

An Introduction to the U.S. Census Post-Enumeration Survey and Applications for Machine Learning

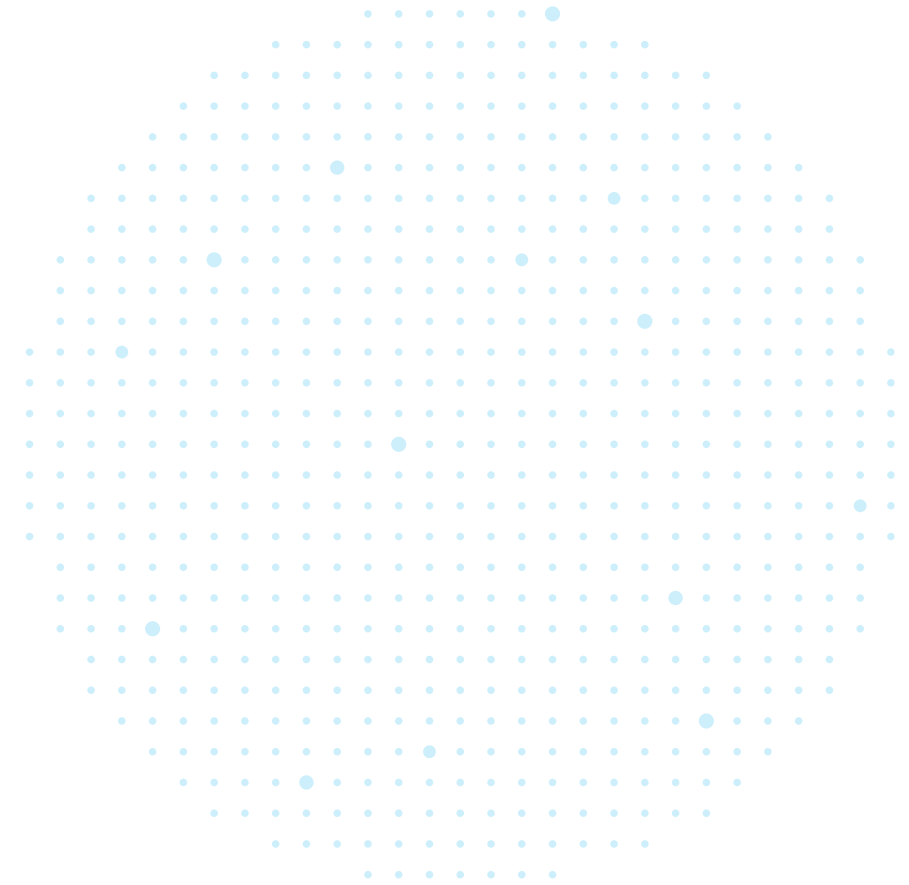
Presented by Shadana Myers, U.S. Census Bureau
AAPOR Conference
May 17, 2024

Disclaimer

- Any opinions and conclusions expressed herein are those of the author and do not reflect the views of the U.S. Census Bureau.
- This presentation does not contain sensitive information including Title 13, Title 26, Title 5, other controlled unclassified information, or administratively restricted information.

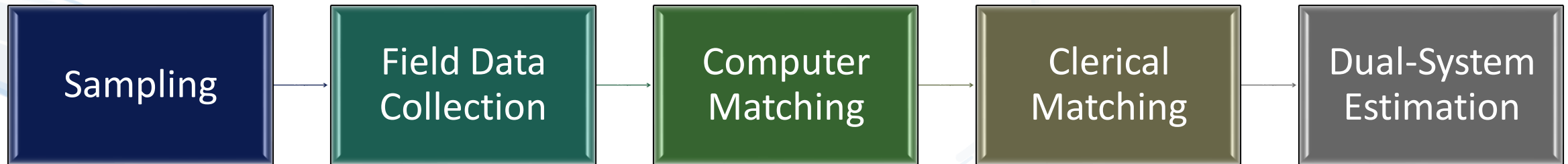
Outline

- What is the Post-Enumeration Survey (PES)?
- Census and PES
- Motivation
- Research Study Aims
- Applications for Machine Learning



What is the Post-Enumeration Survey (PES)?

- A single survey that conducts similar operations to the Decennial Census and is used to measure coverage of the Census. The PES operations are independent of Census and performed on a sampled population.
- We use this information to estimate how many people are in the U.S. and Puerto Rico, and how many people were correctly counted in the census.
- Major operations in the PES:



Census and PES

| | DECENNIAL CENSUS | POST-ENUMERATION SURVEY (PES) |
|-------------------------|--|--|
| First occurrence | 1790 | 1950 |
| Size | Entire Population | Sampled Areas |
| Living Quarters | All addresses in U.S. and P.R., including group quarters and Remote Alaska | Housing Units in U.S. and P.R., excluding group quarters and Remote Alaska |
| Individuals | All people | People rostered in PES sampled housing units |
| Data Collection Methods | In-office, Internet, Telephone, Field | Field |
| Data Processing | Deduplication, Imputation | Matching, Deduplication, Imputation |
| Data Products | Produces official population counts | Produces dual-system estimates |

[Post-Enumeration Surveys \(census.gov\)](https://www.census.gov)

Motivation

PES 2020 Production Person Matching Operation

- 1 year, 1 month (*from October 7, 2020 – November 3, 2021*)
- 250 clerical matchers (*technicians and analysts*)
- Eleven matching activities (**each clerical activity includes 2 stages*)
 - Automated residence status coding (RSC)
 - Automated geocoding (CGC)
 - Clerical CGC*
 - Clerical RSC*
 - Automated person computer matching
 - Person before followup (BFU) clerical matching*
 - Person after followup (AFU) clerical matching*

Research Study Aims

- Investigate whether utilizing machine learning models would result in making more high-quality computer links to reduce the overall clerical workload
- Determine to what extent can machine learning be used to reproduce the codes assigned by clerical matchers
- Discover which data variables contribute the most to the assignment of classification codes

Applications for Machine Learning in Person Matching Operations

- Clerical Residence Status Coding, in which an individual collected in the 2020 PES is assigned a code to categorize whether they reside at the sampled residence
- Name sufficiency determination, in which an individual's first, middle and last name were categorized as sufficient or insufficient for subsequent person data collection and matching operations
- Person matching process, in which individuals enumerated in the 2020 Decennial Census were matched to individuals collected in the 2020 PES



SHADANA.R.MYERS@CENSUS.GOV



WWW.CENSUS.GOV