

Evaluating Race and Hispanic Origin in Administrative Records Relative to the 2010 Census

Sonya Rastogi
U.S. Census Bureau

Federal Committee on Statistical Methodology
Research Conference
November 4, 2013

This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on technical, statistical, or methodological issues are those of the author and not necessarily those of the U.S. Census Bureau.

Overview

- Motivation for research
- 2010 Census Match Study
demographic response analysis
- Administrative records Hispanic origin
and race assignment and comparison
to 2010 Census
- Conclusion and next steps

Motivation

- Implement use of administrative records (AR)
- Address missing race and Hispanic origin data
- Previous research – low quality race and Hispanic AR for minorities
- Goal – Understand individual sources and data quality issues

2010 Census Match Study

Demographic Response Analysis

Research Question #1

- What is the agreement of Hispanic origin and race responses between the 2010 Census and AR sources?

Data

- 2010 Census data (as reported, no proxy)
- Three different files from Housing and Urban Development (HUD)
- Indian Health Service (IHS)
- Medicare (MEDB)
- Numerical Identification System (Numident)
- Previous Census Records
- Temporary Assistance for Needy Families (TANF)
- 4 Commercial Files

Methodology

- Record Linkage - Unique identifiers called Protected Identification Keys (PIKs) were used to link persons in AR and the 2010 Census
- Response Match Ratios calculated as:

$$\frac{\text{Individuals with matching AR and 2010 Census demographic response}}{\text{Individuals in 2010 Census for whom AR provides demographic data}}$$

Agreement of Hispanic Origin Response

Administrative Records Source File	Percent in Agreement: Non-Hispanic	Percent in Agreement: Hispanic
Federal Files		
Previous Census Records	99.4	93.1
Numident	N/A	54.2
HUDCHUMS	98.5	80.0
HUDPIC	98.1	86.0
HUDTRACS	98.6	78.6
TANF	98.3	70.7
MEDB	N/A	29.4
Commercial Files		
Commercial File 1	98.2	83.5
Commercial File 2	98.0	77.3
Commercial File 3	98.0	80.2
Commercial File 4	97.9	77.1

Agreement of Race Response

2010 Census Race	Range in Percent Agreement by AR Source Files	
	Federal	Commercial
White Alone	96.1 to 99.1	94.7 to 97.9
Black Alone	87.4 to 98.3	37.3 to 61.1
Asian Alone	58.0 to 94.1	73.6 to 85.2
American Indian or Alaska Native Alone (AIAN)	24.6 to 97.6	6.4 to 13.2
Native Hawaiian or Other Pacific Islander Alone (NHPI)	37.0 to 76.0	14.3 to 19.2
Some Other Race Alone	14.1 to 54.9	1.1 to 3.4
Two or More Races	3.6 to 36.3	N/A

Source: 2010 Census and Administrative Records

AR and 2010 Comparisons after assignment of Hispanic Origin and Race Response from all AR Sources

Research Question #2

- What are the characteristics of people whose Hispanic origin and race responses match, do not match, or are missing in AR?

Data

- 2010 Census data (as reported, no proxy)
- AR Hispanic
 - All files used in Census Match Study
 - Medicaid Statistical Information System (MSIS)
 - Texas Supplemental Nutrition Assistance Program (SNAP)
- AR Race
 - All files used in Census Match Study except commercial data

Method to Assign AR Hispanic Origin Response

Case #	AR Source 1	AR Source 2	AR Source 3	AR Source 4	AR Hispanic
1	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic	Non-Hispanic
2	Missing	Non-Hispanic	Missing	Missing	Non-Hispanic
3	Hispanic	Hispanic	Missing	Hispanic	Hispanic
4	Missing	Hispanic	Missing	Missing	Hispanic
5	Hispanic	Non-Hispanic	Missing	Hispanic	Hispanic
6	Non-Hispanic	Non-Hispanic	Non-Hispanic	Hispanic	Hispanic
7	Missing	Missing	Missing	Missing	Missing

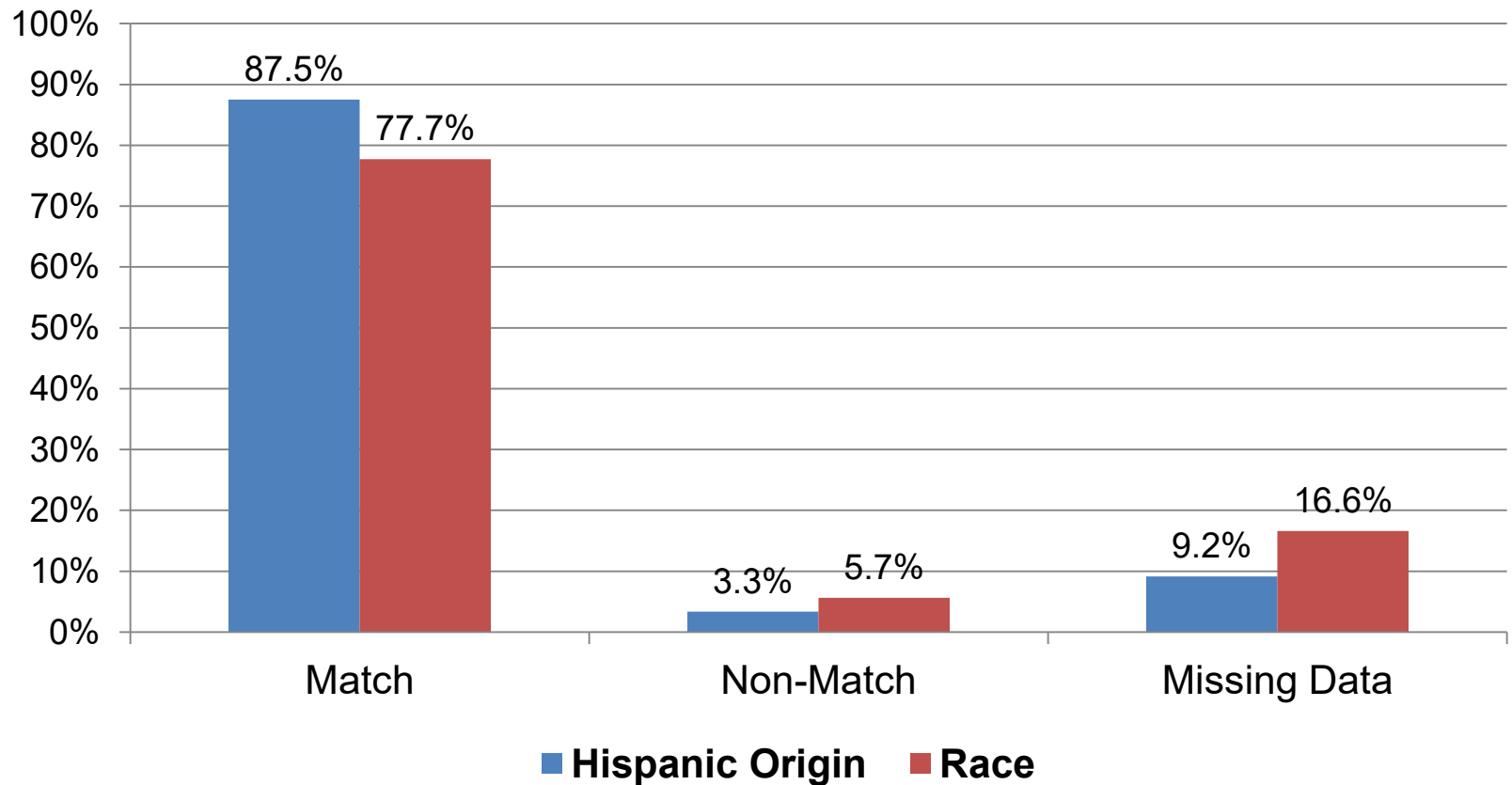
Method to Assign AR Race Response

Case #	AR Source 1	AR Source 2	AR Source 3	AR Source 4	AR Race
1	Missing	White alone	White alone	White alone	White alone
2	Black alone	Missing	Black alone	Black alone	Black alone
3	NHPI alone	White alone	Missing	NHPI alone	NHPI alone
4	NHPI alone	AIAN alone	White alone	Missing	NHPI alone
5	AIAN alone	White alone	Missing	AIAN alone	AIAN alone
6	Asian alone	AIAN alone	Missing	AIAN alone	AIAN alone
7	Asian alone	Missing	Black alone	Asian alone	Asian alone
8	Missing	Missing	Missing	Missing	Missing

Methodology

- Multinomial logistic regression
 - Dependent variable has 3 categories
 - Matched response (Reference)
 - Unmatched response
 - Missing AR response
 - Separate regressions for Hispanic origin and race

Distribution of Dependent Variables, Census-AR Linked Data



Source: 2010 Census and Administrative Records

Select Multinomial Logistic Regression Results, Odds Ratios

	Hispanic Origin Model		Race Model	
	Non-matching responses	Missing in AR	Non-matching responses	Missing in AR
Ethnicity in Census (Not Hispanic omitted)				
Hispanic	1.61***	0.67***	43.74***	11.37***
Race in Census (Non-Hispanic White alone omitted)				
Non-Hispanic Black alone	0.68***	0.88***	1.70***	0.70***
Non-Hispanic AIAN alone	2.21***	0.95***	13.87***	0.41***
Non-Hispanic Asian alone	2.48***	3.03***	6.27***	10.45***
Non-Hispanic NHPI alone	3.48***	1.77***	38.75***	9.08***
Age (18-44 years old omitted)				
0-17 years old	1.20***	14.87***	1.88***	16.13***
45-64 years old	0.96***	0.37***	0.99***	0.49***
65+ years old	0.85***	0.21***	0.91***	0.19***

Regressions include variables for gender, region, household type, household size, tenure, rural, census mode, median income in tract, and percent tract non-Hispanic White.

Hispanic Model n: 253,639,590

Race Model n: 256,138,867

***p<.001

Source: 2010 Census and Administrative Records

Conclusion and Next Steps

- Similar findings relative to work in 2000, AR responses for minority groups lower relative to non-Hispanic whites
- Pursue additional federal, state, commercial data to fill gaps where AR race and Hispanic origin agreement is low relative to the 2010 Census
- Research additional business rules
- Tease out racial and Hispanic origin fluidity from data quality issues

Thank You!

sonya.rastogi@census.gov