

What Do We know About the Presence of Young Children in Administrative Records

By
William P. O'Hare

The Annie E. Casey Foundation

Abstract

The U.S. Census Bureau is planning to use administrative records in conducting the 2020 U.S. Decennial Census. In that context, it is important to examine how the groups that are undercounted at the highest rates in the Census are represented in administrative records. The Census Bureau's Demographic Analysis found a net undercount of nearly a million young children in the 2010 Census which amounts to 4.6 percent of this age group. The net undercount of young children (under age 5) in the 2010 U.S. Decennial Census was twice as high as any other age group. This paper reviews what we know about the presence of young children in the set of administrative records often used by the Census Bureau. Data show that the youngest children (ages 0 to 2) have the lowest match rates between the Census and administrative records. Implications for use for administrative records in the Decennial Census and major Census Bureau surveys are discussed.

What Do We know About the Presence of Young Children in Administrative Records

By
William P. O'Hare

Introduction

The costs of the U.S. Decennial Census have been growing for several decades and congress has told Census Bureau to reduce costs in 2020. Administrative records have the potential to reduce costs without compromising quality and the Census Bureau operational plans (2015, page 26) call for “increased use of administrative records and third-party data to impute response data (in whole or in part).” Expanded use of administrative records is part of an effort to save \$5 billion in the 2020 compared to what it would have cost to repeat the 2010 methodology (Blummerman 2015; O'Hare and Lowenthal 2015). The increased use of administrative records by the U.S. Census Bureau is consistent with an international trend related to new and different ways to take a census (Pfeffermann 2015).

This paper first looks at the undercount of young children in the U.S. Decennial Census and then examines data related to how well young children are reflected in administrative data. Some of the implications of the findings are also discussed.

Census Undercounts

If administrative records are going to be used in the context of the decennial census, it is important to examine which groups are currently missed in the Census at the highest rates. Figure 1 shows coverage rates from the 2010 Census by five-year age groups. The net undercount of young children (ages 0 to 4), at 4.6 percent, is substantially higher than any other age group. There is growing recognition that young children (ages 0 to 4) have a high net undercount in the 2010 Census (O'Hare 2015, Griffin 2014).

Moreover there has been an increase in the net undercount of young children in recent decades. O'Hare (2014) shows the net undercount rate for young children has increased from 1.6 percent in 1980 to 4.6 percent in 2010.

Most of the Census Bureau studies that have matched administrative records to census data have broken out children age 0 to 2 separately (Rastogi and O'Hara 2012; Luque and Bhaskar 2014; Massey 2015). The undercount rates for the population ages 0 to 2 and ages 3 & 4 are shown separately in Table 1 and the data indicate the net undercount rate in the 2010 U.S. Decennial Census for the population age 0 to 2 was 4.8 percent compared to 4.2 percent for those age 3 or 4.

Because young children have a higher net undercount rate than any other group, it is important to see how well they are represented in the kinds of administrative records that are being considered for use in the 2020 Census. The remainder of this paper focuses on the representation of young children in administrative records.

Young Children in Administrative Records

Soon after the 2010 Decennial Census, staff at the U.S. Census Bureau's Center on Administrative Records Research and Applications (CARRA) undertook a study to match people counted in the Census to people in several administrative data sets they had available. This study is described in a Census Bureau publication by Rastogi and O'Hara (2012). The administrative records files used in the Rastogi and O'Hara study are shown below.

Initial Administrative Records Files

HUD Computerized Homes Underwriting Management System
Public and Indian Housing Information Center
Tenant Rental Assistance Certification System
Indian Health Service
Internal Revenue Service Form 1040
Internal Revenue Service Form 1099
Medicare
National Change of Address file
Selective Service System
Supplemental Security Income
Temporary Assistance for Needy Families

Some of the key results from the Rastogi and O'Hara study are shown in Figure 2. The data in Figure 2 show that only 81 percent of the people age 0 to 2 counted in the census were also identified in the administrative records, which is a lower match rate than any other age group. Likewise when there was an attempt to match both the person and the address the match rate for the population age 0 to 2 was only 58 percent which is lower than any other age group except age 18 to 24.

Figure 3 shows match rates for age 0 to 2 by race and Hispanic Origin Status. The major point in Figure 3 is that Hispanics in this age group are less likely to be matched than most other groups. Note that young Hispanics also have a very high net undercount rate in the 2010 Census. O'Hare (2015) shows that Hispanic age 0 to 4 had a net undercount rate of 7.5 percent in the 2010 U.S. Census.

A study by Luque and Bahskar (2014) found similar match rate patterns when administrative records were matched against persons in the Census Bureau's 2010 American Community Survey. The results, shown in Figure 4, indicate on 76 percent of young children (age 0 to 2) were matched which is much lower than any other age group.

The bottom line from these two studies is that the youngest children were the most difficult to match to administrative records.

New Administrative Data

Because of the relatively low match rate between young children in the Decennial Census and administrative records available in 2010, the Census Bureau looked for some additional administrative records that might be used to produce a higher match rate for young children. Three such files were found and used for an updated 2010 Census match study. The results were published in 2015 (Massey 2015).

The new files included in the 2015 study were the file of Medicaid recipients, the Social Security Numident file, and the file of people in New York State who received Supplemental Nutritional Assistance

Program (SNAP). SNAP was formerly known as Food Stamps. The Numident file is a files of all people who have applied for a social security number.

Figure 4 shows the match rate for those age 0 to 2, in the new files compared to the files used in the 2010 Census match test. Relative to most of the files used in the 2010 Census match study, the new files look promising. In particular, there was a 94 percent match rate with the Numident file. However, it is unclear how much new coverage the new files provide compared to the older files. It is also unclear how much overlap there is in the coverage offered by new files. So while the new administrative data looks positive it still unclear exactly how much addition coverage they will add.

Also, there are couple of caveats or limitations regarding these new files. One was already mentioned; the SNAP data is held by the states rather than the federal government. So the Census Bureau must negotiate state by state to gain access to these administrative records. So far, the Census Bureau has found it difficult to negotiate access to these files. Only the file from New York State was available for the 2015 test. It is extremely unlikely that all, or even most states, will provide the SNAP data to the Census Bureau in time for use in the 2020 Census.

As seen in Figure 4, there is an extremely high match rate with the Numident file. In recent years, nearly every birth in the U.S. has been accompanied by a submission to the Social Security Administration for a Social Security card which puts the newborn in the Numident file. However, the Numident files does not have address information, which a key piece of information in matching people from administrative records and the census. It seems possible that the people least likely to apply for a social security card for their child are also likely to be among the more difficult to count in the census. This may limit the usefulness of this file. The Medicaid file has a similar limitation in terms of not having an address.

Discussion

The relatively low match rate between administrative records and young children raises questions about the wisdom of relying on administrative records in the 2020 Census. While use of administrative record to replace sending out census enumerators would save money, it is not clear that it would not exacerbate the already high net undercount of young children. For example young Hispanic children have a very high net undercount rate in the Decennial Census and a very low match rate in administrative records. Further research needs to be done on exactly how administrative data can be used for young children without lowering the quality of the census data.

There is also an issue of timing in potential use of administrative records. It is not clear how soon updated files could be delivered to the Census Bureau after the birth of a child. For the administrative data to be useful in conducting the 2020 Census, the file must be available be late spring 2020. For some administrative data the lag time is a couple of years. For example, when the Census Bureau released its Demographic Analysis estimates of the population in December of 2010, the birth data for 2008 and 2009 were not available.

If getting updated administrative records file in a timely manner is a problem, such files may provide biased estimates of the population. The chance for older people (older than 2) being in the files is higher than the chance of a young person (age 0 to 2) being in the files.

To date, studies matching administrative records and census files have focused on which people counted in the Census are also in administrative records. This makes sense if one is considering using administrative data in place of a census. However, the Census Bureau now has an opportunity to see which people not in the Census are in administrative files. The findings from such a study would help identify the characteristics of the young children missed in the Census. This path of investigation is particularly important for some of the groups of young children with the highest net undercounts, such as young Hispanic children, to see if administrative records could be used not just to save money but to do a better job of counting young children.

In addition to research into the characteristics of those missed, administrative records may also be useful in conducting the 2020 Census. Using administrative records to determine who has not been included on returned census questionnaires, could be used to trigger further investigation by enumerators.

Summary

The data examined in this paper lead to several conclusions. First, match rates between administrative records and the 2010 Decennial Census are lower for children age 0 to 2 than any other age group. Second, match rates between administrative records and the 2010 American Community Survey are lower for children age 0 to 2 than any other age group. Third, match rates for young Hispanic children ages 0 to 2 are lower than for other major race/Hispanic group in this age range. Fourth, new administrative records have the potential to improve overall match rates for young children but the usability and collective impact is unclear at this point.

Administrative records can be used to shed light on which people in administrative records are not included the census. Knowing more about the characteristics of the people missed in the census could help improve coverage in the 2020 Census.

Finally, the relatively low match rate between administrative records and young children raises questions about the wisdom of relying on administrative records in the 2020 Census

References

- Blumerman, L.M (2015) "Planning for the 2020 Census: A New Design for the 21st Century," AMSTATNEWS, Issue #462. The American Statistical Association, Washington, DC.
- Griffin, D.H. (2014). "Final Task Force Report: Task Force on the Undercount of Young Children," Memorandum for Frank A. Vitrano, U.S. Census Bureau, Washington, DC. February 2
- Luque, A. and Bhaskar, R. (2014) 2010 American Community Survey Math Study, CARRA Working Paper #2014-03, Center for Administrative Records Research and Application, U.S. Census Bureau, Washington DC.
- Massey, C.G. (2015) Coverage of Children in Administrative Records and the 2010 Census, internal Census Bureau Document, Center for Administrative Records Research and Application, U.S. Census Bureau, Washington DC.
- O'Hare, W.P. (2014). "Historical Examination of Net Coverage Error for Children in the U.S. Decennial Census: 1950 to 2010." *Center for Survey Measurement Study Series (Survey Methodology #2014-03)*. U.S. Census Bureau. Available online at <http://www.census.gov/srd/papers/pdf/ssm2014-03.pdf>.
- O'Hare, W.P. (2015). *The Undercount of Young Children in the U.S. Decennial Census*. Springer Publishers
- O'Hare, W. P. and Lowenthal, T. A. (2015). "The 2020 Census: The Most Difficult in History?" Applied Demography, *Committee on Applied Demography Newsletter*, Page 8-10, Population Association of America, Washington, DC.
- Pfefferman, D. (2015). "Methodological Issues and Challenges in Production of Official Statistics," *Journal of Survey Statistics and Methodology*, Vol 3, No.1. pp425-467.
- Rastogi, S. and O'Hara, A. (2012) 2010 Census Match Study, 2010 Census Planning Memorandum Series, No. 247, U.S. Census Bureau, Washington, DC.

Figure 1. Net Undercount Rates in 2010 Census by Age

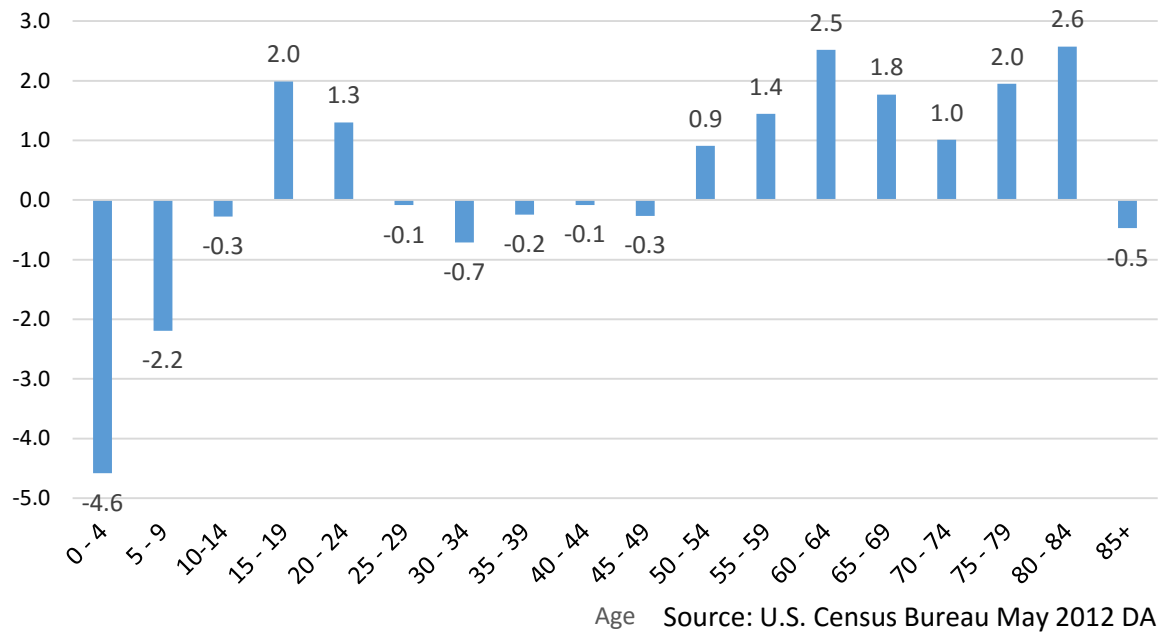
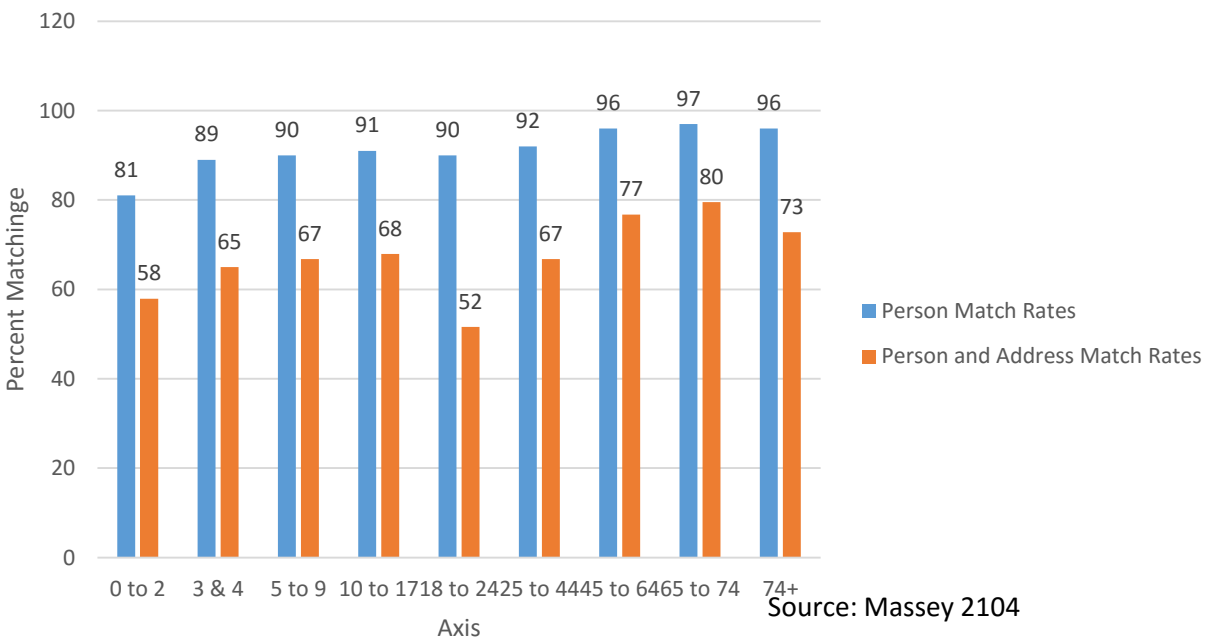


Table 1. Difference Between 2010 Census Counts and DA Estimates for Children Ages 0 to 2 by Race and Hispanic Origin

	Numeric Difference (rounded to nearest 1000)	Percent Difference
Age 0 to 2	-612,000	-4.8
Age 3 & 4	-358,000	-4.2
Black Alone or in Combination Age 0 to 2	157,000	-6.7
Hispanic Age 0 to 2	246,000	-7.4

Source: U.S. Census Bureau, Demographic Analysis December 2010 and May 2012 Releases

Figure 2. Administrative Records Coverage of 2010 Census by Age



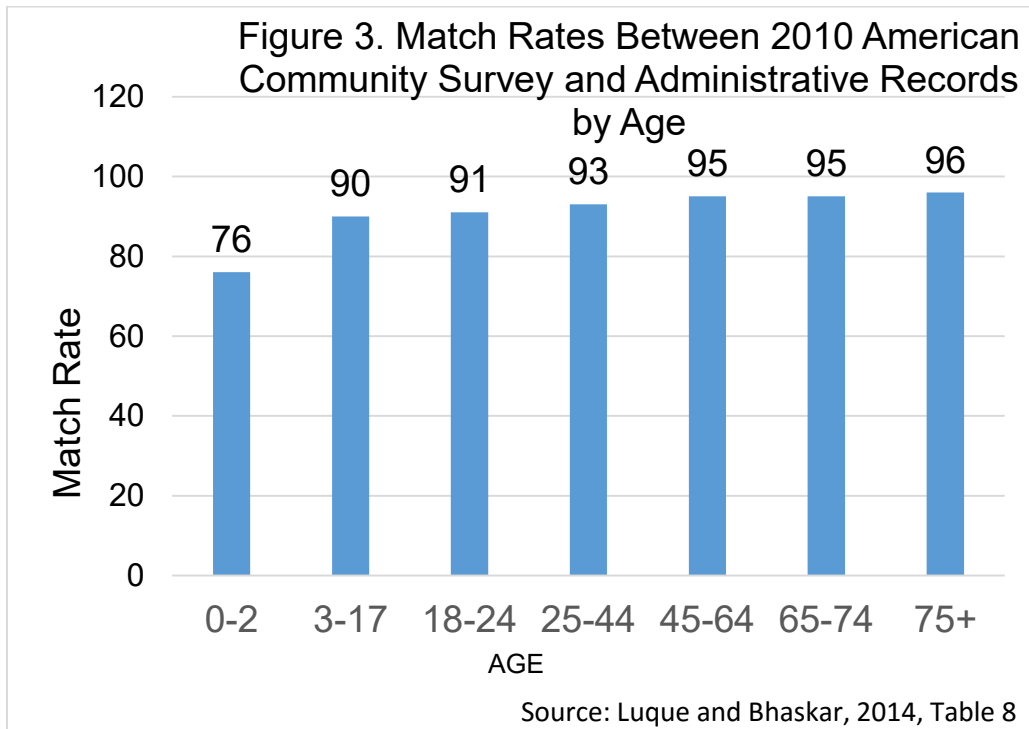


Figure 4. Match Rate Age 0-2 in Various Administrative Records Files

