

Demographic Frame

Leveraging Person-Level Data to Enhance Census and Survey Taking

Joint Statistical Meetings | August 7, 2023



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).

What You Will Learn Today

- What is the Demographic Frame?
- How are we building the Demographic Frame?
- What does the current extract include?
- How is the Demographic Frame being evaluated?
- What have we learned about the current extract?



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
- Linking information about individuals across time and space



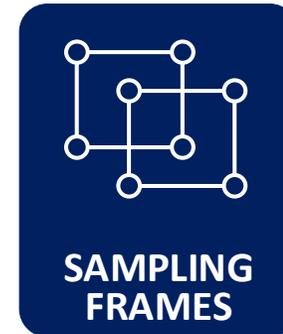
Demographic Frame Extracts

Dates available: April 1, 2020 and July 1, 2021

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



Ongoing Evaluations

Person- and Address-Level Comparisons

- 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

Gap Analysis for the LEHD

Person (PIK) and Address (MAFID)-level comparisons

Gap Analysis for the LEHD

- The Demographic Frame is expected to provide information about the **residential location of individuals** for the LEHD program
 - The LEHD program currently obtains this information from the Residence Candidate File (RCF)
- The Gap Analysis evaluated how well the Demographic Frame overlaps with the RCF through **PIK and MAFID-level comparisons**
Do we find the same people and do we put them at the same address?
- This study is a first step in determining feasibility of replacing the RCF with the Demographic Frame and identifying what improvements may be needed to meet the needs of the LEHD program

Gap Analysis: PIK Coverage

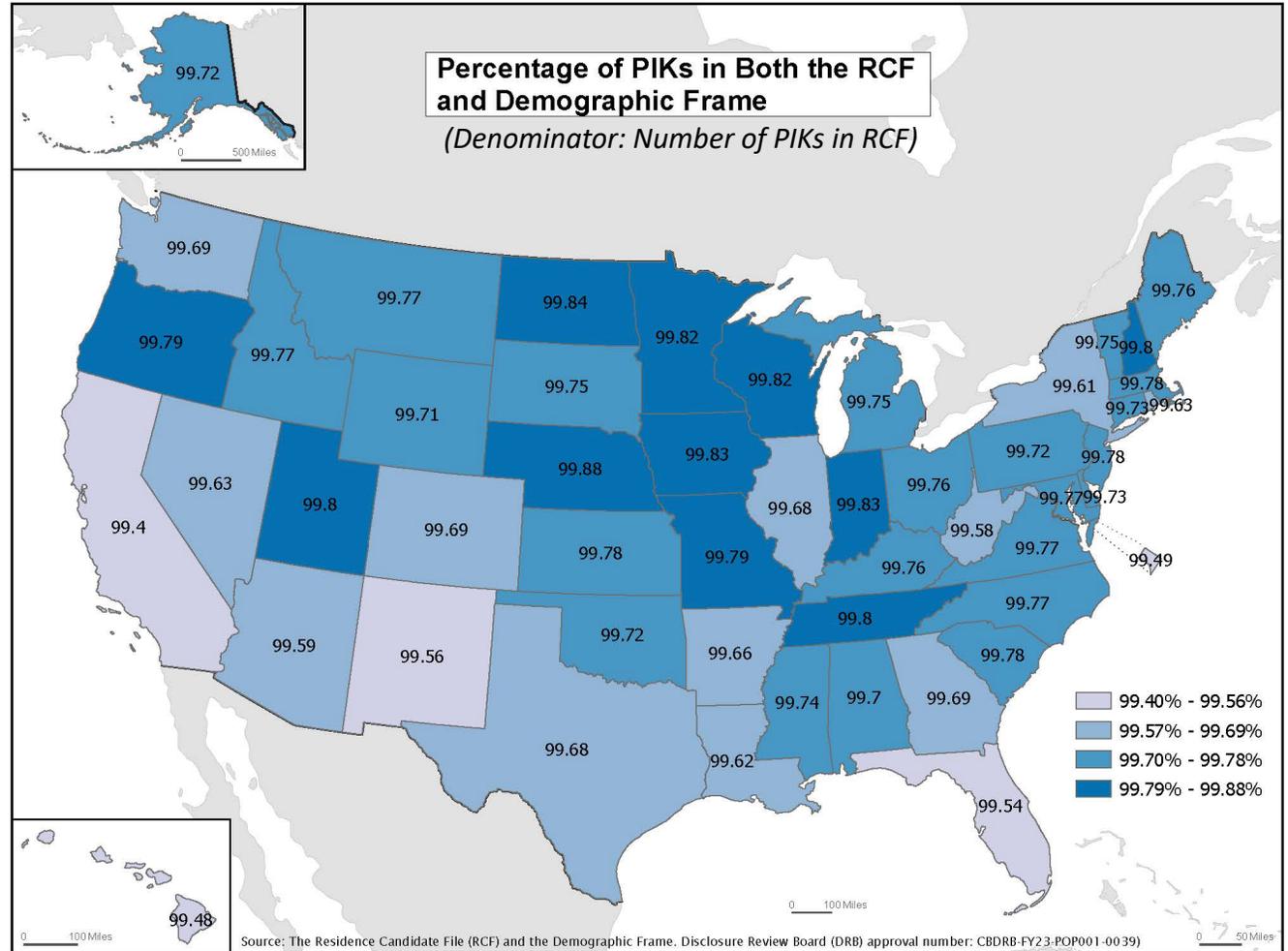
Do we find the same people in the Demographic Frame and Residence Candidate File?

Total Number of PIKs by File

Total in RCF	Total in DF	PIKs in both files	Only in the RCF	Only in the DF
313,100	319,100	311,900	1,112	7,113
<i>Numbers in thousands and rounded to 4-significant digits</i>				

99.6% of PIKs in the RCF were also found in the Demographic Frame

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

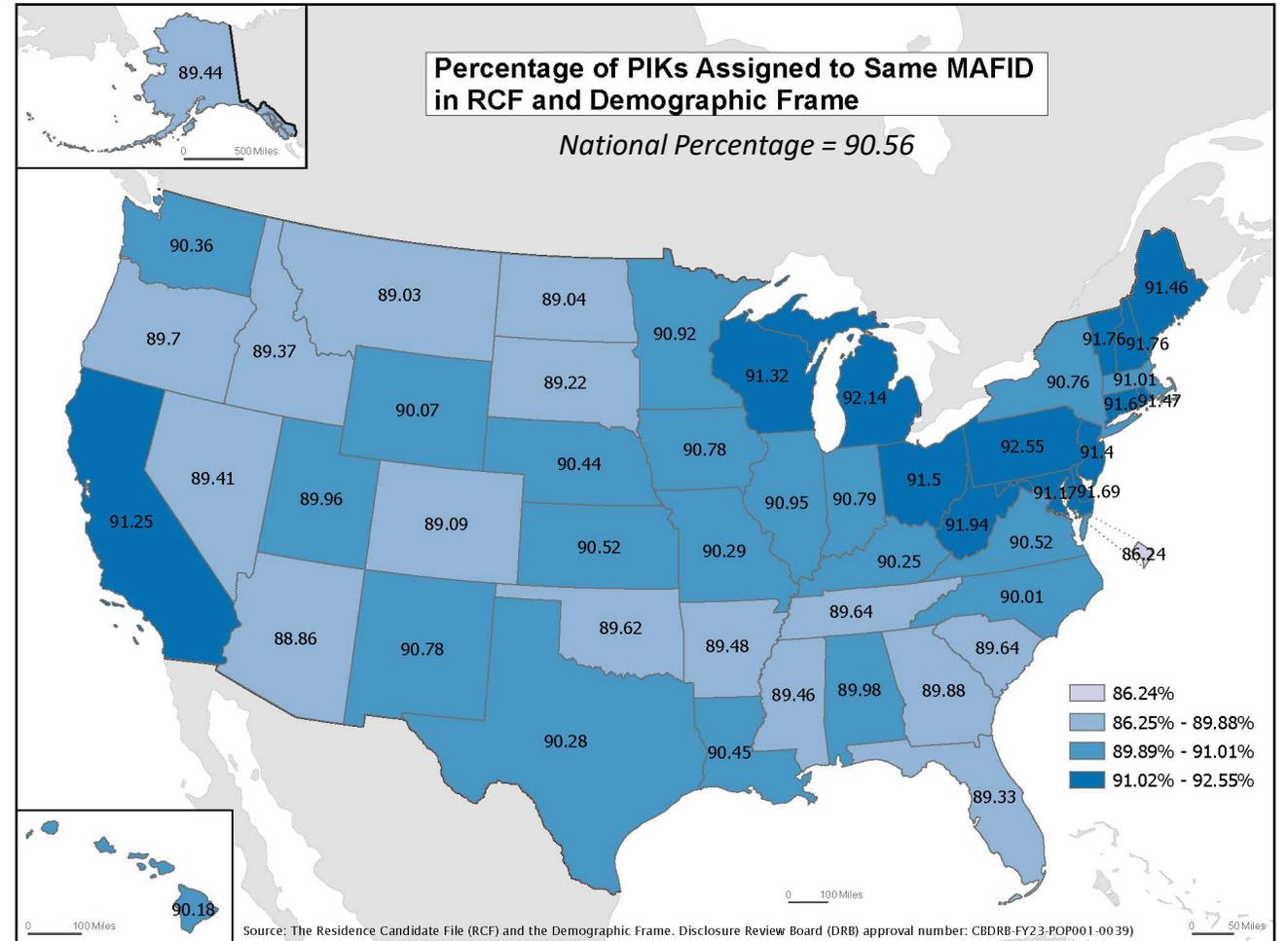


Gap Analysis: MAFID Assignment

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

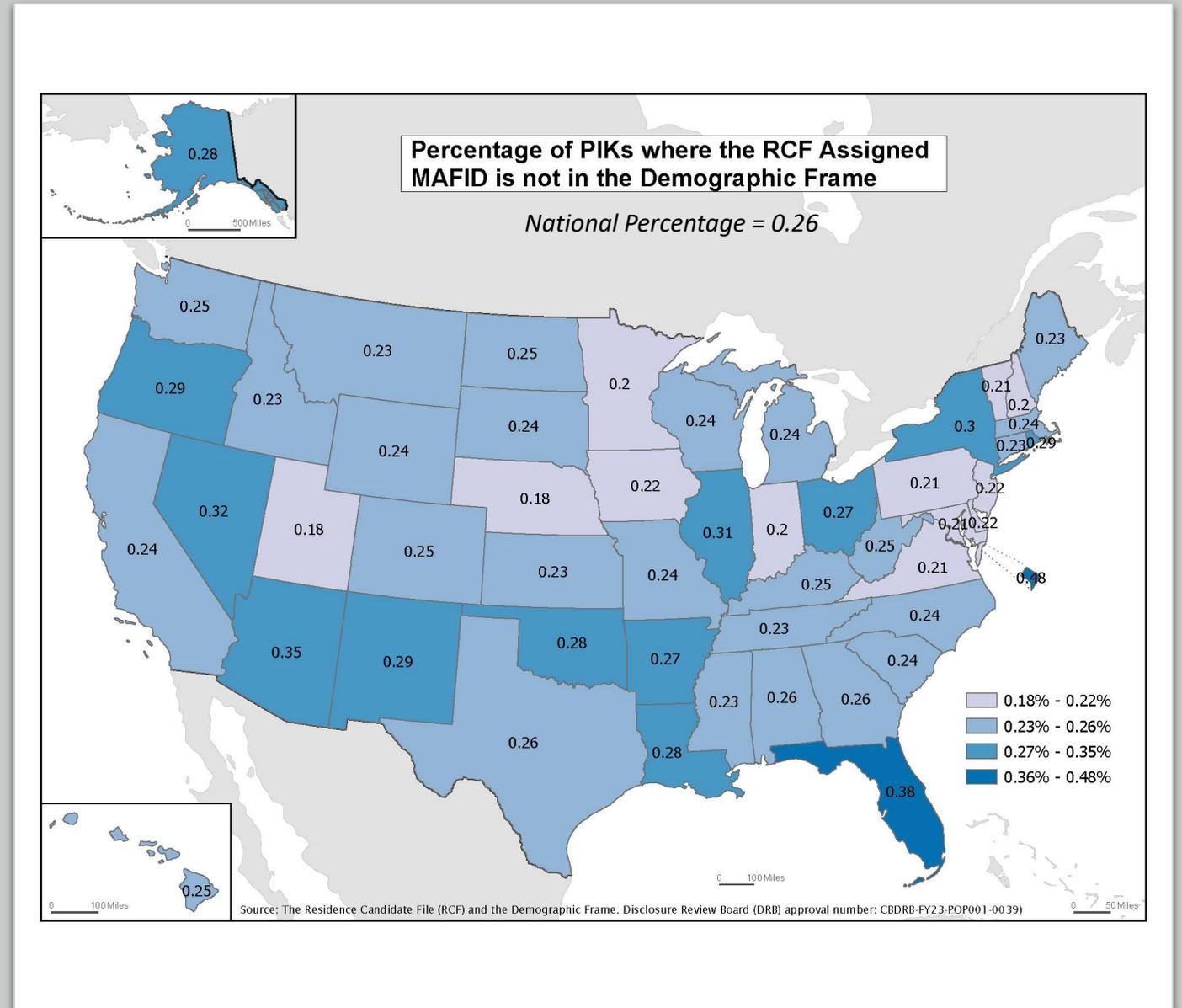
90.56 percent of PIKs were assigned to the same MAFID in the RCF and Demographic Frame.

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



0.26 percent of PIKs were assigned to a different MAFID in the RCF that was not a MAFID associated with that PIK in the Demographic Frame.

This ranges from **0.2 to 0.5 percent** by state.



Benchmarking the FY2022 Demographic Frame Prototype

Aggregate comparisons – number of people by geography

Benchmarking the FY2022 Demographic Frame Prototype

- Compared **aggregate number of people** by geography to assess the quality of the Demographic Frame – in terms of **coverage and accuracy** – relative to two benchmark data sources:
 - 2020 Census – April 1, 2020
 - 2020 Population Estimates – July 1, 2020 (2010 Census Base Population)

$$\text{Percent Net Coverage Error} = 100 * \left[\frac{\text{Benchmark} - \text{Demographic Frame}}{\text{Benchmark}} \right]$$

where: *Benchmark* = 2020 Census count or 2020 Population Estimate

(+) → “undercoverage” (-) → “overcoverage”

Interpretation: relative amount that the Demographic Frame is less (or more) than the benchmark

Percent Net Coverage Error of the Demographic Frame by Geography

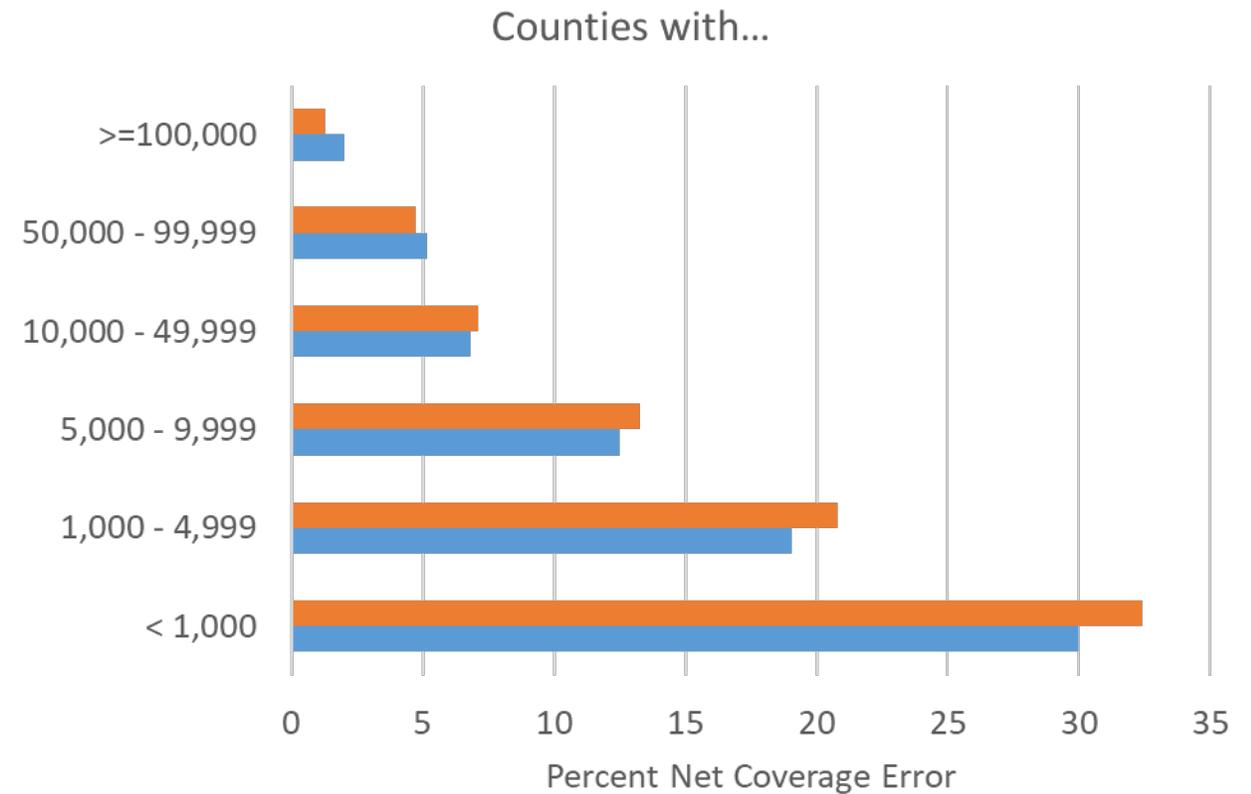
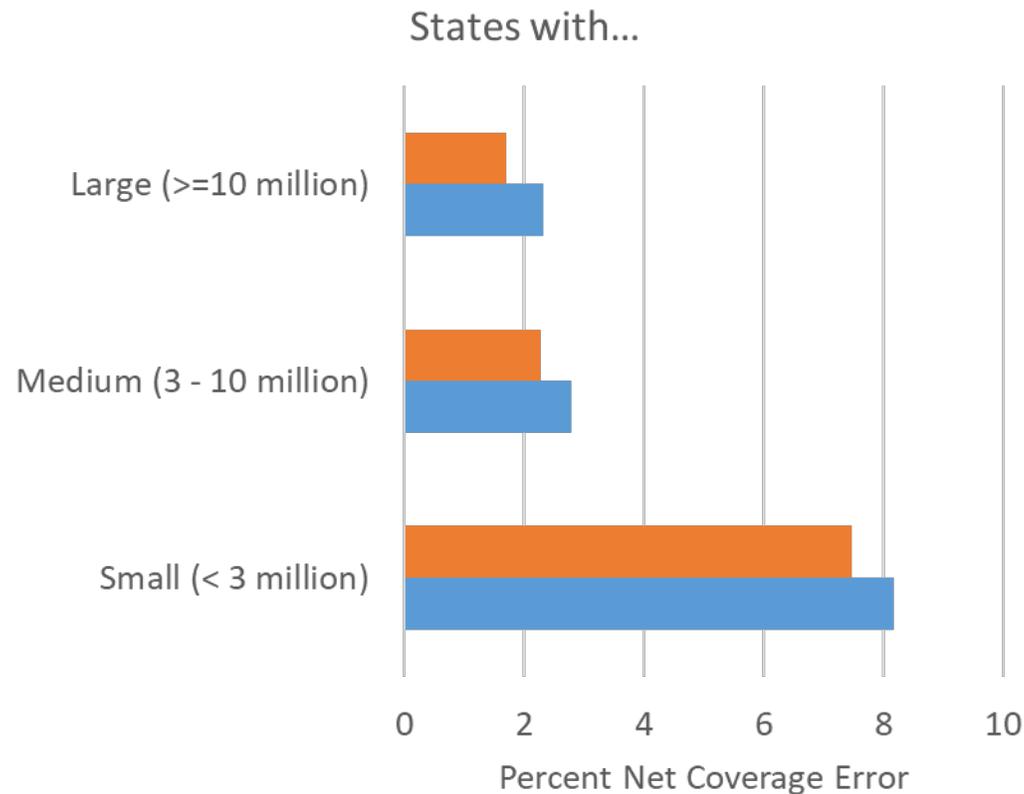
Geography	Demographic Frame		2020 Census		2020 Population Estimates		Demo Frame Net Coverage Error (%)	
	Number	Percent	Number	Percent	Number	Percent	Relative to Census	Relative to Pop Est.
United States	323,300	100.0	331,500	100.0	329,500	100.0	2.45	1.87
<i>Numbers in thousands and rounded to 4-significant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

The Demographic Frame is:

2.45% less than the 2020 Census

1.87% less than the 2020 Population Estimates

Percent Net Coverage of the Demographic Frame by Population Size



For More Information



jennifer.m.ortman@census.gov

Appendix

Frames Program

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

Data Input Categories



ADMIN RECORDS



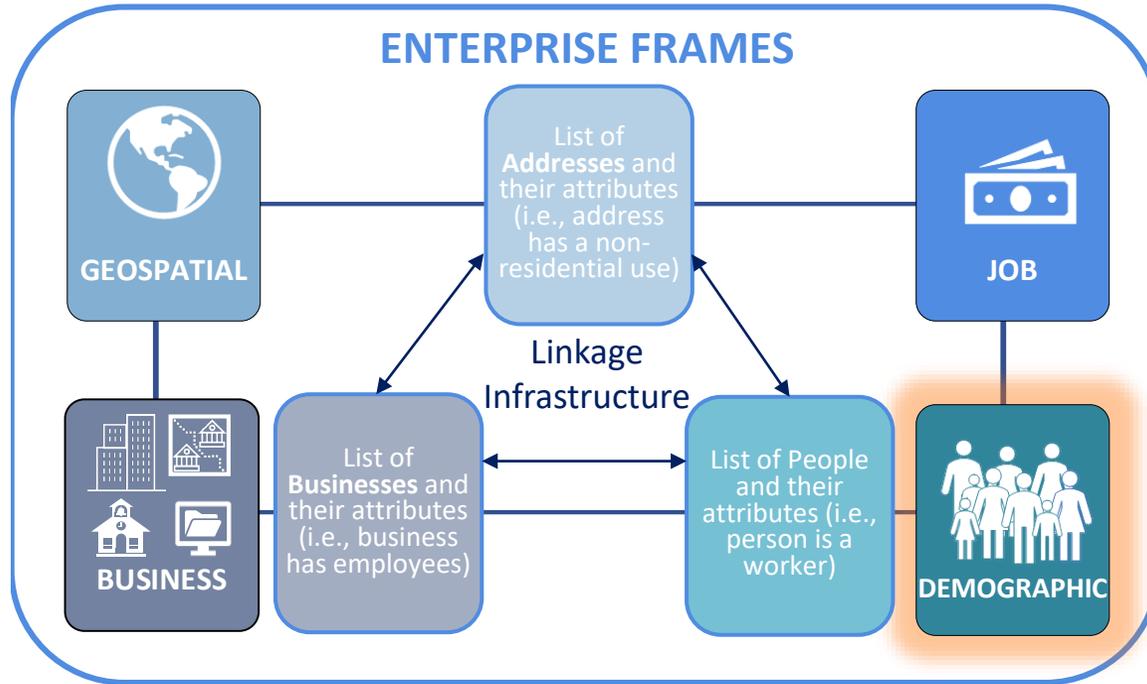
CENSUS/SURVEY



PUBLIC RECORDS



THIRD PARTY



Programmatic and Research Activities



PRODUCTS/SERVICES



ENTERPRISE INITIATIVE SUPPORT



SAMPLING FRAMES



QUALITY IMPROVEMENT RESEARCH

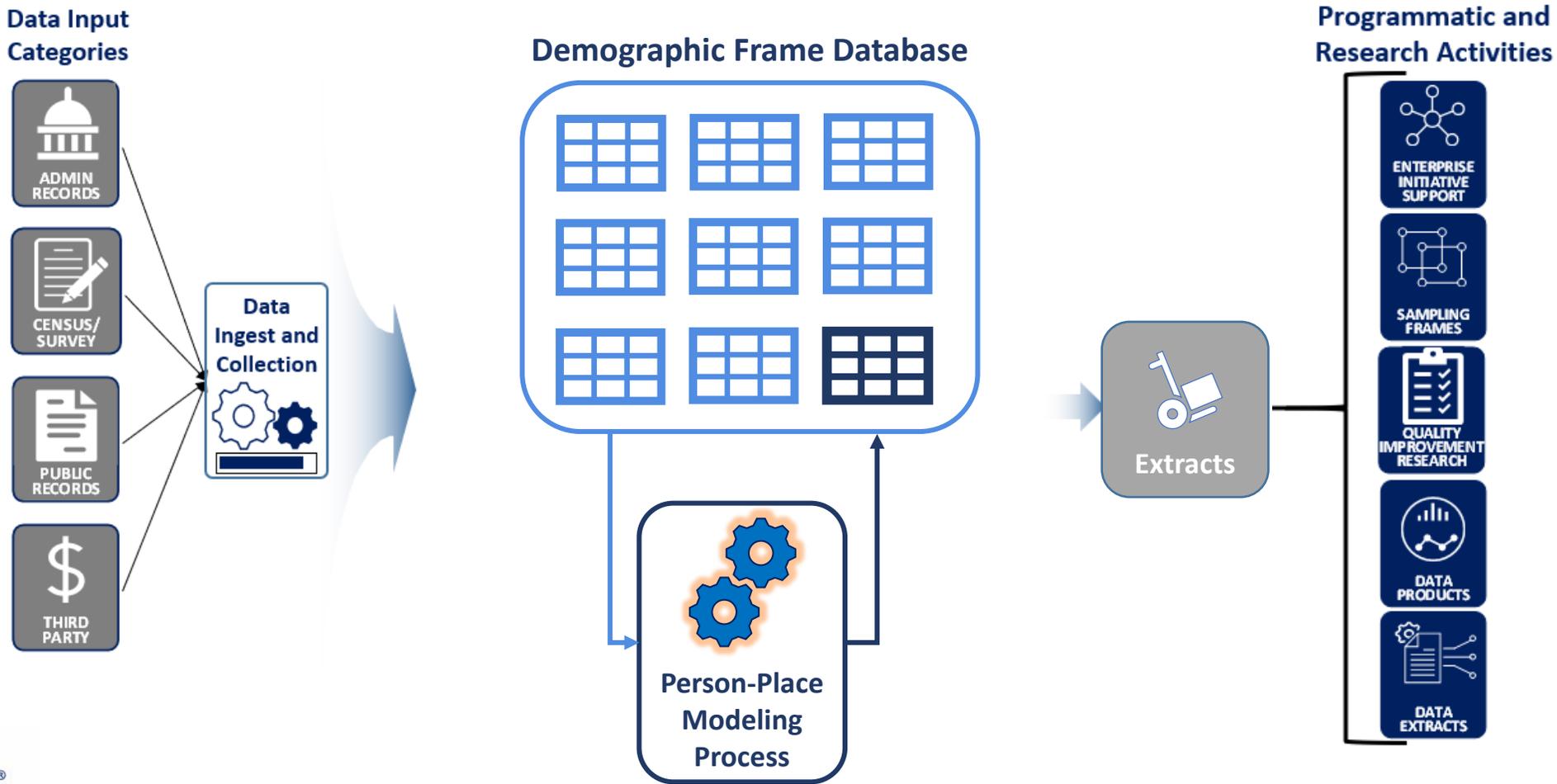


DATA PRODUCTS



DATA EXTRACTS

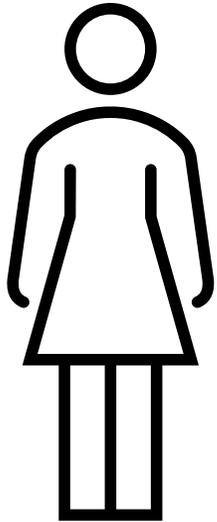
Building the Demographic Frame



2023 Demographic Frame Prototype Inputs

American Community Survey (ACS)	Master Address File (MAF)	IRS (1040, 1099, 1099-R)	USPS National Change of Address (NCOA)	CMS Medicare Database
CMS TMSIS file (Medicaid/CHIP)	Selective Service System (SSS) Registration Data	VSGI Consumer Reference Data	HUD (FHA, PIC & TRAC)	Census Household Composition Key File (CHCK)
Modified Best Race/Ethnicity File	Census Numident	SSA Master Beneficiary Record (MBR)	SSA Supplemental Security Record (SSR)	State SNAP, TANF, and WIC files*

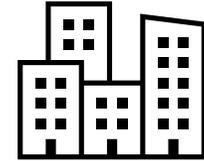
Where does this person live?



THIS HOUSE?



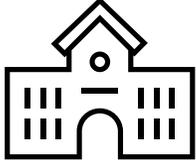
THIS DORM?



THIS APARTMENT?

The Person-Place Model (PPM)

- A process for assigning probabilities to PIK-MAFID 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

Current and Prospective Demographic Frame Customers



- 2030 Census In-Office Enumeration
- Continuous Count Study
- Self-Response Quality Assurance (SRQA) for Special Census
- American Community Survey (ACS)
- Census Household Panel Survey
- Non-Employer Statistics – Demographics (NES-D)
- Longitudinal Employer-Household Dynamics (LEHD)

Resources

[Real-Time 2020 Administrative Record Census Simulation](#)

The Demographic Frame is leveraging methodology from the 2020 Administrative Record Census (Adrec Census) simulation. This report describes the methodology used to produce the Adrec Census along with findings and conclusions.

[Developing a Residence Candidate File for Use With Employer-Employee Matched Data \(census.gov\)](#)

The Demographic Frame is leveraging methodology from the Residence Candidate File (RCF) to generate modeled probabilities that each address associated with a person is their current address on a specified reference date. This report describes the methodology used to produce the RCF.