

The U.S. Census Undercount of Native-Born Children: Estimates, Correlates, and Implications

Janna E. Johnson, University of Minnesota

Introduction and Background

With Census 2020 fast approaching, there has been significant interest in the media and among policymakers in the ability of the census to accurately measure the U.S. resident population. The Census Bureau has estimated *undercount* – the percent of the population missed in the census – for decades using the method of Demographic Analysis (DA) (Robinson 2011). This method relies on Vital Statistics birth and death records and measures of in- and out-migration to calculate an independent measure of the U.S. population, which is then compared to the population count in the decennial census. The Census Bureau reports measures of undercount for each decennial census by sex, age, and for two racial groups: black and non-black.

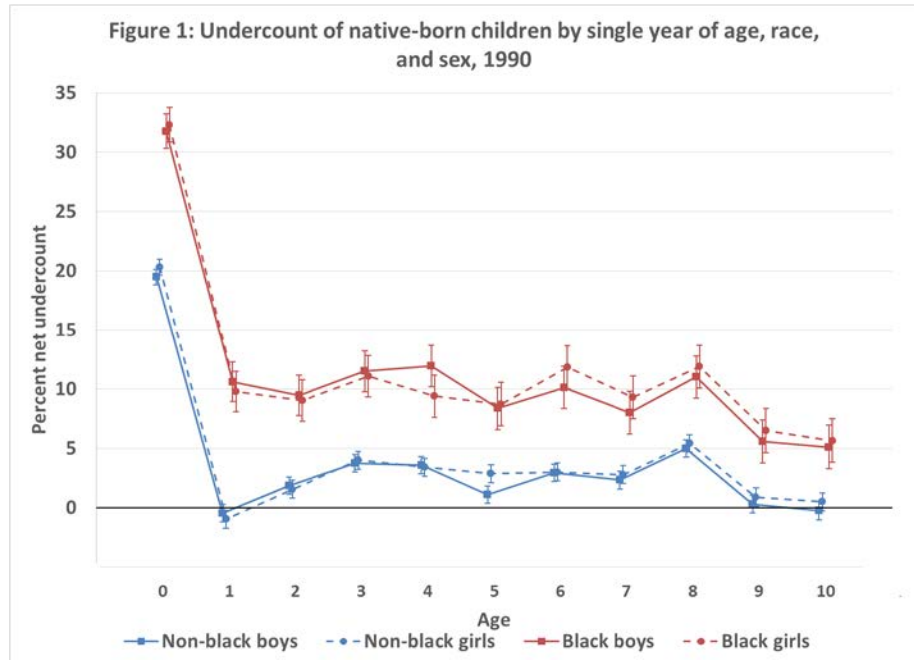
This paper explores the undercount of one particular group in much more detail than previously: children under the age of 5. Young children have consistently been among the highest undercounted groups for decades, and the exact reasons for their under-enumeration remain a mystery (O’Hare 2009, 2015). DA estimates of individuals under age 5 are also most reliant on accurate, complete birth records, due to the relatively low overall death and net migration rates of the youngest population. In my analysis, I ignore measures of migration completely, and the aggregate undercount estimates produced are virtually identical to those reported by the Census Bureau. I move beyond the results reported by Census to calculate undercount for additional racial and ethnic groups, and by state of birth for the 1980, 1990, and 2000 censuses. My results show significant variation in the undercount rate of children by these characteristics.

The goal of this paper is to not only document patterns in undercount by characteristics not previously explored, but also to show how undercount can affect empirical research using census data. I therefore use the individual-level PUMS data to form my census counts for my undercount calculations, as the PUMS is much more widely used among social scientists than the full aggregate counts used by Census in their undercount calculations. Using my undercount estimates by state of birth, sex, race, and age, I plan to adjust the weights that accompany the PUMS data for the under-enumeration of children, and use the adjusted weights to investigate several empirical questions. Comparing the results from this analysis to those produced using the unadjusted weights will demonstrate the effect undercount has on research reliant on census data.

Undercount in the PUMS

I begin by replicating Census’ results for undercount of children by age, sex, and race (black and non-black). My results for 1980, 1990, and 2000 using the 5 percent PUMS as the census population largely correspond with those reported by Census (who use the 100 percent census count), with one major exception. Figure 1 shows my undercount results for Census 1990. Note the large undercount of children under the age of 1, which does not appear in the 1990 Census Bureau undercount estimates (Robinson et al. 1993). According to Census documents, the age

question in 1990 was designed in a way that made it difficult for individuals to enter the age of someone who had not yet reached a full year of age, leading to many parents to misreport their infant's age as 1, 9, or 10 years, or to not report them at all. The Bureau realized this error and corrected the overall 100 percent counts for this pattern (U.S. Census Bureau 1992), but the individual-level PUMS are left in their original state. Therefore, the 1990 data most used by researchers is missing nearly one-third of all black infants, and 20 percent of non-black infants, with some of these infants misclassified as 1-, 9-, or 10-year-olds.



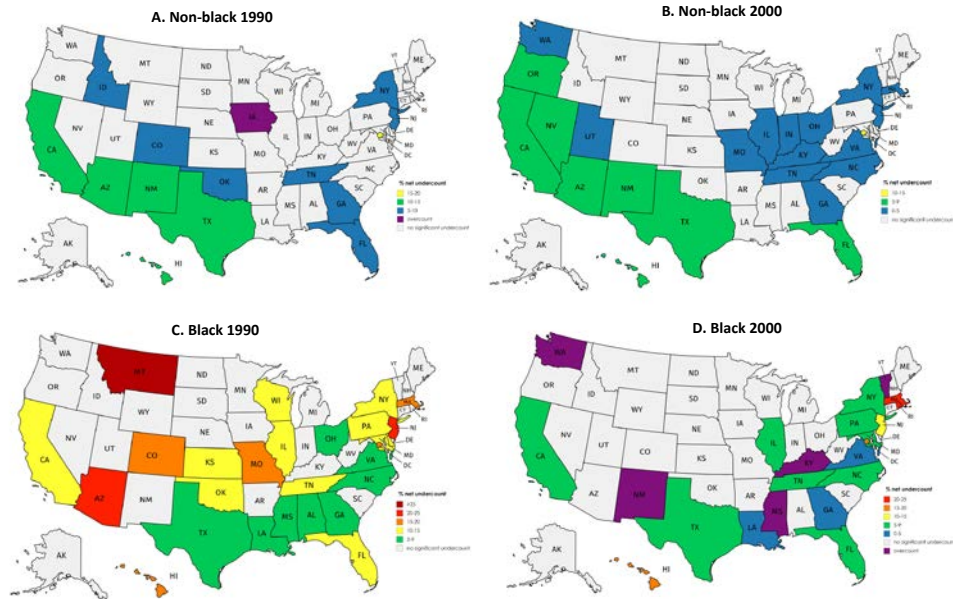
Undercount by State of Birth

I proceed to for children under 5 by state of birth, a statistic not reported by the Census Bureau. Although young children have not been alive long enough for a large

fraction of them to move outside their state of birth, I calculate undercount by state of birth, rather than state of residence, to eliminate the

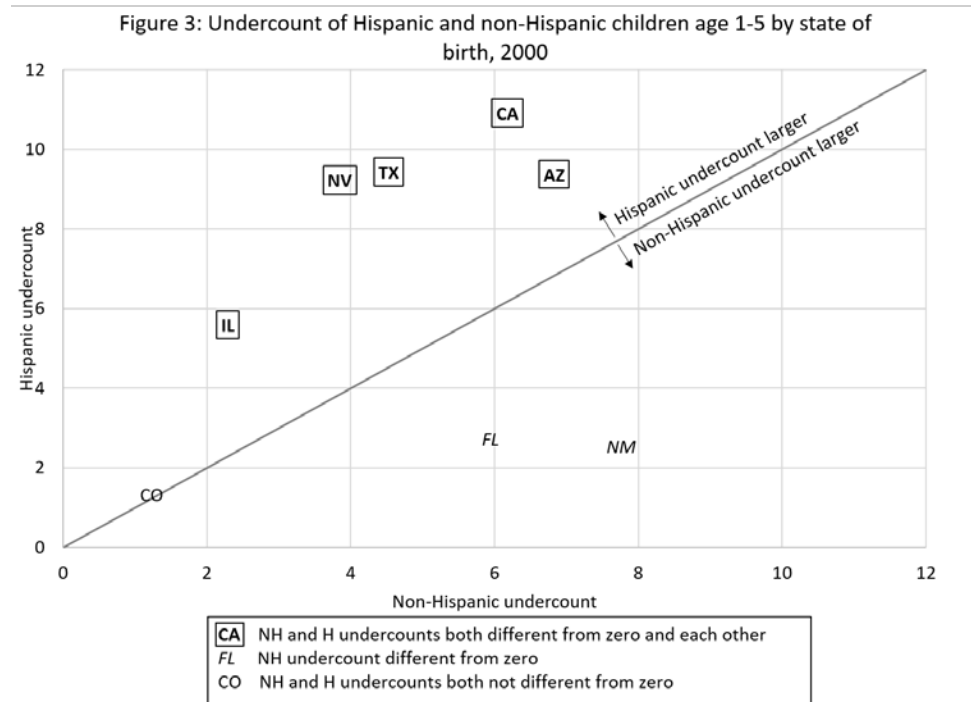
effect of movement across states on measures of undercount. Results for the 1990 and 2000 censuses for non-blacks and blacks aged 1-5 are shown in Figure 2. (I cannot calculate

Figure 2: Undercount by state of birth and race, age 1-5



undercount by state of birth for the 1980 census due to the state of birth not being available on the death certificate prior to 1979.) Two patterns emerge from this figure. First, undercount of non-black children is highest in states in the South and Southwest. Second, undercount of black children is high in many of those same states (like California, Texas, and Florida), as well as in the Northeastern states of New York, Massachusetts, and New Jersey. Despite the difference in overall undercount rates of these groups across the two census years, the patterns of undercount across states are similar.

The state-of-birth undercount pattern for non-black children could be explained by high undercount rates of Hispanic non-blacks, as the states that have the highest non-black undercount rates are also those with large Hispanic populations. I investigate this possibility by calculating undercount by Hispanic status for non-blacks. The



Census Bureau does not report undercount by Hispanic status in its official undercount estimates. Unfortunately, although Hispanic status was first reported by some states on the birth certificate in 1984, it was not expanded widely until 1989, and not fully adopted by all states until 1997. I therefore calculate undercount by Hispanic status for children under 5 in the 2000 census only, and for the 34 states that began Hispanic status reporting in 1989. (Fortunately, these earlier adopters of Hispanic status reporting include most high-Hispanic-population states.) Figure 3 compares the undercount of Hispanic and non-Hispanic non-black children age 1-5 in the 2000 census for 8 states with Hispanics accounting for more than 10 percent of their total populations in 2000. (Note that New Jersey and New York also fall in this category, but did not report Hispanic status on their birth certificates until 1997.) In five of these eight states, the Hispanic and non-Hispanic undercounts are significantly different from each other and from zero, but non-Hispanic non-black children are still undercounted at a high rate. The overall undercount rate of non-blacks age 1-5 in 2000 is 3.9 percent, and all of these states except Illinois have undercounts of non-Hispanic non-blacks higher than this level. The states of Florida and Illinois undercount non-Hispanics born in their states at a higher level than non-blacks, and Colorado's undercount of these populations is not significantly different from zero. Therefore, it does not appear that disproportionately high undercounts of Hispanic non-blacks explain the state of birth patterns shown in Figure 2.

Next Steps

To show the implications of the undercount of children for research done using Census data, I plan to adjust the sampling weights for children under age 5 provided with the PUMS for the patterns I document. These adjustments will occur by single year of age, race (black and non-black), and state-of-birth for the 1980, 1990, and 2000 censuses. I will then perform several empirical analyses using the original and adjusted weights, examining such questions as

- The variation and predictors of teen motherhood by state
- The labor market effects on mothers of having an additional child (i.e., Angrist and Evans 1996)
- The likelihood a young child attends school, and how this varies by family characteristics, including state

Comparing the results of these analyses across the adjusted and unadjusted data will not only show the implications of undercount for research done using the decennial census, but for other federal data products such as the American Community Survey, Current Population Survey, etc., as these surveys base their sampling design on the decennial census, and are also likely to suffer from similar undercount patterns as those I document.

Progress and Timeline

The adjustment of the PUMS weights and the analyses described above will be completed in Fall 2018. Initial results were presented at the Census Bureau as part of the Summer at Census program in June 2018, and results including the weight adjustment will be presented at the American Economic Association meetings in January 2019. The paper will be submitted to *Demography* soon after the PAA meetings in April 2019.

References

- Angrist, Joshua D., and William N. Evans. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review* 88.3 (1998): 450-477.
- O'Hare, William P. "Why Are Young Children Missed so Often in the Census?" KIDS COUNT Working Paper, Annie E. Casey Foundation (2009).
- O'Hare, William P. *The undercount of young children in the US Decennial Census*. Springer International Publishing (2015).
- Robinson, J. Gregory. "Coverage of population in census 2000 based on demographic analysis: The history behind the numbers." US Census Bureau, Population Division (2011).
- Robinson, J. Gregory, Bashir Ahmed, Prithwis Das Gupta, and Karen A. Woodrow. "Estimation of population coverage in the 1990 United States Census based on demographic analysis." *Journal of the American Statistical Association* 88, no. 423 (1993): 1061-1071.
- U.S. Census Bureau. 1990 Census of Population and Housing Modified Age/Race, Sex, and Hispanic Origin Files (MARS) Technical Documentation (1992).