

# Administrative Data and the 2030 Census

Amy O'Hara & Beth Jarosz

# Overview

---

- Administrative data sources and methods
- (Brief) history of administrative data use
- Data quality considerations
- Privacy & confidentiality considerations
- Consequences of differential participation

# What is administrative data?

---

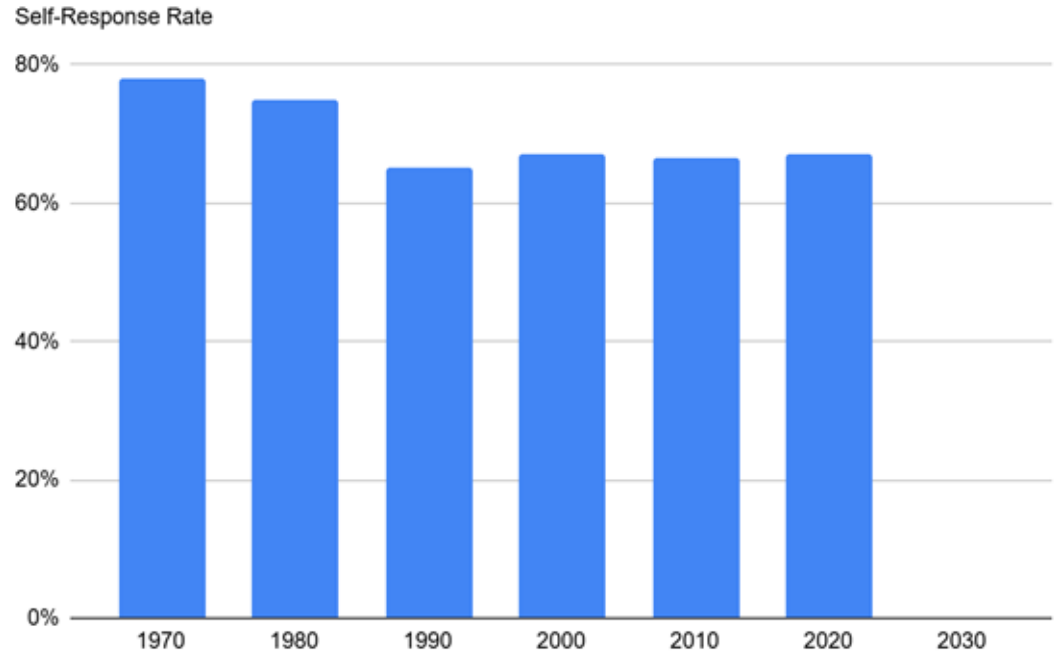
Administrative data is information that is collected and used for operational purposes.

These data are a byproduct of operations, rather than purpose-built for the census.

Data offer a robust way to augment self-response, but with three important considerations: (1) mismatch in data quality relative to census needs, (2) questions about allowable use, (3) consequences of differential participation.

# Background: Why Administrative Records (AdRec)?

- Declining self-response 1970–2000
- AdRecs reduce cost of Nonresponse Follow-Up (NRFU)
- Potentially allow for enumeration without contact attempt(s)
- Cost-reduction & modernization pressures



# (Very Brief) History of Census Operations

---

- 1790 first U.S. census, conducted by U.S. Marshals
- 1880 initial use of trained enumerators to conduct census
- 1920 redistricting failed to pass Congress
- 1960 first “mail-out” census (beginning of widespread self-response)
- 2000 more formal use of postal records
- 2010 transition to “short form”
- 2020 first internet self-response (also wider use of administrative records)

# History of Administrative Data in Census

---

Some sources have a long history of use, including:

- USPS delivery sequence file
- Group quarters administrator lists
- Defense Manpower Data Center

More sources were incorporated into 2020 Census, and in new ways, including:

- IRS
- Medicare Enrollment Database
- Supplemental Nutrition Assistance Programs (SNAP)

Many others have not yet been used, including:

- School enrollment
- Head Start
- Medicaid / CHIP

# How Administrative Data Can Support an Accurate Census

## Frame Development

Before the census, records can help identify the addresses (houses and group quarters facilities) that need to be counted

## Enumeration

Records can help identify locations that are probably vacant, reducing number of contact attempts

Records can help support accurate imputation for non-responding households or facilities

## Evaluation

After the census, records can help understand who may have been missed and those who may have been overcounted

# Administrative Record Data Quality

## Universe and Coverage

Administrative data reflect participation, not the population as a whole.

## Linkage Quality

Messiness in real world data means that records linked together for census use may miss cases (false negatives), or may introduce incorrect linkages (false positives).



## Definitions

Reference date, residence rules, household composition, race/ethnicity, and other details may differ between administrative data and census needs.

## Characteristics

Typically the census requires information about a person including their age, sex, race/ethnicity, relationship to householder, and housing tenure. Many administrative sources lack one or more of these.

# Who's Historically Undercounted

---

Who's historically undercounted?

- Renters
- Young children (under age 5)
- Black, Hispanic/Latino populations
- American Indian/Alaska Native populations living on reservations

Who's historically harder to count? All of the above, plus...

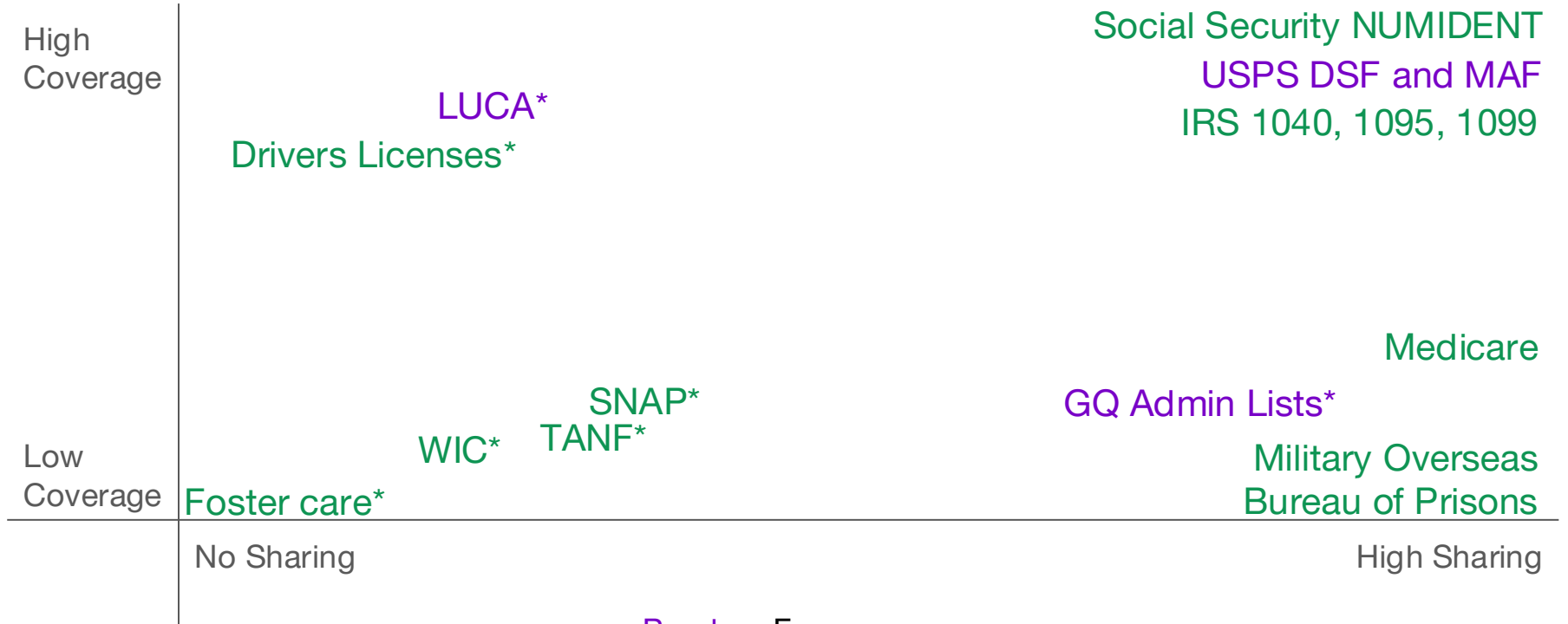
- People who speak languages other than English
- Low-income populations
- Undocumented immigrants
- People with disabilities
- People who do not live in traditional housing

# Key Quality Question: Is It Better Than the Alternative?

---

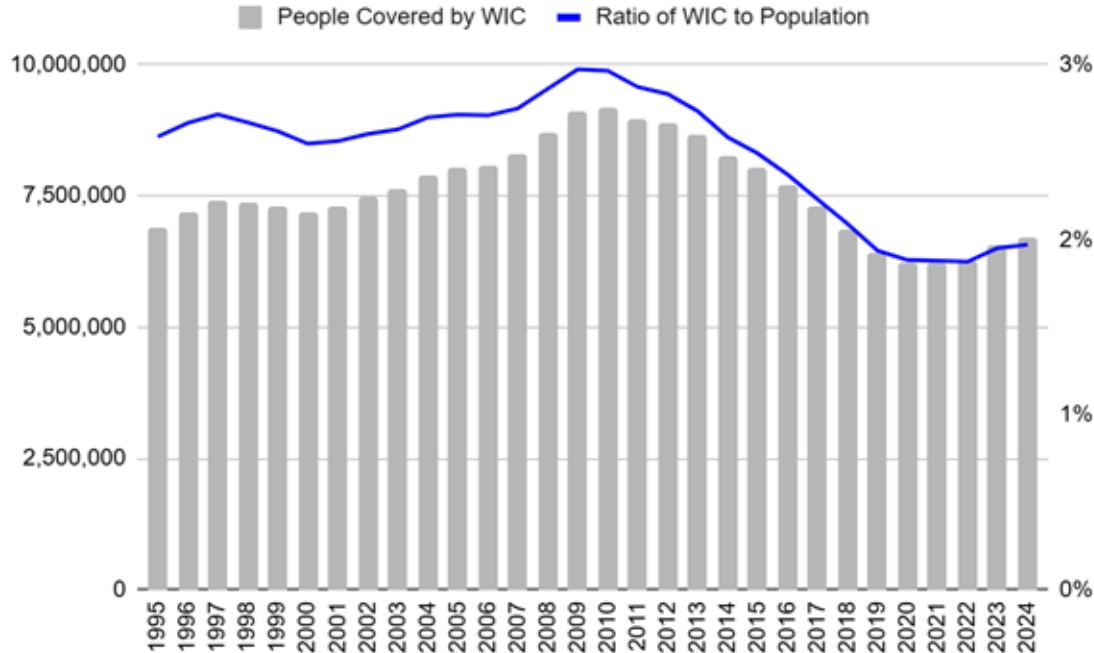


# Differential Sharing and Differential Coverage



Purple = Frame  
Green = Enumeration  
\* = State or Local Data

# Example: WIC Participation (All Ages)



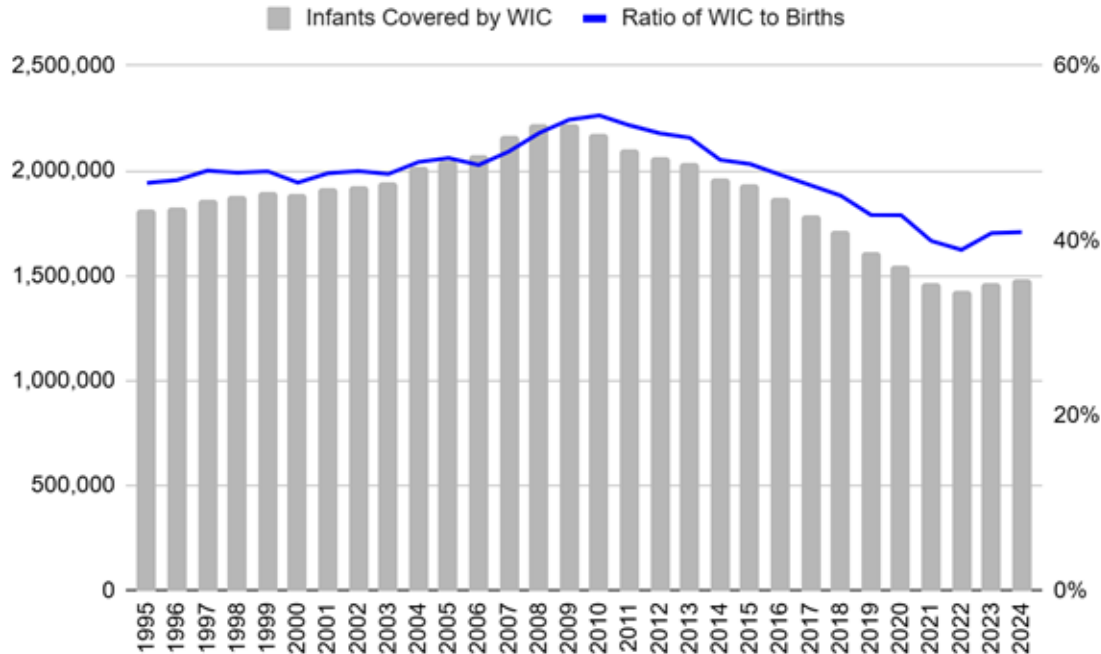
Women, Infants, and Children (WIC) participation changes over time due to economic conditions and policy changes.

As a share of total population, WIC covers less than 3% of people.

And in 2020 Census only 14 states shared their data with the Census Bureau.

Calculations based on USDA Economic Research Service, WIC Program; U.S. Census Bureau, Population Estimates

# Example: WIC Participation (Infants)



Calculations based on USDA Economic Research Service, WIC Program; CDC WONDER, Natality online databases

Young children are one of the most undercounted demographic groups.

WIC covers 40-50% of infants.

WIC records could help reduce the persistent undercount of young children.

But in 2020 Census only 14 states shared WIC data with USCB. Concern about privacy and confidentiality is a barrier.

# Data Privacy, Confidentiality, and Security

---

**Privacy:** Ask only about what's needed

**Confidentiality:** Protect all data collected or acquired, ensuring that data products prevent re-identification

**Security:** Protect from corruption, thefts, and unauthorized access

# Privacy

---

Privacy considerations for census and census-related data include

- Evaluating necessity
- Openness/transparency about how the data will be used
- Respectful treatment of respondents
- Confidentiality

# Confidentiality

---

Confidentiality protections include limits on access to data internally at the Census Bureau and limits on published data

- Historically used aggregation and data suppression to protect confidentiality
- “Differential privacy” applied to published tables in 2020 to reduce the possibility of reconstruction and reidentification

# Security (briefly)

---

Data security protections include

- Access controls
- Encryption
- Physical security
- Digital security, threat monitoring, detection, and response
- Backup and recovery

# “5 Safes” Bridges Data Privacy, Confidentiality, Security

Safe Data	Data is treated to protect any confidentiality concerns	Is there a disclosure risk in the data itself?
Safe Projects	Research projects are approved by data owners for the public good	Is this use of data appropriate, lawful, ethical and sensible?
Safe People	Those with access are trained and authorized to use data safely	Can the user(s) be trusted to use the data appropriately?
Safe Settings	Environment prevents unauthorized use	Does the infrastructure (physical and digital) limit unauthorized access?
Safe Outputs	Screened and approved outputs are not disclosive	Do the published results permit any singling-out of individuals or groups?

Adapted from: [UK Data Service](#) and [fivesafes.org](#)

# Key Points

---

- Administrative data may help ensure an accurate census count, particularly given declining public trust
- Differential sharing across states may result in a higher risk of undercounts in states where data are not shared
- Any administrative data sharing must be paired with robust privacy, confidentiality, and security protections