Administrative Records Coverage of Demographic Response Data in the American Community Survey

Renuka Bhaskar and Adela Luque
Center for Administrative Records
Research and Applications (CARRA)

FedCASIC
March 20, 2013

Overview of Presentation Topics

- Background and Research Objectives

- Data and Methodology

- Results

- Discussion and Future Research

Background

- In an effort to reduce costs, the U.S. Census Bureau is researching ways in which to use administrative records (AR) in decennial census and survey operations

- AR are currently used in many Census programs

- Examples of current research efforts
  - 2010 Census Match Study
  - American Community Survey (ACS) Match Study
  - 2020 Census Research Projects

Research Objectives

- Evaluate AR coverage of demographic response data
  - age, sex, race, and Hispanic origin
- Evaluate how coverage varies
  - by ACS mode of data collection and AR data source
  - for persons with allocated ACS demographic information

Data and Methodology

AR Data

- Centers for Medicare and Medicaid Services Medicare Enrollment Database

- Housing and Urban Development (HUD) Public and Indian Housing Information Center

- HUD Tenant Rental Assistance Certification System

- HUD Computerized Homes Underwriting Management System

- Indian Health Service Patient Registration File

AR Data

- Selective Service System Registration File
- Health and Human Services Temporary Assistance for Needy Families
- Social Security Administration (SSA) Supplemental Security Income Record
- SSA Numerical Identification File (Numident)
- Commercial data (5 vendors, 9 datasets)

Additional Data

- Previous Census Records
  - Census 2000
  - American Community Survey 2001-2009

- 2010 ACS

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Record Linkage - Person Identification Validation System (PVS)

- Assigns unique person identifiers called Protected Identification Keys (PIKs)

- Assigns unique address identifiers called Master Address File Identification Numbers (MAFIDs)

- Identifiers used to unduplicate records

- Identifiers facilitate record linkage

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Evaluating Coverage of Demographic Response Data

- 2010 ACS data merged with AR files by PIK
  - Not all persons in the ACS are assigned a PIK

- For each person in the ACS, determined if AR provided data on age, sex, race, and Hispanic origin

- Focus on presence of demographic data in AR, future work will look at quality of response

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Results

Coverage of each demographic item

- There were 4.47 million people in the 2010 ACS
  - 4.21 million were assigned PIKs (94 percent)
  - 4.18 million matched to AR (93 percent)

- Of the 2010 ACS sample, AR provided:
  - Hispanic origin data for 4.16 million, or 93 percent
  - Race data for 3.61 million, or 81 percent
  - Age data for 4.17 million, or 93 percent
  - Sex data for 4.16 million, or 93 percent

Demographic characteristics

- AR provided higher coverage of Hispanic origin data for non-Hispanics in the ACS (94 percent) compared to Hispanics (83 percent)

- Coverage of race data varied by race group
  - White alone and Black alone: 83 percent
  - American Indian or Alaska Native alone: 82 percent
  - Asian alone and Native Hawaiian and Other Pacific Islander alone: 76 percent
  - Two or More Races: 67 percent
  - Some Other Race alone: 46 percent

- Coverage of age and sex data high (over 90 percent) for all age groups and both sexes

Mode of response

- AR provided higher coverage of demographic response data for persons who responded via mail

- Coverage of Hispanic origin, race, age, and sex data
  - Mail: 86 to 98 percent
  - Telephone: 69 to 82 percent
  - Personal Visit: 67 to 85 percent
  - Group Quarters Visit: 85 to 89 percent

Mode by Demographic Characteristics

- Hispanic origin
  - For all modes, coverage of Hispanic origin data is higher for non-Hispanics than Hispanics
  - Lowest coverage is for Hispanics who responded via telephone (74 percent)
  - Highest coverage is for non-Hispanics who responded via mail (98 percent)

- Race
  - Coverage ranges from 36 percent (for Some Other Race alone in personal visit) to 88 (for Black alone and White alone in mail)
  - For all modes, coverage was lowest for those of Some Other Race alone

Administrative Records Data Source

- For Hispanic origin, sex, and age, federal data provided better coverage relative to commercial data
  - Federal data: over 90 percent coverage
  - Commercial data: under 70 percent coverage

- Federal data coverage for race was notably lower than for other demographic items

- Federal and commercial data provided similar coverage for race data
  - Federal data: 80 percent coverage
  - Commercial data: 81 percent coverage

Federal AR files

- Federal files that provided the highest coverage of demographic data
  - Numident
    - Highest coverage of Hispanic origin, sex, age data
    - Highest coverage of race data for certain groups (Black alone, Asian alone, Native Hawaiian Pacific Islander alone)
  - Previous Census Records
    - Second highest coverage of Hispanic origin, sex, and age data for some age groups (ages 3-64)
    - Highest coverage of race data for certain groups (White alone, Two or More Races, Some Other Race alone)
  - Centers for Medicare and Medicaid Services Medicare Enrollment Database
    - High coverage of age data for older population

Allocated data in the ACS

- The previous sections presented information on demographic coverage for all respondents in the 2010 ACS

- For some of these cases, the Census Bureau had to allocate demographic characteristics because the respondent did not provide a response for a particular demographic variable or because the reported response was incomplete or invalid

- Next section focuses on these allocated cases, where demographic response data was either not reported or was changed through the Census editing procedures

Administrative Records Coverage for Persons with Allocated Data

Demographic Item	Number for which Item was Allocated	Percent of Allocated Cases for which Demographic Response Data was Available in AR
Number of people in the 2010 ACS Sample	4.47 million	-----
….with allocated AGE	42,304	48.7
…with allocated SEX	6,501	69.4
…with allocated RACE	76,095	55.5
…with allocated HISPANIC ORIGIN	96,651	87.0

Coverage of allocated data

- AR coverage of demographic items is notably lower for those whose ACS response was allocated, compared to the total ACS sample

- Similar patterns by mode and data source
  - Those with allocated responses who responded via telephone or personal visit, have lower coverage than those who responded by mail
  - Federal data provides higher coverage relative to commercial data for those with allocated demographic responses
  - Numident and Previous Census Records provide the most information

Discussion

- AR provide demographic response data for the majority of respondents in the ACS

- Coverage varies by response item, mode, and data source

- Lower coverage for those with allocated ACS demographic response

- Further work is needed to evaluate how this can be used in survey operations

Future research

- Evaluate the quality of demographic response data in AR

- Evaluate coverage and quality of additional ACS items

- Continue analysis throughout the decade

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

Questions?

Thank you!