISI 2024, Mexico City - Improving Decision Making for All

Wednesday, May 15, 2024

# The Idea

- First nationally-representative panel to be built by a federal statistical agency primarily for federal statistical agency use, and represents a step forward in providing infrastructure for timely evidence-based decision making across the federal government

- Nationally-representative, address-based, probability-based internet panel (includes non-internet households)

- Our frames and access to administrative data allow a wealth of data available nowhere else

# Census Household Panel Goals

- High-quality panel to improve representativeness and significantly reduce burden on households in the interests of collecting high-frequency data.
- Households carefully selected and recruited by the Census Bureau to reflect the diversity of our Nation's population.
- Approximately 15,000 panelists linked to the Census Bureau's gold standard Master Address File, available for linkage to administrative records securely maintained and curated by the Census Bureau.

# Census Household Panel Goals

- This Panel will become integral to rapidly providing insight on national events that may impact social, economic, or demographic characteristics of the population.
- The Panel will also help us research questions related to surveys:
  - Nationally representative field tests to test content changes and modernization in various surveys.
  - Alternative methods for enhancing data with administrative and other external data sources and developed modeled data.
  - Platform for developing adaptive design procedures that use auxiliary data sources.

# Demographic High-Frequency Surveys (DHFS) Program


- Household Pulse Survey
- Census Household Panel
- Census Military Panel
- School Pulse Survey

United States® Census Bureau

# Demographic High-Frequency Surveys (DHFS) Program Scope and Key Features


- A **streamlined survey lifecycle** facilitated by a **lighter infrastructure** (e.g., minimal processing, data product limitations) for instrument development, data processing, and dissemination.
- Results in surveys having at least one of these characteristics:
  - **Rapid Response:** rapidly developed, deployed and produces data close to real-time in response to emergent events that affect the social and/or economic well-being of the U.S. public (e.g., the COVID-19 pandemic, the 2008 housing market collapse).
  - **High Frequency:** produces data released on a high-frequency, cyclical basis when there is a need to measure rapidly changing circumstances.
- Meets specific criteria for **urgency**, **compelling need**, and are **not collected elsewhere** for the needs and uses of the Census Bureau.
- Produces data on a **time-limited basis**; once the events prompting the needs have passed, the content is moved off the survey; not intended as a permanent data collection vehicle.
- **Limitations** will be **clearly documented** and made **publicly available.**
- Accepting these characteristics also means these data may not meet regular program quality standards.

United States® Census Bureau

# Census Household Panel - Methods

- Online using Qualtrics as the data collection platform:
  - Self-response online.
  - Inbound TQA.
  - Outbound nonresponse follow-up.
- Recruitment consisted of 3 mail contacts and telephone follow-up.
- Participants were enrolled via a baseline questionnaire.
- Panelists are invited to complete monthly topical surveys.
- Incentives are provided to respondents upon completion of each survey.

# Census Household Panel - Collection Structure

EXPERIMENTAL DATA

## Recruitment/Baseline

- $5 visible prepaid incentive.
- $20 contingent incentive.
- Questions from ACS and other national surveys.
  - To facilitate non-response bias analysis.
  - Frame and baseline data to allow for targeted sampling for topical surveys.

## Monthly Topical Surveys

- $10 contingent incentive.
- SMS/Email invitation.
- Web completion.
- ~20 min.

United States® Census Bureau

# Census Household Panel - Baseline Topics

- Roster
- Demographics
- Employment
- Transience and Housing
- Language
- Internet access and digital affinity
- Health and health insurance
- Food security
- News Sources
- Volunteering
- Privacy and confidentiality
- Trust in government
- Contact information

# Census Household Panel - Topical Content


- Census Barriers, Attitudes and Motivators Survey (CBAMS).
- Household Pulse Survey content.
- Content development and testing for Census Bureau-conducted household surveys.

United States® Census Bureau

# Uses of Administrative and Alternate Data Sources


- Various entry points for administrative and alternate data sources:
  - Pre-data collection (sampling).
  - During data collection (adaptive design in contact and incentive structure).
  - Post processing (weighting).

United States® Census Bureau

# Pre-Data Collection: Sampling

EXPERIMENTAL DATA

- Master Address File/Geospatial Frame.
- Demographic Frame.
- Planning Database (PDB).

United States® Census Bureau

# Recruitment Sampling

EXPERIMENTAL DATA

- Stratified systematic sample of 75,000 addresses from all eligible units on MAF, including all 50 states + DC.
- Auxiliary data from Demographic Frame and PDB to form 16 strata in each Census Region:
  - Demographic Frame used to stratify households into Hispanic/Race strata based on 3 race strata (Black Alone, White Alone and Other) and 2 Hispanic strata (Hispanic/Non-Hispanic).
  - Households not matched to the Demographic Frame will use PDB strata (Hispanic, Non-Hispanic Black Alone, Non-Hispanic White Alone and Non-Hispanic Other).
  - Final strata include households missing information on either the Demographic Frame or PDB.

United States® Census Bureau

# Sampling

EXPERIMENTAL DATA

Formed Stratum for MAFIDs Matching the Demographic Frame

| Ethnicity | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Hispanic (HP) | | | | | | Non-Hispanic (NH) | | | | | |
| Race | | | | | | Race | | | | | |
| Black Alone (BK) | | White Alone (WH) | | Other (OT) | | Black Alone (BK) | | White Alone (WH) | | Other (OT) | |
| With Children KIDHHFLG = 1 | Without Children KIDHHFLG = 0 | With Children KIDHHFLG = 1 | Without Children KIDHHFLG = 0 | With Children KIDHHFLG = 1 | Without Children KIDHHFLG = 0 | With Children KIDHHFLG = 1 | Without Children KIDHHFLG = 0 | With Children KIDHHFLG = 1 | Without Children KIDHHFLG = 0 | With Children KIDHHFLG = 1 | Without Children KIDHHFLG = 0 |
| HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L | HH Size S to L |

Formed Stratum for MAFIDs Matching the Planning Database

| Ethnicity | | | |
|---|---|---|---|
| Hispanic (HP) | Non-Hispanic (NH) | | |
| | Race | | |
| | Black Alone (BK) | White Alone (WH) | Other (OT) |
| PCT_KID_UNDER_18 | PCT_KID_UNDER_18 | PCT_KID_UNDER_18 | PCT_KID_UNDER_18 |

United States® Census Bureau

# Adaptive Design During Data Collection

- Focused outbound calling.
  - Using data from PDB, sampled cases were quartiled according to self-response propensity to the American Community Survey.
    - Lowest and highest quartiles received no outbound calling.
    - Middle quartiles received 1-2 calls.
- Plans to evaluate and re-assess for replenishment sampling.

United States® Census Bureau

# Additional Use of Alternate Data Sources

- Data Quality.
  - Comparing survey data to frame and administrative data.
  - Respondent and nonrespondent frame data.
  - Research on weighting strategies for current and future data collections.
- Targeted Survey Design.
  - Targeting data collection based on frame data.
  - Tailoring survey questions based on previously collected information and frame data (dependent data).
- Targeted Recruitment.
  - Differential contact strategies, case prioritization, incentives, etc. based on frame data.
  - Replenishment.

# First Replenishment Operation

- We conducted our first sample replenishment in March 2024.
  - Sample Size – 30,000.
  - Used same methods as initial recruitment (3 mail contacts and telephone follow-up).
  - New Panel Members Recruited – ~5,700.

# Panel Overview

EXPERIMENTAL DATA

- Initial recruitment operation yielded over **12,000** respondents to the Baseline survey.
- Monthly Topical Operations - ~7,500 respondents per month.
- Refreshment Operation - ~5,700 respondents.
- Current Panel Size - ~17,700.
- New panel members expected to receive their first topical survey in May.

United States® Census Bureau

# THANK YOU!

**Contact:**
jason.m.fields@census.gov

**Census Household Panel respondent page:**
<https://www.census.gov/programs-surveys/census-household-panel.html>

**Household Pulse Survey main page:**
<https://www.census.gov/householdpulsedata>

**U.S. Census Bureau Frames Program:**
<https://www.census.gov/about/what/transformation/maximizing-operational-efficiency/data-centric-business-ecosystem/frames-overview/frames.html>

**U.S. Census Bureau Statistical Quality Standards:**
<https://www2.census.gov/about/policies/quality/quality-standards.pdf>