

Using Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census

Leticia Fernandez, Rachel Shattuck and James Noon
Center for Administrative Records Research and Applications

FCSM 2018

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.

Background

- In many countries, censuses and surveys undercount young children (Anderson 2004; Griffin 2014, O'Hare 2015, 2017)
- Children under age five have been undercounted in U.S. decennial censuses for decades
 - In 2010 Census the 4.6 percent net undercount is larger than for any other age groups
 - Higher for Hispanic and black children
- The persistent undercount of young children has implications
 - Federal/state/local funding for child-related programs
 - Indicators of child well-being

Previous Research

- Researchers with the Census Task Force on the Undercount of Young Children have reviewed and conducted several studies to understand its causes
- Factors that increase young children's risk of being omitted from Census household roster include (O'Hare, Griffin & Konicki 2017, forthcoming)
 - Related to the householder as grandchild, other relative or nonrelative
 - Large or complex households
 - Renter-occupied multi-unit buildings
 - Other characteristics associated with hard-to-count households

Research Questions

What can we learn using administrative records about reasons for the undercount and the characteristics of children under five that are not in 2010 Census?

- Does age misreporting contribute to the undercount?
- Are children missed within housing units covered in Census or is their whole housing unit missed?
- Linking AR to ACS, what can we learn about the characteristics of undercounted children and their households?

Frequently Used Acronyms

- **AR – Administrative records** are collected by federal and state governments in the course of providing services to program participants
 - May supplement Census data collection efforts
 - Children under five are not covered as well as adults
- **PIK – Unique Protected Identification Key** assigned to each individual based on personal identifiers using probability record linkage techniques
 - PIKs not assigned to individuals with insufficient information
- **MAFID – Master Address File Identification** number is an address identifier assigned to each housing unit. A housing unit may contain unrelated individuals or more than one family
 - Some AR files do not have address information and cannot be assigned a MAFID

Administrative Records Composite

- Two different files from Internal Revenue Services (IRS)
- Three files from Housing and Urban Development (HUD)
- Medicare (MEDB) and Medicaid (MSIS)
- Indian Health Service (IHS)
- National Change of Address (NCOA)
- Temporary Assistance for Needy Families (TANF)
- Numerical Identification System (Numident)
- Previous Census Records
- Third party data from four vendors
- 2011 Master Address File (MAF) extract

Methodology

- Children in AR under age five as of April 1, 2010, linked to:
 - 2010 Census by PIK and by MAFID
 - American Community Survey (ACS) 2006-2010 five-year file by PIK
- Bivariate comparisons and logistic regression models exploring factors associated with risk of not matching to Census
 - Variables in the analysis include child characteristics, household-level variables, and tract-level demographic composition

Limitations

Findings from this study cannot be generalized to the U.S. child population

- Children in Census without a PIK are excluded from analysis
- Only a small fraction of children in AR can be linked to the ACS, with unknown biases
- AR and ACS undercount children under age five (Jensen & Hogan 2017; Rastogi & O'Hara 2012)

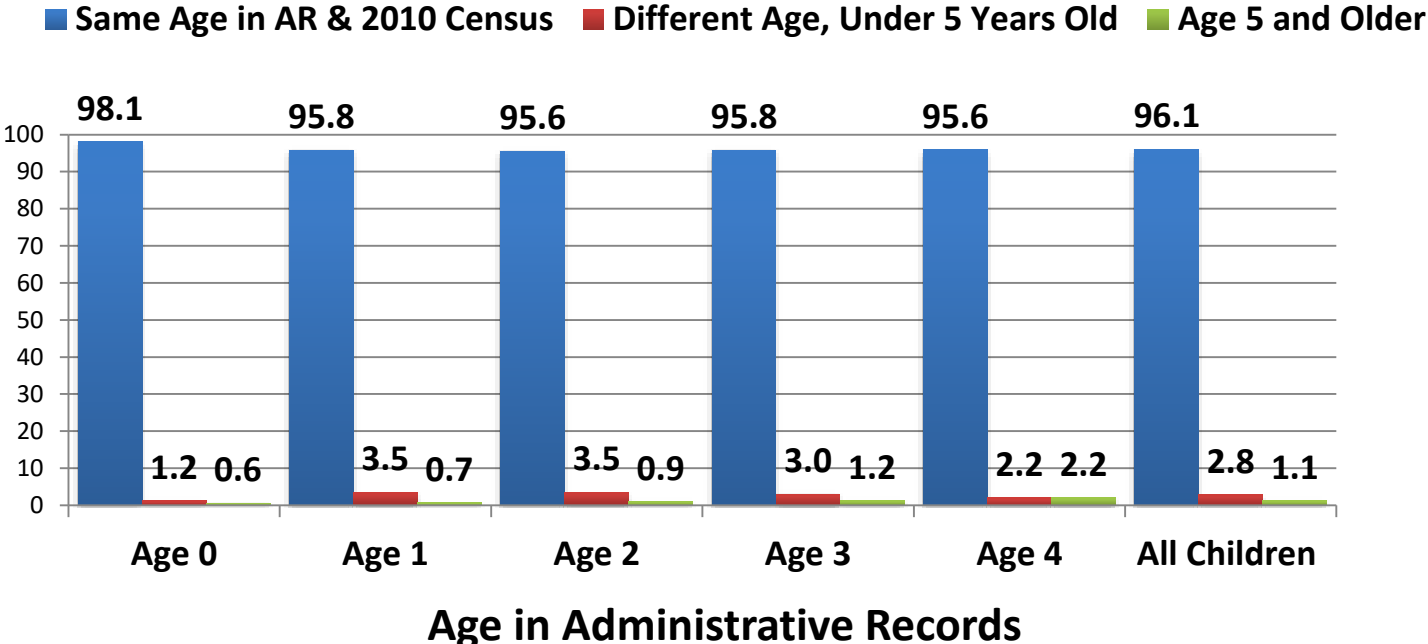
Preliminary Findings

- In AR, all children have a PIK; 77.5 percent have a MAFID
- In Census, all children have a MAFID; 90 percent have PIK
- Compared to children assigned a PIK, unPIKed children in 2010 Census are
 - More likely to be under one year old, racial minorities or Hispanic
 - Less likely to be reported as son/daughter of the reference person or live in a single family home
 - Less likely to be in a self-responder household

Does age misreporting contribute to the undercount?

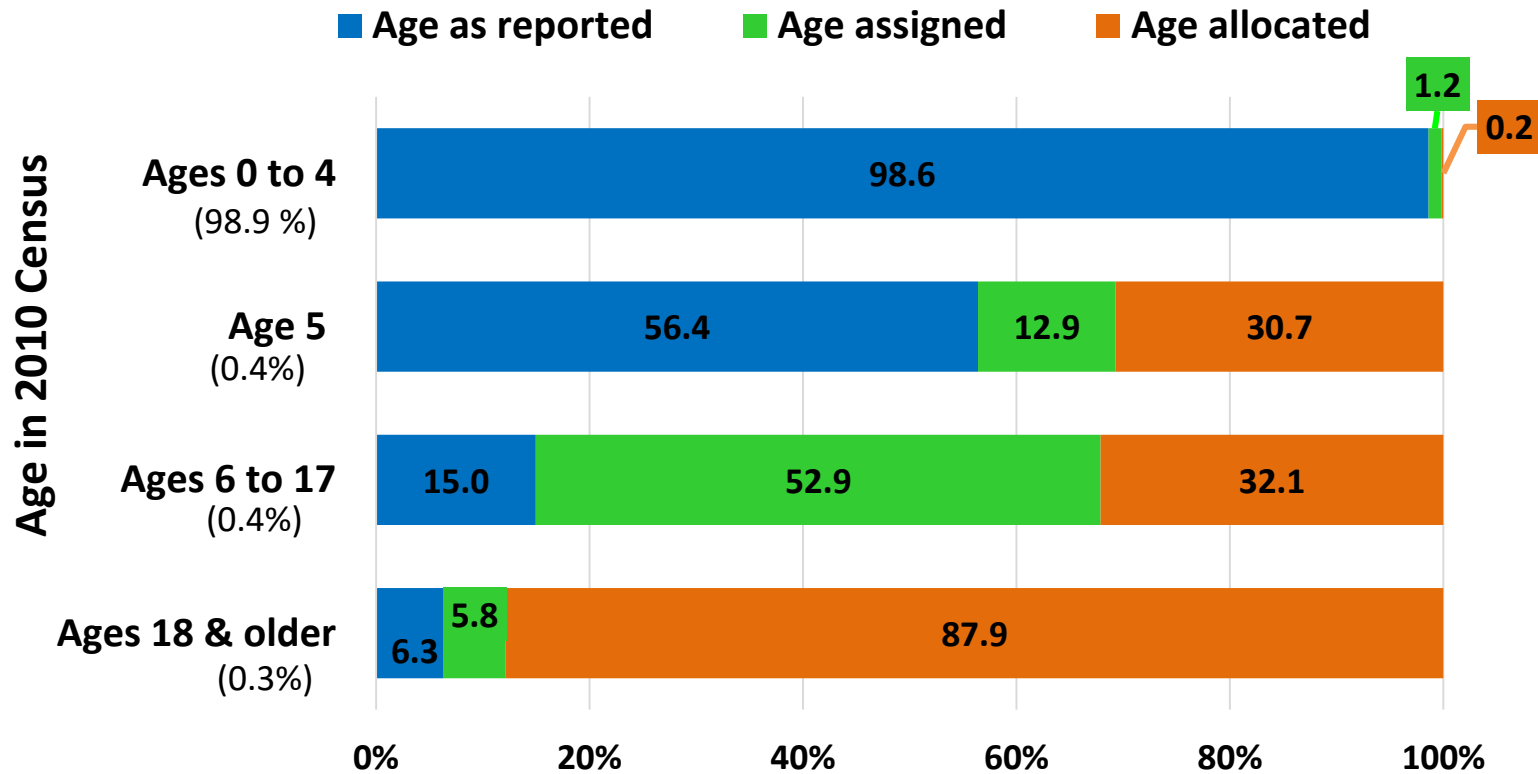
In General, Children in AR who Matched to 2010 Census Have the Same Age

- Out of a total of 20.1 million children in AR, 80 percent were found in 2010 Census (16.2 million)
- 96.1 percent of the children who are in both AR & Census have the same age; 98.8 percent are under age 5 in Census
- 1.1 percent (about 177,000 children) erroneously reported as ages 5 and older



Source: Authors' computations, 2010 Census and AR composite. Numbers rounded to nearest multiple of five to meet disclosure avoidance requirements.

Age Differences for AR Children who Matched to 2010 Census are Mostly from Edit/Allocation Procedures

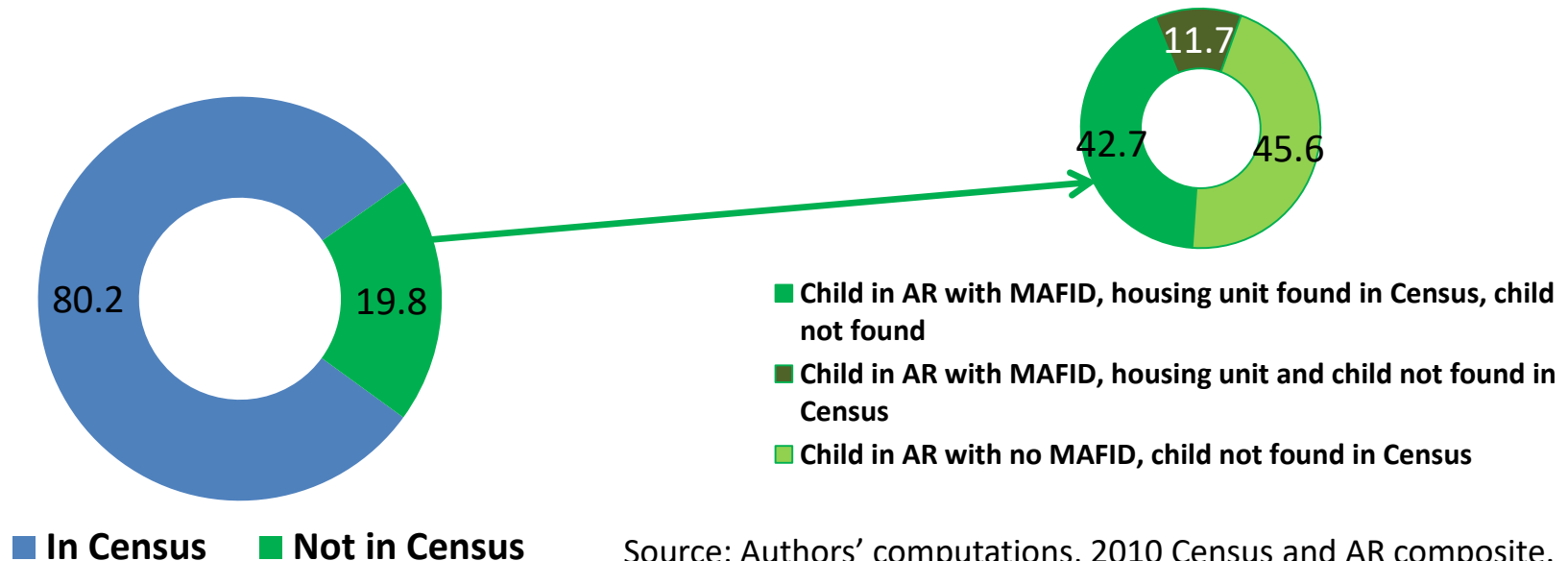


- **As reported** are ages provided by the household respondent.
- **Assigned** ages refer to cases with inconsistent age and DOB.
- **Allocated** values are used when no age is available for a person. Census imputes an age based on nearby persons with similar characteristics.

Are children missed within housing units covered in Census, or are they missed because they are in housing units that were not covered?

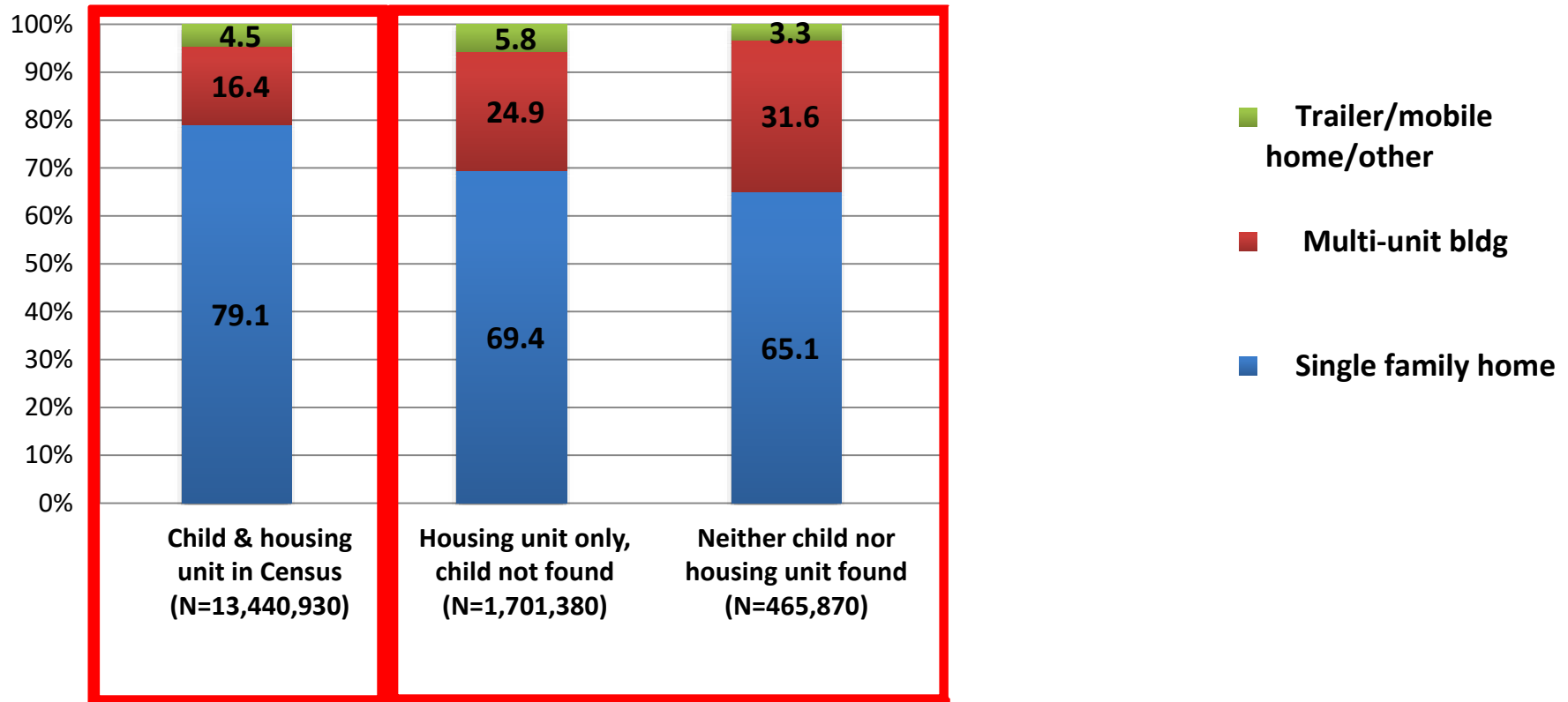
AR Children Ages 0-4 in the 2010 Census

- Out of a total of 20.1 million children in AR, 20 percent were not matched to 2010 Census (about 4 million children)
- 45.6 percent of the children in AR not found in Census could not be assigned a MAFID
- Of those with a MAFID, for 78.5 percent their housing unit was in Census, and for 21.5 neither the child nor the housing unit were found in Census
- Evidence of both, missing children in households covered by Census and missing the whole household



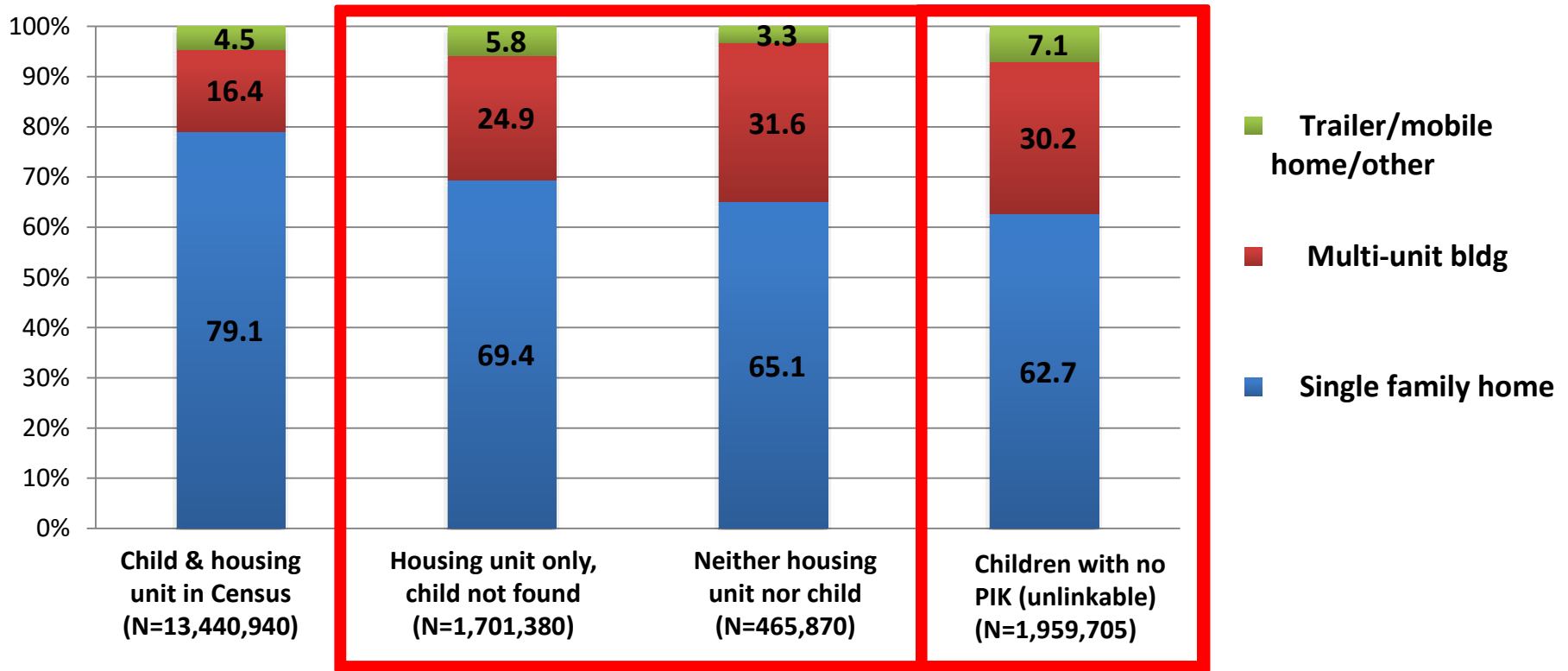
Source: Authors' computations, 2010 Census and AR composite.

Are Undercounted Children Living in Different Housing Unit Types than Children in Census?



- Children in AR with MAFID that were not matched to Census were less likely to live in single-family homes and more likely to live in multi-unit buildings than children in AR with MAFID matched to Census

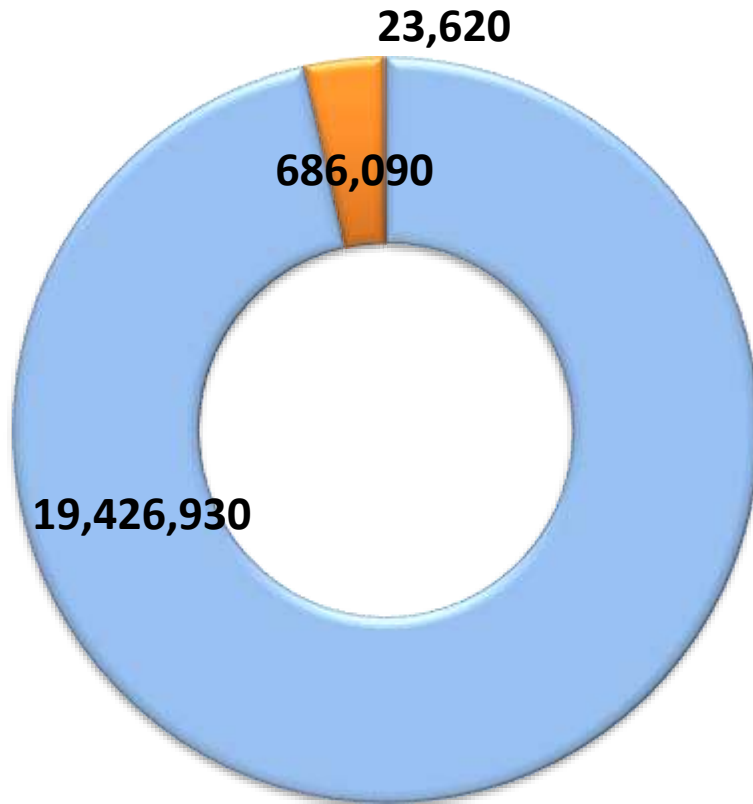
Are Undercounted Children Living in Different Housing Unit Types than Children in Census? - What about Census Children without PIK?



- Children with no PIK in Census (unlinkable) have a similar housing unit type distribution as those who are in AR but did not match to Census
- If all the unlinkable children had a PIK that matched to AR, the percent of AR children found in Census would increase from 80 to 90 percent

Linking AR to ACS, what can we learn about the characteristics of undercounted children and their households?

AR Children Ages 0-4 in the 2006-2010 ACS



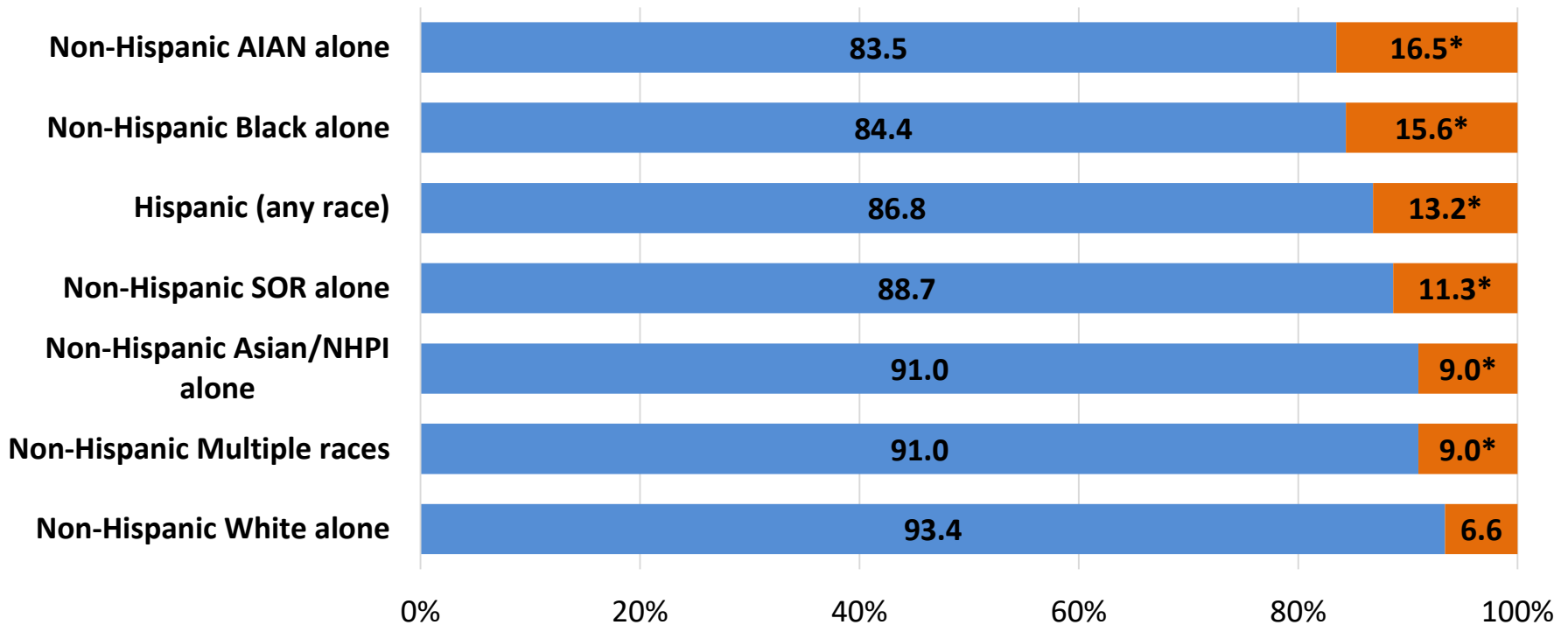
- Child in AR, not in ACS (96.5%)
- Child in AR & ACS, likely matches (3.4%)
- Child in AR & ACS, unlikely matches (0.1%)

- Out of 20.1 million children in AR, 3.4 percent (N=709,710) match to the ACS
- After removing unlikely matches (relationship and age) the AR-ACS sample size is N=686,090
- 90.9 percent (N=623,810) are also in Census and 9.1 percent (N=62,280) are not

Source: Authors' computations, 2010 Census, AR composite and 2006-2010 ACS 5-year file. Numbers rounded to nearest multiple of five to meet disclosure avoidance requirements.

AR-ACS children reported as racial minority or Hispanic are less likely to be found in Census than non-Hispanic white children

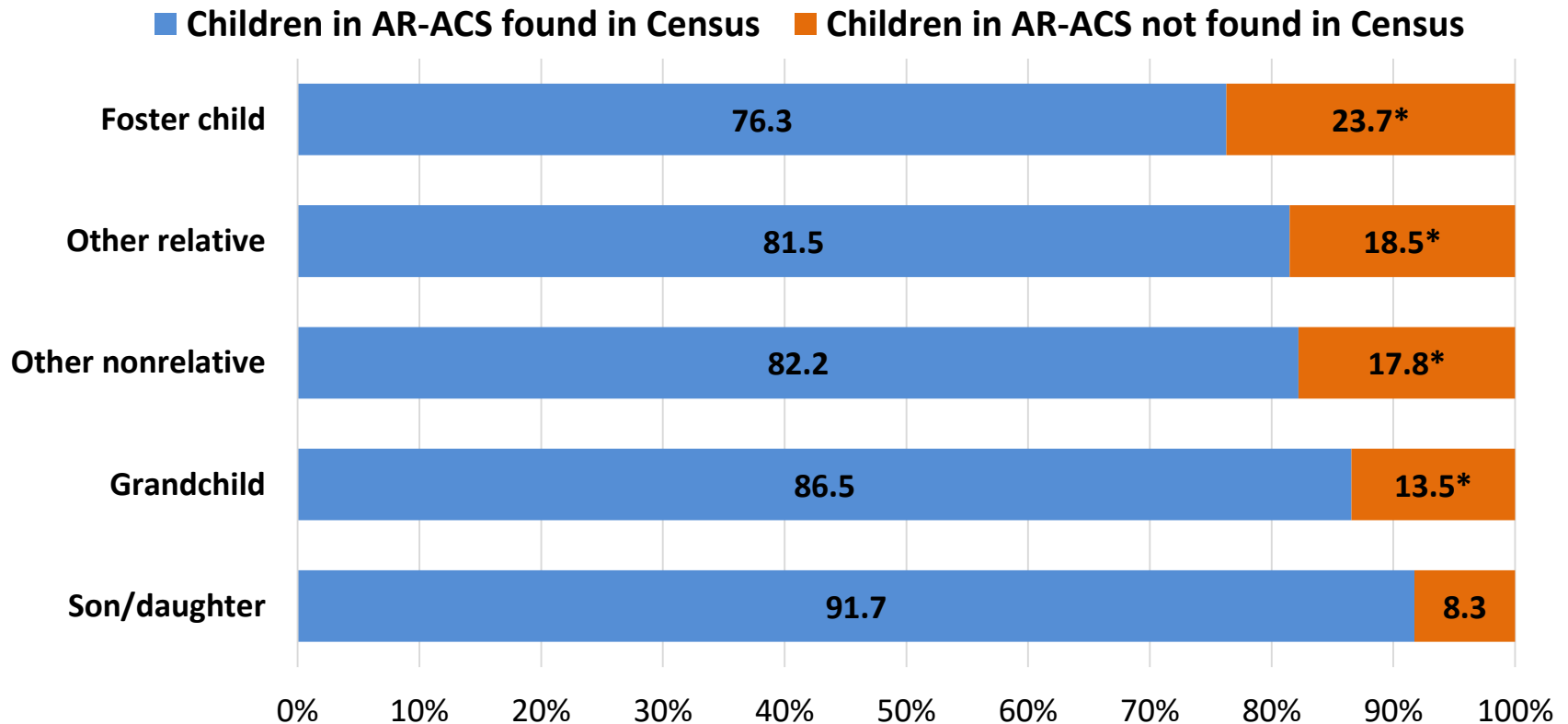
■ Children in AR-ACS found in Census ■ Children in AR-ACS not found in Census



* Statistically significantly higher than for non-Hispanic White alone children.

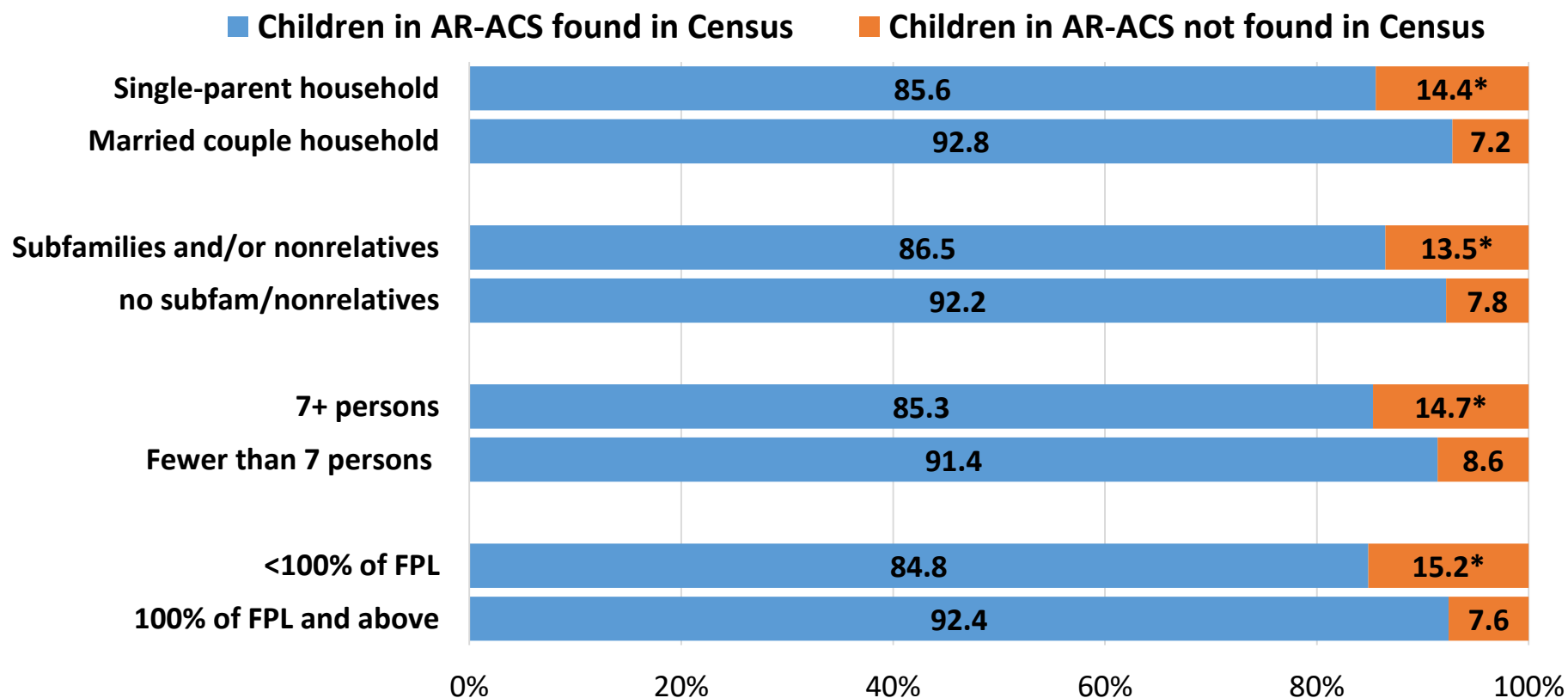
AIAN = American Indian or Alaska Native; SOR = Some Other race; NHPI = Native Hawaiian or Pacific Islander

AR-ACS children less likely to be found in Census if they are foster children, other relative, other non relative or grandchildren than children reported as sons or daughters



* Statistically significantly higher than for children reported as son/daughter of the reference person.
 Other relative: brother/sister, other. Other nonrelative: Roomer/boarder, housemate/roommate, other.

AR-ACS children less likely to be found in Census if they live in single-parent households, complex households, large households or households in poverty than if they live in smaller married-couple households with incomes above poverty level



* Statistically significantly higher than for children in the comparison group

Logistic Regression of the Likelihood of an AR-ACS Child not Matching to 2010 Census, Odds Ratios

| Selected Variables in the Analysis | Hispanic | Non-Hispanic Black | Non-Hispanic White |
|---|----------|--------------------|--------------------|
| Relationship to Reference Person -- Son/daughter (omitted) | | | |
| Grandchild | 1.27** | 1.23** | 1.62** |
| Other (relatives or nonrelatives) | 1.57** | 1.32** | 1.64** |
| Family Type -- Married couple (omitted) | | | |
| Female reference, no spouse | 1.28** | 1.30** | 1.29** |
| Male reference, no spouse | 1.34** | 1.36** | 1.30** |
| Household Size – Fewer than 7 people (omitted) | | | |
| 7 or more people | 1.09** | 1.20** | 1.09** |
| Education of People 25 & Older in the Household | | | |
| At least one adult completed college or higher (omitted) | | | |
| No adult in the household completed college | 1.24** | 1.29** | 1.22** |

*<=.05, **<=.01

Variables in the models also include child age, race/ethnicity, housing unit type, presence of subfamilies and/or nonrelatives, mode of data collection, and tract-level demographic information.

Source: Linked administrative records composite, 2010 Census, and 2006-2010 ACS 5-year file.

(Continued) Logistic Regression of the Likelihood of an AR-ACS Child not Matching to 2010 Census, Odds Ratios

| Selected Variables in the Analysis | Hispanic | Non-Hispanic Black | Non-Hispanic White |
|--|----------|--------------------|--------------------|
| Household Income | | | |
| 300% or above of Federal Poverty Line (FPL) (omitted) | | | |
| Income less than 100% FPL | 1.39** | 1.28** | 1.47** |
| 100% to less than 200% FPL | 1.16** | 1.15** | 1.21** |
| 200% to less than 300% FPL | 1.15** | 1.09 | 1.06** |
| Unemployed People 16 & Older in the Household | | | |
| No Person in the household in the labor force is unemployed (omitted) | | | |
| One or more persons in the labor force are unemployed in the household | 1.05** | 1.08** | 1.07** |
| English Proficiency of People 17 & Older in the Household | | | |
| At least one person speaks English “well” or better (omitted) | | | |
| No one speaks English “well” or better | 1.17** | n/a | n/a |

* $\leq .05$, ** $\leq .01$

Variables in the models also include child age, race/ethnicity, housing unit type, presence of subfamilies and/or nonrelatives, mode of data collection, and tract-level demographic information.

Source: Linked administrative records composite, 2010 Census, and 2006-2010 ACS 5-year file.

Conclusions

- AR can be helpful in identifying the characteristics of children missed by Census
- The undercount of young children seems to arise from multiple factors, including age misreporting, age imputation, housing and household characteristics
- Children in AR & ACS are less likely to match to Census if they are racial/ethnic minorities, if they are reported as grandchildren, other relatives or nonrelatives or if they live in large low-income complex households

Thank You!

leticia.esther.fernandez@census.gov

rachel.m.shattuck@census.gov

james.noon@census.gov