

The Frames Program

Lori B. Zehr, PhD

October 30, 2024

Presented to New York State Data Center

*Any opinions and conclusions expressed herein are those of the author(s)
and do not reflect the views of the U.S. Census Bureau.*

What You Will Learn Today

- What is the Frames Program?
- What are the benefits of the Frames Program?
- What is the current state of the Enterprise Frames?
- Where are we headed with the Frames Program?

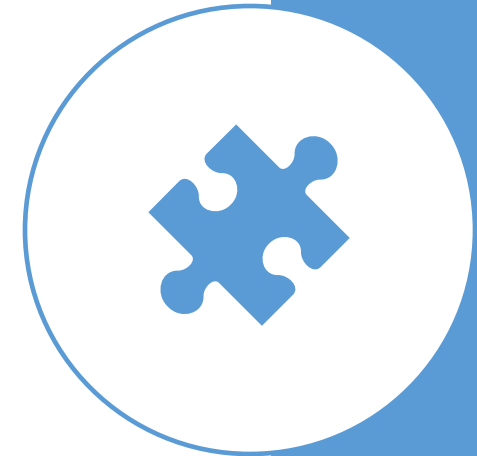


What is a Frame?

A frame refers to a collection of datasets organized around one of four foundational units: Person, Business, Location, Job.

A frame is supported by a collection of standard methods, code, processes, documentation, and subject matter expertise.

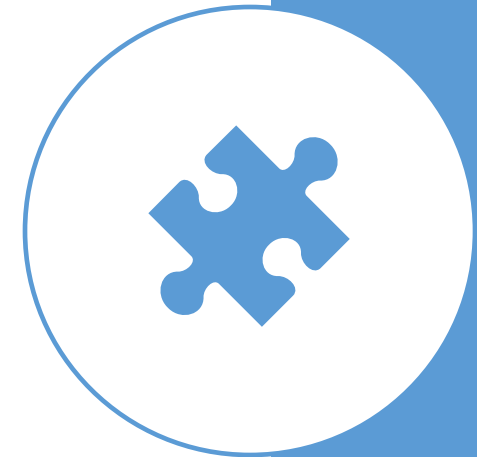
A sampling frame often refers to a subset of the larger foundational dataset.



Frames Program

The Challenge

- Various frame-like datasets are in use at the Census Bureau.
- Contain information about addresses, businesses, jobs, and people.
- Some information is unique within a particular frame, while other information is duplicative.
 - e.g., group quarters exist in both the Master Address File and Business Register.
- Frames exist in an uncoordinated and unintegrated environment.
- No systematic process exists that allows for the direct linkage of information contained in one frame with information in any other frame.



Frames Program

The Vision

- *To create Enterprise-wide frames linkable in nature, agile in structure, accessible for production or research on a need-to-know basis, and that adhere to best practices in terms of technology usage, data management, and methodology.*
- Linkable in Nature
 - Each frame will include the necessary unique identifiers and keys for linkage to each other.
 - Location information will be available on each person, job, and business record.
 - Person records will be linkable to job records.
 - Job records will be linkable to business records.



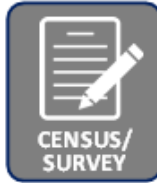
Frames Program

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

Data Input Categories



ADMIN RECORDS



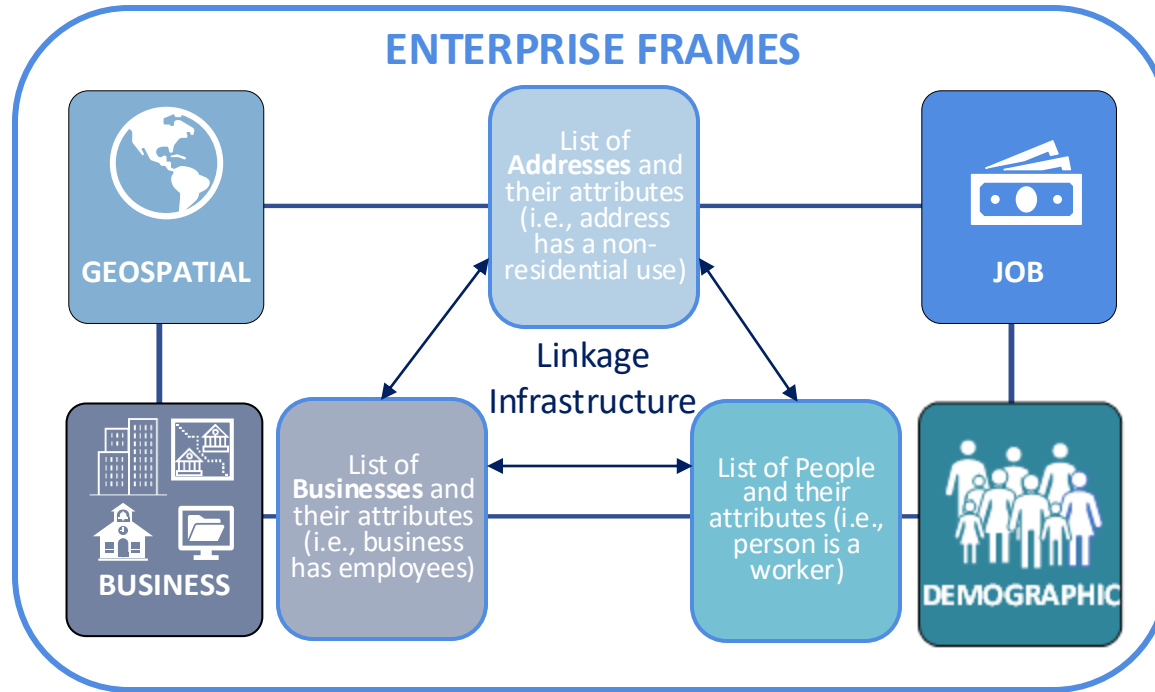
CENSUS/SURVEY



PUBLIC RECORDS



THIRD PARTY



Programmatic and Research Activities



PRODUCTS/SERVICES



ENTERPRISE INITIATIVE SUPPORT



SAMPLING FRAMES



QUALITY IMPROVEMENT RESEARCH



DATA PRODUCTS



DATA EXTRACTS



Geospatial Frame

Current State

- Master Address File/Topologically Integrated Geographic Encoding and Referencing (TIGER) System. Contains:
 - National address list of more than 200 million addresses prioritizing residential addresses.
 - Over 14 million unique geographic areas.
 - Approximately 7 million miles of roads.
- Serves as the source of address, location, and boundary information for the Decennial Census, American Community Survey, and other demographic surveys.
- Updated using information from the U.S. Postal Service; tribal, federal, state, and local governments; and Census data collection operations.

Future State

- Stores a comprehensive set of residential, **business**, and **mixed-use** addresses.
- Continues to provide “Master Address File IDs” (MAFIDs) for all addresses in the frame.
- Provides enhanced information about addresses; e.g., parcel, building footprint, tenure, and seasonality.
- Will be integrated with Address Matching Tools (AMT) to enhance assignment of MAFIDs to datasets with addresses.
- Linkable to other Enterprise Frames.



Business Frame

Current State

- National list of 30M+ businesses.
- Serves as the enumeration list for the Economic Census and the source of sampling for Economic Surveys.
- Updated using data from the Internal Revenue Service (IRS) tax records; the Bureau of Labor Statistics; the Social Security Administration; the Economic Census; and some Census surveys.
- Updated using geographic information from the Census Bureau's Master Address File.

Future State

- Collection of rich, harmonized business data stored in a central location.
- Incorporates the Longitudinal Business Database, resulting in consistent methodology.
- Connects the Government's Master Address File, ensuring improved quality for the Government sector.
- Linkable to other Enterprise Frames.



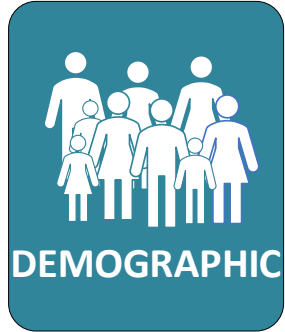
Job Frame

Current State

- An inventory of 14+ billion “job records” as part of the Longitudinal Employer-Household Dynamics Program (LEHD).
- Each “job record” is a relationship between a person and a firm in which at least \$1 of wages were paid in a given quarter.
- Covers >97% of private employment and most federal, state, and local government jobs.
- Supports public-use data products.
- Additionally, the LEHD Program maintains a business frame that is independent of the BR, known as the Employer Characteristics File (ECF).

Future State

- Expands on the job frame that was built as part of the LEHD.
- Contains two linked databases of jobs:
 - Quarterly Job Frame, derived from quarterly unemployment insurance wage records supplied by states
 - Annual Job Frame, derived from W-2 tax records
- Adds self-employment jobs.
- Focuses explicitly on job records.
- Linkable to other Enterprise Frames.



Demographic Frame

Current State

- Several disparate person frames exist within the Census Bureau.
- Each was developed to serve a specific purpose:
 - Evaluate administrative data coverage
 - Support 2020 Decennial Census

Future State

- Consists of demographic, social, and economic characteristics of individuals derived from census, survey, administrative, and third-party data sources.
- Includes unique person-level identifiers used to link individuals across datasets.
- Links information about individuals across time and space.
- Can be aggregated to family and household units.
- Linkable to other Enterprise Frames.

The Frames Program: Enabling Transformation through Linked Data



Frames Program

Benefits

- Fosters an environment to do more work virtually.
- Reduce burden on respondents (individuals and businesses) by re-using data that already exist in one or more frames.
- Reduce duplication, manual efforts, and one-off data cleaning projects.
- Allows more time for production and research.
- Allows for ability to more easily trace data and code through the linked frames, facilitating reproducibility.
- Expands the Census Bureau's ability to report on changes and trends affecting the Nation's population, economy, and communities.





lori.b.zehr@census.gov

geo.frames.pgmo@census.gov

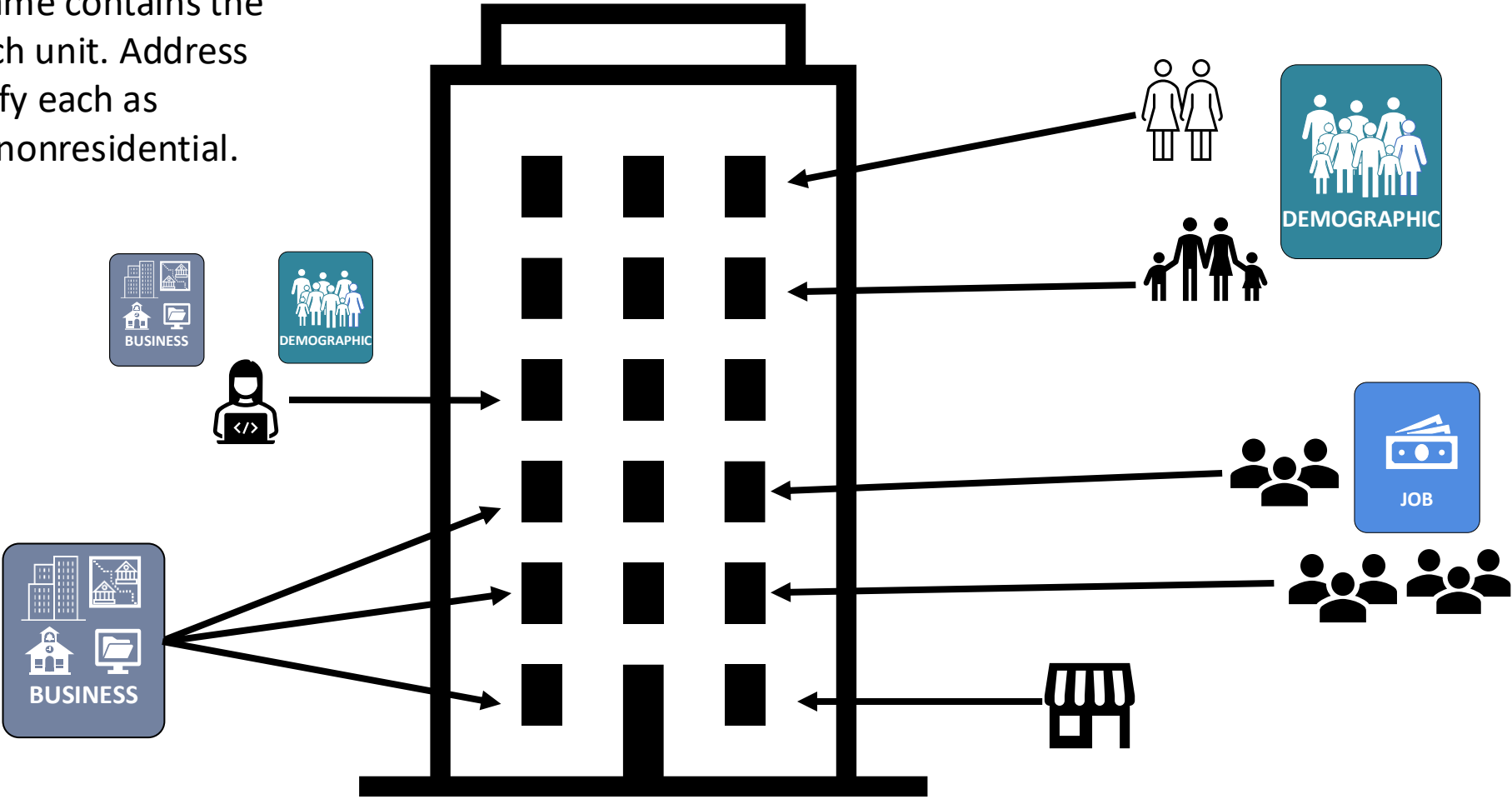


census.gov/frames

Basic Street Address: 1234 Example Street, Anywhereville

Building contains six units. What can the frames tell us about them?

Geospatial Frame contains the address of each unit. Address sources identify each as residential or nonresidential.



Geospatial Frame Update

April Avnayim

New York State Data Center Meeting

October 30, 2024

How does the Geospatial Frame differ from the MAF/TIGER System*?

*Master Address File/Topologically Integrated Geographic Encoding and Referencing System

Frames Program



The Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) System is the **Geospatial Frame**. The MAF/TIGER System, along with other linkable geospatial data, such as parcels, building footprint files, grids, and imagery underpins the Census Bureau's mission.



The **Business Frame** is a collection of rich, harmonized business data across multiple sources, including (but not limited to) the Longitudinal Business Database, Governments Master Address File, and third-party sources, linked to the Business Register, and stored in a central location.



The **Job Frame** expands on the job frame that was built as part of the Longitudinal Employer-Household Dynamics (LEHD) program. It contains the Quarterly Job Frame, derived from quarterly unemployment insurance wage records supplied by states, and the Annual Job Frame, derived from W-2 tax records. It adds self-employment jobs data.



The **Demographic Frame** consists of demographic, social, and economic characteristics of individuals derived from Census Bureau surveys, administrative records, and third-party data sources. It includes unique person-level identifiers used to link individuals across datasets and over time. It can be aggregated to family and household units.

The MAF/TIGER System Today

Geospatial information in our MAF/TIGER System includes...



Over 14 million unique geographic areas, including legal, administrative, and statistical areas (e.g., census tracts, block groups, blocks)



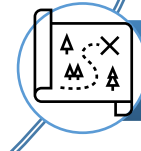
Legal boundaries for approximately 40,000 units of government



Approximately 7 million miles of roads

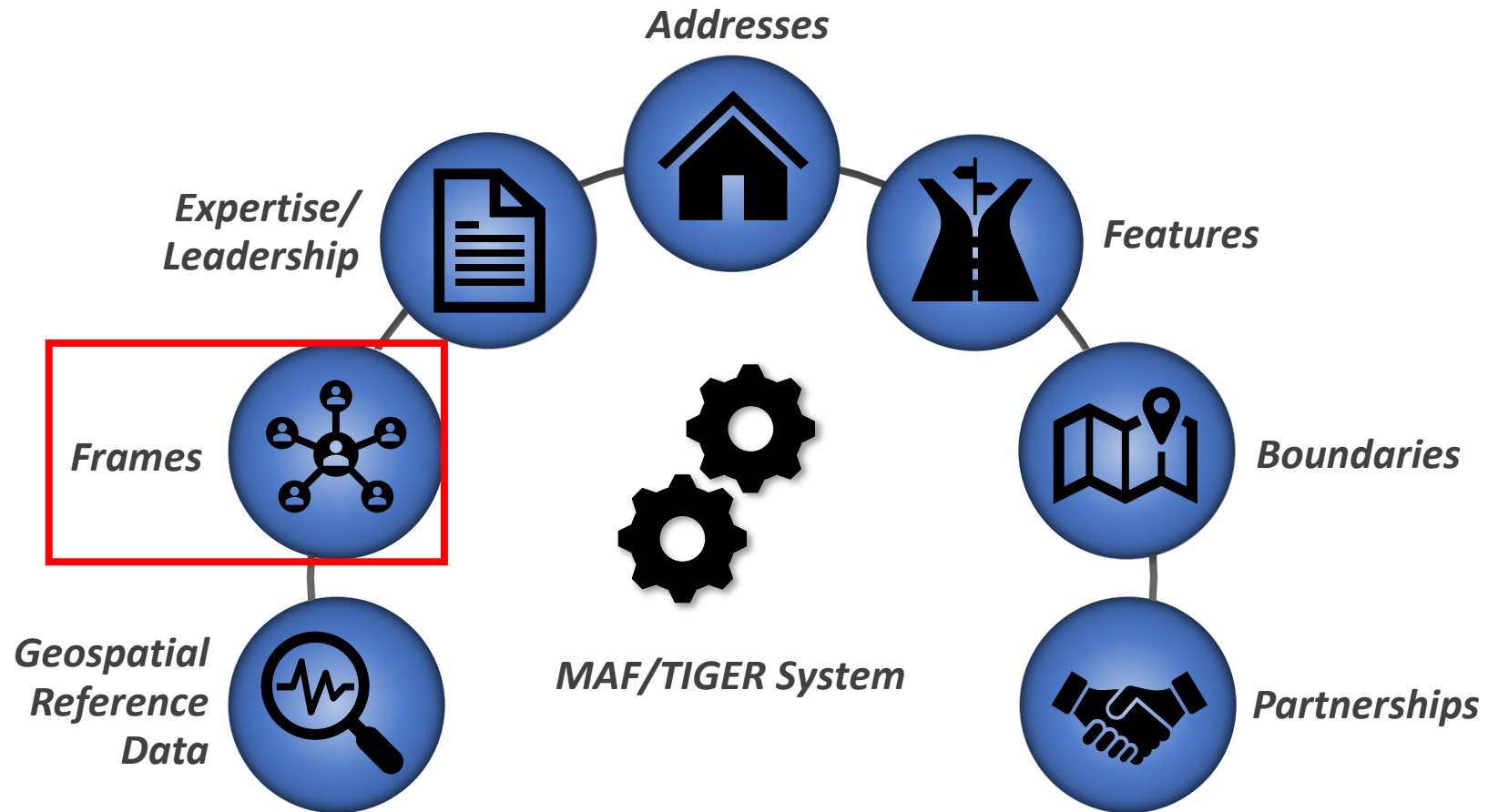


More than 144.5 million housing units



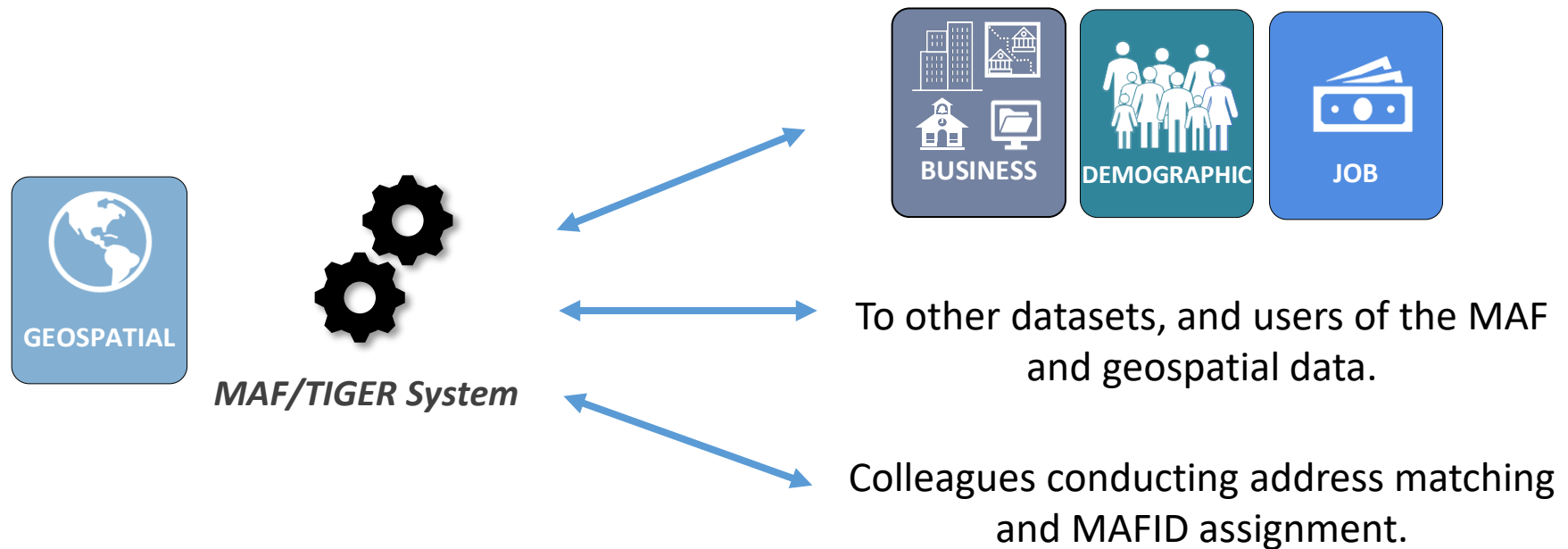
Structure points for approximately 94% of those housing units

Census Bureau's Geographic Support Program and the Frames Program



Geospatial Frame Team's role

Improve and enhance linkages using addresses and geospatial information



Improve and enhance linkages using addresses and geospatial information

Data Linkages

- Business Frame to Geospatial Frame address matching and linking.
- Updating MAF units with enterprise frame information.
- Frame MAFID-based research, evaluation, and collaboration.

Communication and Information Linkages

- Facilitating cross-directorate discussions on address matching and linking to the MAF.
- Developing materials to inform the best use of MAF data.

Business Frame – Geospatial Frame matching

Business Register-MAF Matching Results

Match Rates, Viable Business Register Addresses

	TOTAL	PERCENT
VIABLE BR ADDRESSES	37,020,000	
VIABLE CITY-STYLE ADDRESSES	34,560,000	93.4
MATCHED	32,120,000	93.0
UNMATCHED	2,440,000	7.0
VIABLE NON-CITY STYLE ADDRESSES	2,460,000	6.6
MATCHED	240,000	9.8
UNMATCHED	2,220,000	90.2

Business Register Addresses by Type of Match

	TOTAL	PERCENT
CITY-STYLE	32,120,000	
EXACT	27,880,000	86.8
PROBABILISTIC	2,640,000	8.2
BSA	1,500,000	4.7
ZIP3	90,000	0.3
NON-CITY STYLE	240,000	
EXACT	230,000	95.8
PROBABILISTIC	10,000	4.2

DRB Approval Number: CBDRB-FY24-0497

Adding enterprise frame information into the MAF

New fields within the MAF (IsBusiness, IsDemographic, IsJob) that will be populated if a MAFID is present on the corresponding Enterprise Frame.

This will allow us to:

- Compare information across the Enterprise Frames.
- Better document mixed-use addresses and gain insights into MAF attributes.
- Learn more about MAFID assignment in the other Enterprise Frames.

Frame MAFID-based research, evaluation, and collaboration



Worked with Demographic Frame Team to get a better understanding of MAFIDs both within and not within the Demographic Frame, and their relationship to 2020 Census housing unit-level population counts and MAF attributes.



Conducting additional research with the Business Frame Team to understand MAFID assignment based on Business Frame (Business Register) attributes, e.g., source, in the Economic Census or other survey universe, payroll, etc.

Communication and Information Linkages

Address matching, MAFID assignment, and MAF information sharing and collaboration



Facilitate regular cross-directorate meetings to:

- Discuss address matching and MAFID assignment experiences and needs.
- Share knowledge about tools and methodology, identify opportunities for collaboration.



Identify tools and methods to enhance MAFID assignment across the Census Bureau

- Goal: Yield more uniform and authoritative MAFID assignment to all datasets.
- MAFIDs often serve as a key to link incoming addressed data to existing data within the Census Bureau and to link Census Bureau datasets to each other.
- Geography is the common denominator for the Enterprise Frames.



Develop materials to inform the best use of MAF data

- MAF data is used in a variety of ways by a variety of individuals throughout the Census Bureau.
- Preparing and disseminating documentation to help users determine the utility of the information in the MAF for their purposes and analyses.

Some address data enhancements

Data Enhancements

- Adding tenure data to the MAF and maintaining up-to-date tenure and seasonal data throughout the decade.
- Conducting research for storing and managing PO box and other non-city style addresses found in other Enterprise Frames and datasets.

For More Information



april.l.avnayim@census.gov



[Census.gov/Frames](https://www.census.gov/frames)

Demographic Frame: An Overview

Jennifer M. Ortman

October 30, 2024

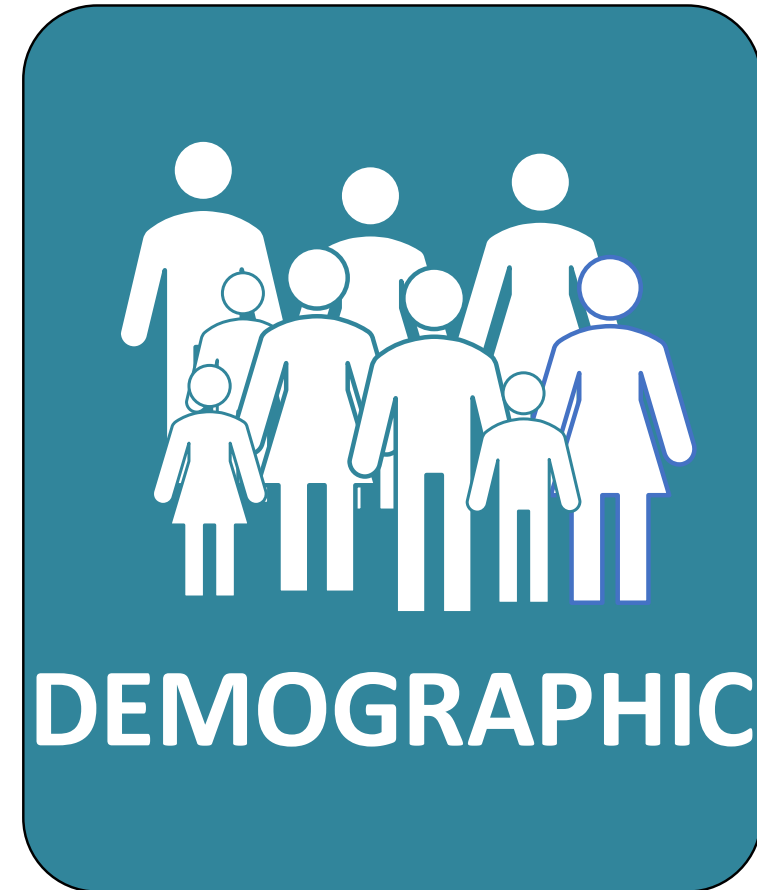
Presentation for the NYS Data Center affiliates Meeting

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-7525174, Disclosure Review Board (DRB) approval number: CBDRB-FY23-POP001-0039).

Any opinions and conclusions expressed herein are those of the author and do not reflect the views of the U.S. 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

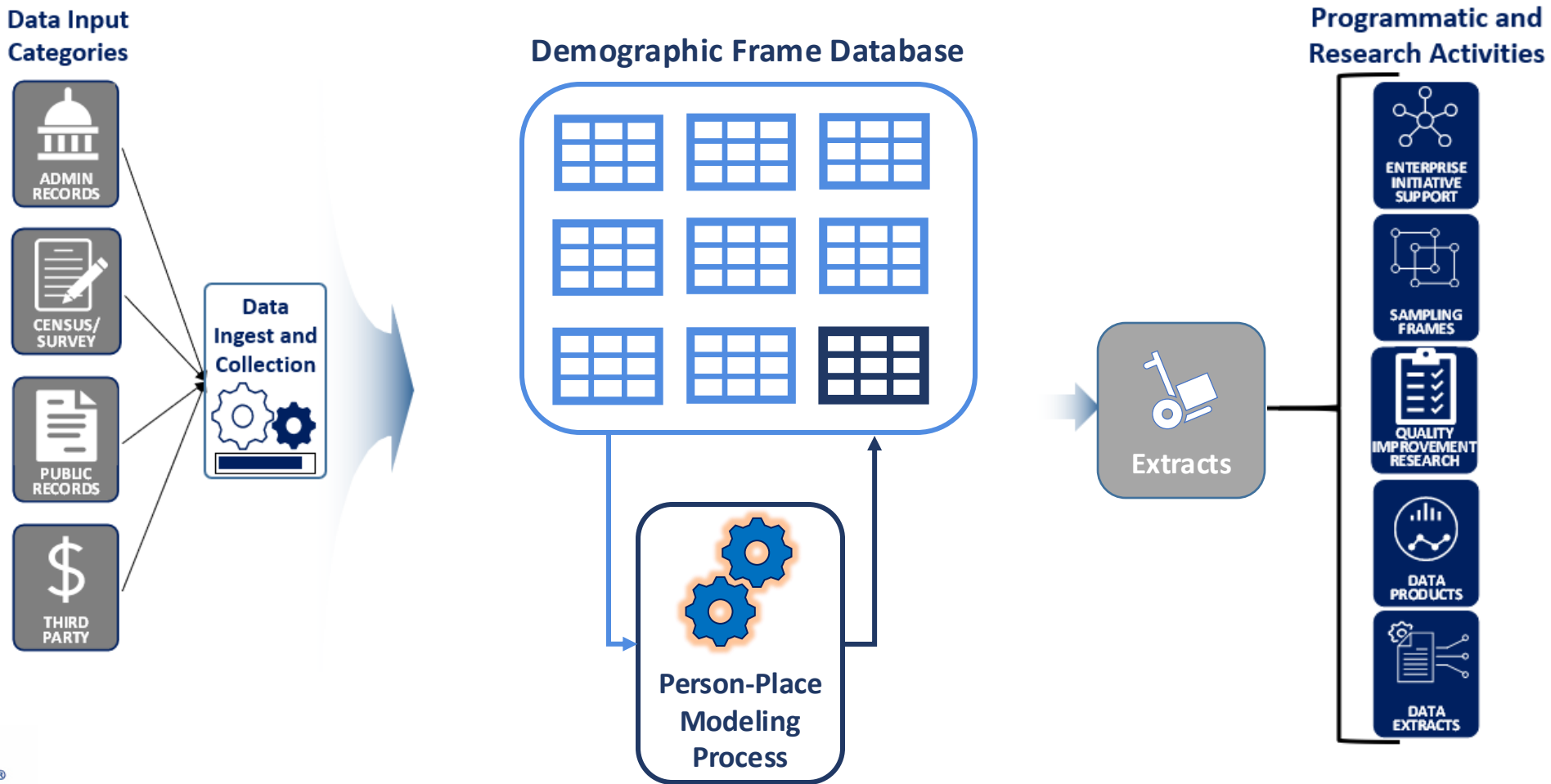


Capabilities under development include:

- 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



Building the Demographic Frame



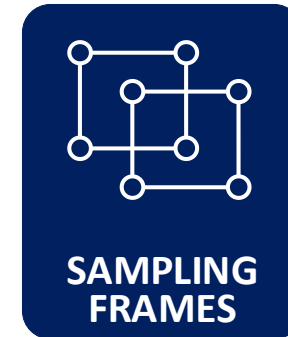
Demographic Frame Extracts

Dates available: 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

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



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)

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

Highlights from Our Evaluations

- 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?

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

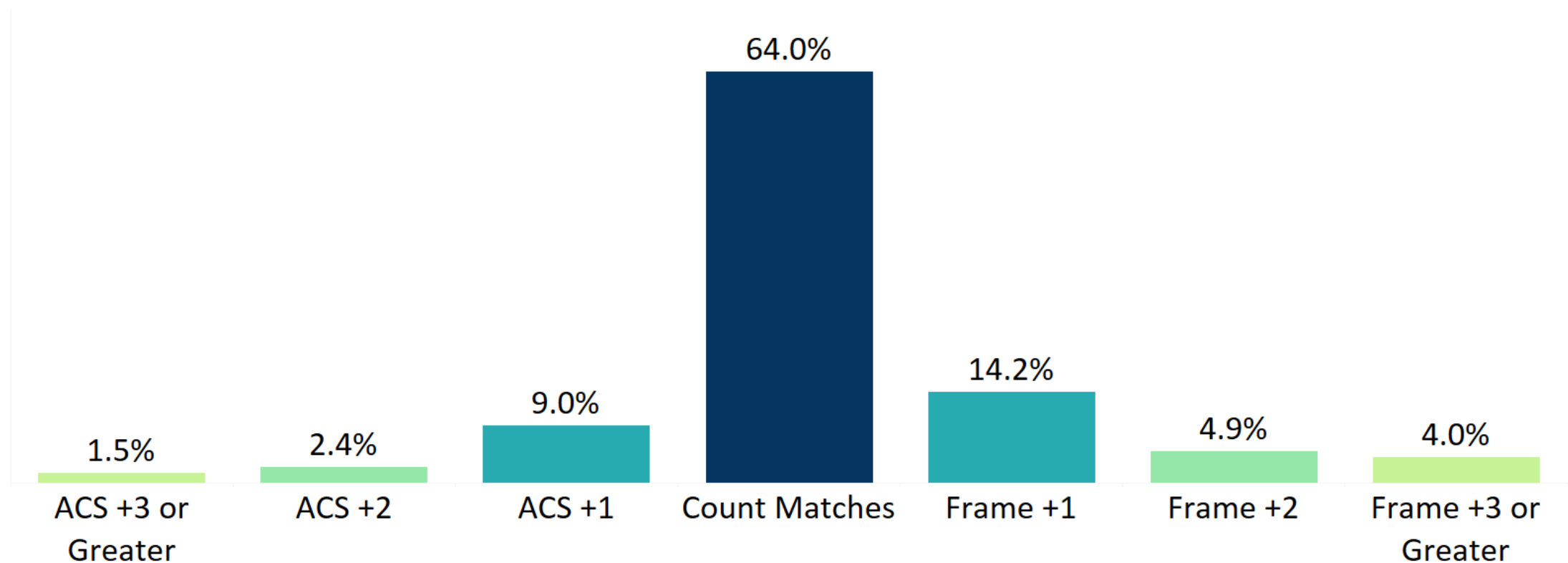
The Demographic Frame is:

2.45% less than the 2020 Census

1.87% less than the 2020 Population Estimates

Geography	Demographic Frame		2020 Census		2020 Population Estimates		Demo Frame Net Coverage Error (%)	
	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
<i>Numbers in thousands and rounded to 4-significant digits</i>								

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



For More Information



jennifer.m.ortman@census.gov



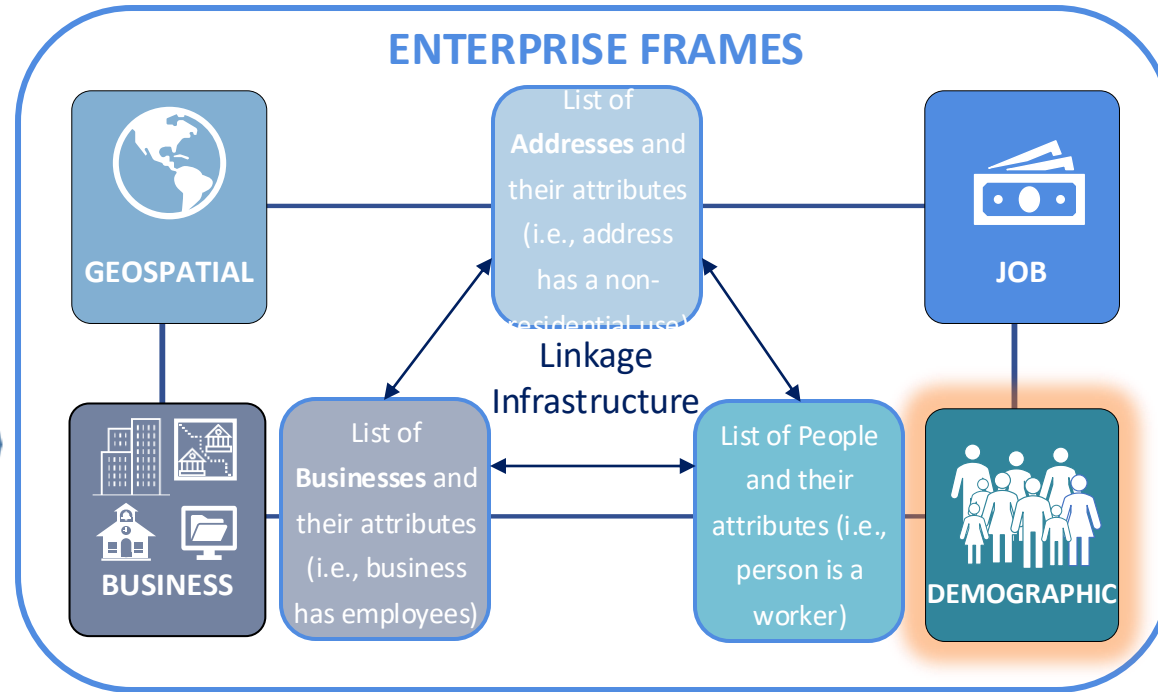
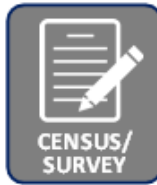
[Census.gov/Frames](https://www.census.gov/frames)

Appendix Slides

Frames Program

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

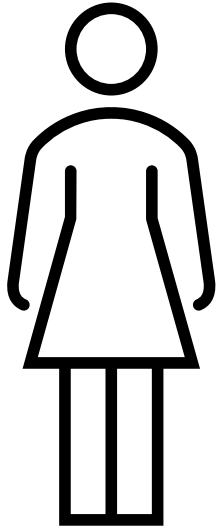
Data Input Categories



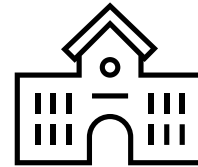
Programmatic and Research Activities



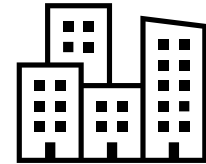
Where does this person live?



THIS HOUSE?



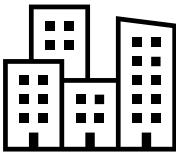
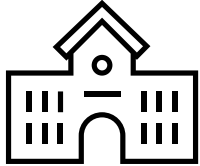
THIS DORM?



THIS APARTMENT?

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

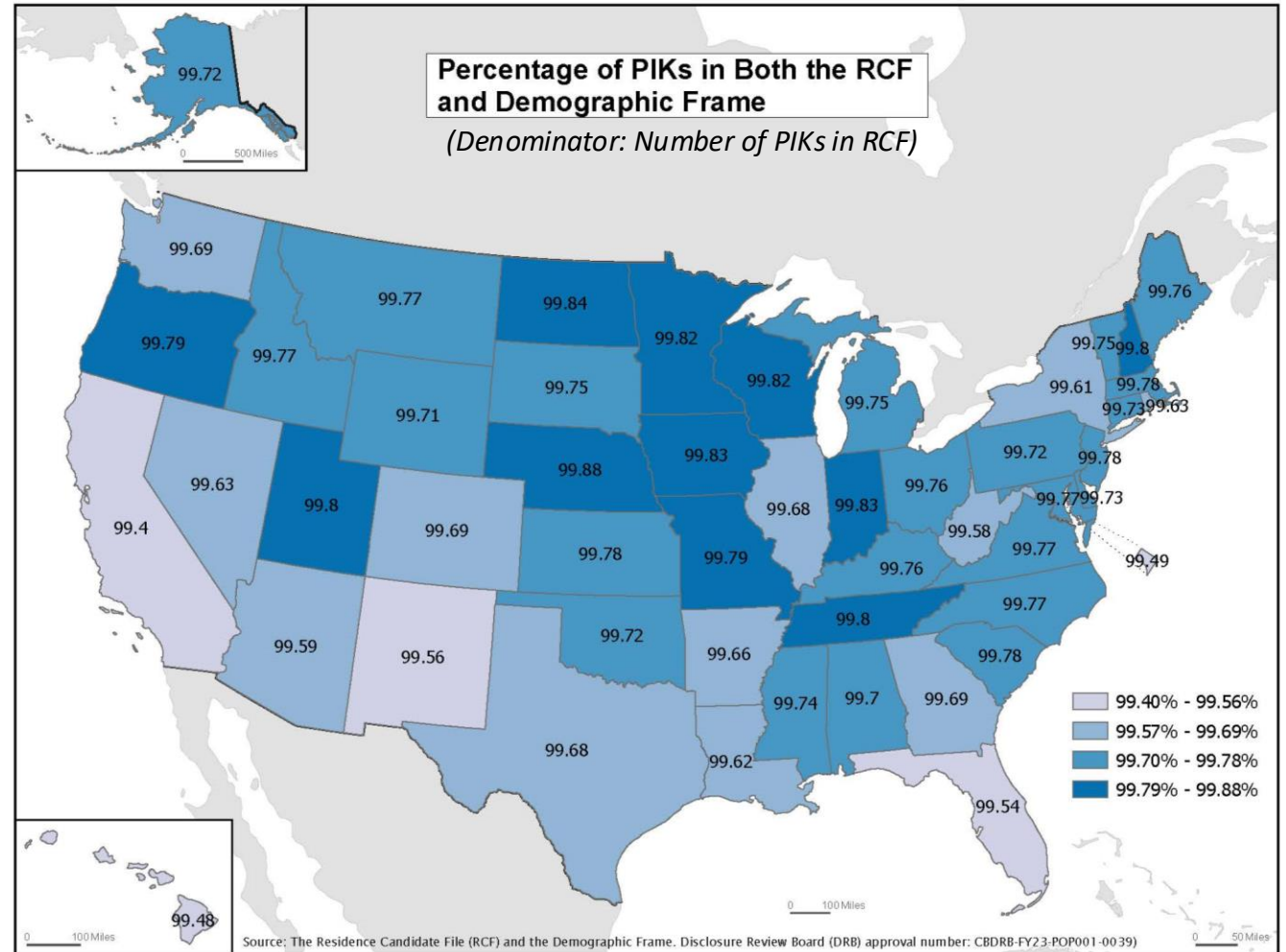
PIK	MAFID	Probability
		0.75
		0.20
		0.05

Additional Evaluation Results

RCF Analysis

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

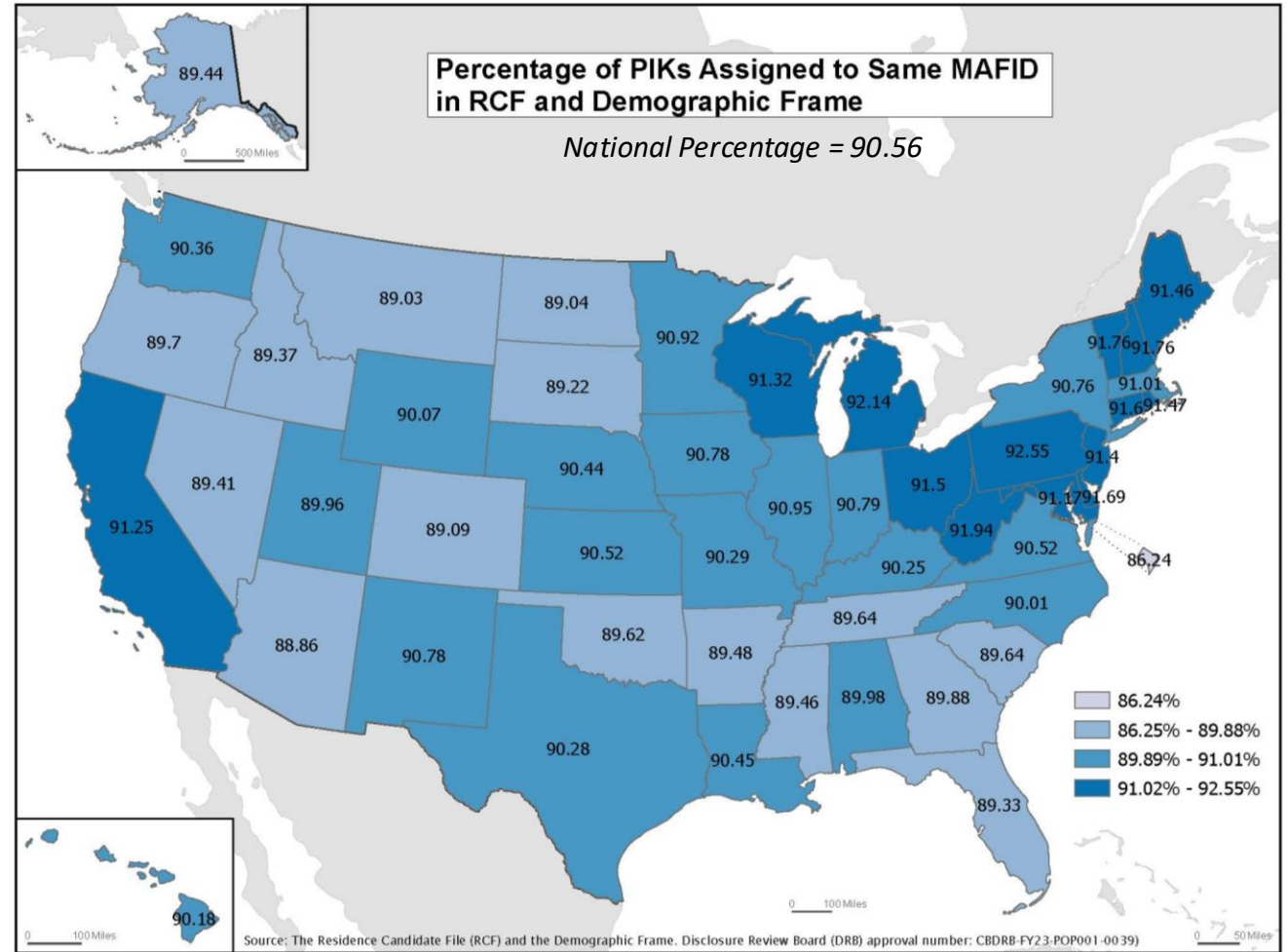
Over **99 percent** of PIKs observed in the RCF were also found in the Demographic Frame at the national level and when looking by state.



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

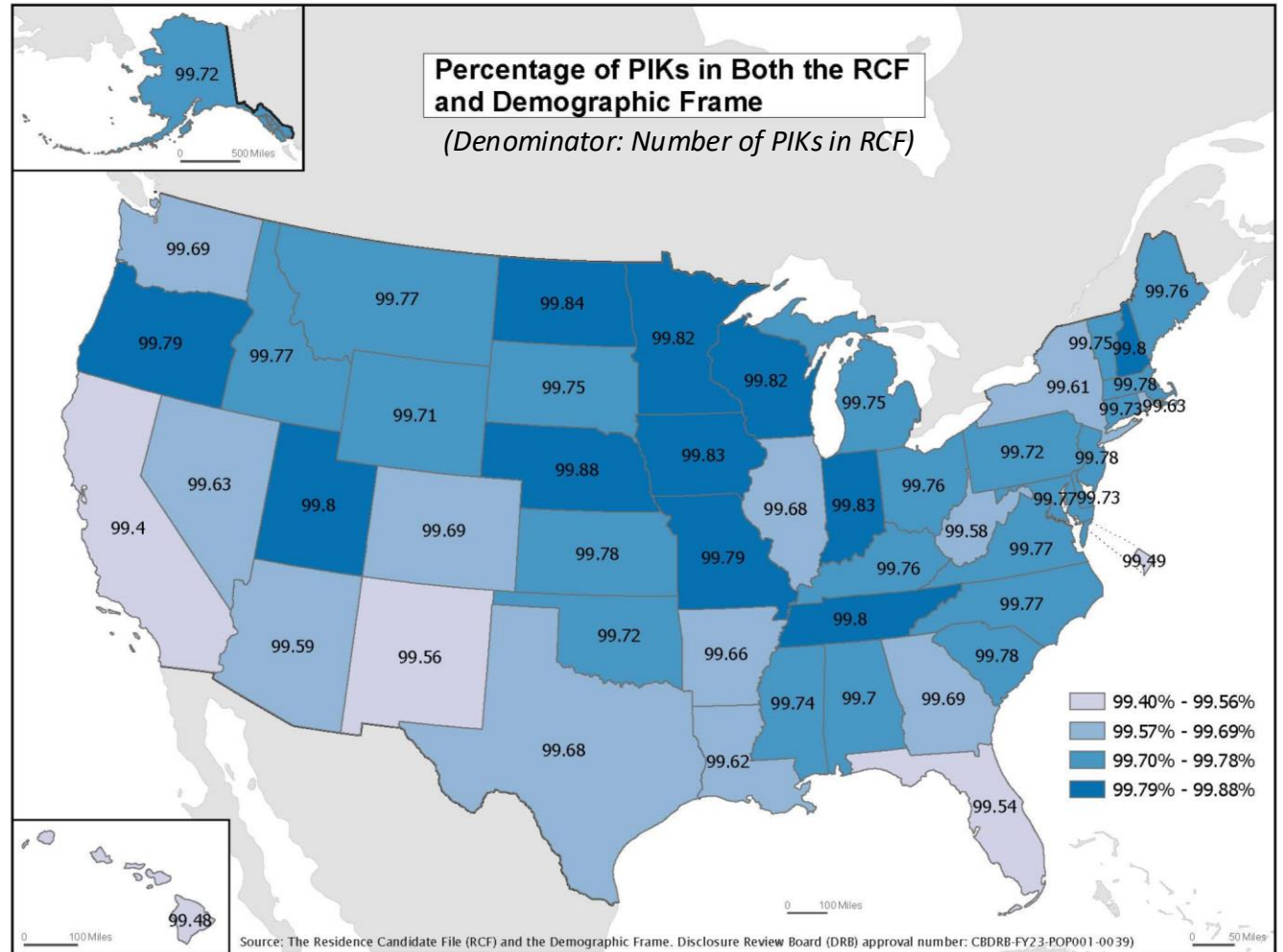
90.56 percent of PIKs were assigned to the same MAFID in the RCF and Demographic Frame.

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



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

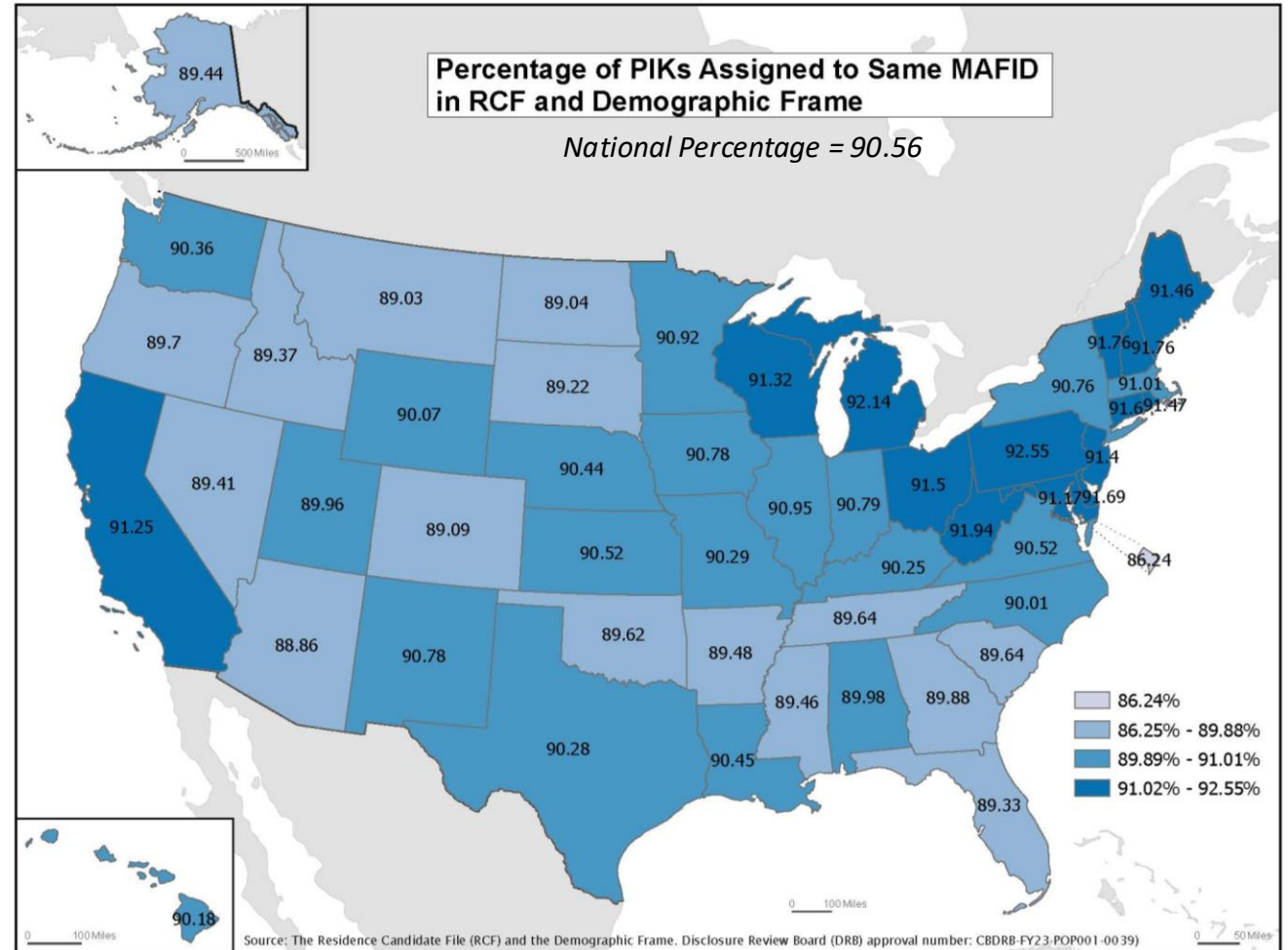
Over **99 percent** of PIKs observed in the RCF were also found in the Demographic Frame at the national level and when looking by state.



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.



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/nescd.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

Nonemployer Business Owners

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

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

Results from Marquette (2023)

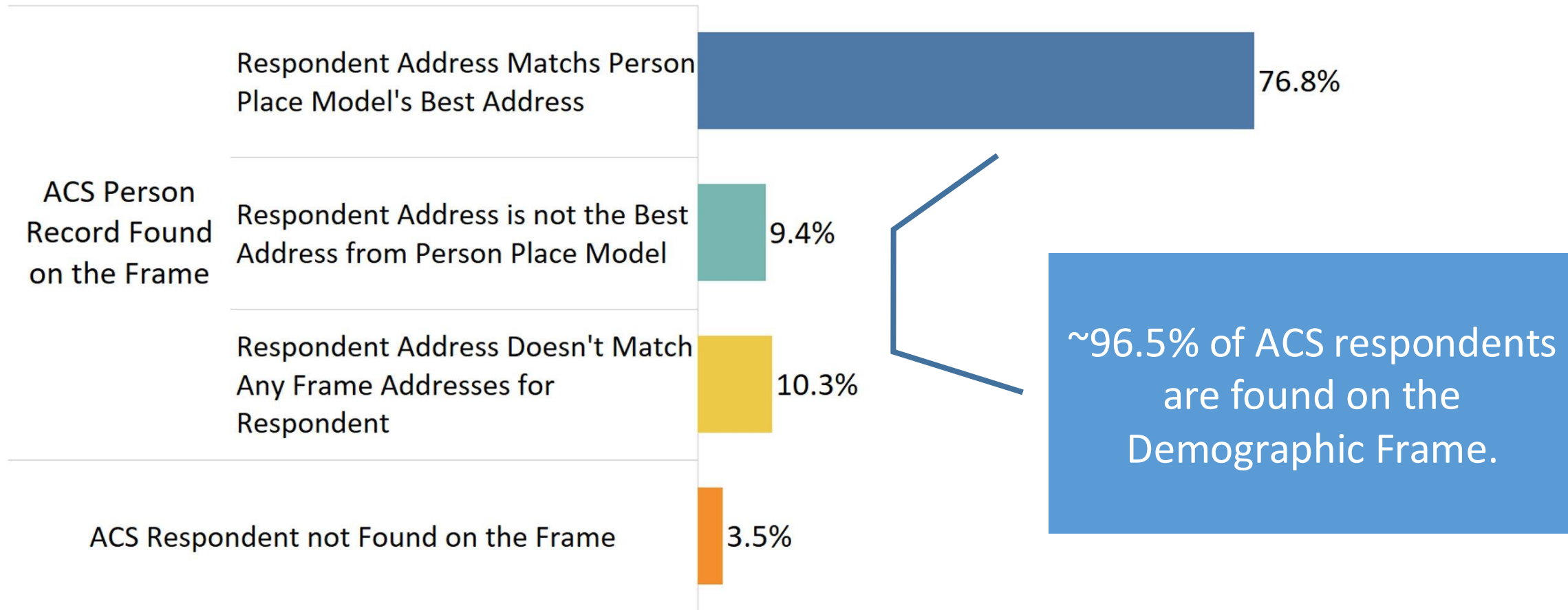
Disclosure Review Board (DRB) approval number: CBDRB-FY23-ESMD002-027

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%

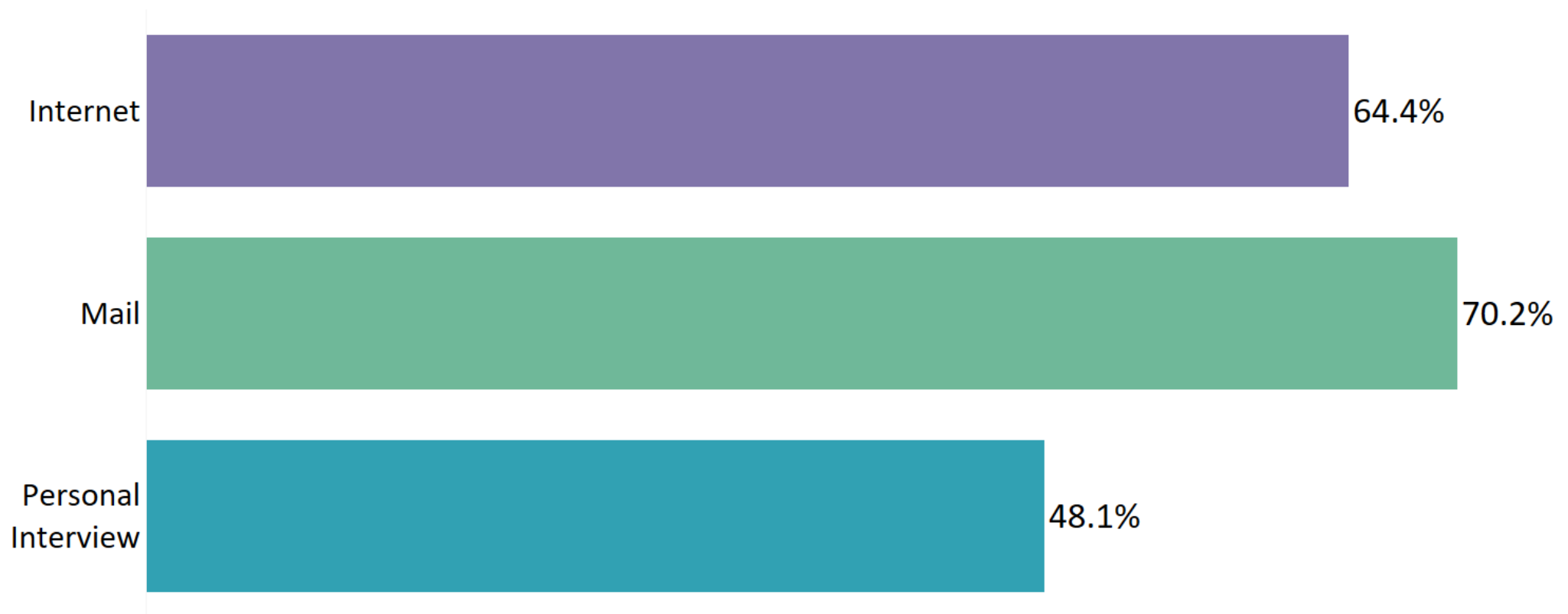
ACS Household Size Evaluation

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



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



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

