



Administrative Records Research and Planning

Presentation to the Census Scientific Advisory Committee

September 15, 2017

Andrew Keller

Decennial Statistical Studies Division

Administrative Records Research and Planning

Administrative Record Usage

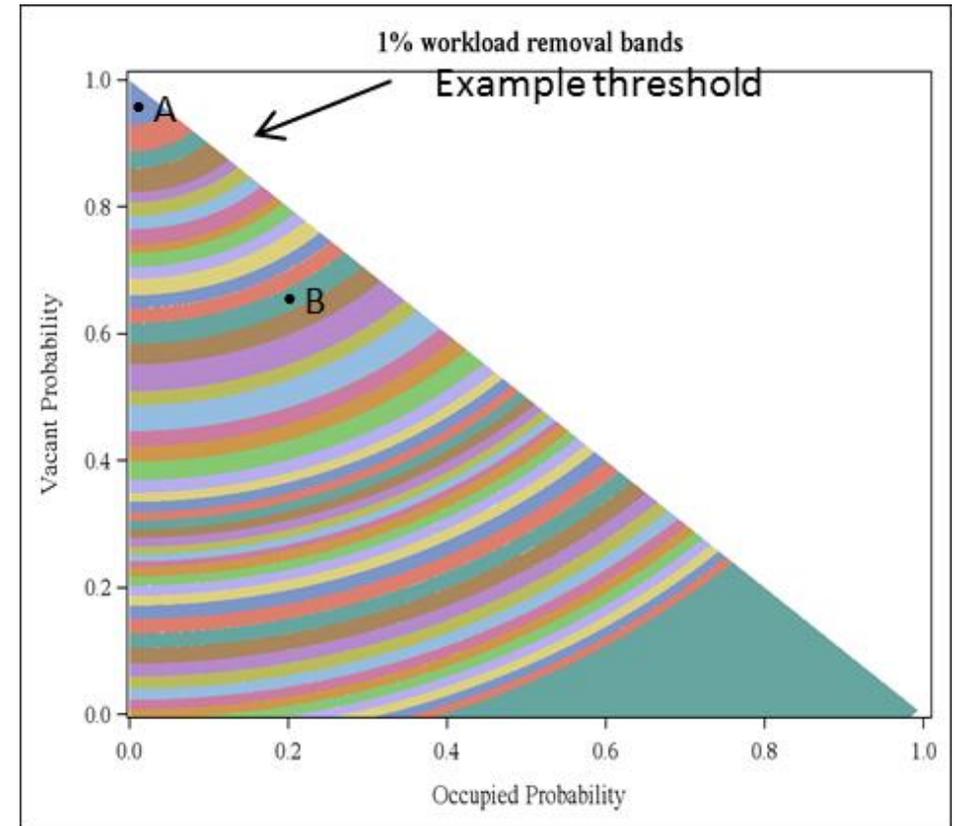
Outline

- Highlight March 2017 Census Scientific Advisory Committee Presentation
- 2016 and 2018 End-to-End Census Test Nonresponse Followup (NRFU) Contact Strategy
- Research Topics
 - Assigning characteristics for administrative record (AR) enumeration
 - Additional AR sources
 - Census Bureau Kidlink
 - Program data from States
 - Multiple Listing Service
 - Standardized distance accounting for covariance instead of Euclidean distance

Administrative Records Research and Planning

Identifying Administrative Records Vacant Units

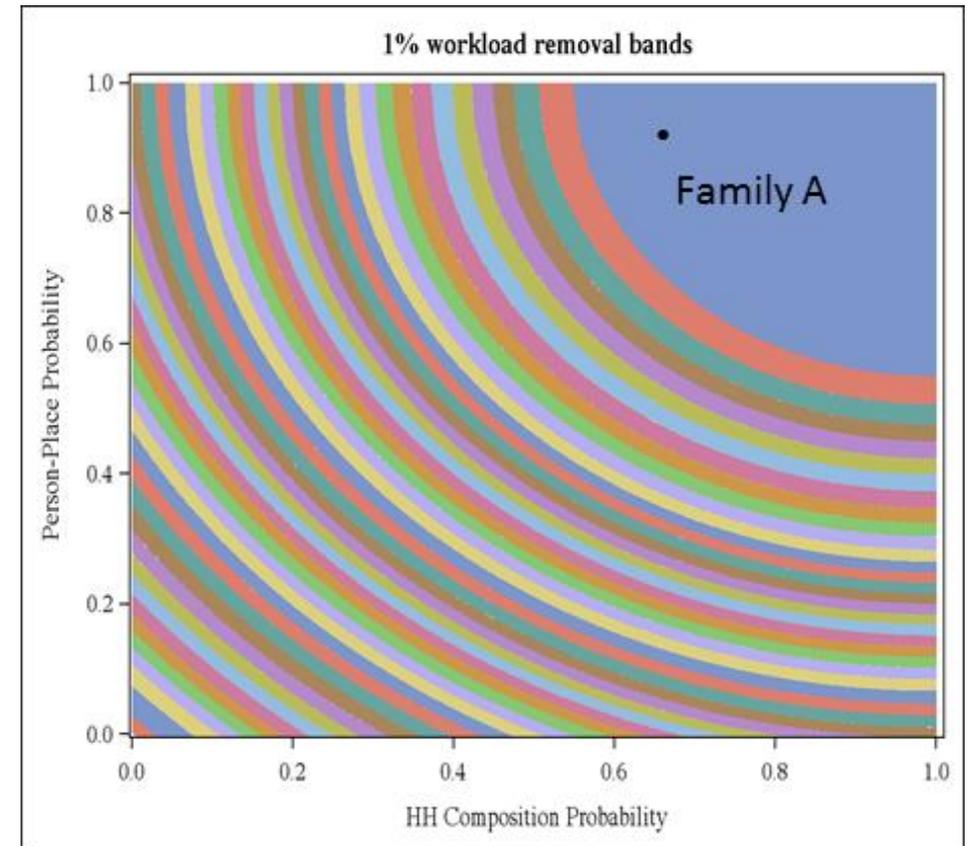
- Use 2010 NRFU data to train national-level housing unit status model
- Estimate predicted probability of occupied, vacant, non-existent (delete)
 - United States Postal Service (USPS) Undeliverable-as-Addressed (UAA) reasons, Delivery Sequence File
 - Internal Revenue Service (IRS) 1040 filings
 - IRS 1099 information returns
 - Centers for Medicare and Medicaid Services Medicare Enrollment database
 - Indian Health Service Patient Database
 - Third-party Veterans Service Group of Illinois (VSGI) files
- Euclidean distance function can be visualized as successive bands of cases emanating from the point (0,1) in the top left corner
- In this example, unit A is identified as AR Vacant while unit B is not
- Similar approach implemented for non-existent or addresses that need to be deleted



Administrative Records Research and Planning

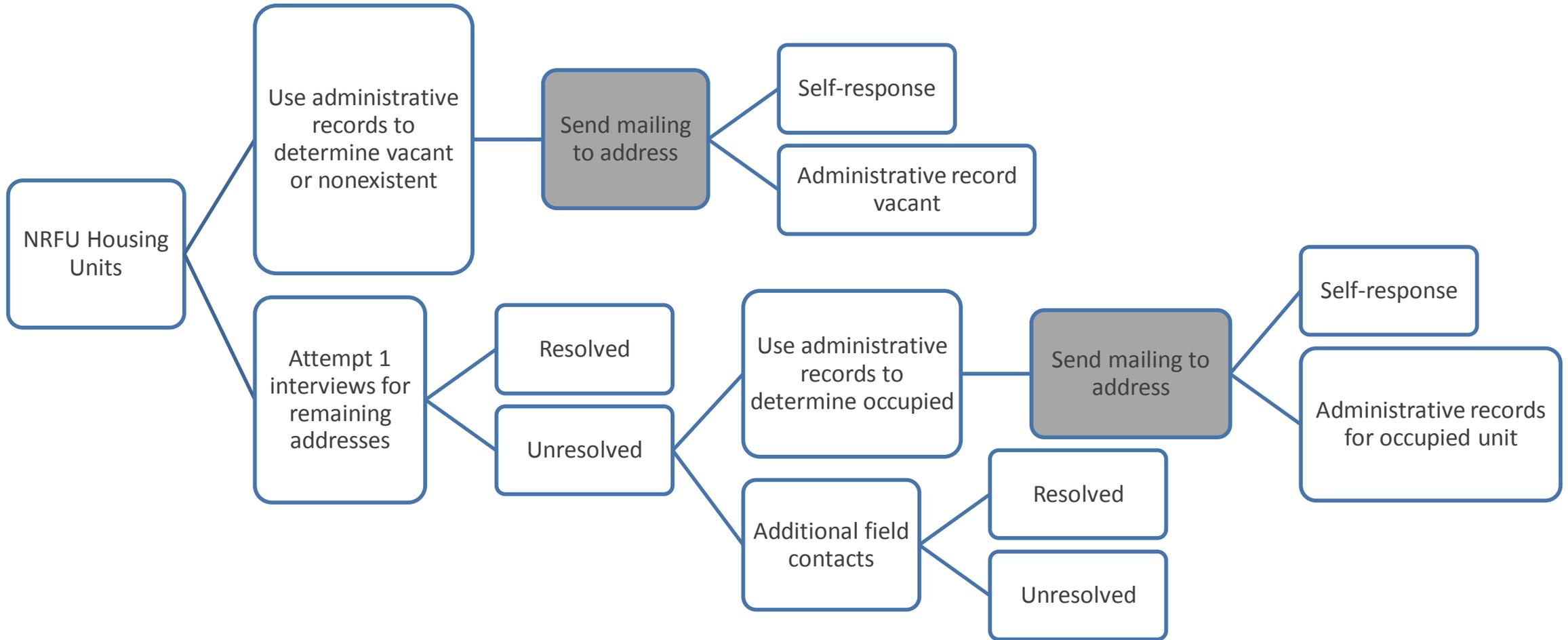
Identifying Administrative Records Occupied Units

- Use 2010 NRFU data to train national-level models
 - Similar set of AR sources as housing unit status models
 - **Person-place model** – How likely is it that we enumerate persons at the right place?
 - **Household Composition model** – How likely is it that the household composition of the Family A matches the Census?
- Estimate predicted probability for person-place and household composition models
- Euclidean distance function can be visualized as successive bands of cases emanating from the point (1,1) in the top right corner
- Each successive band represents an additional 1 percent of the NRFU workload
- In this example, Family A is in our top 1 percentile of administrative record rosters



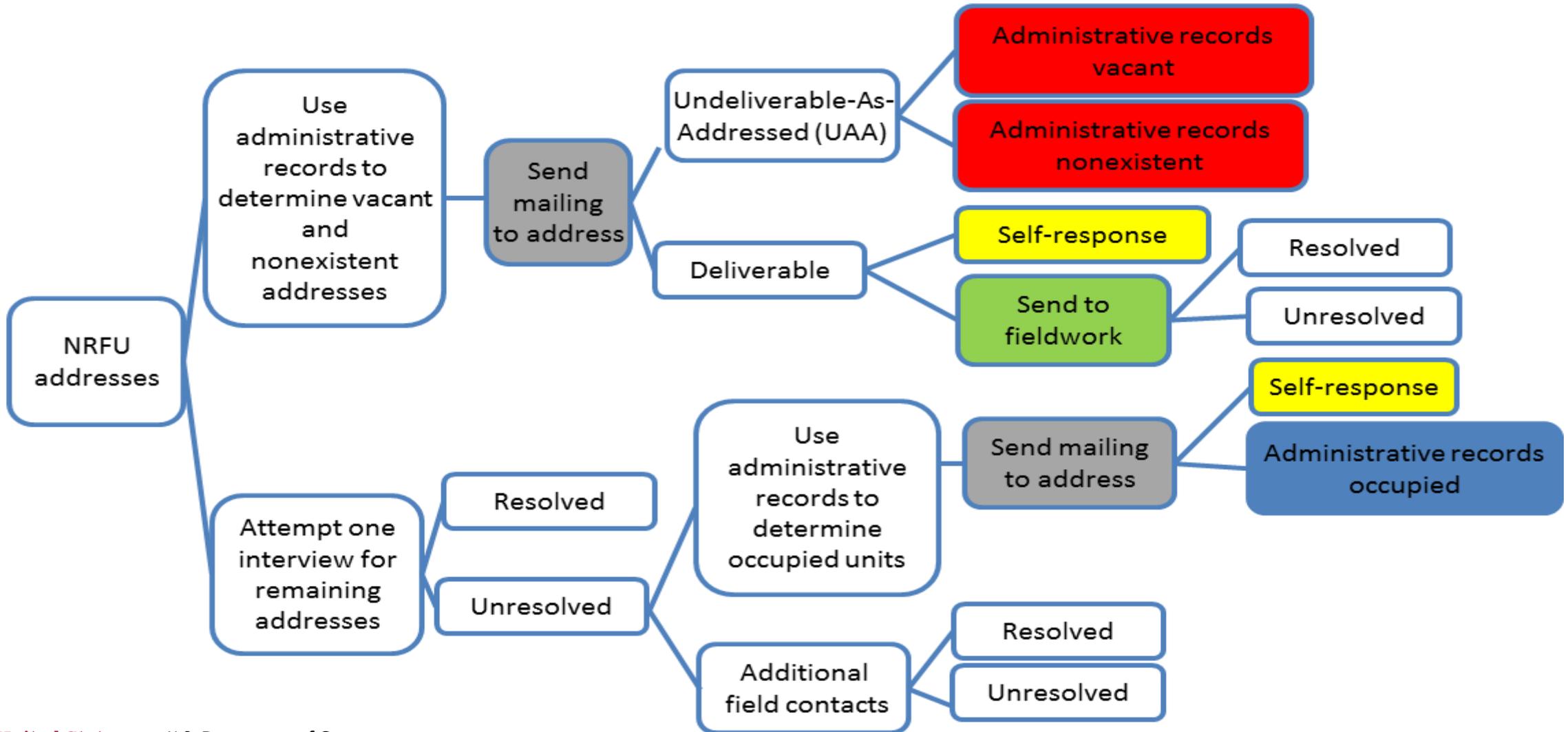
Administrative Records Research and Planning

2016 Contact Strategy



Administrative Records Research and Planning

2018 End-to-End Census Test Proposed Contact Strategy



Administrative Records Research and Planning

How do we assign characteristics to administrative records enumerations?

- No characteristic data from respondent
- Supplement AR enumerations with previously collected Title 13 data
- AR could provide additional data points

Age & Date of Birth / Sex

- Census Numident

Race / Hispanic Origin

- Country of Origin (Census Numident)
- Census Bureau's Center for AR Research and Application Race and Hispanic Origin file

Relationship to Householder

- Census Bureau KIDLINK file

Tenure

- Housing and Urban Development files
- Commercial Files

Administrative Records Research and Planning

How do we assign characteristics to administrative records enumerations? (Continued)

Use Census reported results to assess AR quality

- Create truth decks using 2010 Census data
- Assign instead with AR sources
- Compare imputations versus reported data
- Apply same idea to mid-decade census test

Operational Challenges

- Coverage – What part of the population does AR cover?
- Temporal concerns – Incorporating AR into characteristic imputation as part of decennial processing
- Mid-decade census tests help inform feasibility of sources
- Incorporating AR within existing edit and imputation procedures

Administrative Records Research and Planning

How do we assign characteristics to administrative records enumerations? (Continued)

2010 Simulation

About 16 million people in simulated AR Occupied units

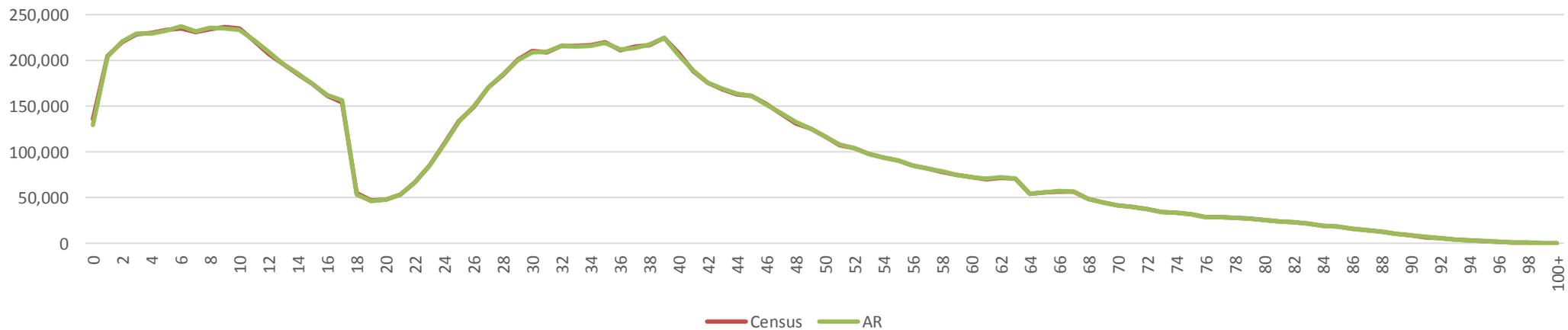
Many were in 2010 Census at same unit

Compare agreement between 2010 Census reports and AR characteristic.

Result

- Graph shows Census and administrative record results are overlapping

2010 Census Age vs. AR Age for persons in AR Occupied Units

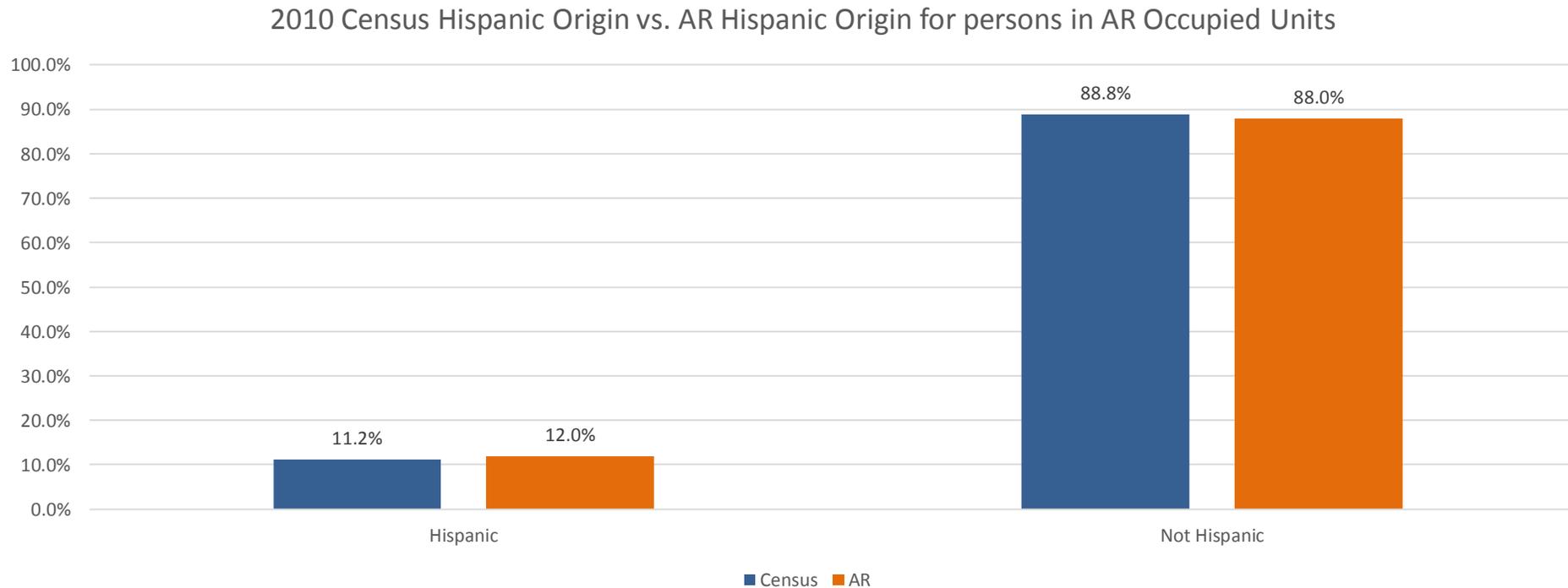


Administrative Records Research and Planning

How do we assign characteristics to administrative records enumerations? (Continued)

2010 Simulation

Hispanic Origin



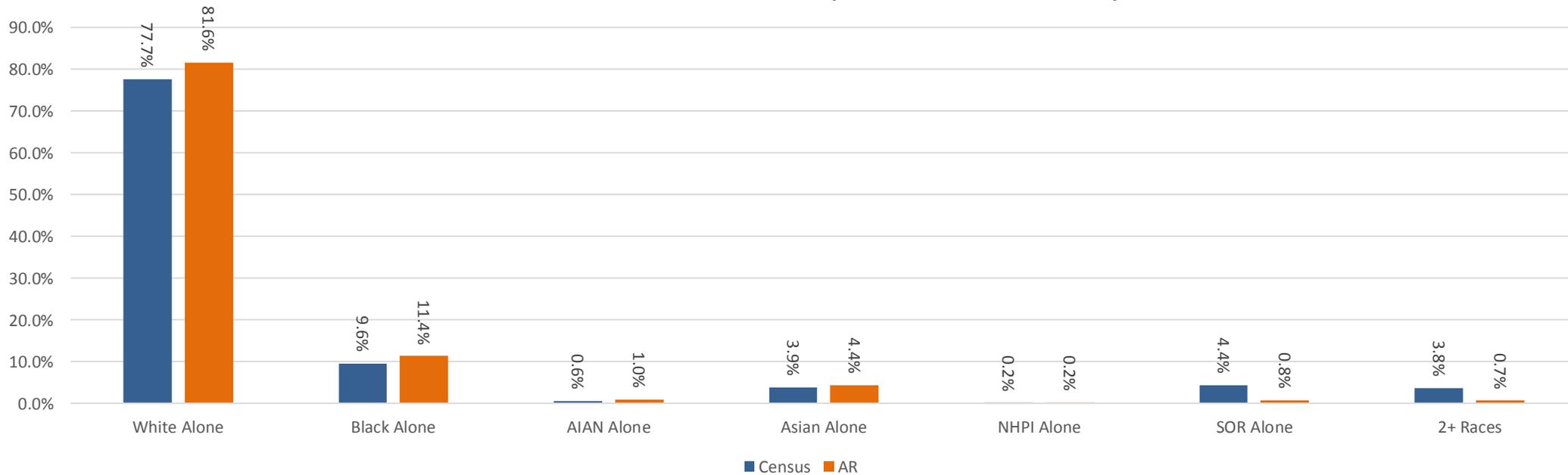
Administrative Records Research and Planning

How do we assign characteristics to administrative records enumerations? (Continued)

2010 Simulation

Race

2010 Census Race vs. AR Race for persons in AR Occupied Units

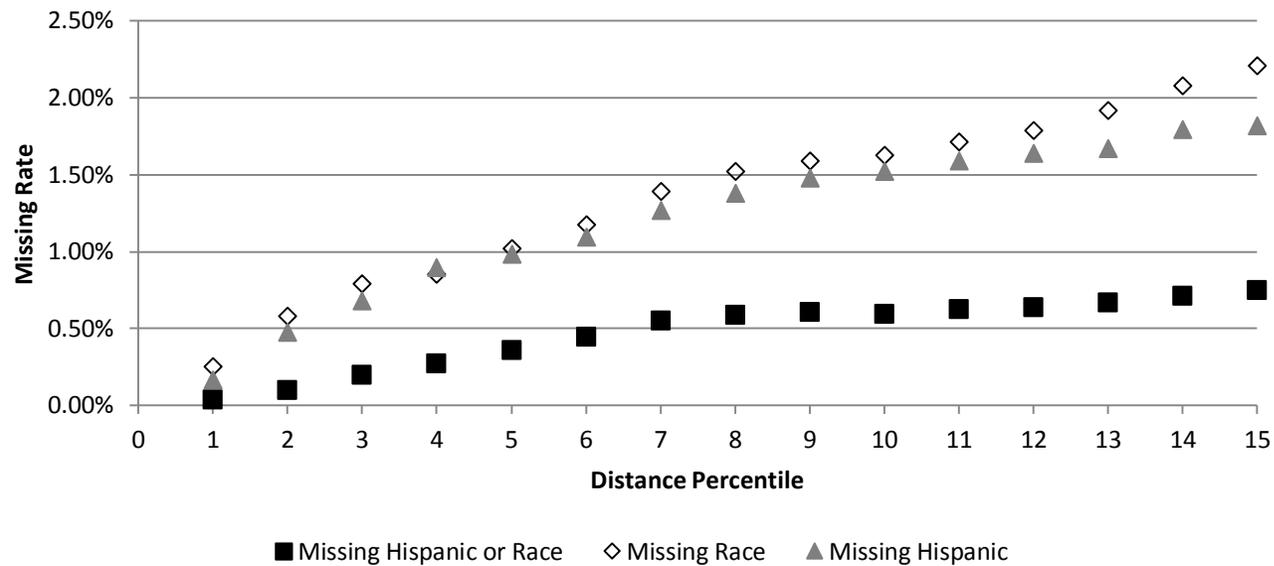


Administrative Records Research and Planning

How do we assign characteristics to administrative records enumerations? (Continued)

2010 Simulation

Marginal effect on race and Hispanic origin household missingness rate for additional 500,000 cases identified as AR Occupied



Administrative Records Research and Planning

Additional Sources

KIDLINK – Definition and Motivation

- Research file created by the Center for Administrative Records Research and Applications (CARRA) at the Census Bureau
 - Different from Social Security Administration (SSA) Kidlink
- Uses parent names from child Social Security Number (SSN) application to assign Protected Identification Key (PIK) for mother and father of child
- Potential undercoverage of children in core AR sources
 - Children appear as dependents on IRS 1040 filings
 - Less likely to appear in other core sources

Administrative Records Research and Planning

Additional Sources (Continued)

KIDLINK - Viewpoints

Pros

- Coverage of all children with SSN
- Single, national-level file

Cons

- No address information
- Links children to parents at the time of SSN application. Child may not be living with these parents at time of census.
- Does not cover children without SSN

Administrative Records Research and Planning

Additional Sources (Continued)

KIDLINK - Implementation

1. **From the Beginning:** Use rosters augmented with CARRA Kidlink to train and apply predictive models
 - Treats CARRA Kidlink as another core AR source
2. **Added After:** Add CARRA Kidlink children to rosters after AR occupied identification
 - Model training and predictions conducted without CARRA Kidlink

Administrative Records Research and Planning

Additional Sources (Continued)

KIDLINK - Results

- CARRA Kidlink can improve coverage of children in core AR sources
- Approach that incorporates CARRA Kidlink information in the modeling process (“from the beginning”) improves count agreement, especially in high poverty areas
- Approach that adds CARRA Kidlink children to units after baseline AR occupied determinations may count children in the wrong location

Administrative Records Research and Planning

Additional Sources (Continued)

State Data

- March 2017 CSAC recommendation to utilize state data as much as possible
- Acquiring
 - Supplemental Nutrition Assistance Program (SNAP)
 - Woman, Infants, and Children (WIC),
 - Temporary Assistance for Needy Family (TANF) data
- Researching Possible Implementation
 - Add to core sources
 - Use state data as a business rule to identify address at which to conduct complete set of NRFU in-person contacts
 - Use for characteristic imputation

Administrative Records Research and Planning

Additional Sources (Continued)

Third-Party

Multiple Listing Service

- Corelogic
- Assessing coverage
- Determining if can help in determination of vacant addresses that might be on market

Administrative Records Research and Planning

Standardized Distance

A recommendation from the March meeting

- Modify Euclidian distance to produce a standardized distance similar to a Wald statistics that can account for the covariance of the predictions
- Moises Yi and Larry Warren of CES were able to do some work on this
- Initial results shared with the CSAC Administrative Record Working group was not showing much difference

Discussion

What are your reactions for assigning characteristics for administrative record enumerations?

As we finish our research, do you have suggestions of these or other things to focus on?