

Use of Administrative Records and Third-Party Data for Enumeration in the Decennial Censuses of the United States: 1970 to 2020

Elizabeth M. Grieco and Howard Hogan
U.S. Census Bureau

Annual Meeting of the Population Association of America
Chicago, Illinois
April 27-29, 2017

Purpose of Presentation

To describe how the U.S. Census Bureau has used administrative records and third-party data to enumerate the population in the decennial censuses since 1970.

- The focus of this presentation is on the enumeration of population in housing units
- It does not address the enumeration of group quarters or how administrative data are used in research

PART I:

Defining Administrative Records and Third-Party Data

What are administrative records and third-party data and what are some examples?

What Are Administrative Records and Third-Party Data?

Administrative records and third-party data refer to microdata records contained in files collected and maintained by agencies and commercial entities for the purpose of administering programs or providing services.

- Administrative records data – information collected by government agencies for the purpose of administering programs or providing services
- Private, or commercial, data – information collected by third parties that may be acquired by the Census Bureau

Note: The Census Bureau has no control over the information included in the administrative data provided to it by agencies and commercial entities. Changes in the law or economic conditions, program changes, etc., can affect data system content.

Federal, State, and Third-Party Administrative Data: Examples

Federal Administrative Data

- Internal Revenue Service (**IRS**) Individual Tax Returns (Form 1040) and Information Returns (Form 1099)
- Social Security Administration (**SSA**) Numident
- Centers for **Medicare and Medicaid** Services (CMS) Medicare Enrollment Database and Medicaid Statistical Information System
- **Indian Health Service** (IHS) Patient Registration File
- U.S. Postal Service (**USPS**) Delivery Sequence File, National Change of Address File, and Undeliverable As Addressed File
- Housing and Urban Development (**HUD**) Computerized Homes Underwriting Management System, Assisted Renters Public and Indian Housing Center Database, and Tenant Rental Assistance Certification System Database
- **Selective Service System** Registration File

State Administrative Data

- U.S. Department of Agriculture (USDA) Food and Nutrition Service Supplemental Nutrition Assistance Program (**SNAP**) and Women, Infants and Children (**WIC**)
- Department of Health and Human Services (DHHS) Temporary Assistance to Needy Families (**TANF**)

Third-Party Data

- Veterans Service Group of Illinois (**VSGI**) Name and Address Resource Consumer File and Tracker File
- **D.A.R. Partners** Database
- **Targus** Federal Consumer File and Pure Wireless File
- **Experian** End-Dated Records File and In-Source File
- **CoreLogic** Property Tax and Deed Data File

PART II: Historic Context

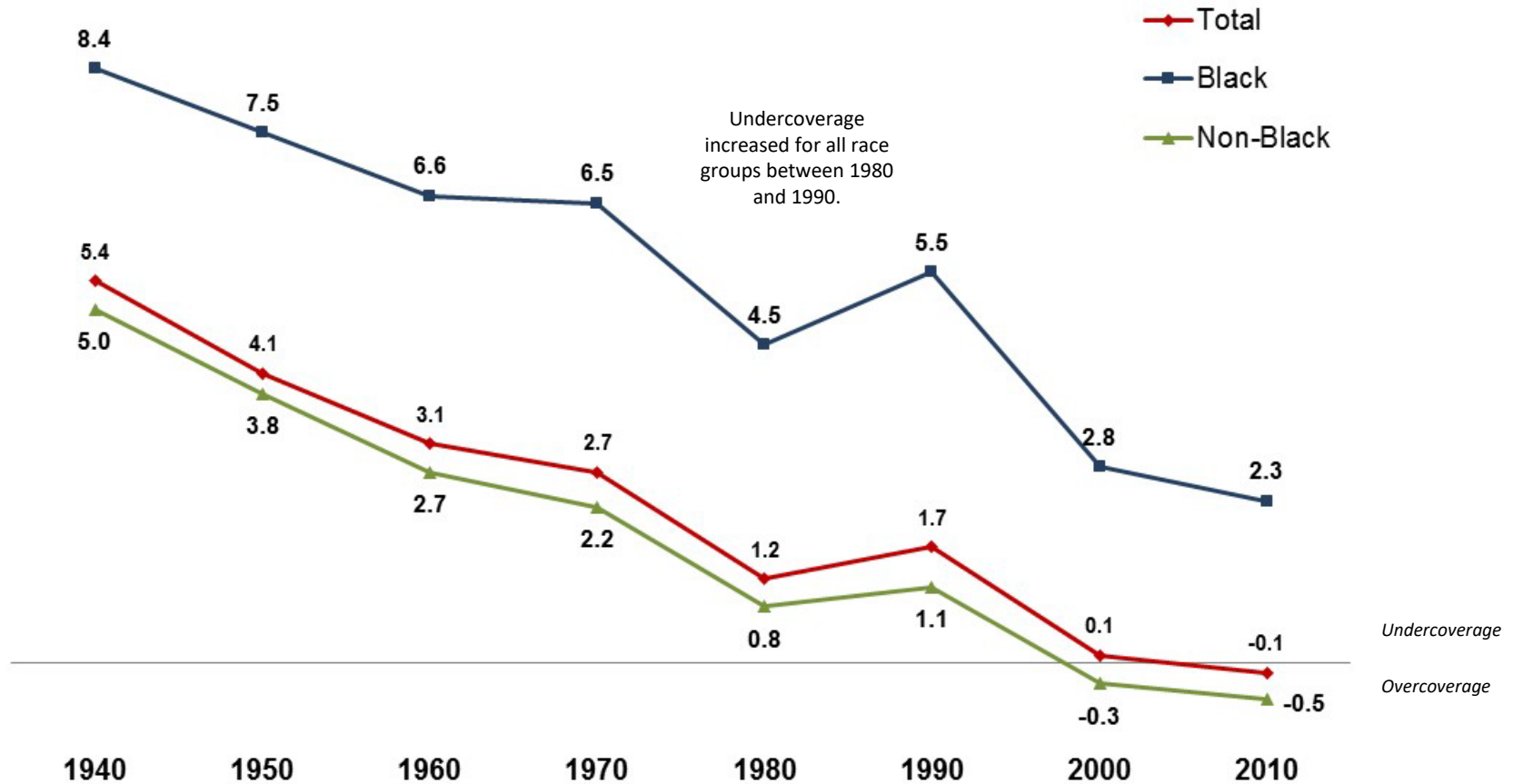
What enabled or motivated the Census Bureau to expand its use of administrative records and third-party data in decennial enumerations?

Increased Opportunities: Technological and Methodological Advancements

Increased use of administrative records for enumeration after 1970 was made possible by advancements in computer technology, data availability, and linkage methodologies.

- Increased use of computerized data files across the federal government
- Increased computing capacity to merge and manage multiple large data sets
- Improved ability to parse, standardize, and sanitize person and address data
- Improved ability to match across data sources without relying on personally identifiable information (PII)
- Developed “best” person, address, and person-at-address algorithms
- Improved speed of data deliveries and currency of person-place matches
- Established policies around confidentiality protection and appropriate use of administrative data

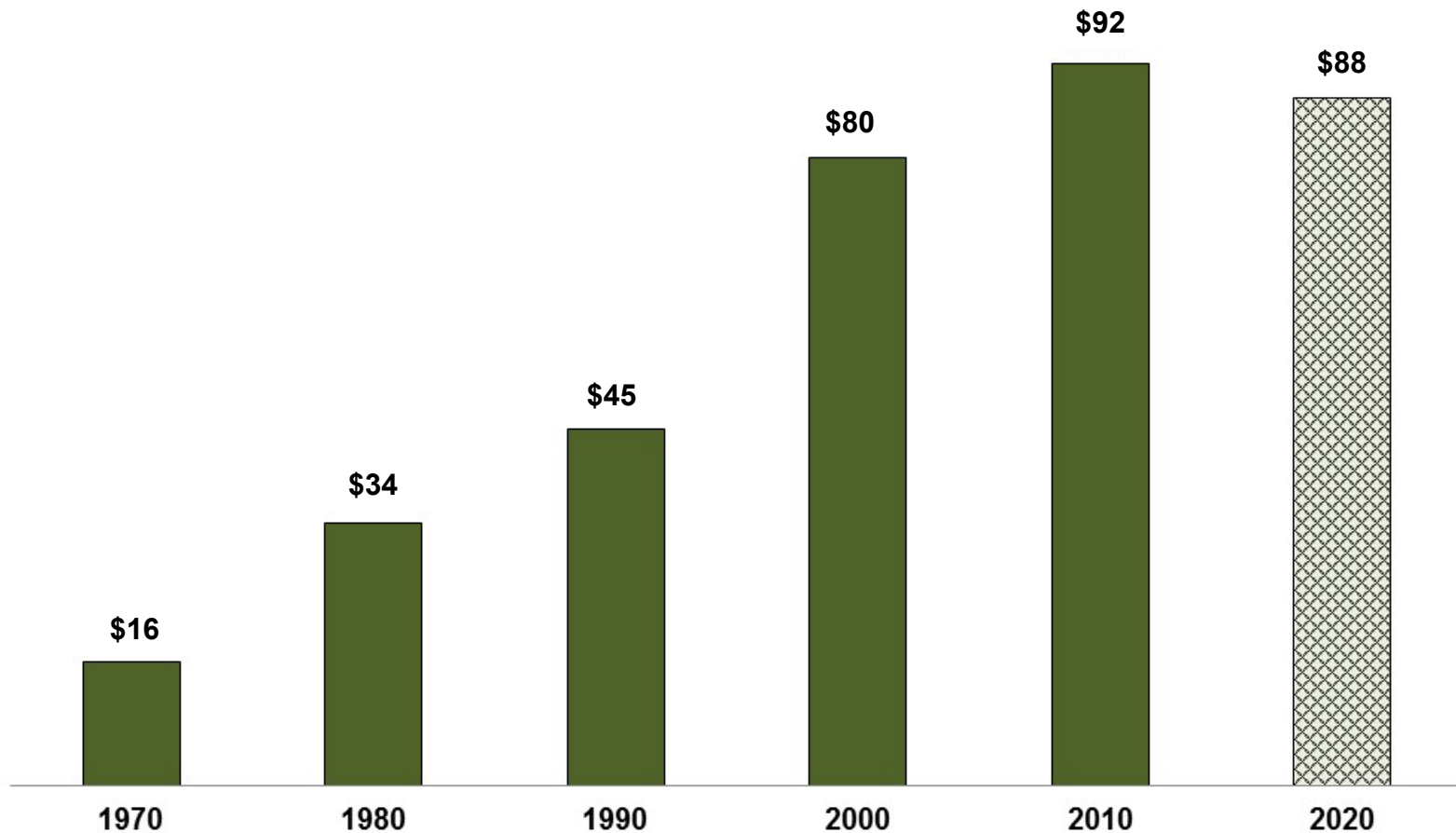
Increased Need: Undercoverage by Race



Percent Net Decennial Census Undercoverage by Race, As Estimated by Demographic Analysis: 1940 to 2010

Source: U.S. Census Bureau, Demographic Analysis, 1940 to 2010.

Increased Need: Rising Cost of Enumeration



Average Cost of Enumeration Per Housing Unit in the Decennial Census: 1970 to 2020

Note: In 2020 dollars. The 2020 average cost is estimated.

Source: U.S. Census Bureau, Decennial Census, 1970 to 2010, and internal estimates.

PART III:

Administrative Records Use 1970 to 2010

How were administrative records data used to enumerate the population in the 1970 to 2010 decennial censuses?

Administrative Records Use 1970 to 1990: Improving Coverage of Specific Populations

Increased use of administrative records for enumeration since 1970 reflected the growing concern over differential coverage by race.

- Coverage differentials seen as a problem that needed to be resolved, or one that could at least be mitigated using the right tools
- Analysts directed their research efforts to finding ways to improve response rates of groups with low response rates
- Special focus on programs aimed at improving minority coverage

Administrative Records Use 1970 to 1990: Early Programs

Movers Check (1970)

Purpose: To improve the count of people who moved and may have been missed during the census-taking period. For the inner-city areas of the 20 largest metropolitan areas, U.S. Postal Service records on address changes were compared clerically with census records in an effort to find missed individuals. Movers found not to have been enumerated were visited by census takers for inclusion in the count.

Nonhousehold Sources Program (1980)

Purpose: To reduce differential undercoverage of urban minority populations. Names and addresses from independent lists (e.g., drivers' licenses) were compared clerically with census records during enumeration in areas with large minority populations. Individuals determined to have been missed in households where a census questionnaire was received were assigned for follow-up and subsequently enumerated.

Post-Enumeration Post-Office Check (PEPOC) (1970 and 1980)

Purpose: To improve coverage in areas enumerated using the conventional door-to-door (i.e., nonmail) procedure. Census records were checked against post-office address lists to determine housing units missed. In 1970, PEPOC covered rural areas in 16 southern states but it covered all nonmail areas in 1980.

Parolee/Probationer Coverage Improvement Program (1990)

Purpose: To ensure complete enumeration of parolees and probationers, with special emphasis on minority populations. Administrative lists of names and addresses of parolees and probationers on Census Day, provided and certified by the States' Departments of Corrections, were matched clerically against census records. Nonmatches were added to the count.

Administrative Records Use 1970 to 2010: Military and Federally Affiliated Employees Overseas

1970: Combined Use of Direct Enumeration and Administrative Records

Department of Defense (DOD) provided records of the Army, Air Force, and Marine Corps personnel stationed overseas as of April 1, 1970. Additionally, DOD distributed and collected questionnaires from land-based Navy and Coast Guard personnel, military dependents, and its civilian employees and their families stationed overseas. The Department of State handled the enumeration of other civilian government employees and other Americans residing abroad through the U.S. embassies and other diplomatic and consular posts.

1990: Congress Reverses Plans to Exclude Military and Federally Affiliated Employees From Count

The Census Bureau did not initially plan to count overseas military personnel, U.S. government civilian employees, or their dependents in the 1990 Census but, based on strong interest from Congress, decided to include them again for apportionment purposes. Counts were obtained from 30 federal departments and agencies and were based on these administrative records.

1980: Shift to Administrative Records Begins

For the 1980 Census, the Census Bureau decided to enumerate Americans overseas through administrative counts of the armed forces and federal employees overseas, together with the number of dependents living with them, provided by DOD and the Office of Personnel Management. The data were not included in the apportionment counts. No attempt was made to enumerate directly the “overseas” population (e.g., private U.S. citizens abroad for extended periods) in 1980 or future censuses.

2000 and 2010: Continued Reliance on Administrative Records

The Census Bureau worked with the DOD and other federal departments and agencies to obtain counts of Armed Forces personnel and civilian employees stationed overseas and their dependents living with them for apportionment purposes. These counts were based primarily on administrative records used by the departments and agencies for payroll and personnel purposes.

Administrative Records Use in 2000: No Special Programs

In the 2000 Census, administrative data were used *only* to include overseas military, federally affiliated employees, and their dependents in the apportionment counts.

- There were no special administrative data programs directed at reducing differential coverage of particular groups, as in the 1970, 1980, and 1990 censuses

Other methodological and outreach improvements stemmed the historical decline in mail response rates.

- Differential coverage declined between 1990 and 2000 for all groups

Using the same approach, the 2010 Census demonstrated equally low coverage differentials.

- Census 2000 marked the end of special administrative data programs aimed at reducing differential response for select minority populations

Research Using Administrative Data: StARS and AREX 2000

Statistical Administrative Records System (StARS)

- The Census Bureau considered several new approaches for Census 2000 during planning in the 1990s, including using administrative records
- As part of this effort, census researchers built StARS, an administrative records database prototype integrating six large federal input files

Administrative Records Experiment (AREX) 2000

- An attempt to use the StARS database to simulate an administrative records census
- For two counties, StARS matched to approximately 85 percent of census addresses and accounted for about 95 percent of the census persons count

AREX demonstrated that, in the near term, administrative records could not supplant a decennial census.

- However, both StARS and AREX provided Census Bureau analysts a foundation for using administrative data in the 2010 and 2020 censuses

Administrative Records Use in 2010: Coverage Followup of Individuals in Housing Units

Administrative data were used to identify housing units for additional enumeration efforts to ensure proper coverage.

- Households were included in coverage followup if the information available from administrative sources suggested someone may have been missed

Respondents were matched to previous census and survey responses to impute missing characteristics (race and ethnicity).

- Federal tax information enabled linkages among data sources through the use of a matching key

2020 Census will considerably expand the use of administrative records for matching and modeling in various production processes.

PART IV:

Administrative Records and Third-Party Data Use in 2020

How does the Census Bureau plan to use administrative records and third-party data to enumerate the population in the 2020 Census?

Administrative Records Use in 2020:

Utilizing Administrative Records and Third-Party Data

Utilizing administrative records and third-party data is a key innovation area for the 2020 Census.

- Use may result in \$1.4 billion in cost avoidance

Administrative data will be used in four broad ways:

- Improve the census address list
- Increase the effectiveness of advertising and contact strategies
- Validate respondent submissions
- Reduce field workload for follow-up activities

Administrative Records Use in 2020: Enumerating Nonresponding Housing Units

- Letters will be sent to every address encouraging self response by internet, mail, or phone
- Vacant and nonexistent housing units will be identified from administrative and third-party data sources and removed from the follow-up workload
- A census taker will make at least one visit to each nonvacant nonresponding address
- After the opportunity to self-respond and after one visit, we will assess the remaining nonresponding addresses to determine which have reliable information from administrative data sources and use it to count people living at those addresses
- For the rest, census takers will continue in-person visits until cases are resolved

PART V: Conclusion

Conclusion

Over the last several decades, the use of federal, state, and third-party data for enumeration has gradually expanded and shifted emphasis.

- From programs aimed at correcting differential coverage of certain groups
- To broader efforts geared toward improving operational efficiencies, reducing total cost, and improving coverage for all nonresponding housing units

The use of administrative data for enumeration in the decennial census is not new.

- The planned use of these data in the 2020 Census represents the continuation of a long tradition of using the best sources of information available to conduct the most complete count possible

Contact Information

Elizabeth M. Grieco

Senior Advisor, Administrative Records and Data Linkage

elizabeth.m.grieco@census.gov

(301) 763-5275

Howard Hogan

Chief Demographer

howard.r.hogan@census.gov

(301) 763-4884