

When Race and Hispanic Origin Reporting are Discrepant Across Administrative Records Sources: Exploring Methods to Assign Responses

Sharon Ennis, Sonya Rastogi, and James Noon
U.S. Census Bureau

Statistics Canada 2016 International Methodology Symposium
March 22, 2016

Disclaimer: This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Overview

- Goals of the research
- Background
- Business rules used to assign administrative records (AR) race and Hispanic origin responses
- Multinomial logistic regression analysis
- Conclusion and next steps

Goals of Research

- Utilize business rules to assign one Hispanic origin and one race response when this information differs across AR sources
- Evaluate characteristics of people whose Hispanic origin and race responses match Census data, do not match Census data, or are missing in AR

Background

- Research on data quality comparing race and Hispanic origin responses in AR to Census responses
 - Previous research – low quality race and Hispanic AR for minorities
- Research on racial fluidity
 - Previous research – race response change is common among the American Indian and Alaska Native (AIAN), Native Hawaiian and Other Pacific Islander (NHPI), multiracial, and Hispanic populations

Data

- 2010 Census data (as reported)

AR Race

- 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)

AR Hispanic

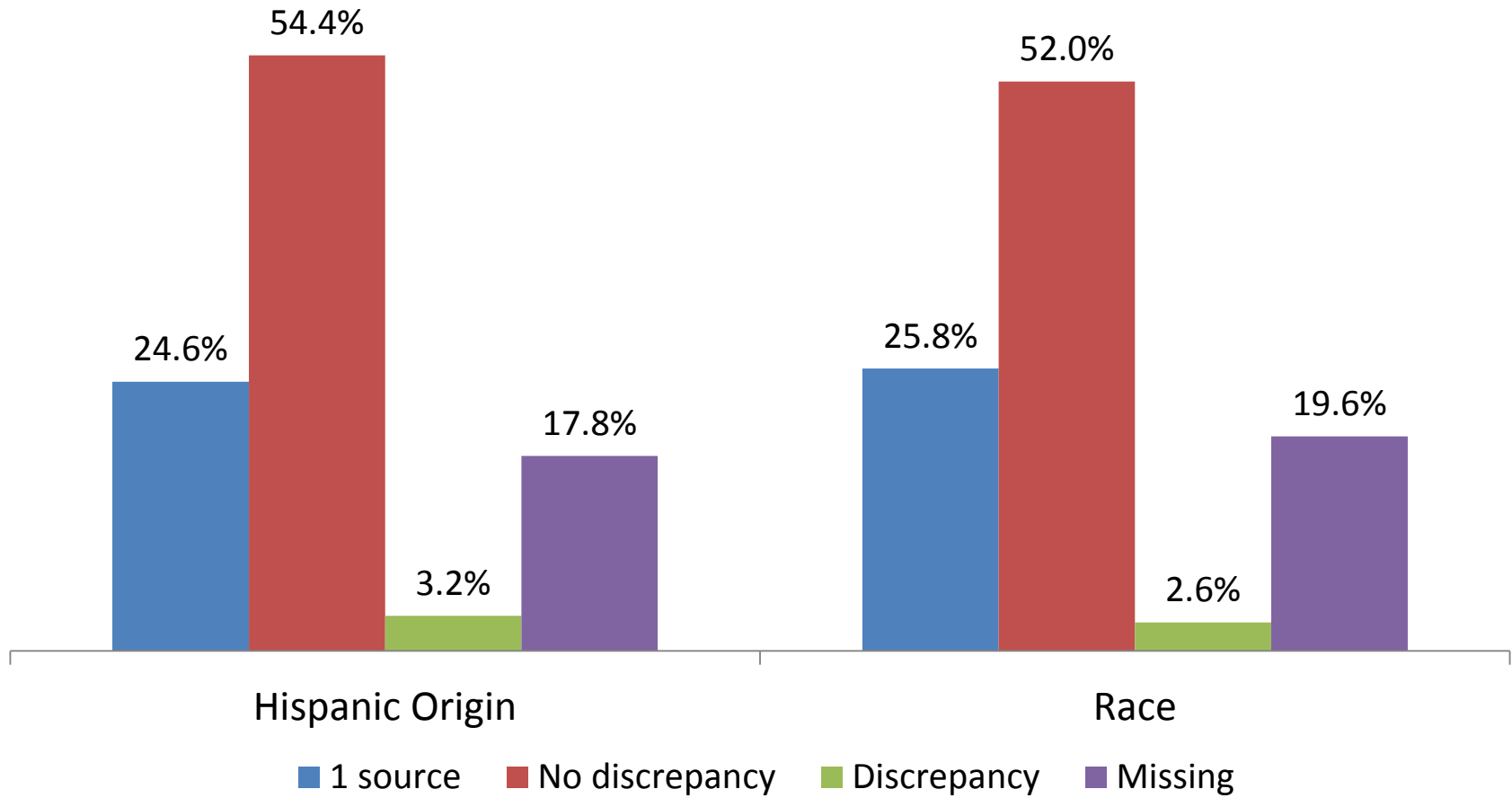
- All files used for AR race
- Medicaid Statistical Information System (MSIS)
- Texas Supplemental Nutrition Assistance Program (SNAP)
- 4 Third party files

Methodology

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

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

Percentage of Records with a Discrepancy Across AR Sources



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

Other Method Used to Assign AR Hispanic Origin

- AR Hispanic method 2
 - Hispanic response assigned from Previous Census Records. If no Previous Census Records response then Hispanic assigned if present in any other AR source.

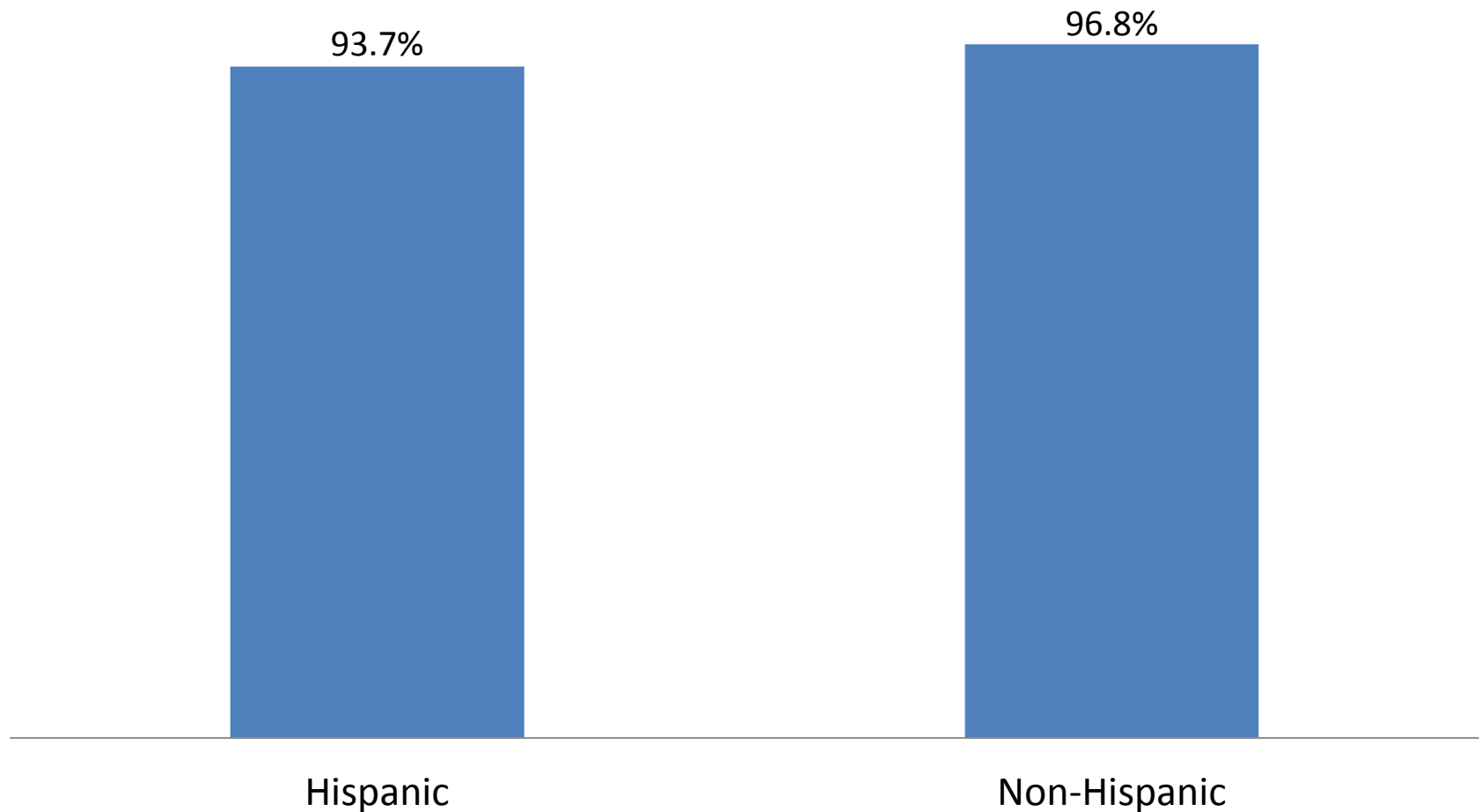
Other Methods Used to Assign AR Race

- AR race method 2
 - Similar to method 1 but preference was given to Two or More Races first.
- AR race method 3
 - Race assigned based on the most frequently reported race among the AR sources. If no most frequent race, race was assigned according to method 1.
- AR race method 4
 - Race assigned based on the most frequently reported race. If no most frequent race, race was assigned according to method 2.

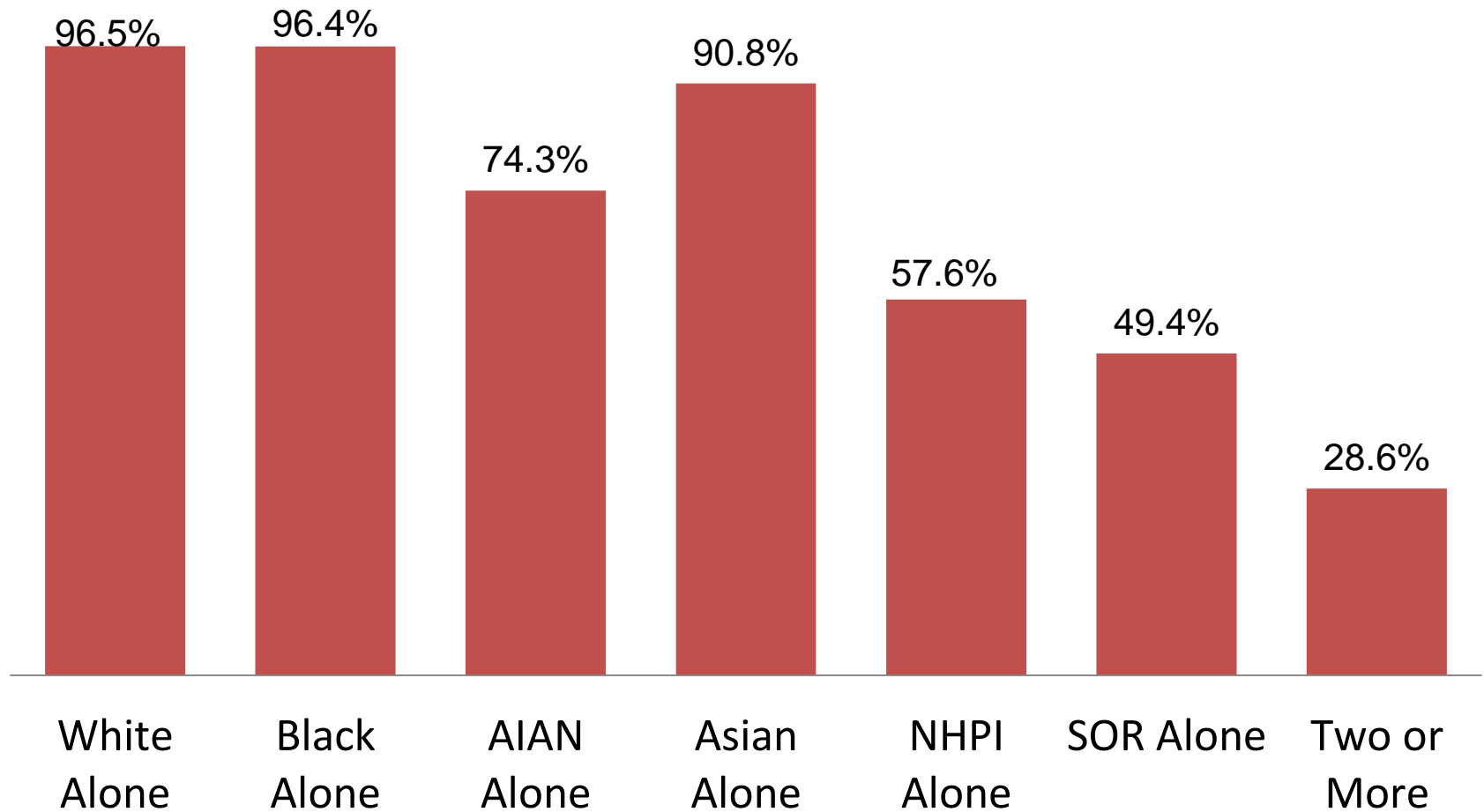
Other Methods Used to Assign AR Race

- AR race method 5
 - Race assigned from Previous Census Records. If no Previous Census Records response then race was assigned according to method 1.
- AR race method 6
 - If AIAN response present in the IHS data, that response was assigned to the AR record. Otherwise, a race response was assigned according to method 5.

Percentage of Census-AR Linked Data Matching on Hispanic Origin



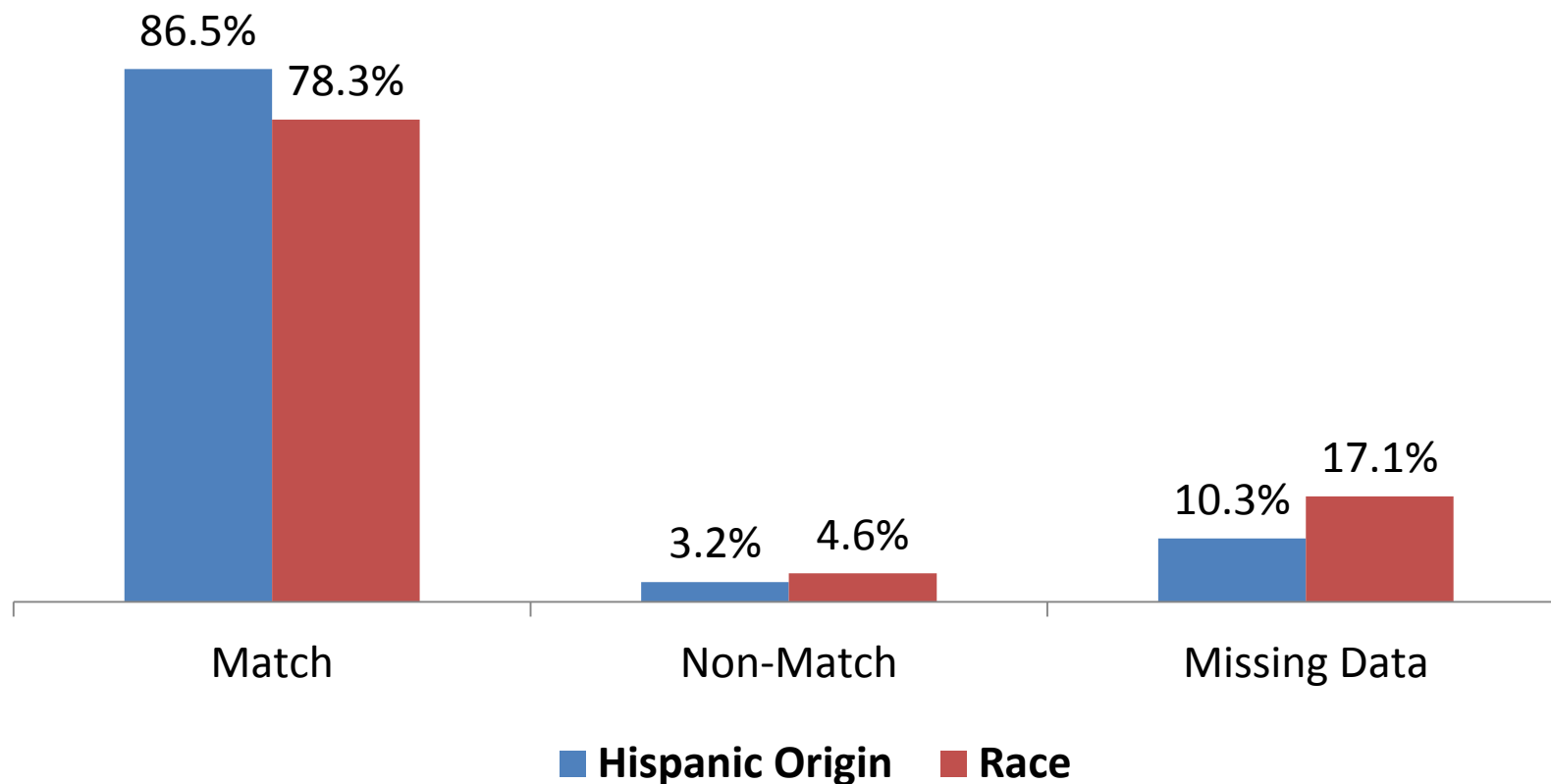
Percentage of Census-AR Linked Data Matching on Race



Methodology

- Multinomial logistic regression
 - Dependent variable has 3 categories
 - Matching response (Reference)
 - Non-matching 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/Race in Census (NH White alone omitted)				
Hispanic	1.58***	0.71***	42.64***	11.24***
NH Black alone	0.67***	0.88***	1.63***	0.71***
NH AIAN alone	2.20***	0.89***	13.67***	0.41***
NH Asian alone	2.51***	2.98***	6.29***	10.88***
NH NHPI alone	3.60***	2.01***	35.95***	9.24***
Age (18-44 years old omitted)				
0-17 years old	1.24***	12.66***	2.19***	16.85***
45-64 years old	0.96***	0.40***	0.95***	0.48***
65+ years old	0.83***	0.24***	0.86***	0.25***

Hispanic Model n: 251,320,952

Race Model n: 253,905,696

NH = Non-Hispanic ***p<.001

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/Race in Census (NH White alone omitted)				
Hispanic	1.58***	0.71***	42.64***	11.24***
NH Black alone	0.67***	0.88***	1.63***	0.71***
NH AIAN alone	2.20***	0.89***	13.67***	0.41***
NH Asian alone	2.51***	2.98***	6.29***	10.88***
NH NHPI alone	3.60***	2.01***	35.95***	9.24***
Age (18-44 years old omitted)				
0-17 years old	1.24***	12.66***	2.19***	16.85***
45-64 years old	0.96***	0.40***	0.95***	0.48***
65+ years old	0.83***	0.24***	0.86***	0.25***

Hispanic Model n: 251,320,952

Race Model n: 253,905,696

NH = Non-Hispanic ***p<.001

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/Race in Census (NH White alone omitted)				
Hispanic	1.58***	0.71***	42.64***	11.24***
NH Black alone	0.67***	0.88***	1.63***	0.71***
NH AIAN alone	2.20***	0.89***	13.67***	0.41***
NH Asian alone	2.51***	2.98***	6.29***	10.88***
NH NHPI alone	3.60***	2.01***	35.95***	9.24***
Age (18-44 years old omitted)				
0-17 years old	1.24***	12.66***	2.19***	16.85***
45-64 years old	0.96***	0.40***	0.95***	0.48***
65+ years old	0.83***	0.24***	0.86***	0.25***

Hispanic Model n: 251,320,952

Race Model n: 253,905,696

NH = Non-Hispanic ***p<.001

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/Race in Census (NH White alone omitted)				
Hispanic	1.58***	0.71***	42.64***	11.24***
NH Black alone	0.67***	0.88***	1.63***	0.71***
NH AIAN alone	2.20***	0.89***	13.67***	0.41***
NH Asian alone	2.51***	2.98***	6.29***	10.88***
NH NHPI alone	3.60***	2.01***	35.95***	9.24***
Age (18-44 years old omitted)				
0-17 years old	1.24***	12.66***	2.19***	16.85***
45-64 years old	0.96***	0.40***	0.95***	0.48***
65+ years old	0.83***	0.24***	0.86***	0.25***

Hispanic Model n: 251,320,952

Race Model n: 253,905,696

NH = Non-Hispanic ***p<.001

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/Race in Census (NH White alone omitted)				
Hispanic	1.58***	0.71***	42.64***	11.24***
NH Black alone	0.67***	0.88***	1.63***	0.71***
NH AIAN alone	2.20***	0.89***	13.67***	0.41***
NH Asian alone	2.51***	2.98***	6.29***	10.88***
NH NHPI alone	3.60***	2.01***	35.95***	9.24***
Age (18-44 years old omitted)				
0-17 years old	1.24***	12.66***	2.19***	16.85***
45-64 years old	0.96***	0.40***	0.95***	0.48***
65+ years old	0.83***	0.24***	0.86***	0.25***

Hispanic Model n: 251,320,952

Race Model n: 253,905,696

NH = Non-Hispanic ***p<.001

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, third party 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!

sharon.r.ennis@census.gov

sonya.rastogi@census.gov

james.noon@census.gov