

Comparing Administrative Records and 2010 Census Persons

James Noon
U.S Census Bureau

2014 Federal CASIC Workshops
March 20, 2014

Disclaimer: 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 authors and not necessarily those of the U.S. Census Bureau.

Overview of Presentation Topics

- Motivation for Research
- Data and Methodology
- Major Findings
- Conclusions and Next Steps

Motivation

- To reduce costs, many countries use administrative records (AR) to assist and replace traditional censuses
- AR are currently used in many Census programs
- Understand AR coverage of the 2010 Census person population

Background:

2010 Census Match Study

- Compare 2010 Census and AR counts and matches:
 - Persons
 - Addresses
 - Persons at addresses
- Assess quality and coverage of AR demographic response data relative to the 2010 Census

Data

- Medicare Enrollment Database
- Three files from Housing and Urban Development (HUD)
- Indian Health Service Patient Registration File
- Internal Revenue Service (IRS) Individual Income Returns 1040 and IRS Information Returns 1099/W2
- Selective Service System Registration File
- Supplemental Security Income Record
- Temporary Assistance for Needy Families
- Commercial data (5 vendors, 9 datasets)
- 2010 Census Data

Methodology

- Record Linkage - Unique identifiers called Protected Identification Keys (PIKs) were used to link persons in AR and the 2010 Census
 - Assigned using information such as SSN, name, date of birth, address
 - Facilitate unduplication and linkage of records
- Person Match Ratio calculated as:
(AR and 2010 Census Match/2010 Census Count)*100

Person Differences in 2010 Census and AR Data

2010 Census

Includes duplicates

Includes household imputations

Includes records reconciled through processing

Includes records with missing or bad name and date of birth (DOB) data

Includes proxy responses

AR Data

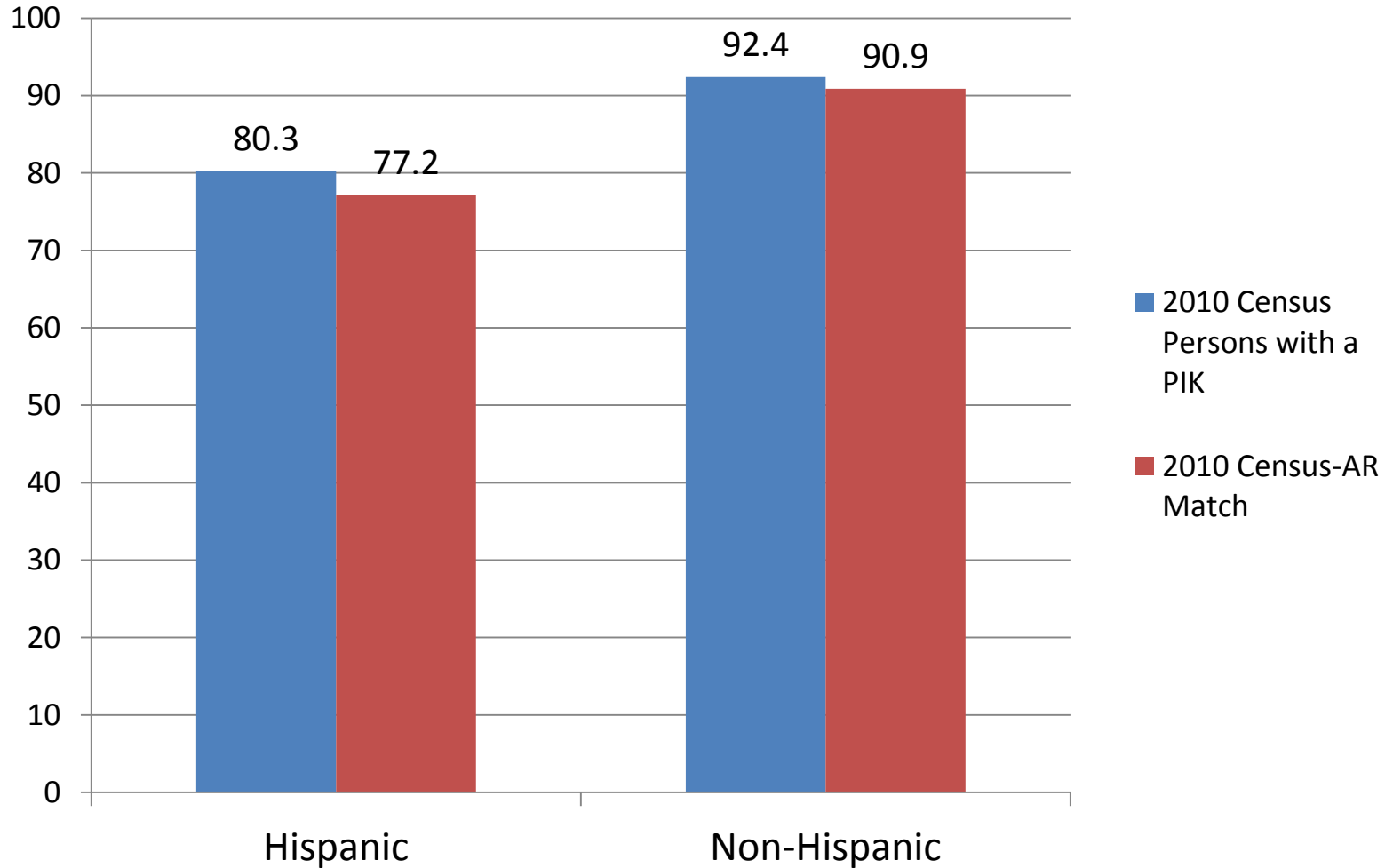
Includes only records with validated identifiers

Lag for small children (tax code effect)

Person Results

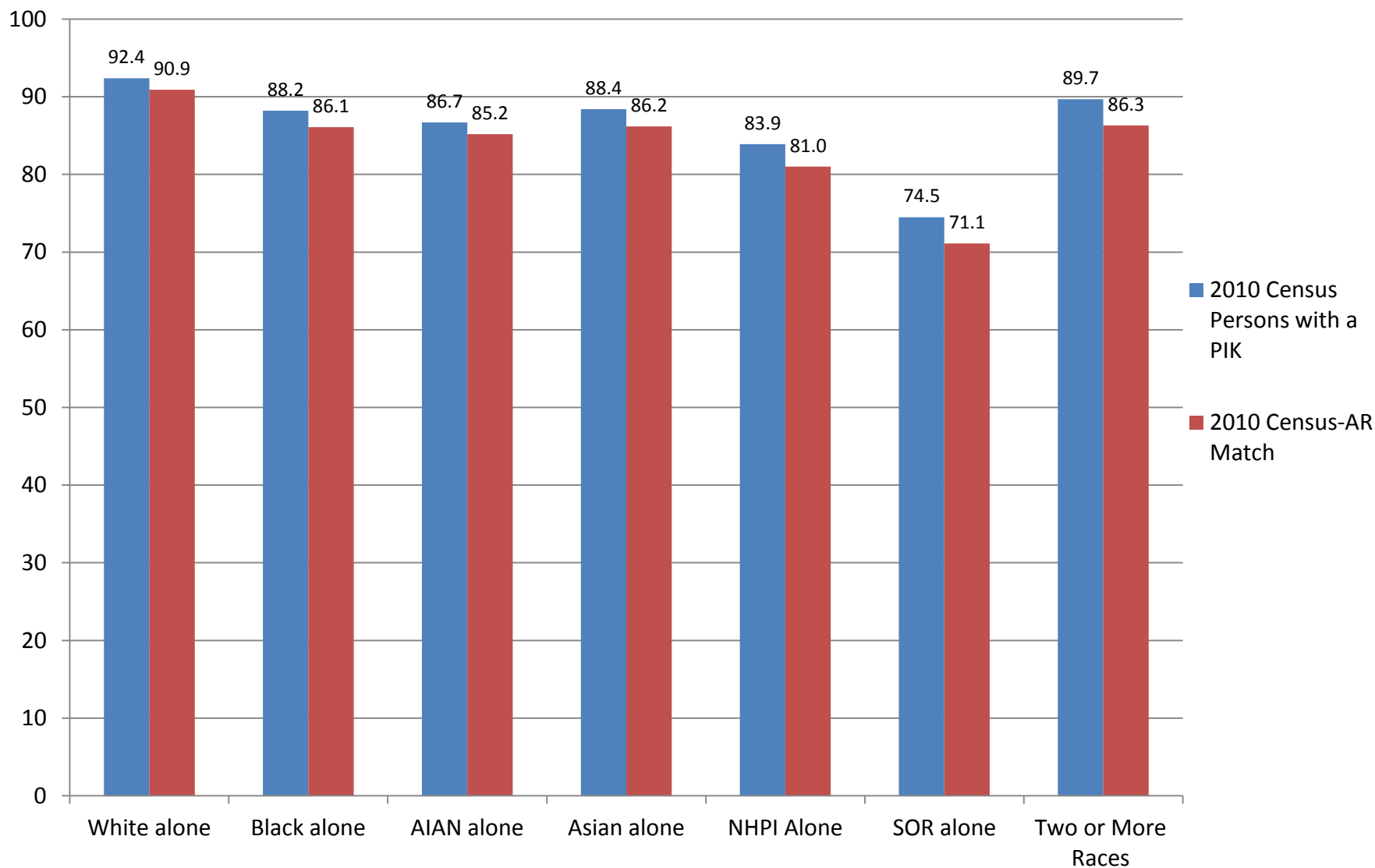
2010 Census Count	2010 Census PIKs	AR Count	AR and 2010 Census Match	Match Ratio
308,745,538	279,179,329	312,214,325	273,643,411	88.6

Percentage of 2010 Census Persons with PIKs and 2010 Census-AR Match Ratios: Hispanic Origin



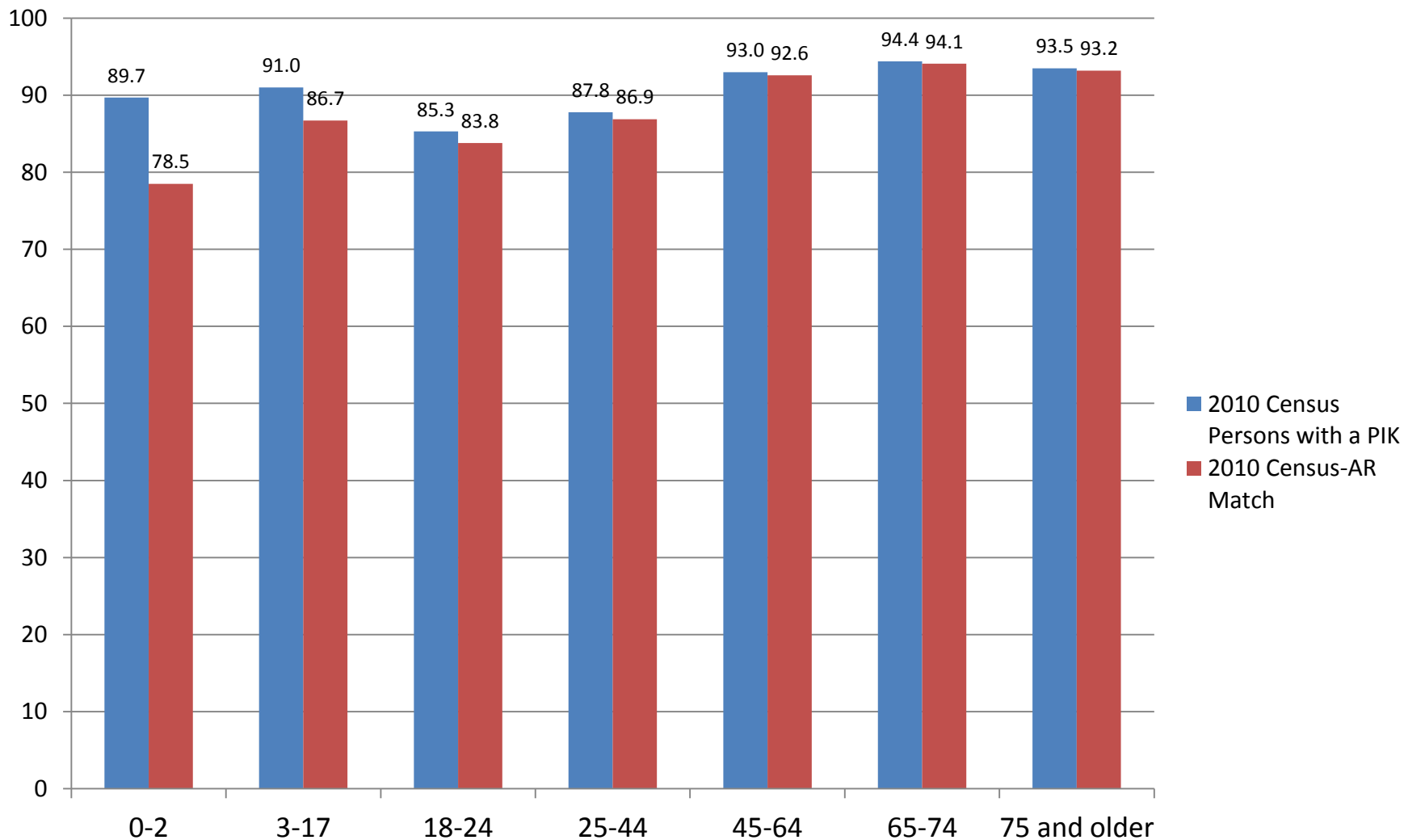
Sources: 2010 Census and 2010 Census Match Study Administrative Records

Percentage of 2010 Census Persons with PIKs and 2010 Census-AR Match Ratios: Race



Sources: 2010 Census and 2010 Census Match Study Administrative Records

Percentage of 2010 Census Persons with PIKs and 2010 Census-AR Match Ratios: Age



Sources: 2010 Census and 2010 Census Match Study Administrative Records

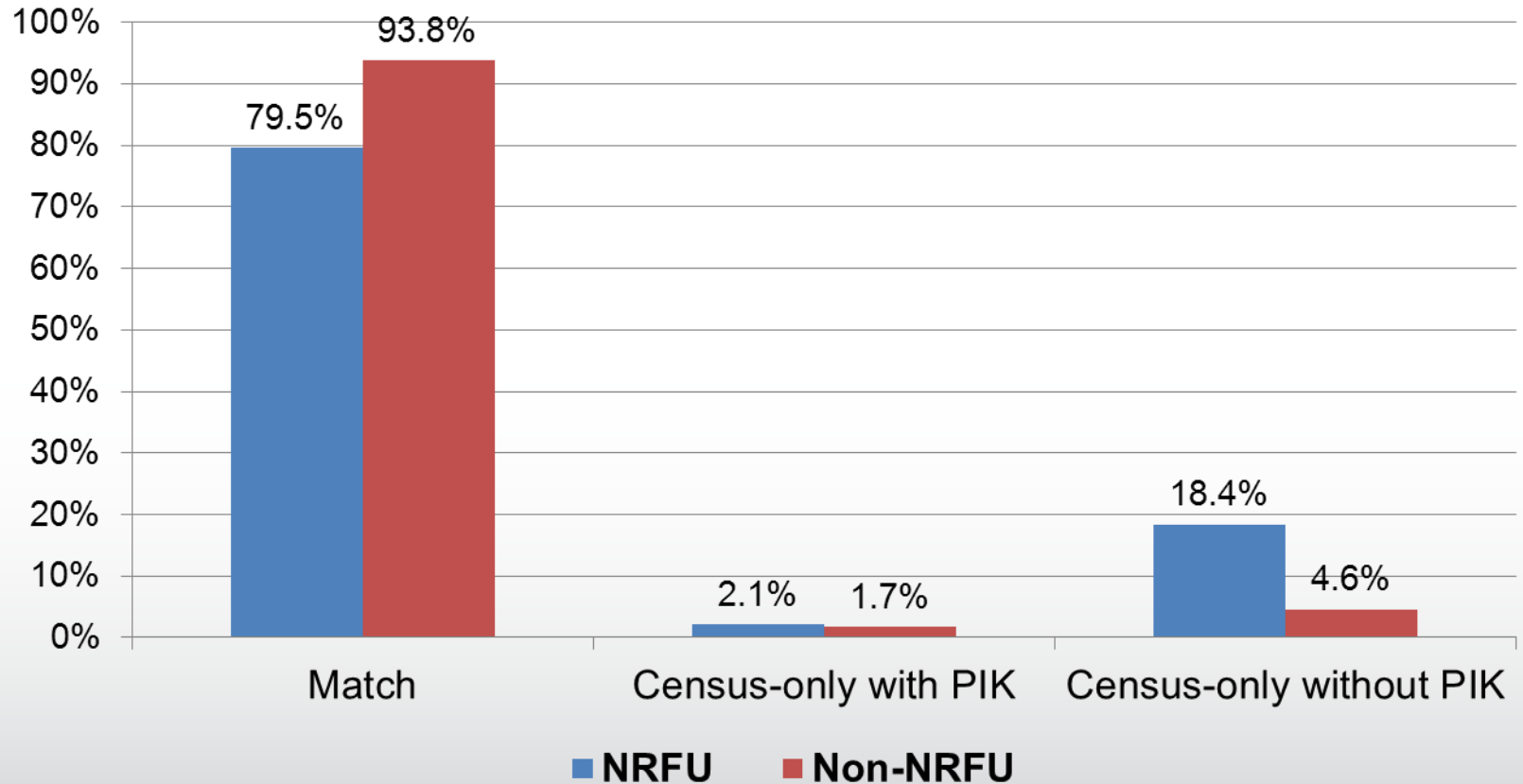
Research Question

- What are the person, household, and contextual characteristics of people who:
 - match to AR
 - are in 2010 Census only with a PIK
 - are in 2010 Census only without a PIK

Methodology

- Multinomial logistic regression
 - Dependent variable has 3 categories
 - Matched response (Reference)
 - 2010 Census only with a PIK
 - 2010 Census only without a PIK
 - Separate regressions for Nonresponse Followup and other modes of response
 - Include county-level measures from 2006-2010 American Community Survey data

Distribution of Dependent Variable, Census-AR Linked Data



Source: 2010 Census and Administrative Records

Select Multinomial Logistic Regression Results, Odds Ratios: Race and Ethnicity

	NRFU Model		Non-NRFU Model	
	Census-only, with PIK	Census only, without PIK	Census-only, with PIK	Census-only, without PIK
	(Census-AR Match Omitted)		(Census-AR Match Omitted)	
Ethnicity in Census (Not Hispanic omitted)				
Hispanic	1.391 ***	2.761***	1.359***	4.173***
Race in Census (Non-Hispanic White alone omitted)				
Non-Hispanic Black alone				
Non-Hispanic AIAN alone				
Non-Hispanic Asian alone				
Non-Hispanic NHPI alone				
Non-Hispanic SOR alone				
Non-Hispanic Two or More Races				
Gender – Male				

Regressions also include variables for proxy response, region, type of enumeration area, and other county-level measures

NRFU Model n: 60,397,468

Non-NRFU Model n: 234,179,674

***p<.001

Source: 2010 Census and Administrative Records

Select Multinomial Logistic Regression Results, Odds Ratios: Age

	NRFU Model		Non-NRFU Model	
	Census-only, with PIK	Census only, without PIK	Census-only, with PIK	Census-only, without PIK
	(Census-AR Match Omitted)		(Census-AR Match Omitted)	
Age (25 to 44 omitted)				
Age Category: 0 to 1 Years	17.875***		23.990***	
Age Category: 2 to 4 Years	5.720***		6.986***	
Age Category: 5 to 17 Years	6.225***		5.873***	
Age Category: 18 to 24 Years	2.044***		1.945***	
Age Category: 45 to 64 Years	0.561***		0.540***	
Age Category: 65 or More Years	0.600***		0.722***	

Regressions also include variables for proxy response, region, type of enumeration area, and other county-level measures

NRFU Model n: 60,397,468

Non-NRFU Model n: 234,179,674

***p<.001

Source: 2010 Census and Administrative Records

Select Multinomial Logistic Regression Results, Odds Ratios: Household and County Characteristics

	NRFU Model		Non-NRFU Model	
	Census-only, with PIK	Census only, without PIK	Census-only, with PIK	Census-only, without PIK
	(Census-AR Match Omitted)		(Census-AR Match Omitted)	
<i>Household Characteristics</i>				
Single Father Household (Married omitted)	1.110***	1.196***	1.259***	1.595***
Single Mother Household (Married omitted)	0.868***	0.698***	0.964***	0.875***
Household size more than four persons (One-person household omitted)				
Household Tenure: Rental Property (Owned omitted)				
Rural Address				
<i>County Characteristics</i>				
Logged Median Household Income in County				

Regressions also include variables for proxy response, region, type of enumeration area, and other county-level measures

NRFU Model n: 60,397,468

Non-NRFU Model n: 234,179,674

***p<.001

Source: 2010 Census and Administrative Records

Conclusions and Next Steps

- Acquire additional federal, state, and commercial data to improve person coverage and fill gaps particularly for groups such as children and Hispanics and persons in rental properties
- Pursue use of AR in NRFU operations to reduce costs
- Record linkage research on data that do not have adequate name and date of birth data

Thank you!

james.noon@census.gov