# Geographic Variation in 2010 U.S. Census Coverage Rates for Young Children:
# A Look at Counties

### William P. O'Hare[1]

[1]President, O'Hare Data and Demographic Services LLC, 11 Randolph Avenue, Cape Charles, VA 23310, United States

Correspondence: William P. O'Hare, President, O'Hare Data and Demographic Services LLC, 11 Randolph Avenue, Cape Charles, VA 23310, United States.

## Abstract

Data from the Census Bureau's Demographic Analysis (DA) show there was a net undercount of almost one million children under age 5 in the 2010 Decennial Census. The net undercount for young children was more than twice as high as any other age group. Given the high net total undercount of young children it would be useful to know how this net undercount is distributed geographically. In this study, the 2010 Decennial Census county-level counts of children age 0 to 4 are compared to corresponding figures from the Census Bureau's Vintage 2010 population estimates, to ascertain census coverage for young children. Results show that net undercount rates for young children are higher in larger counties and counties with high percentages of Black or Hispanic children but county growth rates do not seem to be related to net undercount rates for young children. Discussion explores how this information may be helpful in the 2020 Census.

**Keywords:** census, undercount, coverage, children, counties

## 1. Introduction

Several studies have shown that young children have had relatively high net undercount rates in the past several U.S. Decennial Censuses (West and Robinson, 1999; O'Hare, 1999; O'Hare 2009, O'Hare 2012a; O'Hare 2014a; O'Hare 2014b, O'Hare 2014d, O'Hare 2014e, U.S. Census Bureau 2014). Data from the Census Bureau's Demographic Analysis (DA) show the net undercount of 970,000 children age 0 to 4 in the 2010 Decennial Census. The net undercount rate for the population age 0 to 4 was 4.6 percent which is more than twice as high as any other age group (O'Hare 2015).

Given the relatively high nationwide net undercount rate for young children, it would be useful to have a better understanding of the geographic differences in census coverage rates for young children. This study addresses the geographic variation in census coverage by examining the net undercount estimates for young children for all the counties or county equivalents in the U.S.

Based on their analysis of 2000 Decennial Census data compared to subnational population estimates, Adlakha, Robinson, West, and Bruce (2003, page v) recommend we, "…expand the current demographic analysis to include subnational benchmarks in the 2010 Census evaluation." Mayol-Garcia and Robinson (2011) also conclude "More studies are needed on the patterns of this population age group compared to the results of the previous censuses." The final report of the U.S. Census Bureau's Task Force on the Undercount of Young Children (U.S. Census Bureau, 2014, page ii) recommends, "Addition research using existing 2010 datasets, such as DA, population estimates, the planning database, census control and response files, and CCM holds promise to provide greater insights into causes and possible solutions." In addition, the report states: "This work must look below the national level to determine if certain areas, populations, or census operations were more likely to have these errors." The present analysis responds to these recommendations.

The Census Bureau's Vintage 2010 State and County Population Estimates for young children provide an opportunity to assess subnational census results using a DA-like methodology for the population age 0 to 4. Like the DA estimates, the yearly population estimates are based on a simple cohort-component demographic accounting equation that uses

number of births, deaths, and net migration. This study examines geographic differences in census coverage rates for young children by comparing the Census Bureau's Vintage 2010 population estimates for the population age 0 to 4 to the 2010 Decennial Census counts across counties in the U.S. The methodology employed here, comparing population estimates to the Census, are only valid for the population under age 10, because the Vintage 2010 population estimates for people age 10 and older are linked to census counts in the 2000 Census so they are not independent of the census.

This study focuses on the population age 0 to 4, because the 2010 DA analysis shows this age group had a larger net undercount than any other age group including age 5 to 9. The 2010 national undercount rate for the population aged 0 to 4 based on Demographic Analysis is 4.6 percent compared to 2.2 percent for age 5 to 9.9 (O'Hare 2015). Therefore, it is important to examine the population age 0 to 4 separately.

First, a description of the population estimation methodology for counties is presented and data sources are identified. Then differences between 2010 Decennial Census counts and Vintage 2010 county population estimates for the population age 0 to 4 are explored. Variations by county size, county population growth and racial composition are explored. This is followed by sections discussing limitations and some of the implications of the findings.

## 2. Background

The high net undercount among young children was discovered in the 1940s, early in the history of DA. Coale (1955) found children age 0 to 4 had a high net undercount rate in the censuses of 1940 and 1950. Additional research by Siegel and Zelnik (1966) also found a significant net undercount of children age 0 to 4 in the 1950 and 1960 Decennial Censuses. Coale and Zelnick (1963) discovered high net undercount rates for young children in the U.S. censuses as far back as 1880. Coale and Rives (1973) found very high undercount rates for young Black children in every decennial census from 1880 to 1970. Genealogical research also shows a pattern of underreporting young children as far back as the 1850s (Adams and Kasakoff, 1991). Despite these early studies, there has been a dearth of studies addressing this topic.

The net undercount rate for the population age 0 to 4 is not only higher than any other age group, it has been increasing over the past several decades. O'Hare (2014e) shows that the net undercount rate for young children rose from 1.4 percent in 1980 to 4.6 percent in 2010. During the same period, the net coverage rate for adults (age 18+) went from and undercount of 1.4 percent to a net overcount of 0.7 percent in 2010. This underscores the importance of examining undercounts of the population age 0 to 4 in more detail (O'Hare 2014e).

Research on subnational assessments of the U.S. Census results are limited. However, following the 1970 Census, Siegel, Passel, Rives and Robinson (1977) examined census coverage for states and for various population groups grouped by race and age. They used several different approaches with varied results.

After the 1990 Decennial Census, Robinson, Bashir, Das Dupta, Woodward (1993) offered a set of undercount estimates for states for the total population (all ages) but the estimates are only evaluated at the regional level. The authors also proposed alternatives for evaluating the 2000 Census at the state and sub-state levels and listed several reasons such an evaluation is needed.

After the 2000 Census, Adlakha et al (2003) utilized U.S. Census Bureau Population Estimates to assess the state-level Decennial Census counts for the population age 0 to 9. However, their analysis did not go below the regional level and did not show data for the population age 0 to 4 separately.

Cohn (2011) compared the state Census Bureau Population Estimates to the 2010 Census counts for the total population (i.e. all ages) but did not break out young children separately. Cohn concludes that the Census counts and the Population Estimates are very similar for most states in terms of total population.

Mayol-Garcia and Robinson (2011) examined state population differences in age 0 to 4 and age 0 to 9 populations in the 2010 U.S. Decennial Census but only provided limited results and did not explore any patterns. Regarding the state-level data on the net undercounts of the population age 0 to 4, Mayol-Garcia and Robinson (2011, page 3) note, "The relatively large differences noted nationally for 0-4 year olds are observed at the state level as well."

O'Hare (2014c) examined 2010 census coverage rates at the state level for the population age 0 to 4 and found substantial variation in coverage rates across the states. He also found several characteristics of states that were statistically significantly correlated with differences in coverage rates.

O'Hare (2013, 2014f) also provides limited preliminary descriptive analysis of net census coverage rates for young children at the subnational level. The short descriptive studies by O'Hare are expanded in this present study.

The present analysis extends the work done by previous researchers by examining 2010 county-level census coverage rates for the population age 0 to 4.

## 3. Methodology and Data Sources

The DA methodology used to assess census accuracy at the national level compares the census results to an independent population estimate to ascertain undercounts and overcounts for the total population and for selected age, sex, and racial/Hispanic groups. One of the major limitations of the Demographic Analysis technique for measuring the Census undercounts for most demographic groups is that it can only be applied at the national level because population estimates for those over age 10 include census errors from the previous census.

The estimation methodology employed by the Census Bureau for post-census population estimates uses data from the previous census in its estimates for all age groups that were included in the previous census, so the 2010 population estimates for groups age 10 and higher are not independent of the Decennial Census. Coverage errors from the previous Censuses are reflected in subsequent population estimates. If the population age 20 to 24 was undercounted by 10 percent in 2000, for example, and the population age 30 to 34 was undercounted by 10 percent in 2010, comparing the 2010 Population Estimates to the 2010 Census counts might make it look as if there were no undercount in 2010, because the 2010 population estimate for age 30 to 34 is based on aging a cohort that was undercounted by 10 percent in 2000.

The DA methodological approach is well-suited for analyzing census coverage of young children. With respect to the results of the 2000 Census evaluation for the count of young children, the U.S. Census Bureau (2003, page v) states:

> *"The Demographic Analysis estimate for this age group is more accurate than those for other age groups because the estimate for young children depends primarily on recent birth registration data which are believed to be highly accurate."*

In comparing the Demographic Analysis results to Dual Systems Estimates results in the 2000 Decennial Census, Zeller (2006. P. 320) also concluded;

> *"Since the Demographic Analysis estimate for young children depended on highly accurate recent birth registration data, The Demographic Analysis estimate is believed to be more accurate."*

The favorable view of the DA methodology is related to the simplicity of the method and the quality of the key data, that is, births and deaths. Overall, nearly all (99.6 percent) of the estimated population from the national DA estimates for those age 0 to 4 in 2010, comes from birth data (U.S. Census Bureau, 2010b). Dependence on birth certificate data and the high quality of birth certificate data provides a strong foundation for county population estimates for the population age 0 to 4. The birth and death data used in the Census Bureau's DA estimates come from the U.S. National Center on Health Statistics (NCHS) and these records are widely viewed as being accurate and complete (Devine, Sink, DeSalvo and Cortes, 2010).

The Vintage 2010 Population Estimates have these same favorable qualities as DA. The biggest difference between DA and county population estimates is the inclusion of migration across counties. Migration between counties is captured in the Census Bureau administrative records technique which uses federal tax records to estimate such migration (U.S. Census Bureau, 2012d).

The Vintage 2010 Population Estimates provide data for everyone under age 10, but only the population age 0 to 4 are examined here. (Vintage 2010 refers to the year of the estimates, not the year in which the estimates were published.)

The county population estimates are derived using the formula in Equation 1 which is taken from U.S. Census Bureau (2013d).

$$P1 = P0 + B - D + NDM + NIM \qquad (1)$$

Where:

P1 = Population at the end of the year

P0 = population at the beginning of the year

B = Births during the year

D = deaths during the year

NDM = net domestic migration during the year

NIM = net international migration during the year.

County-level census coverage rates for young children are derived by comparing the U.S. Census Bureau's Vintage 2010 Population Estimates for the population age 0 to 4 to the 2010 U.S. Decennial Census counts for this age group. This methodology for examining census coverage at the state and local level has been used by several analysts in the past including several demographers at the Census Bureau (O'Hare 2014c; Siegel et al 1977; Robinson et al 1993; Adlakha et al 2003; Mayol-Garcia and Robinson 2011; U.S. Census Bureau 2014; Cohn 2011).

*3.1 The Data*

The Vintage 2010 Population Estimates used here are taken from the Census Bureau's file labeled "Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin: April 1, 2000 to July 1, 2010." The file is also labeled "CC-EST2010-ALLDATA." The file was released March 2012 and it is available on the Census Bureau's website at http://www.census.gov/popest/research/eval-estimates/eval-est2010.html. These estimates incorporate the results of special censuses and successful local challenges during the previous decade.

This file contains yearly estimates for 2000 through 2010, but only the estimates from April 1, 2010, are used in this study. These estimates have a great deal of demographic detail such as sex and racial/Hispanic groups, but only the figures for the total population age 0 to 4 are used here.

The data from the 2010 Decennial Census are taken from Table QT-P1 in Summary File 1. The data were obtained through American Factfinder available on the Census Bureau's website. The data for the population age 0 to 4, were taken from this file.

Only counties that had data available from three sources (2000 Census, 2010 Census and Vintages 2010 Population Estimates) were used in the analysis. Three small Counties (Kalawao County, Hawaii; Loving County, Texas; and Bedford City, Virginia) were not included in the analysis because there were problems with the data. For example, there was no population age 0 to 4 in Kalawao County, Hawaii, in the 2010 Census. There also a handful of small counties that had no comparable data in both the 2000 and 2010 Census so they were removed. There were 3,133 counties or county equivalents used in the analysis. The District of Columbia is treated as a county in this analysis.

The estimate of the net undercount rate for ages 0 to 4 from the DA analysis is 4.6 percent (O'Hare 2015). Based on the county estimates the net undercount rates for ages 0 to 4 was 5.0 percent. This difference is related to when final data on births and deaths statistics become available from the National Center for Health Statistics for the two sets of estimates.

For showing coverage in this paper, I utilize the model used by Velkoff (2011). differences are assessed by subtracting the Vintage 2010 population estimate from the corresponding 2010 Decennial Census figure. If the Vintage 2010 Population Estimate is larger than the 2010 Decennial Census count, that is referred to here as a net undercount and denoted with a negative sign and if the Vintage 2010 Population Estimate is smaller than the 2010 Decennial Census count that is referred to here as a net overcount and denoted with a positive sign. Other studies sometimes report net undercounts as positive numbers (Mule, 2012; Robinson and Adlakha, 2002).

The terms undercounts and overcounts are used here to reflect differences in a particular direction. When the Census population count is lower than the population estimate, it is referred to as an undercount. When the Census count is higher than the population estimate it is referred to as an overcount. This language is not precise because the differences include estimation errors as well as census coverage, but it greatly simplifies identifying and communicating patterns because it identifies the direction of differences.

Table 1.Difference Between Vintage 2010 Population Estimates, May 2012 DA Estimates, and 2010 U.S. Decennial Census Counts by Age

| (Figures in 1,000s) | All Ages | Age 0-4 | Age 5 and older |
|---|---|---|---|
| **Population Figures** | | | |
| Vintage 2010  Population Estimates* | 308,450 | 21,263 | 287,187 |
| May 2012 DA Estimates ** | 308,346 | 21,171 | 287,175 |
| 2010 Census Counts *** | 308,746 | 20,201 | 288,545 |
| **Numeric Differences** | | | |
| Census - Population Estimates | 296 | -1,062 | 1,358 |
| Census – DA | 400 | -970 | 1,370 |
| Population Estimates- DA | 104 | 92 | 12 |
| **Percentage Differences** | | | |
| (Census – Population estimates)/Population  Estimates | 0.1 | -5 | 0.5 |
| (Census - DA)/DA | 0.1 | -4.6 | 0.5 |
| (Population  Estimates- DA)/ DA | 0 | 0.4 | 0 |
| *Vintage 2010 Population Estimate Program (PEP) results for 4/1/2010  http://www.census.gov/popest/research/eval-estimates/eval-est2010.html | | | |
| **U.S. Census Bureau, Population Division, 2010 Demographic Analysis Released May 2012 http://www.census.gov/popest/research/demo-analysis.html | | | |
| ***2010 U.S. Decennial Census Summary File 1, Table DP-1 Profile of General Population and Housing Characteristics. | | | |

## 4. Results

Table 1 shows national data from the 2010 Decennial Census, the May 2012 Demographic Analysis (DA) results, and the Vintage 2010 Population Estimates. For the total population, the figures from the three sources are nearly identical. That similarity has been taken as a sign of a high-quality census (Groves, 2011). In reality, the similarities across all three sources mask big counter-balancing differences across age groups.

For the population age 0 to 4, the DA estimates and the Vintage 2010 Population Estimates are very similar (21,263,000 for the population estimate and 21,171,000 for the May 2012 revised DA estimate). More important than the small difference between the Population Estimates and the DA estimates is the fact that both are substantially higher than the 2010 Decennial Census count (20,201,000). The difference between the DA estimates and the Census counts is -4.6 percent for the population age 0 to 4 and the difference between the Vintage 2010 Population Estimates and the Census counts is -5.0 percent.

The consistency of the Vintage 2010 Population Estimates and the DA estimates at the national level underscores the suitability of using the Vintage 2010 Population Estimates to examine the subnational geographic distribution of the net undercount of young children.

*4.1 Tabular Analysis*

Results are shown for groups of counties rather than for individual counties. This is consistent with the advice of Adlakha et al. (2003, page 34) "In general, the coverage analysis has been carried out for aggregations of counties, because benchmark estimates have certain unmeasured deficiencies, the effect of which is dampened when data are aggregated for higher geographic levels." Groupings of counties was also used by Yowell and Devine (2013) as well as Davis (1994) in examining the quality of total population estimates for counties. When counties with random estimation error are combined, the random errors often cancel each other out. In this study, aggregated data from the Census counts and the Vintage 2010 Population Estimates are compared to detect differences in groups of counties.

Table 2. Summary of County Differences Between Vintage 2010 Population Estimates and 2010 Census Counts for the Population Age 0 to 4

|                                              | Age 0 to 4 |
|----------------------------------------------|------------|
| Number of Counties with Net Overcounts       | 1484       |
| Number of Counties with Net Overcounts       | 1633       |
| Mean Numerical Difference                    | -339       |
| Mean Percent Difference                      | 1.1        |
| Mean Absolute Numeric Difference             | 419        |
| Mean Absolute Percentage Difference          | 7.4        |

Based on 3133 counties with complete data.
Sources: U.S. Census Bureau, Vintage 2010 County Population Estimates (CC-EST2010-ALLDATA.csv)
U.S. Census Bureau, Table P12 (Total Population) Sex by Age, 2010 Census Summary File 1 (DEC_10_SF1_P12_with_ann.csv).

Data in Table 2 show the counties are almost evenly split in terms of having an overcount or an undercount of young children, although it should be recognized that many of these overcounts and undercounts were very small and probably not different than zero in any meaningful way. In sixteen counties the Population Estimate and the Census Count were exactly the same so they do not show up an overcount or an undercount in Table 2.

Table 2 provides several measures of differences between the Vintage 2010 Population Estimates and the 2010 Decennial Census counts for counties for the population age 0 to 4. The average county had an overcount of 1.1 percent for the population age 0 to 4. Since this is quite different that the national undercount rate of 5.0 percent for this age group, it indicates that the national rate is driven by high undercount rates in large counties. The mean numerical difference between Decennial Census Counts and the Vintage 2010 Population Estimates for the population age 0 to 4 is -339. This reflects the relatively large undercounts in the counties with undercounts compared with the relatively small overcounts in the counties with overcounts.

Some of the errors cancel each other out in calculating the mean, so it is important to examine the size of absolute differences. From that perspective, the mean difference between the Vintage 2010 Population Estimate and the Census Counts was 7.4 percent and the mean absolute numeric difference between the Vintage 2010 Population estimate for age 0 to 4 and the 2010 Census count for this age group is 419.

*4.2 County Size*

Table 3 shows differences between 2010 Vintage Population Estimates and Census Counts by county size based on total population in the 2010 Census. The size categories employed here are those used by Yowell and Devine (2013, Table 7).

Table 3. Difference Between 2010 Census Counts and Vintage 2010 Population Estimates for Population Age 0 to 4 by County Size

| Total Population Size of County | Number of Counties | Aggregate Population Estimate | Aggregate Census Count | Difference (Census - Estimate)* | |
|---|---|---|---|---|---|
| | | | | Number | Percent |
| Less Than 5,000 | 297 | 47,507 | 49,881 | 2,374 | 5.0 |
| 5,000- 9,999 | 392 | 179,300 | 181,606 | 2,306 | 1.3 |
| 10,000-19,999 | 607 | 546,501 | 547,977 | 1,476 | 0.3 |
| 20,000-49,999 | 862 | 1,778,415 | 1,755,431 | -22,984 | -1.3 |
| 50,000-99,999 | 397 | 1,772,397 | 1,752,806 | -19,591 | -1.1 |
| 100,000-249,999 | 317 | 3,225,631 | 3,150,422 | -75,209 | -2.3 |
| 250,000 -499,999 | 133 | 3,154,991 | 3,027,675 | -127,316 | -4.0 |
| 500,000+ | 128 | 10,552,587 | 9,729,390 | -823,197 | -7.8 |
| Total | 3133 | 21,257,329 | 20,195,188 | -1,062,141 | -5.0 |
| Sources: U.S. Census Bureau, Vintage 2010 County Population Estimates (CC-EST2010-ALLDATA.csv) | | | | | |
| U.S. Census Bureau, Table P12  (Total Population) Sex by Age, 2010 Census Summary File 1 (DEC_10_SF1_P12_with_ann.csv). | | | | | |

The pattern across size categories is pretty clear. The aggregate undercount rate is higher in larger counties and lower in smaller counties. In the three smallest size categories used here there is an overcount of young children. The patterns is seen visually in Figure 1.



Figure 1. Net Aggregate Undercount Rate in the 2010 U.S. Census for the Population Age 0 to 4 by County Size
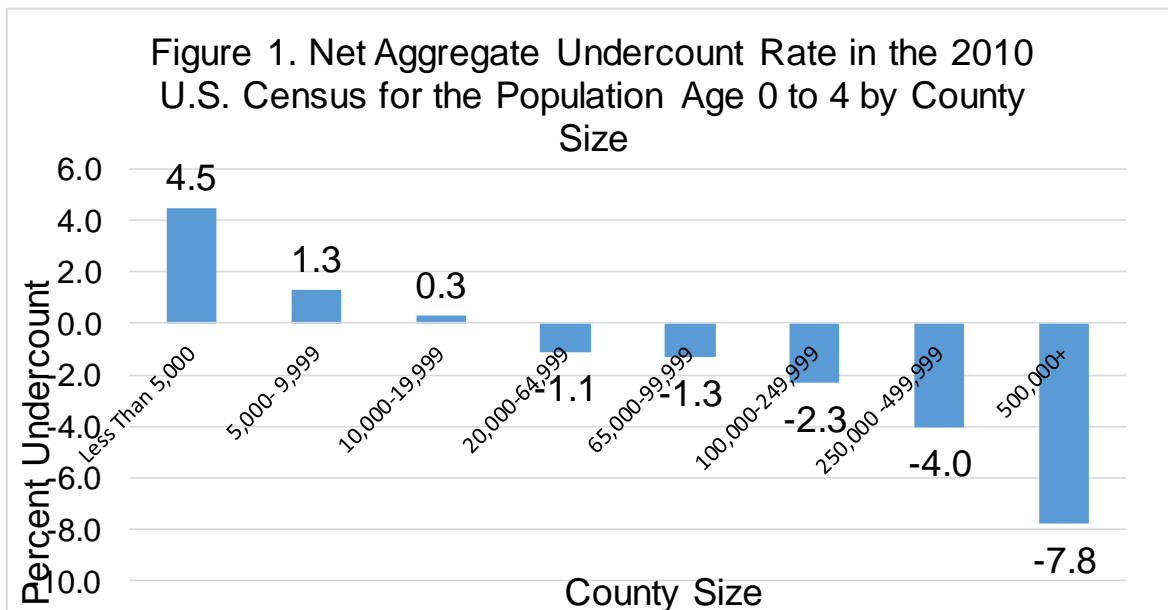
Table 3 shows that 128 counties with half a million more people had a net undercount rate of 7.8 percent for the population age 0 to 4. In the ten largest counties in the country the mean net undercount rate was 11.3 percent (data not shown here). .

Data in Table 3 show that the net undercount of the population age 0 to 4 is highly concentrated. The 128 largest counties (those with total populations of 500,000 or more) account for more than 80 percent of the national net undercount for people age 0 to 4. The concentration of undercounted young children in these 128 counties is a product to the large number of young children who live in those counties and the high net undercount rate in those counties. About half of the young child population in the country live in the largest 128 counties and the net undercount rate for these counties (7.8 percent) is more than 50 percent higher than the nationwide undercount rate for age 0 to 4.

The cumulative net undercount by county size is shown graphically in Figure 2.

Table 4. Difference Between 2010 Census Counts and Vintage 2010 Population Estimates for Population Age 0 to 4 by Count Growth Rates 2000 to 2010

| Total Population Size of County | Number of Counties | Aggregate Population Estimate | Aggregate Census Count | Difference (Census - Estimate) | |
|---|---|---|---|---|---|
| | | | | Number | Percent |
| Decrease by 10% or more | 216 | 250,158 | 233,687 | -16,471 | -6.6 |
| Decreas between 0 and 10% | 878 | 2,528,138 | 2,427,129 | -101,009 | -4.0 |
| Increase between 0 and 9.9% | 1235 | 8,989,307 | 8,495,200 | -494,107 | -5.5 |
| Increase between 10 and 24.9% | 583 | 6,474,341 | 6,147,641 | -326,700 | -5.0 |
| Increase between 25 and 49.9% | 186 | 2,472,155 | 2,357,346 | -114,809 | -4.6 |
| Increase of 50% or more | 35 | 543,230 | 534,185 | -9,045 | -1.7 |
| Total | 3133 | 21,257,329 | 20,195,188 | -1,062,141 | -5.0 |

Sources: U.S. Census Bureau, Vintage 2010 County Population Estimates (CC-EST2010-ALLDATA.csv)

U.S. Census Bureau, Table P12 (Total Population) Sex by Age, 2010 Census Summary File 1 (DEC_10_SF1_P12_with_ann.csv).



Figure 2. Cumulative Undercount in 2010 U.S. Census for age 0 to 4 Across Counties
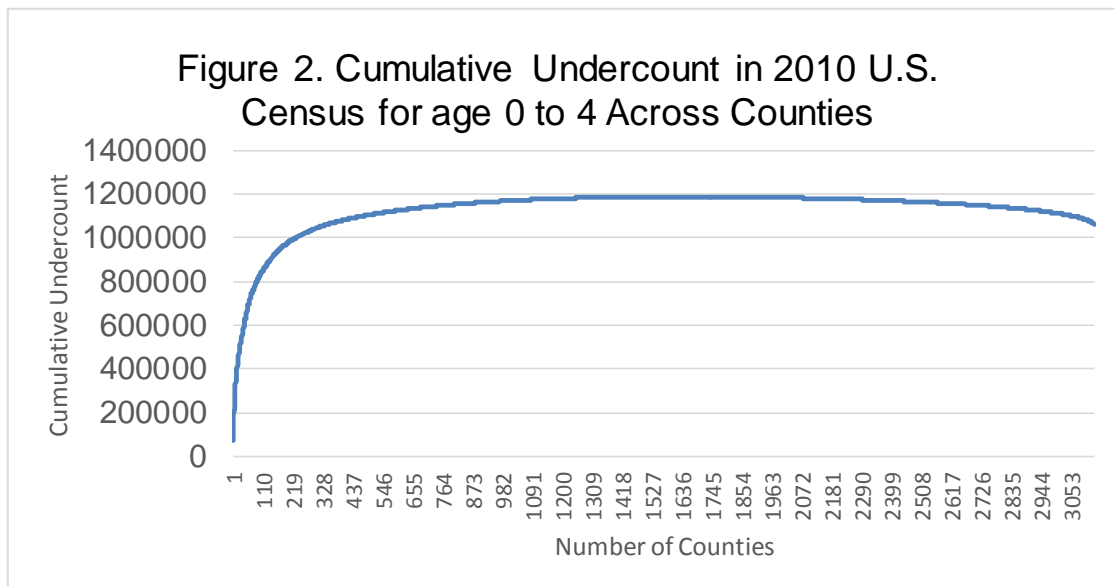
*4.3 County Population Change*

It would be useful to know if the net undercounts for the population age 0 to 4 are higher in counties that are growing more rapidly. Table 4 shows aggregate 2010 Census coverage of children age 0 to 4 for county groups based on total population growth rates from 2000 to 2010. The categories used here are those used by Yowell and Devine (2013, Table 8).

All categories of counties experienced a net undercount of young children. Some groups of counties had higher net undercount rates than others but it is difficult to detect a pattern related to county growth patterns. It may be noteworthy that the 35 counties the experienced growth rates of 50 percent of more between 2000 and 2010 had a net undercount rate of only 1.7 percent which is much lower than any other group of counties. It is not clear why this group of counties had a markedly lower net undercount rate than other counties.

*4.4 Race and Hispanic Origin Status*

The DA results for 2010 show that there was a net undercount of 7.5 percent of Hispanics age 0 to 4 and a net undercount rate of 6.3 percent for Blacks Alone or in Combination in this age group (O'Hare 2015). A recent update (Jensen, Benetsky, & Garrow 2016) showed the net undercount of Hispanic children age 0 to 4 in the 2010 Census was 6.5 percent. Consequently, one would expect counties with relatively large numbers of Hispanics and/or Blacks Alone or in Combination would have higher net undercount rates for the population age 0 to 4.

The Census Bureau did not produce Vintage 2010 Population Estimates of Black Alone or in Combination age 0 to 4 at the county level so I use the data for Black Alone age 0 to 4, taken from the Vintage 2010 Population Estimates to measure the percent of the children in each county who are Black. The vast majority of those categorized as Black Alone or in Combination in the 2010 Census are in the Black Alone category. So the distribution of Black Alone should be highly correlation with the distribution of Black Alone or in Combination. The percent Hispanic is also taken from the Vintages 2010 Population Estimates.

In the next section of this report which shows correlational analysis, it is clear that combining Black and Hispanic into one group is a more powerful predictor than using either separately. Therefore, counties are categorized based on the combined population of Blacks and Hispanics in this part of the analysis. This results in a slight double counting of black Hispanics, but it is doubtful this would have much impact on the outcome of the analysis.

Table 5. Difference Between 2010 Census Counts and Vintage 2010 Population Estimates for Population Age 0 to 4 by Percent Black and Hispanic

| Black and Hispanics as a Percentage of County Population | Number of Counties | Aggregate Population Estimate | Aggregate Census Count | Difference (Census - Estimate) | |
|---|---|---|---|---|---|
| | | | | Number | Percent |
| Less than 5% | 750 | 1,021,418 | 1,041,047 | 19,629 | 1.9 |
| 5-9.9% | 525 | 1,171,286 | 1,185,690 | 14,404 | 1.2 |
| 10-24.9% | 738 | 4,209,072 | 4,152,831 | -56,241 | -1.3 |
| 25-49.9% | 654 | 6,714,682 | 6,400,922 | -313,760 | -4.7 |
| 50%+ | 466 | 8,140,871 | 7,414,698 | -726,173 | -8.9 |
| Total | 3,133 | 21,257,329 | 20,195,188 | -1,062,141 | -5.0 |
| Sources: U.S. Census Bureau, Vintage 2010 County Population Esimates (CC-ET2010-ALLDATA.csv) | | | | | |
| U.S. Census Bureau, Table P12 (Total Population) Sex by Age, 2010 Census Summary File 1 (DEC_10_Sf1_p12_with_ann.csv). | | | | | |

The percent Black or Hispanic are used to sort counties into categories based on the percent of the county population age 0 to 4 that is Black or Hispanic. Table 5 shows aggregate census coverage rates for these groups

Table 5 shows there is a clear pattern with respect to relative minority population size in a county and the net undercount rate for young children. The higher the percent of Blacks and Hispanics in a county, the higher the net undercount rate. For counties where the percent Black and Hispanic is more than 50 percent, the net undercount rate was 9.1 percent. For counties with relatively small percentages of Blacks or Hispanics there were net overcount rates for age 0-4 in the 2010 Census. Figure 3 shows the patterns graphically.

### 4.6 Correlational Analysis

All the factors examined in this section are continuous variables and therefore amenable to correlational analysis. Pearson Product-Moment Correlation Coefficients are shown first and multiple regression analysis follows.

The demographic literature indicates that population estimates are usually more accurate for larger counties than for smaller one (Felton, 1986; Davis, 1994, O'Hare, 1988). Moreover, when one is trying to estimate a relatively small subpopulation like those age 0 to 4, the challenge is even more difficult. Consequently, it is reasonable to expect that the differences between the 2010 Decennial Census counts and the Vintage 2010 population estimates for many small counties are fraught with random error. Therefore, data for individual small counties are not used in this study.



Figure 3. Net Undercount for Age 0 t 4 by County Racial Composition
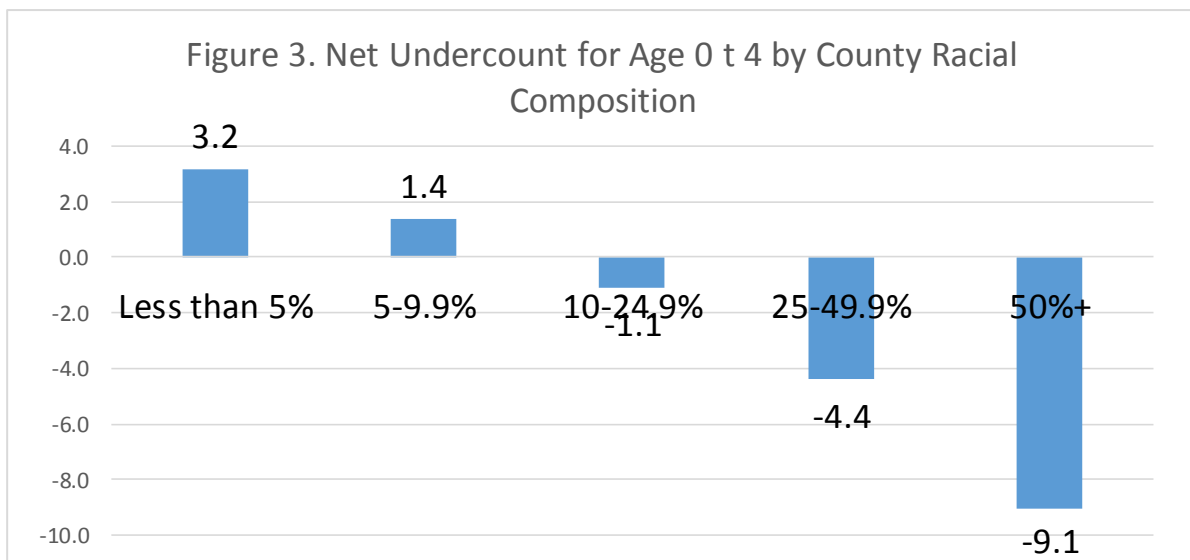
First the analysis includes all counties then the analysis is repeated with only the largest counties. Given the level of estimation errors for the population age 0 to 4 in smaller counties, it is reasonable to believe the correlations for all counties may be confounded. Analysis of larger counties (250,000 or more in total population) where estimates are more accurate may help sort out relationships.

Table 6 shows the correlations between county size, county growth, and county racial composition with net undercount rates for the population age 0 to 4. All correlations shown in Table 6 are statistically significant.

Using all counties in the analysis, the correlation between census coverage rate for the population age 0 to 4 and size of county population (total population) is -0.14. Since undercount rates are expressed as negative number this means larger counties are correlated with bigger undercounts for the population age 0 to 4. But the overall correlation may be depressed by significant estimation errors in smaller counties. Among 261 counties with a total population of 250,000 or more the correlation between the undercount rate and county population size is -0.32. Larger counties are associated with bigger undercounts for the population age 0 to 4.

**Table 6. Zero-Order Correlations of Net Undercount for the Population Age 0 to4 with Population Size, Population Growth and Minority Populations**

| | All counties | | Large counties (Total Population 250,000+) | |
|---|---|---|---|---|
| Total Population Size | -0.14 | ** | -0.32 | ** |
| Total Population Change 2000 to 2010 | -0.08 | ** | 0.15 | * |
| Percent of Population age 0 to 4 that are Black | -0.24 | ** | -0.34 | ** |
| Percent of the Population age 0 to 4 that are Hispanic | -0.18 | ** | -0.43 | ** |
| Percent of the Population age 0 to 4 that are either Black or Hispanic | -0.31 | ** | -0.64 | ** |
| ** statistically significant at .01 level | | | | |
| * statistically signficant at the .05 level | | | | |

Table 6 shows the correlation between county growth rates and census coverage rates for young children. The correlation coefficients for county growth rates, are relatively small (-.08 for all counties and +.015 for large counties). Also, the coefficient changes signs between the analysis of all counties and the analysis based only on large counties. This suggests that the relationship shown for all counties may be confounded by estimates errors in the small counties.

Three different measures of county racial composition are used here to look at Black and Hispanics separately and collectively. Table 6 shows that for all counties the correlations are in the expected direction (negative correlations for every minority group) but they are relatively modest in size. Correlations across the groups range from -0.18 to -0.31.

For large counties (those of 250,000 or more people) the correlations are in the predicted direction but much higher in magnitude than those for all counties. Consistent with findings from O'Hare (2014c) there is a higher correlation when Blacks and Hispanics are combined (-0.64) into one measure of minority population than for either Blacks (-0.34 for Non-Hispanic Black alone) or Hispanics (-0.43) although all the correlations are statistically significant. It is possible that the higher correlation of the combined Black and Hispanic variable reflects the fact that Blacks are the dominant minority population in most of the Southeastern states and Hispanics are the dominant minority population in most Southwestern states. Consequently, combining the two groups together captures the dominant minority population in more counties.

Table 7. Inter-correlations among County Size, County Growth Rate, and Percent Black/Hispanic

| | Percent Black and Hispanic | 2010 Total Population Size |
|---|---|---|
| 2010 Total Population Size | 0.21 | |
| Growth Rate 2000 to 2010 | 0.12 | 0.14 |
| All Correlations are significant at the 0.01 level. | | |

Finally, the main independent variables examined above are used in a multiple regression analysis to try and estimate their independent relationships with net undercounts for the population age 0 to 4. Table 7 shows the inter-relationships among the main independent variable and indicators that none are highly correlated so multicollinearity is not a problem. The model is run once with all counties and again with only the largest counties, that is those with a population of 250,000 or more.

**Table 8. Multiple Regression Analysis with Percent Net Undercount age 0-4 as Dependent Variable and Population Size, Population Growth and Percent Minorities and Independent Variables**

| | All Counties | | | Large Counties (Total Population 250,000+) | |
|---|---|---|---|---|---|
| | Standardized Coefficients | | | Standardized Coefficients | |
| Total Population Size 2010 | -0.07 | ** | | -0.12 | ** |
| Total Population Growth 2000 to 2010 | -0.04 | ** | | 0.17 | ** |
| Percent of Population Age 0-4 Black or Hispanic | -0.29 | ** | | -0.61 | ** |
| Adjusted Multiple R -squared | 0.1 | | | 0.45 | |
| ** statistically significant at .01 level, * statistically significant at .05 level | | | | | |

The results of the multiple regression analysis are shown in Table 8. All of the standardized coefficients are statistically significant, but there is substantial variation in the magnitude of the coefficients.

The results indicate that racial composition is a more powerful predictor of net undercount for age 0 to 4 at the county level than county size or county growth rate. Again the relationships seen in the analysis including only the larger counties are more pronounced than those for all counties.

Examining the Adjusted Multiple R-squared ta the bottom of Table 8 indicates that the independent variables employed here are much more predictive of net undercount rates for young children in larger counties than all counties. The set of independent variables explains 45 percent of the variation in net undercount rates for young children in larger counties but only 10 percent in all counties.

It is also interesting that the relationship between county growth and net undercount rates changes sign between the analysis of all counties and the analysis of larger counties. This reinforces that idea that determining relationship based on all counties may be difficult because of the estimates errors in small counties.

## 5. Limitations

The results of this study rely on the comparison of Vintage 2010 Population Estimates with 2010 Census counts for counties. As noted earlier, the difference between the Census counts and the Population Estimates reflect estimation errors as well as census coverage. This is a notable problem for individual counties, particularly smaller counties, but it is mitigated when counties are grouped together. While the estimates presented here involve some estimation error, these figures are the best data available to understand the geographic distribution of the young children missed in the 2010 Census.

The higher levels of estimation error for small counties reinforces the importance of examining larger counties separately from all counties. Devine and Yowell (2013, Table 7) for that for counties of 250,000 to 500,000 people the Mean Absolute Percent Error (MAPE) is 1.47 and for counties of 500,000 or more people the MAPE is 1.50. Consequently more of the difference between the Population Estimates and the Census count for counties in these size categories is accounted for by the net undercount not estimation error.

## 6. Discussion and Implications

This study extends a set of analyses that address one of the biggest problems in terms of the accuracy in the U.S. Decennial Census, namely the high net undercount of young children. Previous studies have identified young children (age 0 to 4) as the age group with the highest net undercount in the 2010 U.S. Decennial Census (O'Hare 2015). Analysis here addresses the geographic distribution of the net undercount of young children.

All of the correlations examined here are more pronounced for larger counties than all counties which probably reflects the impact estimates errors in smaller counties (see Tables 6 and 8). This suggest that further research examining county characteristics and net undercounts for age 0 to 4 should be focused on larger counties.

The analysis indicates that the net undercount of young children is concentrated in the largest counties in the country. The concentration of the net undercount of young children in a relatively small number of large counties is a product of the large number of young children who live in those large counties and the relatively high net undercount rate for young children in those counties.

The findings here suggest that planning for the 2020 Census should focus on households with young minority children located in large counties. The potential impact of such a focus can be shown in the following way. If the counties with total populations of more than 500,000 had the same undercount rate for age 0 to 4 as all counties in the 2010 Census, the net undercount of young children would have been reduced by 300,000 and the net undercount rate would have been a little over 3 percent.

The evidence presented here also indicates that the net undercount rate for children age 0 to 4, is higher in counties with relatively large minority populations. But it is not clear if this relationship is direct or spurious. For example, it is likely that counties with high percentages of Blacks or Hispanic also have high poverty rates. According to Schwede et al (2014, p 293-294), " Though there is no reason to believe that race or ethnicity in and of itself leads to coverage error, it seems that some underlying variables associated in past studies with undercounting may also be correlated with race ( e.g. mobility, complex living situations, and language isolation)."

In future research it would be useful to try and untangle the web of relationships between minority status and hard-to-count characteristics to better understand what is driving the high net undercount rate of young children in counties with high percentages of Black or Hispanics. However, such research is likely to be complicated because of the high correlations between percent minorities and other characteristics related to undercounts such as poverty rates.

## 7. Summary and Conclusions

The data examined here indicate that the national net undercount rate for the population age 0 to 4 varies substantially across counties. About half of all counties had a net undercount and half had a net overcount of the population age 0 to 4. Moreover, the data show that larger counties account for the vast majority of the national net undercount for the population age 0 to 4. In the 128 largest counties based on total population, there was a net undercount of 823,000 persons age 0 to 4 which accounts for more than 80 percent of the nationwide net undercount for this age group. This information about where the net undercount rates for young children are the highest should help the U.S. Census Bureau prepare for the 2020 Decennial Census. The data presented in this study will help the Census Bureau pinpoint the places that deserve special attention in the 2020 Census.

## References

Adams, J. W., & Kasakoff, A. B. (1991). Estimates of U.S. Decennial Census Underenumeration Based on Genealogies. *Social Science History, 15*(4), winter, 527-543.

Adlakha, A. L., Robinson, J. G., West, K. K, & Bruce, A. (2003). Assessment of Consistency of Census Data with Demographic Benchmarks at the Subnational Level. Census 2000 Evaluation 0.20 U.S. Census Bureau, August 18.

Coale, A. J. (1955). The Population of the United States in 1950 Classified by Age. Sex and Color-a Revision of Census Figures. *Journal of the American Statistical Association, 50,* 16-54.

Coale, A. J, and Rives, N. W. (1973). A Statistical Reconstruction of Black Population of the United States: 1880 to 1970: Estimates of True Numbers by Age and Sex, Birth Rates, and Total Fertility. *Population Index, 39*(1), 3-36. https://doi.org/10.2307/2733759

Coale, A. J., & Zelnick, M. (1963). *New Estimates of Fertility and Population in the United States,* Princeton University Press, Princeton NJ.

Cohn, D. (2011). State Population Estimates and Census 2010 Counts: Did they match? Pew Social and Demographic Trends, Pew Research Center, Washington, DC. January 12.

Davis, S. T. (1994). Evaluation of Post Census County Estimates for the 198. *Current Population reports*, Series P-25, No. 963, U.S. Census Bureau, Washington DC.

Devine, J., Sink, L., DeSalvo, B., & Cortes R. (2010). The Use of Vital Statistics in the 2010 Demographic Analysis Estimates. Census Bureau Working Paper, No. 88, available online at http://www.census.gov/population/www/documentation/twps0088/twps0088.pdf

Felton, G. T. (1986). Evaluation of Population Estimation Procedures for Counties: 1980, U.S. Census Bureau, Current Population Reports, P- 25, No. 984, U.S. Government Printing Office

Groves, R. M. (2011). Census Learning: Lessons from 2010, Planning for 2020. Prepared Statement Testimony before Committee on Homeland Security and Governmental Affairs Subcommittee on Federal Financial Management, Government Information, Federal Services, and International Security, U.S. Senate, April 6, 2011.

Jensen, E. B., Benetsky, M., & Garrow, S. (2016). Updating the 2010 Demographic Analysis Estimates of the Hispanic Population, Paper presented at the Southern Demographic Association Conference, October, Athens, GA.

Mayol-Garcia, Y., & Robinson, J. G. (2011) "Census 2010 Counts Compared to the 2010 Population Estimates by Demographic Characteristics," Poster presented at the      Southern Demographic Association Conference, October , Tallahassee, FL.

Mule, T. (2012). Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States. DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01, Table 12.

O'Hare W. P. (2012a). The Net Undercount of Children in the 2010 Decennial Census. Paper presented at the Applied Demography Conference at the University of Texas, San Antonio, January.

O'Hare, W. P. (2013). *Difference Between 2010 Census Counts and Vintage 2010 Population Estimates for Age 0-4 at the State and County Level*. Poster presented at the 2013 A Annual Conference of the Population Association of America, New Orleans, LA.

O'Hare, W. P. (1999), The Overlooked Undercount: Children Missed in the Decennial Census, The Annie. E. Casey Foundation, Baltimore, MD.

O'Hare, W. P. (2009), Why Are Young Children Missed So Often in the Census, KIDS COUNT Working Paper, The Annie E. Casey Foundation, Baltimore MD. December.

O'Hare, W. P. (1988). How to Evaluate Population Estimates. *American Demographics, 10*(1), January.

O'Hare, W. P. (2014a). Assessing Net Coverage for Young Children in the 2010 U.S.Decennial Census, *International Journal of Population Research*, 2014.

O'Hare, W. P. (2014b). The Net Undercount of Children the 2010 U.S. Decennial Census. in *Emerging Techniques in Applied Demography*, (edited by M. Hoque and L. Potter). *Applied Demography Series, 4,* 39-51, October.

O'Hare, W. P. (2014c). State-Level 2010 Census Coverage Rates for Young Children. *Population Research and Policy Review, 33*(6), 797-816. https://doi.org/10.1007/s11113-014-9335-8

O'Hare, W. P. (2014d). Assessing Net Coverage Error for Young Children in the 2010 U.S. Decennial Census. *Center for Survey Measurement Study Series (SurveyMethodology #201402).* U.S. Census Bureau. Available online at http://www.census.gov/srd/papers/pdf/ssm2014-02.pdf

O'Hare, W. P. (2014e). 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. (2014f). *Estimating the Net Undercount of Young Children in the 2010 U.S. Decennial Census at the County Level*, Poster presented at the 2014 Population Association of America Conference, Boston, MA. April.

O'Hare, W. P. (2015). *The Undercount of Young Children in the U.S. Decennial Census. Springer Publishers*. https://doi.org/10.1007/978-3-319-18917-8

Robinson, G. J., Bashir, A., Das Dupta, P., & Woodward, K. A. (1993). Estimates of Population Coverage in the 1990 United States Census Based on Demographic Analysis. *Journal of the American Statistical Association, 88*(423), 1061-1071. https://doi.org/10.1080/01621459.1993.10476375

Robinson, J. G., & Kobilarkic, E. L. (1995). Identifying Differential Undercounts at LocalGeographic Levels: A Targeting Database Approach. Paper presented at theAnnual Meeting of the Population Association of America, San Francisco.

Robinson, J. G., &Adlakha, A. (2002). Comparison of A.C.E. Revision II with Demographic  Analysis. DSSD A.C.E. Revision II Memorandum Series #PP-41 U.S. Census Bureau.

Schwede, L., Terry, R., & Hunter, J. (2014). Ethnographic evaluations on coverage  of Hard-to-Count minority in the US Decennial Censuses," in *Hart-to-Surve Populations*, Edited by Tourangeau, R., Edwards, B., Johnson, T.P., Wolter, K.M., and Bates, N. Cambridge University Press, Cambridge, England, 293-315.

Siegel, J. S., & Zelnik, M. (1966). An Evaluation of Coverage in the 1960 U.S. Census of Population by Techniques of Demographic Analysis and by Composite Methods." In *Proceedings of the Social Statistics Section of the American Statistical Association*, 71-85. Washington, D.C.: American Statistical Association.

Siegel, J. S., & Zelnik, M. (1966). An Evaluation of Coverage in the 1960 U.S. Census of Population by Techniques of Demographic Analysis and by Composite Methods. In *Proceedings of the Social Statistics Section of the American Statistical Association*, 71-85. Washington, D.C.: American Statistical Association.

Siegel, J. S., Passel, J. S., Rives, N. W., & Robinson, J. G., (1977). Developmental Estimates of the Coverage of the Population of States in the 1970 Census: Demographic Analysis. *Current Population Reports*, Special Studies, Series P-23, No.65, Dec.

U.S. Census Bureau. (2003). *Technical Assessment of A.C.E. Revision II*, U.S. Census Bureau, Washington, DC.

U.S. Census Bureau. (2010b) Tables released at December 2010 Conference Table 8 available online at http://www.census.gov/coverage_measurement/demographic_analysis/

U.S. Census Bureau. (2012a) Estimation Methodology, U.S. Census Bureau, Washington DC. http://www.census.gov/popest/methodology/2009-stco-char-meth.pdf

U.S. Census Bureau. (2012b) Documentation for the Revised 2010 Demographic Analysis; Middle Series Estimates., U.S. Census Bureau, Washington, DC.

U.S. Census Bureau. (2013).

U.S. Census Bureau. (2014). Final Task Force Report on the Undercount of Young Children, U.S. Census Bureau, Washington, DC.

U.S. Census Bureau (no date). Methodology for the United States Resident Population Estimates by Age, Sex, Race and Hispanic Origin and (Vintage 2009): April 1, 2000 to July 1, 2009. U.S. Census Bureau, Available on the Census Bureau' website at http://www.census.gov/popest/methodology/2009-stco-char-meth.pdf

U.S. Census Bureau. (2010a), The Development and Sensitivity Analysis of the 2010 Demographic Analysis Estimates, Population Division Background paper of DA Conference Dec 6, 2010. 11/29/2010, Table 2.

Velkoff, V. (2011), Demographic Evaluation of the 2010 Census, Paper presented at the 2011 PAA annual Conference, Washington, DC, March.

West, K., & Robinson, J. G. (1999), What Do We Know About the Undercount or Children? U.S. Census Bureau, Population Division working paper.

Yowell, T., & Devine, J. (2013). Evaluating Current and Alternative Methods to Produce 2010 County Population Estimates, "Population Division Working Paper 100, U.S. Bureau of the Census. Washington DC, July.

Zeller, A. (2006). Inconsistency between Accuracy and Coverage Evaluation Revision II and Demographic Analysis Estimates for Children 0 to 9 Years of Age, Paper presented at the American Statistical Association annual conference.