

CARRA Working Paper Series

Working Paper Series #2018 - 05

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

Leticia Fernandez
U.S. Census Bureau

Rachel Shattuck
U.S. Census Bureau

James Noon
U.S. Census Bureau

Center for Administrative Records Research and Applications
U.S. Census Bureau
Washington, DC 20233

Paper Issued: May 25, 2018

Disclaimer: This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. The information in this paper has been approved for release, with Disclosure Review Board number: CBDRB-FY18-102

The Use of 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

U.S. Census Bureau

Abstract

Children under age five are historically one of the most difficult segments of the population to enumerate in the U.S. decennial census. The persistent undercount of young children is highest among Hispanics and racial minorities. In this study, we link 2010 Census data to administrative records from government and third party data sources, such as Medicaid enrollment data and tenant rental assistance program records from the Department of Housing and Urban Development, to identify differences between children reported and not reported in the 2010 Census. In addition, we link children in administrative records to the American Community Survey to identify various characteristics of households with children under age five who may have been missed in the last census. This research contributes to what is known about the demographic, socioeconomic, and household characteristics of young children undercounted by the census. Our research also informs the potential benefits of using administrative records and surveys to supplement the U.S. Census Bureau child population enumeration efforts in future decennial censuses.

Acknowledgements: We wish to thank Lisa M. Blumerman, Sonya R. Porter, Renuka Bhaskar, Joseph C. Powers, and Scott M. Konicki for their valuable comments and suggestions. We are also indebted to Deborah Griffin and Bill O'Hare for their insightful comments throughout this study.

Introduction

National censuses often cover some groups better than others due to various factors, such as differences in response rates, high residential mobility in some groups, difficulties accessing some types of housing units, and other challenges to enumeration. Comparisons of census counts to independent population estimates based on demographic analysis methods that use births, deaths and net migration show that in the United States children under age five have been undercounted for several decades (Bruce et al. 2002; Hacker 2012; O’Hare 2009, 2014, 2015; U.S. Census Bureau 2014). Most recently, the 2010 Census enumerated 20.2 million children ages 0 to 4, but independent estimates based on demographic analysis estimated there should be 21.2 million children in that age range (Konicki 2016; O’Hare et al. 2012; U.S. Census Bureau 2012). This means that the 2010 Census had an estimated net undercount of about one million children or 4.6 percent, higher than for any other age group. The estimated net undercount was larger for Hispanic children (7.5 percent) and black children (6.3 percent) than for all children combined (O’Hare 2015, 2017).

Incomplete enumeration of young children has consequences for federal funding of various programs that serve children’s needs, such as Head Start, the Special Supplemental Program for Women, Infants, and Children (WIC), the Child Care and Development Block Grant, and Foster Care Title IV-E, among others (O’Hare 2014; O’Hare et al. 2016). In addition, the undercount of young children can affect communities’ estimates of their need for programs for children, such as childcare services and schools. The census undercount of young children may also bias certain indicators if the characteristics of children left out of the decennial enumeration differ from those who are in the census counts. Statistics about racial/ethnic distribution, poverty rate, or even living arrangements of young children—all of which reflect various dimensions of child wellbeing—may be inaccurate as a result of the child undercount (Hernandez and Denton 2001).

This study uses administrative records in conjunction with survey data to generate a profile of the characteristics of children who are undercounted in the decennial census. Administrative records are those collected by federal and state agencies in the course of administering programs or providing services to program participants, and include data collected by commercial third parties. Our findings suggest that young children who are Hispanic or racial minorities, reported as other than the son/daughter of the household reference person, and living in large multigenerational or complex low-income households are more likely to be missed by census than their counterparts. Our research provides evidence of the potential utility of administrative records and surveys to assess and supplement child population census enumeration efforts.

Evidence to Date on the Child Undercount

Comparisons of decennial census counts to independent population estimates based on births, deaths and net migration suggest that the census net undercount of the total U.S. population has declined for several decades and even reversed to a net overcount of 0.1 percent in the 2010 Census, which is about 400,000 people (Velkoff 2011). The net undercount is the difference between census omissions and erroneous enumerations (e.g., not counting some individuals and double counting others) in the total population. Independent estimates by age group, however, show that net undercounts vary by age and are more common among younger groups, while net overcounts are more likely at older ages. In particular, O'Hare (2014) reports that for children ages 0 to 4, the net undercount grew from 1.4 percent in 1980 to 4.6 percent in 2010, when it was larger than for any other age group.

Researchers studying the undercount of young children suggest that it results from multiple overlapping factors (U.S. Census Bureau 2014). In particular, young children tend to concentrate, more so than older ones, in certain types of housing units, living arrangements, and communities that may be difficult to reach and more likely missed, such as low-income multi-generational households living in multi-unit buildings in densely populated urban areas (O'Hare 2009, 2015; U.S. Census Bureau 2014; West and Robinson 1999). It may be, for example, that enumerators are more likely to miss households in multi-unit buildings with difficult access. Alternatively, the respondent may omit some children in the census form. Children may be omitted from the roster, for example, because they have more than one usual residence (e.g., parental shared custody) or are in a temporary living arrangement, especially if they are unrelated to the household respondent (Hogan and Robinson 1993; Konicki 2016; O'Hare 2009, 2015; O'Hare et al. forthcoming; Robinson et al. 1993; U.S. Census Bureau 2014). In addition, examination of data from the American Community Survey (ACS) coverage study suggests that impoverished young mothers who are unmarried and not living in their own household may be at higher risk of being missed by the census or ACS than older mothers in better socioeconomic circumstances. This suggests that young children might be missed along with their mothers (Konicki 2016).

There may also be situations in which respondents do not report children in the household because they are concerned about negative repercussions, such as fines for exceeding the number of people allowed in the unit, issues with immigration status, or unwelcomed visits from social service agencies (U.S. Census Bureau 2014; West and Robinson 1999). An additional explanation may be that the coverage prompts in the census forms may be unclear, such that respondents in complex or multi-family households are not sure whether to include a child who sometimes stays at a different housing unit or who is not part of the immediate or nuclear family (O'Hare 2009, 2015).

O'Hare (2009) reports that the undercount of young children varies by state and county size. Arizona, California, Florida, Texas, Georgia and Nevada have some of the highest estimated undercounts of young children, and the undercount is higher in larger populous

counties than in smaller ones. In four of these states, Arizona, California, Texas and Nevada, 25 percent or more of the children ages 0 to 17 live in hard-to-count tracts¹.

A recent report by the Census Bureau Task Force on the Undercount of Young Children calls for additional research using untapped data sources such as administrative records (U.S. Census Bureau 2014). Administrative records matched to decennial census data may be helpful to investigate whether some children may be in the census but reported as older, and thus may erroneously contribute to the undercount of children under age five. In addition, administrative records matched to census can explore whether children are missed because they were omitted in an enumerated household or because the housing unit was missed altogether. Furthermore, by matching administrative records to census and surveys, they may provide new insights or confirm previous findings regarding the type of households and situations where young children are most at risk of being missed.

Some shortcomings of administrative records, however, have become more evident in recent years. The Census Bureau conducted two studies to assess the quality and coverage of persons and addresses in administrative records compared to the 2010 Census and the American Community Survey. These studies included evaluating coverage in administrative records by age. Although differences in coverage across age groups were not statistically tested in these studies, both the 2010 Census Match Study and the 2010 ACS Match Study found that administrative records covered older age groups better than younger ones. In the 2010 Census Match Study, 11.1 percent of (linkable) children ages 0 to 2 and 4.3 percent of children ages 3 to 17 in the 2010 Census were not found in administrative records, higher than for all older age groups. In the ACS Match Study, 6.7 percent of the children ages 0 to 2 and 7.3 percent of children ages 3 to 17 in the 2010 ACS were not found in administrative record, higher than for individuals ages 45 and older (Luque and Bhaskar 2014; Rastogi and O'Hara 2012).

The undercount of young children, however, is not limited to administrative records and decennial counts. A recent analysis of demographic surveys also found that compared to estimates using demographic analysis methods, young children have statistically significantly lower coverage rates in the ACS than older age groups. In addition, the undercount is higher for Hispanic and non-Hispanic black young children in the ACS compared to non-Hispanic white children of the same age. For example, the undercount of young children in the 2009 ACS (90 percent confidence interval in parentheses) was estimated at 14.0 percent (12.8, 15.2) for non-Hispanic black children and 15.0 percent (14.1, 15.9) for Hispanic children, compared to 10.0 percent (9.5, 10.5) for non-Hispanic white children (Jensen and Hogan 2017).

The evaluation studies mentioned here suggest that administrative records, the ACS, and decennial census do not cover young children as well as older ages. It might be that some children are covered in some data sources while omitted in others, depending on factors such as

¹ The Census Bureau developed a hard-to-count index based on twelve tract-level sociodemographic, economic and housing variables associated with low mail response rates. The index combines factors such as the percent of people below poverty, percent of linguistically isolated households, and percent of multi-unit buildings (Bruce et al. 2001; Bruce and Robinson 2003). The hard-to-count index has been used to identify tracts that are difficult to enumerate. In many states, young children are more concentrated in hard-to-count tracts than older individuals (O'Hare 2009).

mode of data collection (e.g., self-response vs. in-person interviewer) or by which member of the household happens to respond to the survey or census. To date, no studies have combined these data to explore the extent to which they may supplement each other and help identify characteristics of undercounted young children in the census. In this study we match 2010 Census data and the 2006-2010 American Community Survey 5-year file to administrative records for children ages 0 to 4 and compare the demographic, socioeconomic and neighborhood characteristics of children who are reported in the 2010 Census to those of children found only in administrative and survey records but not in the census.

Research Questions

We focus on three research questions that have been suggested by previous researchers as salient to understanding the causes and nature of the child undercount in decennial census (U.S. Census Bureau 2014):

1. Are young children present in the census but with an incorrect (older) age, such that they are erroneously considered part of the undercount?
2. Are young children missed because their whole housing unit is missing in census, or are they missed within households that were covered in census?
3. What can we tell about the characteristics of the children and households with children that are at risk of being missed?

Data and Methods

In recent years, the Census Bureau has been researching whether the use of administrative records can supplement survey and census data. The administrative records (AR) composite file used in this study was developed for the 2010 Census Match Study and it is ideal for our study because it combines records from various programs that serve a broad range of the population, including disadvantaged families with young children. The AR composite file combines records from the Internal Revenue Service, Housing and Urban Development, Centers for Medicare and Medicaid, Indian Health Service, the National Change of Address file, Temporary Assistance to Needy Families, and the Social Security Administration Numerical Identification System (Numident) (Rastogi and O'Hara 2012). The AR composite file also includes data from four third-party data vendors. We supplemented the file with two additional years of Medicaid records to increase the likelihood that the youngest children will be in the file (2009 and 2010).

For this study, we restrict the AR composite file to children who are alive and of ages 0 to 4 as of April 1, 2010 ('Census Day'). The age of the children comes from the Numident file, which is the most accurate measure of age since it comes from a child's birth certificate. For 8.1 percent of the children in the sample we use age as reported in Medicaid files because they did not have an age in other AR. Race/ethnicity information comes from the Best Race file created by Census Bureau researchers using an algorithm to resolve inconsistent race and Hispanic

origin responses across various sources of administrative records. The method that was favored selected the smaller race group as the best race response and any report of Hispanic ethnicity as evidence of Hispanic identification. This method was chosen over several others because it yielded the highest rate of consistency in race and Hispanic origin for individuals in AR who matched to the 2010 Census (Ennis et al. 2015). In 6.1 percent of the cases with missing race information in the Best Race file, we were able to use the race that was reported in the Medicaid files, but 13.3 percent of the AR children records had missing race data.

In the first part of the analysis, we linked the AR children file to the 2010 Census at the person level and at the housing unit level to obtain as much information as possible about the characteristics of AR children not found in the 2010 Census. At the person level, we linked AR children to children in the 2010 Census using a unique Protected Identification Key (PIK) based on personally identifiable information and probability record linkage. To protect the confidentiality of responses and respondent anonymity, all personally identifiable information was removed from the files prior to statistical analysis. At the housing unit level, AR records with an address were assigned a housing unit Master Address File identification number (MAFID) from the 2011 Master Address File extract and linked to matching MAFIDs in the 2010 Census housing units. The MAFID is a unique identifier assigned to a housing unit address or location description and contains geographic information of every address known to the Census Bureau.

Since AR contain a limited number of variables, in the second part of the analysis we linked the AR children to the American Community Survey five-year file for 2006-2010 using the same person level identifiers mentioned above (PIKs). This linkage allowed us to explore additional demographic and socioeconomic characteristics of AR children who may have been missed by the 2010 Census. We will refer to this sample as the ‘AR-ACS children’ to distinguish it from the larger AR children file.

There are several limitations in this study. First, at the individual level, records with insufficient personal information such as those lacking date of birth, Social Security Number or name and address are less likely to be assigned a unique identifier (PIK) necessary for linking an individual across files. This is important for our analysis because previous studies find that records that cannot be assigned a PIK are more common among vulnerable and hard-to-count groups, such as Hispanic, non-citizens, and individuals who report low English language proficiency, low levels of education or lower income than their counterparts (Bond et al. 2014).

Second, although only potentially linkable children (those assigned a unique personal identifier or PIK) are included in our AR children sample, not all AR records contain an address that will allow the assignment of a MAFID. In particular, records from Medicaid and Numident files do not have address information. This is not an issue for AR children identified in the 2010 Census by their PIK since we know they were not missed. However, for AR children not found in 2010 Census, not having a MAFID means we are unable to ascertain whether a child was missed with their entire household or omitted in a household that was covered in census.

Third, only a very small fraction of AR children is also in the ACS, and the bias or selectivity associated with their presence is unknown. As mentioned earlier, ACS and other Census Bureau surveys have lower coverage for young children than for older age groups, and

these deficits are larger for ethnic and racial minorities (Jensen and Hogan 2017). These limitations suggest that most likely our analysis underestimates the socioeconomic disadvantages of the children who are left out of surveys and censuses.

Only children with a PIK are included in the AR children file, and preliminary analysis shows that 77.5 percent of these children also have a MAFID. In contrast, in the 2010 Census, all children ages 0 to 4 have a MAFID, and 90.3 percent have a PIK. That is, from the start 9.7 percent of the children ages 0 to 4 in the 2010 Census were excluded from this analysis because they cannot be linked across files. Again, these children are likely to be members of the most undercounted groups (i.e., immigrant or racial minority children, low-income children, and those in unstable residential situations). Appendix Table A shows comparisons between children who were and were not assigned a PIK using the final edited and imputed characteristics for the census records. The children in the 2010 Census who are excluded from the analysis because they could not be assigned a PIK are more likely to be Hispanic (33.3 percent compared to 24.5 percent of those with a PIK) and to report a race other than white (48.6 percent compared to 35.4 percent among those with a PIK). In addition, compared to those with a PIK, children who were not assigned a PIK in the 2010 Census tend to be younger, and less likely to be reported as the child of the household reference person, to live in a single family home, and to live in a household that self-responded by mail. Over 40 percent of the children who did not receive a PIK were enumerated in households that used update/enumerate methods and over one third were enumerated as part of non-response follow up operations. The update/enumerate operations are conducted in communities where housing units do not have conventional mailing addresses or are in remote or rural areas; for example, some American Indian reservations, unincorporated communities along the U.S.-Mexico border (also known as “*colonias*”), and sparsely populated Alaska native villages that do not have street addresses or house numbers. Since these housing units do not receive the mailback census questionnaire, enumerators must collect data in person from these communities (Fallica et al. 2012). The non-response follow up operations involve in-person enumerator interviews of housing units that did receive the mail-back census questionnaire but did not mail it back with their responses (Walker et al. 2012). All frequencies in this paper are rounded to the nearest multiple of five to meet Census Bureau’s disclosure avoidance guidelines.

Findings

The AR composite file contains 20,136,640 children ages 0 to 4 as of April 1, 2010. Linking files by PIK, we find that 80.2 percent (about 16 million) of the AR children are also in the 2010 Census, and 19.8 percent (about 4 million) are not. It is likely that some of the children that were excluded from the analysis because they do not have a linkable identifier (PIK) in the 2010 Census would match to those in the AR children file, but we do not have sufficient information to link them. In a later section, we will compare some of their characteristics to those of the AR children not found in the 2010 Census.

Research question 1: Are some young children included in the census but with an incorrect (older) age?

Age misreporting might be a reason to find fewer children under age 5 in the 2010 Census than in AR. In Table 1, we compare children ages 0 to 4 in AR to the children with a matching PIK in the 2010 Census. Overall, 96.1 percent of the AR children found in the 2010 Census have the same age in both files, and an additional 2.8 percent are under five years old even though their ages differ from those in AR. Therefore, 1.1 percent (about 180,000 children) of the 16 million AR children found in the 2010 Census are older than five. Most likely, although it is outside the scope of this paper, errors in age reporting and edits or imputations also occur at similar rates among children ages 5 and older who may be imputed as ages 0 to 4. If so, the errors may cancel each other out.

It is relevant to take a closer look at the sources of age discrepancies for children under age five in AR to understand whether these are differences due to respondents misreporting the age of the young children in the household (e.g., reporting as age five children who are age four) or due to census edit/allocation protocols. There are three types of edit flags:

- (a) Non imputed values, meaning that the response was given by a household member or by a proxy respondent. In the decennial census, households that do not respond through the mail are visited by an enumerator. If the enumerator is unsuccessful in reaching a household respondent, or if the household refuses to participate, the enumerator may find a neighbor, landlord, building manager, or someone who recently moved into the address, to serve as a proxy (U.S. Census Bureau 2009). In the 2010 Census, proxy responses accounted for more than 13 million cases, about 3 to 4 percent of the full census data (Porter et al. 2015).
- (b) Assigned values, which involve cases in which age and date of birth are inconsistent for a person. Census resolves these types of inconsistencies by assigning an age that takes into account the ages of other individuals in the household; and,
- (c) Allocated values, for cases in which no age is available for a person. Census uses allocation techniques to impute an age based on nearby persons or households with similar characteristics.

The top section of Table 2 shows the source of age consistencies and discrepancies for AR children ages 0 to 4 found in the 2010 Census by age group. For the group with ages 0 to 4 in the 2010 Census, 98.6 percent had their age reported by a household or proxy respondent, not edited or imputed. Only 1.2 percent had inconsistent age and date of birth, which had to be resolved by census assignment procedures; and, less than a quarter of one percent had no age information that resulted in age being allocated by census.

In contrast, for the group of children listed as age five in the 2010 Census, over half (56.4 percent) came from household or proxy respondents' answers, suggesting that rounding up does make a small but positive contribution to age misreporting and to the undercount of young children. Only 12.9 percent of the children reported as age five had their age assigned by census due to inconsistent age and date of birth; but 30.7 percent were (erroneously) allocated by census due to missing information. The percent of children with allocated age was higher for those older

than five. The last column shows that 87.9 percent of the ages erroneously reported as 18 or older in the 2010 Census were the result of allocation.

The bottom section of Table 2 shows another way to look at these data, by the percent contribution of each edit/allocation category. The first row shows that nearly all (99.7 percent) ages “as reported” by household or proxy respondents were under five. The second row shows that among the ages assigned by census due to inconsistent age and date of birth, 79.9 percent were assigned under age 5 and the remaining 20 percent were erroneously resolved as older ages. The third row shows that only 26.9 percent of the census allocated ages were under age 5, and the remaining 73.1 percent of cases were erroneously assigned older ages, suggesting that age allocation may be missing the mark in the youngest age groups with missing information. As we already mentioned, we do not have the full age distribution in AR to examine the extent to which the net result may be that these erroneous age assignments cancel out as older individuals in AR are recorded as under age five in the 2010 Census.

One suggestion would be to use administrative records when available, as supplementing information in decennial census collections may reduce the percent of cases that need allocation of age and other variables, possibly increasing accuracy. A study that compared the consistency of race and ethnicity from both editing/imputation techniques and administrative records to the 2010 Census found higher matching rates for the latter (Rastogi et al. 2014).

Research question 2: Are young children missed by census because their housing unit is missed, or are they missed within households that were covered in census?

Table 3 shows that 22.5 percent of the young children in AR cannot be assigned a MAFID because they don’t have address information. However, there are differences in the percent that do not have a MAFID depending on whether there was a match to a child in the 2010 Census. Among the slightly over 16 million AR children found in the 2010 Census, 16.8 percent did not have a MAFID in AR. In contrast, 45.6 percent of the roughly 4 million AR children that we could not find in the 2010 Census could not be assigned a MAFID. Without a MAFID, we are unable to ascertain whether a child’s housing unit is in the census universe. For the rest of the AR children that did not match to the 2010 Census but do have a MAFID, 1.7 million matched to a housing unit in the 2010 Census and 0.47 million could not be matched to a census housing unit. That is, if we consider only the AR children not found in the 2010 Census but with a MAFID, 78.5 percent may have been missed in households covered by census, and 21.5 percent may have been missed with their whole household. This suggests that both factors may contribute to the undercount of young children.

Research question 3: What can we tell about the characteristics of the AR children that were not matched to the 2010 Census?

Administrative records provide a limited number of variables, such as age, sex, race, ethnicity and housing unit type. The first two columns in Table 4 compare the characteristics of children under age five in the 2010 Census with those of children in administrative records. The percent of children age 0 in the AR file is low relative to the percent of children of the same age

in the 2010 Census and, also, relative to older ages in administrative records and census. It might be that young children are found in administrative records with a time lag because their parents do not apply for services for them until they are somewhat older. Similarly, children born on or after January 1, 2010, are likely to be found in IRS forms the following year when tax forms are filed. Adding more years of AR files in the future may help smooth out their age distribution.

Table 4 shows the comparison of selected characteristics between children under age five in the edited/imputed 2010 Census and AR files. The percent Hispanic is similar in the 2010 Census and AR files, with about one in four of the children reported as Hispanic. Among AR children, however, 6.9 percent have missing Hispanic origin, such that only 66.9 percent are non-Hispanic compared to 74.7 percent in the 2010 Census. In terms of race, a higher percent of children are white alone in the 2010 Census than in AR (63.3 percent and 55.6 percent, respectively); and, while other race groups have a similar distribution in both files, 13.3 percent of the AR children are missing race information. In the unedited census file (not shown), 8.2 percent of the children under five had Hispanic origin and race edited or allocated. The missing rate for Hispanic origin is comparable between AR and the 2010 Census, but the missing rate for race is higher in AR. There are also differences between census and AR in terms of housing unit distribution, mostly because as mentioned earlier, in AR 22.5 percent of the children do not have a MAFID.

The last two columns in Table 4 show differences between AR children who were found and not found in the 2010 Census. The AR children that were not found in the 2010 Census (last column) were younger, more likely to be Hispanic, less likely to be white alone, and more likely to be missing race responses than AR children found in the 2010 Census. The percent of AR children with missing information for race, ethnicity and MAFID is much higher among AR children not found in the 2010 Census than among their counterparts. That is, information about the children not matched to the 2010 Census is less complete even in the AR composite file, suggesting that they are harder to reach for reasons that are not obvious given the limited variables we have in the AR composite file.

In Table 5, we show the distribution of housing unit type only for AR children whose records have a MAFID. The first column shows that 79.1 percent of the children in both AR and the 2010 Census live in a single-family home and 9.1 percent live in large (10 or more) multi-unit buildings. The second and third columns show that children not found in the 2010 Census are less likely to live in a single-family home and more likely to live in large multi-unit buildings than those found in the census, with small differences between AR children whose housing unit was and was not matched to the 2010 Census..

The last column in Table 5 shows the distribution of housing unit type for the children in the 2010 Census who were excluded from the analysis because they could not be assigned a linkable unique identifier (PIK). Their housing unit distribution is similar to that of children in AR who were not matched to census. It is likely that if we could assign a PIK to all children in census, many of these children would match across the two files. However, even if all the unlinkable children in the 2010 Census were matched to AR, there would still be about two million children in AR not matched to census.

In the next section, we explore a broader range of characteristics associated with children under five in AR who are and are not in the 2010 Census. We do this by linking the AR

composite file to the ACS five-year file for 2006-2010. Because only a small fraction of the AR-census records can be linked to ACS, our findings are only suggestive of the differences between AR children found and not found in census.

There were N=709,710 children in AR who matched to a child in the ACS. After removing unlikely matches (i.e., children who are older than six, children that are likely to be adults because of their relationship to the householder, and a small number of children in non-family households), the remaining sample contains N=686,090 children. We refer to these children as 'AR-ACS children' to distinguish them from the 20 million children in AR. About ninety-one percent of AR-ACS children are found in the 2010 Census, and about nine percent are not. This is higher than the 80.2 percent of all AR children who are found in the census, which suggests that families who participate in sample surveys are also more likely to participate in the census.

Table 6 shows individual, household and community differences between AR-ACS children by whether they match to a child in the 2010 Census. Compared to the AR-ACS children found in census, those not found were younger and more likely to be members of race/ethnic minority groups, specifically Hispanic, black alone, or American Indian or Alaska Native. They were also less likely to be reported as son/daughter of the household reference person, and more likely to be reported as grandchild, other relative, foster child or other non-relative, suggesting they are in temporary and/or doubled-up housing situations. There were also differences in the housing unit type, with AR-ACS children found in census more likely to live in a single family home and less likely to be missing MAFID information than those not matched to census.

At the household level, our findings suggest that AR-ACS children who are not found in the census are more disadvantaged than those who are in the census. AR-ACS children not found in census were more likely to live in households with incomes that place them in the groups for 'less than 100 percent' and '100 to less than 200 percent' of the federal poverty guidelines (FPG). They were also more likely to live in single-parent households, and to have non-relatives or subfamilies living in the household than those found in census. In addition, AR-ACS children not matched to census were more likely to live in large households (7 or more persons) and households where no adult completed college or higher education, more likely to have unemployed adults in the household, and more likely to live with adults with limited English proficiency.

In addition, the AR-ACS children not found in census were less likely to live in households that self-responded to the ACS by mail (43.5 percent, compared to 65.6 percent among children found in census), and more likely to have responded during follow up operations, either in a telephone interview or a personal visit from a field interviewer. This is consistent with other studies that find that Hispanic, black and American Indian or Alaska Native are less likely than other racial/ethnic groups to self-respond by mail to the initial ACS paper questionnaire (National Academy Press 2015).

At the census tract-level, AR-ACS children not matched to census were more likely to live in racially diverse neighborhoods with a lower percent of college-educated adults, and a higher percent of unemployed individuals. Table B in the appendix shows that the differences

between AR-ACS children who are and are not in census are similar within Hispanic, non-Hispanic black and non-Hispanic white children.

To see if these associations remain significant once other factors are taken into account, we modeled the likelihood that an AR-ACS child will not match to the 2010 Census. The logistic regressions included the full sample and also three separate groups, Hispanic, non-Hispanic black and non-Hispanic white children. The dependent variable in each of these regressions was coded '1' if the child was not found in the 2010 Census and '0' otherwise.

Table 7 shows findings from the full AR-ACS linked sample and the separate regressions by race and Hispanic origin. The coefficients in the first regression show that the odds that a child in the AR-ACS sample is not matched to the 2010 Census are higher for Hispanic children, racial minority children, and children under age 1 than for non-Hispanic white and older children. In addition, children listed as grandchildren or other relatives are less likely to match to census than those reported as son/daughter of the household reference person. Foster children and other non-relatives are not significantly less or more likely to be in the census than children reported as sons/daughters.

In the separate regressions for Hispanic, non-Hispanic black and non-Hispanic white children, we combined all children other than those reported as son/daughter (the reference category) and grandchildren because of the small numbers in the other categories. The coefficients in these regressions show that the odds of not being in the 2010 Census are higher for any children not reported as son/daughter of the household reference person.

Net of other factors, AR-ACS children are less likely to be in the 2010 Census if they live in multi-unit buildings of any size or in large households rather than in single-family housing units and smaller household sizes. In addition, children in poverty and those living in a single-parent household are less likely to be found in the 2010 Census. The presence of nonrelatives and subfamilies in the household does not seem to increase the risk of omitting sons/daughters of the reference person, but increases the risk of omitting other children in the household.

In terms of other household characteristics, children in households with one or more unemployed individuals, and in households where none of the adults have a college degree or higher education were less likely to be found in the census than their counterparts. The models also include the percent of foreign-born individuals in the household, and whether at least one individual age 17 or older spoke English well or very well. For Hispanic and non-Hispanic white children, but not for non-Hispanic black children, living in households where half or more of the residents are foreign born is associated with lower odds of matching to census. Among Hispanics only, children in households where all persons age 17 or older speak English not well or not at all were less likely to be found in the 2010 Census. Combined, these findings suggest that for our AR-ACS sample, children not found in census are more likely to live in hard-to-count households than children matched to census. Note that AR-ACS children in households interviewed through CATI/CAPI² are less likely to match to census. One possible explanation is

² Prior to the introduction of an internet mode in 2013, the ACS data collection operation consisted of three modes: mail, computer-assisted telephone interview (CATI), and computer-assisted personal interview (CAPI). In the first phase, the ACS questionnaire was mailed to households in the sample with a request to complete and return it by

that in the case of households not self-responding by mail, decennial enumerators may be less likely to obtain a complete household enumeration from the household than ACS field representatives or they may have to obtain information from proxy respondents (i.e., neighbors or landlords). Census enumerators tend to be short-term personnel with fewer hours of training. ACS interviewers, in contrast, tend to have more training and experience.

Finally, at the tract level, increases in the percent of the population that are Hispanic or non-Hispanic black are associated with lower odds of a child being in census, but the magnitude of the coefficients vary by the race and Hispanic origin of the child. In particular, increases in the percent Hispanic in a tract is not associated with the odds of not finding a Hispanic child in the census, but increases the odds that non-Hispanic white and non-Hispanic black children are not found in census. Similarly, increases in the percent black in a tract has the smallest coefficient for black children, and the largest increase in the odds of not finding non-Hispanic white children in census.

Conclusions

Persistent differences in census coverage based on age are common in many countries (O'Hare 2017). Children and young adults, particularly males, are more likely to be undercounted than other groups. However, in the United States, the net undercount of children under age five has increased since 1980, and it was larger than for any other age group in the 2010 Census. Incomplete enumeration of children under age five may result in reduced federal, state and local funding for programs that serve young children. In addition, indicators of wellbeing, such as the percent of young children living in poverty, may be biased if unreported children differ from those accounted for in the census. There is evidence from demographic analysis methods that the undercount is larger for Hispanic and black children than for all children combined (O'Hare 2015).

Our comparison of AR with the 2010 Census finds that a non-trivial percent of children who are in both AR and census are recorded with older ages in census, which may explain some of the undercount. These age discrepancies arise from both household respondents misreporting a child's age as well as from census edits and allocations. Since we do not compare the full age distribution in AR to the ages reported in the 2010 Census, we cannot ascertain whether the net result may be that erroneous age assignments cancel out by older children being reported as younger.

In terms of the young children in AR that are not found in the 2010 Census, we find that they are less likely to live in single-family homes and more likely to live in multi-unit buildings than those found in census, which can be more challenging for enumerators' access. We find some evidence that some of the children in the AR population that are not in census may have been omitted from households that were covered in census, while others may have been missed with their whole household.

mail. In the second phase, if no response was received within some weeks, the Census Bureau followed up with a telephone interview if a number was available. In the last phase, a sample of households that could not be reached by phone and households that refused participation were selected for an in-person interview.

One of the limitations in using administrative records to study the characteristics of children not found in census is the limited number of variables. In addition, in many cases the percent of missing values (for example in race and Hispanic origin) is higher for AR children not found in census than for those in both AR and census.

In the AR-ACS linked sample, children who are reported as grandchildren, other relatives or nonrelatives of the household reference person, those living in large households or households in poverty, and those living with a single parent are less likely to be found in the 2010 Census than their counterparts. In addition, children in households with unemployed individuals, adults with low-levels of education or with low English language proficiency are less likely to be found in the 2010 Census. These are also some of the factors that have been identified as characteristics of hard-to-count households (Bruce et al. 2001; Bruce and Robinson 2003).

Our analysis should be taken with caution because of the small sample size resulting from combining AR and ACS data, which likely contains biases inherent in the process of assigning unique identifiers used for linking. However, our findings are consistent with other studies using a variety of different methods (O'Hare et al., forthcoming; U.S. Census Bureau 2017). Therefore, we find that linking administrative records and surveys is useful to identify demographic, socioeconomic, and neighborhood characteristics of young children that may be missed in the census. Moreover, the bias inherent in our sample likely understates the extent to which children who are missed by the census come from disadvantaged groups.

The consensus expressed by researchers focusing on the undercount of young children is that there is no single underlying cause, and that there will be no single solution (O'Hare et al. 2017, forthcoming; U.S. Census Bureau 2014, 2017). In cases of unintentional omissions, it may be that complex living situations, temporary living arrangements or shared custody of a child result in ambiguous household membership, such that the household or proxy respondent may not be sure whether to include a particular child (Martin 2007).

Qualitative research is needed to understand the various reasons children may be unreported, as well as to test strategies to encourage the complete enumeration of household members. Such research should be conducted in various languages in addition to English, and focus on low-income neighborhoods with Hispanic and racial minority populations.

In future research, linking census and administrative records to the Census Planning Database, which contains demographic and socioeconomic information at the tract level of aggregation, may help identify communities where children may be at higher than average risk of being undercounted. In these areas, census could develop partnerships with local agencies and organizations serving young children and their families to distribute materials encouraging participation in the census and the reporting of all children in the household (O'Hare 2009). Examples of promising partnerships include childcare centers, elementary schools and after-school programs, the Women, Infants and Children (WIC) program, children's clinics, public libraries, and community centers.

References

- Bond, Brittany, J. David Brown, Adela Luque and Amy O'Hara. 2014. The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey. Working Paper CARRA-WP-2014-08. U.S. Census Bureau.
- Bruce, Antonio, J. Gregory Robinson and Monique V. Sanders. 2001. Hard-to-Count Scores And Broad Demographic Groups Associated With Patterns of Response Rates In Census 2000. In *Proceedings of the Social Statistics Section*, American Statistical Association.
- Bruce, Antonio, Arjun Adlakha, Peter Johnson and J. Gregory Robinson. 2002. U.S. Historical Profile of Demographic Analysis and Population Estimates: Components across Time, 1935 to 2000. Population Association of America, Conference Poster Session, Atlanta, GA.
- Bruce, Antonio and J. Gregory Robinson. 2003. The Planning Database: Its Development and Use as an Effective Tool in Census 2000. Paper Presented at the Annual Meeting of the Southern Demographic Association, Arlington, VA.
- Ennis, Sharon R., Sonya R. Porter, James M. Noon and Ellen Zapata. 2015. When Race and Hispanic Origin Reporting Are Discrepant across Administrative Records and Third Party Sources: Exploring Methods To Assign Responses. Working Paper CARRA-WP-2015-08. U.S. Census Bureau.
- Fallica, Heather, Sarah Heimel, Geoff Jackson and Bei Zhang. 2012. 2010 Census Update Enumerate Operations Assessment. Update Enumerate Production, Update Enumerate Quality Control, Remote Update Enumerate, and Remote Alaska. 2010 Census Program for Evaluations and Experiments. U.S. Census Bureau.
- Hacker, J. David. 2012. New Estimates of Census Coverage in the United States: 1850-1930, *Social Science History*, 37(1): 71-101.
- Hernandez, Donald and Nancy Denton. 2001. Census Affects Children In Poverty. Memo Prepared for the U.S. Census Bureau Monitoring Board. U.S. Census Bureau.
- Hogan, Howard and J. Gregory Robinson. 1993. What the Census Bureau's Coverage Evaluation Programs Tell Us about Differential Undercount. Paper Presented at the Research Conference on Undercounted Ethnic Populations, May 5-7, Richmond, VA.
- Jensen, Eric B., and Howard R. Hogan. 2017. The Coverage of Young Children in Demographic Surveys. *Statistical Journal of the IAOS*, 33: 321-333.
- Konicki, Scott. 2016. The Undercount of Young Children in the Decennial Census. Presentation for the 2020 Census Program Management Review. U.S. Census Bureau.
- Luque, Adela and Renuka Bhaskar. 2014. 2010 American Community Survey Match Study. Working Paper CARRA-WP-2014-03. U.S. Census Bureau.

- Martin, Elizabeth. 2007. Strength of attachment: Survey Coverage of People with Tenuous Ties to Residences. Research Report Series in Survey Methodology #2007-26. U.S. Census Bureau.
- National Academy Press. 2015. Realizing the Potential of the American Community Survey: Challenges, Tradeoffs, and Opportunities. Panel on Addressing Priority Technical Issues for the Next Decade of the American Community Survey, Committee on National Statistics, Division of Behavioral And Social Sciences And Education. Washington, DC: The National Academies Press.
- O'Hare, William P. 2009. Why Are Young Children Missed So Often In The Census? Kids Count Working Paper. Baltimore, MD: The Annie E. Casey Foundation.
- O'Hare, William P. 2014. Historical Examination of Net Coverage Error for Children in the U.S. Decennial Census: 1950 To 2010. Center for Survey Measurement Working Paper #2014-03. U.S. Census Bureau.
- O'Hare, William P. 2015. *The Undercount of Young Children in the U.S. Decennial Census*. New York City, NY: Springer.
- O'Hare, William P. 2017. An International Perspective on the Undercount of Young Children in the U.S. Census. *Statistical Journal of the IAOS*, 33: 289-304.
- O'Hare, William P., Yeris Mayol-Garcia, Elizabeth Wildsmith and Alicia Torres. 2016. The Invisible Ones: How Latino Children are Left Out of Our Nation's Census Count. Child Trends Hispanic Institute.
- O'Hare, William P., J. Gregory Robinson, Kirsten West, Thomas Mule. 2012. Comparing Demographic Analysis and Dual-Systems Estimates, Results for Children. Paper Presented at the Southern Demographic Association Conference, Williamsburg VA, October 11-12.
- O'Hare, William P., Deborah Griffin and Scott Konicki. (Forthcoming). Investigating the 2010 Undercount of Young Children – Summary of Recent Research. Report from the Task Force on the Undercount of Young Children. U.S. Census Bureau.
- Porter, Sonya R., Carolyn A. Liebler and James M. Noon. 2015. An Outside View: What do Observers Say about Others' Races and Hispanic Origins? Working Paper CARRA-WP-2015-05. U.S. Census Bureau.
- Rastogi, Sonya and Amy O'Hara. 2012. Census Match Study. 2010 Census Program for Evaluations and Experiments, 2010 Planning Memoranda Series No. 247, U.S. Census Bureau.

- Rastogi, Sonya, Leticia Fernandez, James Noon, Ellen Zapata and Renuka Bhaskar. 2014. Exploring Administrative Records Use for Race and Hispanic Origin Item Non-Response. Working Paper CARRA-WP-2014-16. U.S. Census Bureau.
- Robinson, J. Gregory, Bashir Ahmed and Edward W. Fernandez. 1993. Demographic Analysis as an Expanded Program for Early Coverage Evaluation of the 2000 Census. Paper Presented at the Annual Research Conference, March 21-24, Arlington, VA.
- U.S. Census Bureau. 2009. *2010 Census Nonresponse Followup (NRFU) Enumerator Manual*. U.S. Census Bureau.
- U.S. Census Bureau. 2012. 2010 Demographic Analysis. Accessible at http://www.census.gov/popest/research/da-estimates/Table_3.pdf.
- U.S. Census Bureau. 2014. *The Undercount of Young Children*. U.S. Census Bureau.
- U.S. Census Bureau. 2017. *Investigating the 2010 Undercount of Young Children – Analysis of Census Coverage Measurement Results*. Decennial Statistics Studies Division.
- Velkoff, Victoria A. 2011. Demographic Evaluation of the 2010 Census. Paper Presented at the 2011 Population Association of America Annual Conference. Washington, DC.
- Walker, Shelley, Susanna Winder, Geoff Jackson and Sarah Heibel. 2012. 2010 Census Nonresponse Followup Operations Assessment. 2010 Census Program for Evaluations and Experiments. U.S. Census Bureau.
- West, Kirsten K. and J. Gregory Robinson. 1999. What Do We Know About The Undercount Of Children? Population Division Working Paper No. 39. U.S. Census Bureau.

| Table 1. Age Comparison Between Children 0 to 4 in Administrative Records and Their Match in the 2010 Census | | | |
|---|---|----------------------------------|-------------------------------------|
| | Age in the 2010 Census Relative to Administrative Records, Percent§ | | |
| Age in Administrative Records | Same Age in the 2010 Census | Different Age, Under 5 Years Old | Different Age, 5 Years Old or Older |
| Age 0 (N=1,852,230) | 98.1 | 1.2 | 0.6 |
| Age 1 (N=3,491,180) | 95.8 | 3.5 | 0.7 |
| Age 2 (N=3,614,770) | 95.6 | 3.5 | 0.9 |
| Age 3 (N=3,630,300) | 95.8 | 3.0 | 1.1 |
| Age 4 (N=3,562,710) | 96.0 | 2.0 | 2.1 |
| Total (N=16,151,190) | 96.1 | 2.8 | 1.1 |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines. | | | |
| Source: Linked administrative records composite and the 2010 Census. Authors' computations. | | | |

| Table 2. Source of Age for Children 0 to 4 in Administrative Records who Matched to the 2010 Census, by Age & Edit/Allocation Category from the Census Edited File, Percent§ | | | | | |
|--|-------------------------------------|-------|--------------|-----------------|-------|
| Source of Age Reported in Census, Distribution by Age | Reported in the 2010 Census as | | | | |
| | Ages 0 to 4 | Age 5 | Ages 6 to 17 | Ages 18 & older | |
| Age as reported (not imputed) | 98.6 | 56.4 | 15.0 | 6.3 | |
| Age assigned, inconsistent age/date of birth | 1.2 | 12.9 | 52.9 | 5.8 | |
| Age allocated from hot deck | 0.2 | 30.7 | 32.1 | 87.9 | |
| Total | 100.0 | 100.0 | 100.0 | 100.0 | |
| Source of Age Reported in Census, Distribution by Edit/Allocation Category | Reported in the 2010 Census as | | | | |
| | Ages 0 to 4 | Age 5 | Ages 6 to 17 | Ages 18 & older | Total |
| Age as reported (not imputed) | 99.7 | 0.3 | 0.1 | <0.1 | 100.0 |
| Age assigned, inconsistent age/date of birth | 79.9 | 3.9 | 15.2 | 1.0 | 100.0 |
| Age allocated from hot deck | 26.9 | 19.8 | 19.8 | 33.5 | 100.0 |
| § Rows & columns may not add to 100 percent due to rounding error. | | | | | |
| Source: Linked administrative records composite and the 2010 Census. Authors' computations. | | | | | |

| Table 3. Children Ages 0 to 4 in Administrative Records by Presence in the 2010 Census§ | | |
|---|-------------------|---------|
| | Number | Percent |
| AR Children Ages 0 to 4 | 20,136,640 | 100.0 |
| With MAFID | 15,605,895 | 77.5 |
| With no MAFID | 4,530,745 | 22.5 |
| AR Children found in the 2010 Census | 16,151,190 | 100.0 |
| With MAFID | 13,440,940 | 83.2 |
| With no MAFID | 2,710,250 | 16.8 |
| AR Children not found in the 2010 Census | 3,985,450 | 100.0 |
| With MAFID - housing unit found in census | 1,701,380 | 42.7 |
| With MAFID - housing unit not found in census | 465,870 | 11.7 |
| With no MAFID | 1,818,200 | 45.6 |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines. | | |
| Source: Linked administrative records composite and the 2010 Census. Authors' computations. | | |

| Table 4. Selected Characteristics Comparing Children 0 to 4 in the 2010 Census and Administrative Records Before and After Linkage, Percent§ | | | | |
|---|-----------------------------|---------------------|--------------------------|------------------------------|
| Variables | Children in the 2010 Census | Children in AR | Children in AR and ... | |
| | | | found in the 2010 Census | not found in the 2010 Census |
| Number Percent | 20,201,275 100.0 | 20,136,640 100.0 | 16,151,190 80.2 | 3,985,450 19.8 |
| Age in AR, Percent | | | | |
| Less than one year old | 19.5 | 12.8 | 11.5* | 18.2 |
| One year old | 19.7 | 21.6 | 21.6* | 21.3 |
| Two years old | 20.3 | 22.1 | 22.4* | 21.0 |
| Three years old | 20.4 | 22.0 | 22.5* | 20.3 |
| Four years old | 20.1 | 21.5 | 22.1* | 19.2 |
| Hispanic Origin in AR, Percent | | | | |
| Hispanic | 25.3 | 26.2 | 25.3* | 29.9 |
| Non-Hispanic | 74.7 | 66.9 | 72.8* | 42.8 |
| Missing | 0.0 | 6.9 | 1.9* | 27.4 |
| Race in AR, Percent | | | | |
| White alone | 63.3 | 55.6 | 62.5* | 27.6 |
| Black alone | 14.4 | 14.5 | 14.0* | 16.6 |
| American Indian or Alaska Native alone | 1.2 | 1.3 | 1.2* | 1.5 |
| Asian alone | 4.4 | 3.7 | 4.2* | 1.8 |
| Native Hawaiian or Pacific Islander alone | 0.2 | 0.3 | 0.3* | 0.6 |
| Some Other Race alone | 9.5 | 6.0 | 7.1* | 1.5 |
| Multiple races | 6.9 | 5.2 | 6.3* | 1.0 |
| Missing | 0.0 | 13.3 | 4.4* | 49.4 |
| Housing Unit Type in 2011 MAF Extract | | | | |
| Single family home | 74.4 | 60.1 | 65.8* | 37.2 |
| Trailer/mobile home/other | 5.7 | 3.6 | 3.8* | 2.8 |
| 2-9 multi-unit building | 9.3 | 6.1 | 6.0* | 6.5 |
| 10-19 multi-unit building | 2.9 | 2.1 | 2.1* | 2.2 |
| 20+ multi-unit building | 7.7 | 5.5 | 5.5* | 5.7 |
| No MAFID | 0.0 | 22.5 | 16.8* | 45.6 |
| * p<=.05 comparing AR children who match and do not match to the 2010 Census. | | | | |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines. | | | | |
| Source: Linked administrative records composite and the 2010 Census. Authors' computations. | | | | |

| Table 5. Housing Unit Type for Children 0 to 4 with MAFID Information, Percent§ | | | | |
|---|-------------------------|--|---------------------------------------|--------------------------------|
| Housing Unit Type | Children in AR & census | Children not in census, housing unit in census | Children & housing unit not in census | Children in census with no PIK |
| Number | 13,440,940 | 1,701,380 | 465,870 | 1,959,705 |
| Percent | 100.0 | 100.0 | 100.0 | 100.0 |
| Single family home | 79.1 | 69.4 | 65.1 | 62.7 |
| Trailer/mobile home/other | 4.5 | 5.8 | 3.3 | 7.1 |
| 2-9 multi-unit building | 7.3 | 10.9 | 15.7 | 14.2 |
| 10-19 multi-unit building | 2.5 | 3.8 | 4.5 | 4.4 |
| 20+ multi-unit building | 6.6 | 10.1 | 11.4 | 11.6 |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines. | | | | |
| Source: Linked administrative records composite and the 2010 Census. Authors' computations. | | | | |

Table 6. Selected Characteristics Comparing Children 0 to 4 in Administrative Records Linked to ACS by Whether they Were Matched to the 2010 Census, Percent§

| Variables in the Analysis | Children in AR-ACS Matched to the 2010 Census | Children in AR-ACS Not Matched to the 2010 Census |
|---|---|---|
| Number | 623,810 | 62,280 |
| Percent | 90.9 | 9.1 |
| Age in AR | | |
| Less than one year old | 3.0* | 6.8 |
| One year old | 13.6* | 14.1 |
| Two years old | 21.6 | 21.4 |
| Three years old | 28.4* | 27.1 |
| Four years old | 33.5* | 30.5 |
| Hispanic Origin and Race in AR[^] | | |
| Hispanic | 18.6* | 28.2 |
| Non-Hispanic White alone | 62.9* | 44.6 |
| Non-Hispanic Black alone | 8.8* | 16.3 |
| Non-Hispanic American Indian or Alaska Native alone | 1.1* | 2.2 |
| Non-Hispanic Asian/NHPI alone | 4.1 | 4.1 |
| Non-Hispanic Some Other Race alone | 0.2* | 0.3 |
| Non-Hispanic multiple races | 4.3 | 4.3 |
| Relationship to Reference Person | | |
| Son/daughter | 87.8* | 79.1 |
| Grandchild | 10.0* | 15.6 |
| Other relative | 1.6* | 3.6 |
| Foster child | 0.3* | 0.9 |
| Other non-relative | 0.4* | 0.8 |
| Housing Unit Type | | |
| Single family home | 74.5* | 53.9 |
| Trailer/mobile/other | 3.4* | 3.8 |
| 2-4 Multi-unit building | 3.2* | 4.9 |
| 5-9 Multi-unit building | 1.5* | 2.4 |
| 10-19 Multi-unit building | 1.5* | 2.3 |
| 20+ Multi-unit building | 3.9* | 5.8 |
| No MAFID in AR | 11.9* | 26.9 |
| Household Income in Terms of Federal Poverty Guidelines (FPL) | | |
| Income less than 100% of FPL | 17.5* | 31.4 |
| 100% to less than 200% FPL | 20.6* | 25.0 |
| 200% to less than 300% FPL | 17.8* | 15.7 |
| 300% or above of FPL | 43.5* | 26.2 |
| Missing | 0.6* | 1.7 |
| * p<=.05 comparing AR-ACS children who match and do not match to the 2010 Census | | |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines. | | |
| [^] Missing race/ethnicity data in AR-ACS sample was supplemented by ACS for 0.4 percent of the cases. | | |
| Source: Linked administrative records composite, the 2010 Census and 2006-2010 ACS 5-year files. Authors' computations. | | |

| (Continued) Table 6. Selected Characteristics Comparing Children 0 to 4 in Administrative Records Linked to ACS by Whether they Were Matched to the 2010 Census, Percent§ | | |
|---|---|---|
| Variables in the Analysis | Children in AR-ACS Matched to the 2010 Census | Children in AR-ACS Not Matched to the 2010 Census |
| Number (%) | 623,810 (90.9) | 62,280 (9.1) |
| Family Type | | |
| Married couple | 75.7* | 58.9 |
| Female reference, no spouse | 18.5* | 31.8 |
| Male reference, no spouse | 5.8* | 9.3 |
| Complex Household | | |
| No subfamilies or nonrelatives in the household | 78.9* | 66.8 |
| Subfamilies in the household | 11.2* | 17.8 |
| Nonrelatives in the household | 10.0* | 15.4 |
| Household Size | | |
| Fewer than 7 people in the household | 92.5* | 87.0 |
| 7 or more people in household | 7.5* | 13.0 |
| Labor Force Participation (Ages 16 & Older) | | |
| One or more people in the labor force are unemployed in the household | 12.3* | 18.2 |
| No one in the labor force is unemployed in the household | 87.7* | 81.8 |
| Education (Ages 25 & Older) | | |
| No one in household completed college | 56.5* | 73.7 |
| One or more people in the household completed college | 43.5* | 26.3 |
| Foreign-Born (All Ages) | | |
| Half or more of the people in the household are foreign born | 7.9* | 12.2 |
| Fewer than half are foreign born | 92.1* | 87.8 |
| English Proficiency (Ages 17 & Older) | | |
| No one speaks English "well" or better | 2.3* | 5.3 |
| One or more speak English "well or better" | 97.7* | 94.7 |
| Mode of Data Collection | | |
| Mail | 65.6* | 43.5 |
| CATI/CAPI/Enumerator visit | 34.4* | 56.5 |
| Tract-Level Variables | | |
| People 25 & older in tract who have a BA degree or higher | 26.7* | 22.1 |
| People in tract who are non-hispanic black | 10.4* | 15.6 |
| People in tract who are hispanic | 14.7* | 19.7 |
| People in tract in the labor force who are unemployed | 7.8* | 9.1 |
| * p<=.05 comparing AR-ACS children who match and do not match to the 2010 Census | | |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines.. | | |
| Source: Linked administrative records composite, the 2010 Census and 2006-2010 ACS 5-year files. Authors' computations. | | |

Table 7. Logistic Regressions of the Likelihood of an AR-ACS Child Not Matching to the 2010 Census, Odds Ratios

| Variables in the Analysis | Odds Ratios that children in AR-ACS are not matched to the 2010 Census | | | |
|--|--|----------|---------------------------|---------------------------|
| | All race/ ethnic groups | Hispanic | Non- Hispanic Black | Non- Hispanic White |
| | (Reference are AR-ACS children matched to the 2010 Census) | | | |
| AR-ACS Race and Hispanic Origin[^] (Non-Hispanic White alone is omitted) | | | | |
| Hispanic (any race) | 1.20** | | | |
| Non-Hispanic Black alone | 1.28** | | | |
| Non-Hispanic AIAN alone | 1.34** | | | |
| Non-Hispanic Asian/NHPI alone | 1.35** | | | |
| Non-Hispanic Some Other Race alone | 1.39** | | | |
| Non-Hispanic multiple races | 1.10** | | | |
| Age in AR (Four years old is omitted) | | | | |
| Less than one year old | 1.46** | 1.40** | 1.57** | 1.49** |
| One year old | 1.09** | 1.14** | 1.12** | 1.03 |
| Two years old | 1.05** | 1.06** | 1.05 | 1.05** |
| Three years old | 1.03** | 1.04 | 1.04 | 1.01 |
| Relationship to Reference Person (Son/daughter is omitted) | | | | |
| Grandchild | 1.39** | 1.27** | 1.23** | 1.62** |
| Other relative [^] | 1.53** | | | |
| Foster child | 5.56 | | | |
| Other non-relative | 3.01 | | | |
| All other (relatives & nonrelatives) | | 1.57** | 1.32** | 1.64** |
| Housing Unit Type (Single family home is omitted) | | | | |
| 2-9 Multi-unit building | 1.41** | 1.25** | 1.44** | 1.56** |
| 10-19 Multi-unit building | 1.38** | 1.15** | 1.24** | 1.58** |
| 20+ Multi-unit building | 1.35** | 1.16** | 1.30** | 1.63** |
| Trailer/mobile/other | 1.15** | 1.06 | 1.11 | 1.18** |
| No MAFID in AR | 2.35** | 2.43** | 1.79** | 2.36** |
| Household Poverty (300% or above of Federal Poverty Line is omitted) | | | | |
| Income less than 100% of FPL | 1.39** | 1.39** | 1.28** | 1.47** |
| 100% to less than 200% FPL | 1.18** | 1.16** | 1.15** | 1.21** |
| 200% to less than 300% FPL | 1.09** | 1.15** | 1.09 | 1.06** |
| Missing poverty information | 0.61 | 1.49** | 1.66** | 1.64** |
| * p<=.05, **p<=.01 | | | | |
| [^] Missing race/ethnicity data in AR supplemented by ACS for 0.4 percent of the cases. | | | | |
| Source: Linked administrative records composite, the 2010 Census and 2006-2010 ACS 5-year files. Authors' computations. | | | | |

| (Continued) Table 7. Logistic Regressions of the Likelihood of an AR-ACS Child Not Matching to the 2010 Census, Odds Ratios | | | | |
|---|--|----------|---------------------------|---------------------------|
| Variables in the Analysis | Odds Ratios that children in AR-ACS are not matched to the 2010 Census | | | |
| | All race/ ethnic groups | Hispanic | Non- Hispanic Black | Non- Hispanic White |
| | (Reference are AR-ACS children matched to the 2010 Census) | | | |
| Family Type (Married couple is reference) | | | | |
| Female reference, no spouse | 1.30** | 1.28** | 1.30** | 1.29** |
| Male reference, no spouse | 1.34** | 1.34** | 1.36** | 1.30** |
| Complex Household (No subfamilies or nonrelatives in the household is reference) | | | | |
| Nonrelatives in the household | 0.98 | 0.99 | 0.99 | 1.01 |
| Subfamilies in the household | 0.91** | 0.91** | 0.86** | 0.96 |
| Household Size (Fewer than 7 people in the household is reference) | | | | |
| 7 or more people in household | 1.11** | 1.09** | 1.20** | 1.09** |
| Labor Force Participation, Ages 16 & Older (Employed in the labor force is reference) | | | | |
| One or more people in the labor force are unemployed in the household | 1.07** | 1.05* | 1.08** | 1.07** |
| Education, Ages 25 & Older (At least one person in the household completed college reference) | | | | |
| No one age 25+ in household completed college | 1.22** | 1.24** | 1.29** | 1.22** |
| Foreign-Born (Fewer than half of the people in the household are foreign born is reference) | | | | |
| Half or more of the people in the household are foreign born | 1.18** | 1.13** | 0.97 | 1.55** |
| English Proficiency, Ages 17 & Older (At least one person speaks English “well” or better is reference) | | | | |
| No one 17 or older in the household speaks English “well” or better | 1.11** | 1.17** | | |
| Mode of Data Collection (Mail only is reference) | | | | |
| CATI/CAPI/Enumerator visit | 1.71** | 1.74** | 1.82** | 1.77** |
| Census Tract-Level Variables | | | | |
| Percent non-Hispanic black in tract x 10 | 1.04** | 1.05** | 1.01** | 1.10** |
| Percent Hispanic in tract x 10 | 1.02** | 1.00 | 1.02** | 1.06** |
| * p<=.05, **p<=.01 | | | | |
| Source: Linked administrative records composite, the 2010 Census and 2006-2010 ACS 5-year files. Authors' computations. | | | | |

| Appendix Table A. Selected Characteristic of Children Ages 0 to 4 in the 2010 Census, by PIK Assignment (Edited Data), Percent§ | | | |
|---|---|----------------|--------------------|
| Variables | Children ages 0 to 4 in the 2010 Census | Assigned a PIK | Not assigned a PIK |
| Number | 20,201,275 | 18,241,570 | 1,959,705 |
| Percent | 100.0 | 90.3 | 9.7 |
| Hispanic Origin | | | |
| Hispanic | 25.3 | 24.5* | 33.3 |
| Not Hispanic | 74.7 | 75.5* | 66.7 |
| Race | | | |
| White alone | 63.3 | 64.6* | 51.4 |
| Black alone | 14.4 | 13.8* | 19.4 |
| American Indian or Alaska Native alone | 1.2 | 1.2* | 1.6 |
| Asian alone | 4.4 | 4.4* | 5.2 |
| Native Hawaiian or Pacific Islander alone | 0.2 | 0.2* | 0.4 |
| Some Other Race alone | 9.5 | 8.8* | 16.0 |
| Multiple races | 6.9 | 7.0* | 6.0 |
| Age | | | |
| Less than one year old | 19.5 | 18.8* | 26.5 |
| One year old | 19.7 | 19.9* | 18.1 |
| Two years old | 20.3 | 20.4* | 18.8 |
| Three years old | 20.4 | 20.6* | 18.6 |
| Four years old | 20.1 | 20.3* | 18.0 |
| Relationship to Household Respondent | | | |
| Biological/adopted/stepchild | 82.9 | 83.5* | 77.1 |
| Brother/sister | 0.1 | 0.1* | 0.2 |
| Grandchild | 12.4 | 12.4* | 12.8 |
| Other relatives | 2.7 | 2.4* | 5.3 |
| Nonrelatives | 1.9 | 1.6* | 4.6 |
| Housing Unit Type | | | |
| Single family home | 74.4 | 75.6* | 62.7 |
| Trailer/mobile home/other/missing type | 5.7 | 5.6* | 7.1 |
| 2-9 multi-unit building | 9.3 | 8.8* | 14.2 |
| 10-19 multi-unit building | 2.9 | 2.7* | 4.4 |
| 20+ multi-unit building | 7.7 | 7.3* | 11.6 |
| Response mode | | | |
| Non-response follow up (NRFU) | 22.9 | 21.4* | 36.9 |
| Mail out-mail back | 59.9 | 64.1* | 21.2 |
| Update/Enumerate | 17.2 | 14.5* | 42.0 |
| * p<=.05 comparing children with and without a PIK in the 2010 Census | | | |
| § May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines. | | | |
| Source: 2010 Census edited file; authors' computations. | | | |

Appendix Table B. Selected Characteristics Comparing Children 0 to 4 in Administrative Records Linked to ACS by Whether they Were Matched to the 2010 Census, Percent*§

| Variables in the Analysis | AR-ACS Hispanic (any race) Children | | AR-ACS Non-Hispanic Black Children | | AR-ACS Non-Hispanic White Children | |
|--|-------------------------------------|----------------------------|------------------------------------|----------------------------|------------------------------------|----------------------------|
| | Matched to 2010 Census | Not Matched to 2010 Census | Matched to 2010 Census | Not Matched to 2010 Census | Matched to 2010 Census | Not Matched to 2010 Census |
| Number | 115,740 | 17,580 | 54,670 | 10,140 | 392,410 | 27,790 |
| Percent | 86.8 | 13.2 | 84.4 | 15.6 | 93.4 | 6.6 |
| Age in AR | | | | | | |
| Less than one year old | 4.0 * | 7.4 | 4.3 * | 7.8 | 2.5 * | 6.2 |
| One year old | 13.6 * | 14.8 | 13.5 | 14.2 | 13.5 | 13.5 |
| Two years old | 21.7 | 21.4 | 21.5 | 21.0 | 21.4 | 21.5 |
| Three years old | 28.3 * | 27.1 | 28.1 * | 26.8 | 28.6 * | 27.2 |
| Four years old | 32.4 * | 29.4 | 32.6 * | 30.1 | 34.0 * | 31.6 |
| Relationship to Reference Person | | | | | | |
| Son/daughter | 81.4 * | 76.3 | 74.7 * | 71.6 | 91.7 * | 84.0 |
| Grandchild | 14.4 * | 16.0 | 20.6 * | 21.8 | 7.1 * | 12.9 |
| Other relative | 3.3 * | 5.8 | 3.7 * | 4.9 | 0.7 * | 1.6 |
| Foster child | 0.3 * | 0.7 | 0.6 * | 1.2 | 0.2 * | 0.9 |
| Other nonrelative | 0.6 * | 1.2 | 0.4 * | 0.6 | 0.3 * | 0.6 |
| Housing Unit Type | | | | | | |
| Single family home | 63.9 * | 46.6 | 64.2 * | 52.3 | 79.7 * | 60.2 |
| Trailer/mobile/other | 3.9 | 3.6 | 2.5 | 2.5 | 3.6 * | 4.7 |
| 2-4 multi-unit bldg | 5.3 * | 5.9 | 6.1 * | 8.0 | 2.1 * | 3.1 |
| 5-9 multi-unit bldg | 2.7 * | 2.9 | 3.3 * | 4.4 | 0.8 * | 1.4 |
| 10-19 multi-unit bldg | 2.6 | 2.7 | 3.6 | 3.8 | 0.8 * | 1.3 |
| 20+multi-unit bldg | 7.5 | 7.6 | 8.8 * | 10.2 | 1.8 * | 2.9 |
| No MAFID in AR | 14.1 * | 30.6 | 11.5 * | 18.8 | 11.3 * | 26.4 |
| Household Income in Terms of Federal Poverty Guidelines (FPL) | | | | | | |
| Income less than 100% of FPL | 28.2 * | 39.0 | 38.3 * | 47.8 | 11.5 * | 21.4 |
| 100% to less than 200% FPL | 29.7 | 30.3 | 24.8 | 25.3 | 17.7 * | 22.2 |
| 200% to less than 300% FPL | 17.2 * | 15.0 | 14.4 * | 12.0 | 18.9 * | 17.8 |
| 300% or above of FPL | 23.9 * | 13.9 | 21.5 * | 13.2 | 51.4 * | 37.1 |
| Missing | 0.9 * | 1.9 | 1.0 * | 1.8 | 0.5 * | 1.5 |
| Family type | | | | | | |
| Married couple | 66.2 * | 55.3 | 39.2 * | 28.0 | 83.6 * | 71.4 |
| Female reference, no spouse | 24.5 * | 31.6 | 54.5 * | 64.2 | 11.6 * | 20.7 |
| Male reference, no spouse | 9.3 * | 13.1 | 6.4 * | 7.8 | 4.8 * | 7.9 |

§ May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines.

* p<=.05

Source: Linked administrative records composite, the 2010 Census and 2006-2010 ACS 5-year files. Authors' computations.

(Continued) Appendix Table B. Selected Characteristics Comparing Children 0 to 4 in Administrative Records Linked to ACS by Whether they Were Matched to the 2010 Census, Percent*§

| Variables in the Analysis | AR-ACS Hispanic (any race) Children | | AR-ACS Non-Hispanic Black Children | | AR-ACS Non-Hispanic White Children | |
|---|-------------------------------------|----------------------------|------------------------------------|----------------------------|------------------------------------|----------------------------|
| | Matched to 2010 Census | Not Matched to 2010 Census | Matched to 2010 Census | Not Matched to 2010 Census | Matched to 2010 Census | Not Matched to 2010 Census |
| Complex Household | | | | | | |
| No subfamilies or nonrelatives in the household | 66.4 * | 58.0 | 67.6 * | 63.9 | 84.6 * | 73.6 |
| Subfamilies in the household | 18.5 * | 21.8 | 21.3 * | 22.8 | 7.2 * | 12.8 |
| Nonrelatives in the household | 15.2 * | 20.1 | 11.0 * | 13.2 | 8.3 * | 13.6 |
| Household Size | | | | | | |
| Fewer than 7 people in the household | 85.4 * | 80.7 | 89.8 * | 87.0 | 95.2 * | 91.5 |
| 7 or more people in household | 14.6 * | 19.3 | 10.2 * | 13.0 | 4.8 * | 8.5 |
| Labor Force Participation (Ages 16 & Older) | | | | | | |
| One or more people in the labor force are unemployed in the household | 16.9 * | 19.7 | 23.6 * | 27.2 | 9.2 * | 13.8 |
| No one in the labor force is unemployed in the household | 83.1 * | 80.3 | 76.4 * | 72.8 | 90.8 * | 86.2 |
| Education (Ages 25 & Older) | | | | | | |
| No one in household completed college | 78.3 * | 86.8 | 75.4 * | 84.5 | 48.9 * | 63.2 |
| One or more people in the household completed college | 21.7 * | 13.2 | 24.6 * | 15.5 | 51.1 * | 36.8 |
| Foreign-Born (All Ages) | | | | | | |
| Half or more of the people in the household are foreign born | 20.3 * | 26.2 | 6.4 * | 5.3 | 1.7 * | 2.9 |
| Fewer than half are foreign born | 79.7 * | 73.8 | 93.6 * | 94.7 | 98.3 * | 97.1 |
| English Proficiency (Ages 17 & Older) | | | | | | |
| No one speaks English “well” or better | 10.8 * | 17.1 | 0.5 * | 0.7 | 0.1 * | 0.2 |
| One or more speak English “well or better | 89.2 * | 82.9 | 99.5 * | 99.3 | 99.9 * | 99.8 |
| Mode of Data Collection | | | | | | |
| Mail | 44.8 * | 27.1 | 45.3 * | 29.9 | 74.4 * | 57.3 |
| CATI/CAPI/Enumerator visit | 55.2 * | 72.9 | 54.7 * | 70.1 | 25.6 * | 42.7 |
| Tract-Level Variables | | | | | | |
| People 25 & older in tract with a BA degree or higher | 20.1 * | 17.7 | 19.9 * | 17.8 | 28.8 * | 25.2 |
| People in tract who are non-Hispanic Black | 9.9 * | 11.6 | 46.9 * | 49.8 | 5.6 * | 7.0 |
| People in tract who are Hispanic | 41.4 * | 44.0 | 12.0 | 12.5 | 7.4 * | 8.4 |
| People in tract in the labor force who are unemployed | 9.3 * | 9.8 | 12.1 * | 12.9 | 6.8 * | 7.4 |

§ May not add to 100 percent due to rounding error. All frequencies rounded to the nearest multiple of 5 to meet disclosure avoidance guidelines.

* p<=.05

Source: Linked administrative records composite, the 2010 Census and 2006-2010 ACS 5-year files. Authors' computations.