Chapter 5
State and County Level 2010 U.S. Census Coverage Rates for Young Children

Abstract The national net undercount rate for young children masks large differences across states and counties. State-level census coverage rates for age 0–4 range from a net undercount of 10.2 % in Arizona to a 2.1 % overcount in North Dakota. State net undercount rates for children age 0–4 are correlated with state population size, racial/Hispanic composition as well as several Hard-to-Count characteristics such as linguistic isolation, education, and employment. The net undercount of young children is concentrated in the largest counties. The 128 largest counties, all with populations of 500,000 or more, account of 77 % of the total net undercount for age 0–4.

To date there has been little information on how the national Census undercount is distributed among sub-national geographic units such as states and counties. This Chapter addresses the net undercount of young children at the state and county level and also examines correlates of variation in net undercount rates across units of geography.

One of the major limitations of the Demographic Analysis (DA) technique for measuring the Census undercounts for most demographic groups is that it can only be applied at the national level. However, young children are an exception to this rule. For the population under age 10, the U.S. Census Bureau's post-Census Population Estimates provide a population estimate that is independent of the Census. This approach cannot be used for the population over age 10 because the 2010 Estimates are derived from the 2000 Census, so the estimates are not independent of the census.

Specifically, the Census Bureau's Vintage 2010 State and County Population Estimates for young children provide an opportunity to assess sub-national Census results. The Vintage 2010 State and County Population Estimates for those under

---

This Chapter draws heavily on an article by O'Hare (2014c).

age 10 are based on births, deaths, and net migration, which is essentially the same demographic accounting equation used in DA.

In this Chapter state and county-level net undercounts of young children are developed by comparing the U.S. Census Bureau's Vintage 2010 Population Estimates for the population age 0–4 to the 2010 U.S. Census counts for this age group. The analysis focuses on the population age 0–4, rather than 0–9, because the 2010 DA analysis shows the net undercount for the 0–4 age group is much higher than that for the age group 5–9 (see Chap. 3). The 2010 national undercount rate for the population age 0–4 based on DA is 4.6 % compared to only 2.2 % for age 5–9. Therefore it is important to examine the population age 0–4 separately from those aged 5–9. At the same time, it is worth noting there is a very high positive correlation across states in the net undercount rates of the population 0–4 and the population age 5–9 (r = +0.97). Consequently, patterns observed for the population age 0–4, are also likely to be seen in the population age 5–9.

The case for developing sub-national estimates of Census coverage was made eloquently more than 30 years ago by Siegel et al. (1977, p. 1),

> The importance of Census counts of the population in determining political representation, in the disbursement of public funds, and in the planning, conduct, and evaluation of various private and public program has aroused considerable interest in the accuracy of Census counts for States and smaller political units and, particularly, in the availability of estimates of coverage for these areas in the last Census.

States are a useful geographic unit to use for this analysis because most of the past work on sub-national Census coverage has focused on states. In addition, the Population Estimates at the state level are more accurate than those for counties or other smaller geographic units so the undercount estimates for states are more robust than those for counties. Yowell and Devine (2013, Table 2) found the mean absolute percentage error for county Population Estimates was three times that of states in assessing the 2010 Population Estimates. This is consistent with the general principal that population estimates for larger places (in population size) are typically more accurate than smaller places (Felton 1986; Davis 1994; O'Hare 1988).

States are also a useful unit of analysis for geo-political reasons. In terms of public policies related to children, states are much more important than counties and their importance has been growing. According to Gormely (2012, p. 100),

> The role of state government in funding and regulating elementary and secondary education has long been of critical importance, and state expenditures on child health through Medicaid and Child Health Insurance Program (SCHIP), have increased significantly in recent years. More than federal government, state governments devote a substantial percentage of their time and their financial resources to children.

5.1 Background

Past research on sub-national assessments of the U.S. Census results are limited. Much of the public and political interest in sub-national undercounts was first generated by Hill and Steffes (1973) who used a synthetic estimation technique to produce state and local area undercount estimates of the 1970 Census results.

Following the 1970 Decennial Census, Siegel et al. (1977) also examined Census coverage for states and for various population groups defined by race and age. Several different approaches were used with mixed results.

Following the 1980 Census Isaki and colleagues (Isaki et al. 1985) examined a couple of ideas for developing estimates of net undercounts at subnational levels. These efforts involved use of both the PES and the DA results.

Following the 1990 Decennial Census, Robinson et al. (1993) offered a set of 1990 Census undercount estimates for states for the total population (all ages). There were no estimates for children and the estimates were only evaluated at the multi-state regional level. The authors also proposed alternatives for evaluating the 2000 Census at the state and sub-state levels and listed several reasons why such an evaluation is needed. Robinson and Kobilarkic (1995) also discuss sub-national evaluations the 1990 Census using a DA-like approach.

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

Based on unpublished Census Bureau data, Darga (1999, p. 32) examined sub-national undercounts of children under age 10 in the 1990 Census. However, his estimates are based on the 1990 Post Enumeration Survey rather than DA and they were only examined for multi-state regions, not states.

Cohn (2011) compared the Census Bureau's state 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 quite close for most states in terms of total population.

Mayol-Garcia and Robinson (2011) examined differences between Population Estimates and Census counts for states for age 0–4 and age 0–9 populations in the 2010 Census but only provided limited results and did not explore any patterns across states. However, regarding the state-level data on the net undercounts of the population age 0–4, Mayol-Garcia and Robinson (2011, p. 3) note, "The relatively large differences noted nationally for 0–4 year olds are observed at the state level as well." O'Hare (2013, 2014a, b) also provides some preliminary analysis of 2010 net Census coverage rates for young children at the sub-national level.

Based on their analysis of 2000 U.S. Census data, Adlakha et al. (2003, p. v) recommended 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 present analysis responds to those recommendations.

The present analysis extends previous research by examining state and county level Census coverage for young children in more detail and examining factors correlated with variations in state differences in net Census coverage rates for young children. First, the state net undercount rates for age 0–4 are developed and examined in relation to state overall population size, the racial/ethnic composition of states, as well other state characteristics thought to be related to Census undercounts. Then a similar analysis is provided for counties.

5.2 Methodology and Data Sources for State-Level Analysis

The methodology used for state Population Estimates is very similar to that used for DA. Both can be described as using a cohort-component approach where each component of population changes (births, deaths and net migration) is estimated separately for each birth cohort. The biggest difference between the national DA and the state Population Estimates is the inclusion of migration across states. Migration between states is captured in the Census Bureau administrative records technique that uses federal tax records to estimate such migration (U.S. Census Bureau 2012b).

Data from the 2010 American Community Survey indicate that 89 % of the population age 0–4 were living in the same state where they were born. Therefore, the overwhelming majority of children age 0–4 estimated in each state come from births in that state. The heavy reliance on birth certificate data and the high quality of birth certificate data provide a strong foundation for state Population Estimates for the population age 0–4.

The state Population Estimates are derived using the formula in Eq. 5.1, which is taken from U.S. Census Bureau (2012a, b);

$$P1 = P0 + B - D + NDM + NIM \tag{5.1}$$

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.

The estimated undercounts and overcounts shown here also include errors in the population estimates. However, since the mean absolute percent error for state Population Estimates is on the order of 1 % (Yowell and Devine 2013) and the average state net undercount for the population 0–4, is around 3.5 %, the bulk of the difference appears to be the net undercount rather than the estimation error. The 1 % error noted above is for the total population, not the 0–4 population, but it is the best estimate available for the likely accuracy for the 0–4 population.

In the remainder of this Chapter, the differences between the Census counts and Population Estimates are shown as the Census count *minus* the estimate. This is consistent with the convention used by Velkoff (2011). This calculation is sometimes labeled "net Census coverage error" in other research. A negative number implies a net undercount and a positive number implies a net overcount. I chose to use the net Census coverage error because I feel having an undercount reflected by a negative number is more intuitive and is consistent with the presentation of 2010 DA analysis by Velkoff (2011).

In converting the differences between Census counts and DA estimates to percentages, the difference is divided by the DA estimate. Estimates are shown rounded to the nearest thousand for readability.

5.2.1 The Data

The Vintage 2010 State Population Estimates used here are taken from the Census Bureau's file labeled "Annual State Resident Population Estimates for 5 Race Groups (5 Race Alone or in Combination Groups) by Age, Sex, and Hispanic Origin: April 1, 2000 to July 1, 2010." The file is also denoted as "SC-EST2010-ALLDATA5." The file was released March 2012 and it is available on the Census Bureau's website.

These estimates include 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, for the population age 0–4 are used in this study.

The data from the 2010 U.S. 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 total population and for the population age 0–4 were taken from this file.

The District of Columbia was not included in the state analysis for two reasons. First, The District of Columbia does not operate like a state. Demographically and governmentally, the District of Columbia is more like a large city than a state. Second, the net undercount rate of young children for the District of Columbia is an outlier with respect to state undercount rates for the population age 0–4. The net undercount rate for the District of Columbia was 16.2 %, while the highest net undercount rate for age 0–4 in any state was 10.2 % in Arizona.

Table 5.1 provides national data from the U.S. Census Bureau's Vintage 2010 Population Estimates, the U.S. Census Bureau's May 2012 Demographic Analysis (DA) release, and the 2010 U.S. Decennial Census. For the total population the figures from the three sources are remarkably similar. In reality, the similarities across all three sources for the total population are the product of large counterbalancing differences among age groups.

For the population age 0–4, 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 importantly for this paper, both the DA estimate and the Population Estimate figures are substantially higher than the 2010 U.S. Census count (20,201,000). The difference between the DA estimate and the U.S. Census count is 4.6 % for the population age 0–4 and the difference between the Vintage 2010 Population Estimate and the U.S. Census count is 5.0 %. Both the DA estimates and the Vintage 2010 Population Estimates indicate there was a net undercount of about one million children age 0–4.

Table 5.1 Difference between vintage 2010 population estimates, May 2012 DA estimates, and 2010 U.S. census counts by age

(Figures in 1000)		
	All ages	Age 0–4
Population figures		
Population estimates[a]	308,450	21,263
May 2012 DA[b]	308,346	21,171
Census counts[c]	308,746	20,201
Numeric differences		
Census—population estimates	296	−1062
Census—DA	400	−970
Population estimates—DA	104	92
Percentage differences		
(Census—population estimates)/population estimates	0.1	−5
(Census—DA)/DA	0.1	−4.6
(Population estimates—DA)/DA	0	0.4

[a]Vintage 2010 Population Estimate Program (PEP) results for 4/1/2010
[b]U.S. Census Bureau, Population Division, 2010 Demographic Analysis Released May 2012
[c]2010 U.S. Census Summary File 1, Table DP-1 Profile of General Population and Housing Characteristics

The consistency between the national population DA estimates and the corresponding Vintage 2010 State Population Estimates at the national level suggests the Vintage 2010 Population Estimates are likely to be useful for estimating the distribution of the national undercount of the population age 0–4 among the states.

The main reason the Vintage 2010 Population Estimates differ slightly from the 2010 DA estimates is the fact that the DA estimates issued in May 2012 used updated vital events data for 2008, 2009, and the first quarter of 2010. When the Vintage 2010 Population Estimates were issued, the Census Bureau had to estimate the number of births and deaths in 2008, 2009 and the first quarter of 2010 because the empirical data was not yet available from the National Center for Health Statistics. By the time the revised DA estimates were issued in May 2012, the final figures for births and deaths in 2008, 2009 and the first quarter of 2010 were available from the National Center for Health Statistics. The fact that the observed figures for births (used in preparing the 2010 DA estimates released May 2012) were lower than the estimated figures used to prepare the Vintage 2010 Population Estimates, results in the DA estimates being slightly lower than the Vintage 2010 Population Estimates. This results in the Population Estimates providing slightly higher national net undercount rates than DA. But this difference is relatively minor; a 4.6 % net undercount for DA compared to a 5.0 % net undercount for the Population Estimates. Either estimate shows young children had the highest net undercount rate of any age group, by far.

5.3 State-Level Results

Table 5.2 provides several summary measures of differences between the Vintage 2010 Population Estimates and the 2010 U.S. Census counts for state populations age 0–4. For the population age 0–4, the mean difference is −21,114. In relative terms, the mean difference was −3.4 % for the population age 0–4.

Some of the positive and negative differences may cancel each other out in calculating the mean, so it is useful to examine the size of absolute differences. This provides a measure of the size of difference between Census Counts and Population Estimates regardless of direction of the difference. The mean *absolute numeric* difference was 21,176 for the population aged 0–4 and the relative absolute difference was 3.5 %. Since 46 of the 50 states had net undercounts for the population age 0–4 it is not surprising that the numeric mean and the mean of absolute values are similar for young children.

The average state had a net undercount rate of 3.4 % for the population age 0–4, which is substantially less than the national net undercount rate (5.0 % for the Vintage 2010 Population Estimates and 4.6 % based on DA). This indicates that the national undercount for the population age 0–4 is not distributed evenly across the states but is driven by larger errors in large states. This point will be examined in more detail later in this Chapter.

Table 5.3 shows the numeric and percent differences between the Vintage 2010 State Population Estimates and the 2010 U.S. Census counts for the population age 0–4 for each state, developing by subtracting the state estimate from the Census counts.

The data in Table 5.3 indicate that the national undercount rate for the population age 0–4 (5.0 %) masks striking differences across the states. Differences between the 2010 U.S. Census counts and the Vintage 2010 State Population Estimates for age 0–4 range from a net undercount of 10.2 % in Arizona to a 2.1 % net overcount in North Dakota. There were 12 states with net undercounts of 5 % or more.

In population terms, Table 5.3 shows the differences between the 2010 U.S. Census counts and the Vintage 2010 State Population Estimates for age 0–4 range from a net undercount of 210,125 in California to a net overcount of 906 in North Dakota. There are 26 states where the difference between the Vintage 2010 Population estimate and the 2010 Census count for the population age 0–4 was more than 10,000.

Table 5.2 Summary table of *state* differences (census minus population estimates) between Vintage 2010 population estimates and 2010 census count for population age 0–4

Mean numerical difference	−21,114
Mean percent difference	−3.4
Number of states with a net overcount	4
Number of states with a net undercount	46
Mean absolute numeric difference	21,176
Mean absolute percentage difference	3.5

The net undercount of young children is geographically pervasive at the state level. Only four states (North Dakota, Vermont, Montana, and Wyoming) had a net overcount.

There are no standard errors or other measures of uncertainty attached to the Population Estimates or the Census counts, so one cannot employ traditional statistical significance testing. However, in the DA release of December 2010 the Census Bureau offered results for five different DA scenarios to illustrate the uncertainty surrounding the DA estimates. The results of the five scenarios for the population age 0–4, ranged from a low of 21,181,000 to high of 21,265,000. In percentage terms, the difference between the lowest estimate and the highest estimate is 0.4 %. This provides at least one guide to expected errors in the DA estimates.

When state differences between Population Estimates and 2010 U.S. Census counts are compared to the national difference (5.0 %) only four states (Colorado, Delaware, Massachusetts and Mississippi) are within 0.4 percentage points of the national rate. Moreover, only eleven states have a net undercount rate within one percentage point of the national net undercount rate for the population age 0–4. There are only two states (Maine and Wyoming) where the net undercount rate for the population age 0–4 are within 0.4 % points of zero. This suggests significant real variation across the states in the net undercount of the population age 0–4. It also indicates that the national net undercount rate for the population age 0–4 tells us very little about the net undercount rate of young children in most states.

It should be noted that the state-wide net undercount rates examined here reflect significant differences across sub-state areas. In many states, the state figure was a product of net undercounts for young children in large counties and net overcounts in smaller counties. There were 13 states where large counties (populations of 250,000 or more) accounted for all of the net undercount for age 0–4 in the state. This point will be pursued later in this Chapter.

5.4 Characteristics Associated with State Net Undercount Rates for Population Age 0–4

Data presented in the previous section make it clear that the net Census coverage rates for the population age 0–4 vary substantially across the states. In this section, I examine several state characteristics to see which ones are most highly correlated with the net Census coverage rates for young children. While correlation is not the same as causation, finding out which characteristics are most highly correlated with state differences in net undercount of age 0–4 will shed light on what are the most likely causes of the net undercount for young children, or perhaps identify which factors are not likely to be causally related to the net undercount of young children.

Table 5.3 State 2010 census counts minus Vintage 2010 population estimates for the population age 0–4

	Vintage 2010 population estimate program	2010 census counts	Numeric difference (census − estimates)	Percent difference [(census − estimates)/ estimates] * 100
Alabama	317,716	304,957	−12,759	−4
Alaska	54,888	53,996	−892	−1.6
Arizona	507,581	455,715	−51,866	−10.2
Arkansas	204,509	197,689	−6820	−3.3
California	2,741,458	2,531,333	−210,125	−7.7
Colorado	362,049	343,960	−18,089	−5
Connecticut	208,901	202,106	−6795	−3.3
Delaware	59,098	55,886	−3212	−5.4
Florida	1,163,423	1,073,506	−89,917	−7.7
Georgia	741,568	686,785	−54,783	−7.4
Hawaii	90,687	87,407	−3280	−3.6
Idaho	122,759	121,772	−987	−0.8
Illinois	887,157	835,577	−51,580	−5.8
Indiana	444,854	434,075	−10,779	−2.4
Iowa	203,842	202,123	−1719	−0.8
Kansas	207,830	205,492	−2338	−1.1
Kentucky	289,924	282,367	−7557	−2.6
Louisiana	323,481	314,260	−9221	−2.9
Maine	69,779	69,520	−259	−0.4
Maryland	381,289	364,488	−16,801	−4.4
Massachusetts	387,055	367,087	−19,968	−5.2
Michigan	603,376	596,286	−7090	−1.2
Minnesota	362,611	355,504	−7107	−2
Mississippi	221,144	210,956	−10,188	−4.6
Missouri	402,489	390,237	−12,252	−3
Montana	62,143	62,423	280	0.5
Nebraska	134,530	131,908	−2,622	−1.9
Nevada	200,843	187,478	−13,365	−6.7
New Hampshire	71,949	69,806	−2143	−3
New Jersey	555,419	541,020	−14,399	−2.6
New Mexico	153,402	144,981	−8421	−5.5
New York	1,228,587	1,155,822	−72,765	−5.9
North Carolina	657,178	632,040	−25,138	−3.8
North Dakota	43,689	44,595	906	2.1
Ohio	738,494	720,856	−17,638	−2.4
Oklahoma	274,800	264,126	−10,674	−3.9

Table 5.3 (continued)

	Vintage 2010 population estimate program	2010 census counts	Numeric difference (census − estimates)	Percent difference [(census − estimates)/ estimates] * 100
Oregon	248,107	237,556	−10,551	−4.3
Pennsylvania	750,821	729,538	−21,283	−2.8
Rhode Island	59,523	57,448	−2075	−3.5
South Carolina	313,334	302,297	−11,037	−3.5
South Dakota	59,998	59,621	−377	−0.6
Tennessee	423,204	407,813	−15,391	−3.6
Texas	2,083,265	1,928,473	−154,792	−7.4
Utah	274,529	263,924	−10,605	−3.9
Vermont	31,699	31,952	253	0.8
Virginia	532,874	509,625	−23,249	−4.4
Washington	457,757	439,657	−18,100	−4
West Virginia	106,985	104,060	−2925	−2.7
Wisconsin	361,741	358,443	−3298	−0.9
Wyoming	40,085	40,203	118	0.3

Sources U.S. Census Bureau (2012a, b)

5.4.1 State Size

Table 5.4 shows the percent difference between 2010 Census counts and Vintage 2010 Population Estimates for population age 0–4 by quintiles of state population size. For all the population size groups, there is a net undercount for the 0–4 population but larger states (in terms of total population) tend to have bigger percentage differences between the 2010 U.S. Census counts and the Vintage 2010 Population Estimates for age 0–4 than smaller states. The collective net undercount for the smallest population quintile was 1.5 % but it was 6.1 % for the largest population quintile.

The five states with the largest net undercounts for children age 0–4, (California, Texas, Florida, New York, and Georgia) had a collective undercount of 579,000 which amounts to 55 % of the total net undercount nationwide for this age group. But only 37 % of the national population age 0–4 live in those five states.

The correlation between the undercount rate for age 0–4 and state population size in the state is −0.54, which underscores the fact that larger states tend to have bigger net undercount rates. Recall undercounts are expressed as a negative number.

The correlation between state population size and net undercount rates for the population age 0–4 is likely related to some of the characteristics of the states that are related to undercounts rather than population size per se. This idea is explored next.

Table 5.4 State differences between 2010 decennial census counts and vintage 2010 population estimates for population age 0–4 by state population quintiles

	Numeric difference between population estimates and census counts (census − estimates)	Percent difference between population estimates and census counts [(census − estimates/ estimates) * 100]
Smallest quintile	−8538	−1.5
Second smallest quintile	−60,414	−3.6
Middle quintile	−88,402	−3.5
Second largest quintile	−193,210	−4
Largest quintile	−705,111	−6.1
Total	−1,055,675	−5

Quintiles are based on total states census counts

5.4.2 Race and Hispanic Origin

Nationally, in 2010, the net undercount rate for Hispanics age 0–4 was 7.5 % and for Blacks Alone or in Combination age 0–4 it was 6.3 % (see Chap. 3). Therefore, one might expect to find that the racial and ethnic composition in a state is related to the net undercount rate for young children. I use the Census counts to measure the distribution of minority groups because the groups are quite small in some states and I feel the counts for Hispanics and Black Alone or in Combination are likely to be more reliable and accurate than Population Estimates for small populations in those states.

Table 5.5 shows correlations between four measures of racial composition and state net undercount rates for the population age 0–4. There is little difference between the correlations based on race/Hispanic origin status of the adult population or the population age 0–4, so only the data for the adult population are shown in Table 5.5.

All of the correlations in Table 5.5 are in the predicted direction, namely, the higher the percentage of minorities the higher the net undercount rate of the population age 0–4. But the magnitudes of the associations vary substantially. All of the correlations in Table 5.5 are statistically significant.

The correlation coefficient between percent Hispanic and the net undercount rates for the population age 0–4 is −0.67, while the correlation between percent Black Alone or in Combination and net undercount rates is −0.35 for the population age 0–4. The higher correlation for Hispanics than for Blacks Alone or in Combination may be due to the fact that Hispanics are a larger population than Black Alone or in Combination and the net undercount rate of young Hispanic children is somewhat higher than that of Black Alone or in Combination. Therefore the impact of Hispanics on a state net undercount rate is likely to be higher than the impact of Blacks Alone or in Combination.

However, when the Black Alone or in Combination population is combined with the Hispanic population to form a broader measure of minority populations,

Table 5.5 Correlations between racial/hispanic composition and net undercount rates for age 0–4 across states

Percent black alone not hispanic age 18+	−0.35
Percent hispanic age 18+	−0.67
Percent black alone not his apnic plus hispanic age 18+	−0.76
Percent minority* age 18+	−0.68

All correlations are statistically significantly different from zero a the 0.05 level
Source Demographic Data from 2010 Census, PL94-171 file
*Minority is defined here as anyone other than Non-Hispanic White Alone

the correlation is higher than either group by itself. For the population age 0–4 the correlation between percent of the population that is Black Alone or in Combination or Hispanic and net undercount is −0.76. I suspect the higher correlation for the combined population of Black Alone or in Combination and Hispanic group 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. So the combined group covers the largest minority populations in more states.

It is worth noting that the correlation between net Census coverage for age 0–4 and the percent of the state population that is any racial/Hispanic minority (i.e. anything other than Non-Hispanic White Alone) is not as high as the correlation using the combination of Hispanic and Black Alone or in Combination. The correlation between Percent Total Minority and net coverage rate of young children is −0.68. This may be due to the concentration of Asians and American Indians/Alaskan Native in a few states such as Hawaii and Alaska where net undercount rates for the population age 0–4 are low relative to other states.

Many observers feel that the racial differences noted above are the product of differences in factors such as housing and living arrangements rather than race or ethnicity per se. In discussing the undercount of minorities, Schwede et al. (2015, p. 293) state, "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)." This idea is examined below.

5.4.3 Hard-to-Count Characteristics

Since the 1990s, there has been a sustained effort at the Census Bureau to build a Planning Data Base at the Census tract level which includes information on Hard-to-Count characteristics (Bruce and Robinson 2003, 2007; Bruce et al. 2001; Bruce et al. 2012). Twelve Hard-to-Count factors were used to construct a Hard-to-Count score for each Census tract in the 2000 Census. The twelve

characteristics used to calculate a Hard-to-Count scores (Bruce and Robinson 2003) are linked to low mail response rates and the likelihood of being missed in the Census. According to Bruce and Robinson (2003, p. 74), "The variables included in the Planning Database (PDB) were guided by extensive research conducted by the Census Bureau and others to measure the undercount and to identify reasons people are missed…".

In 2014, the Census Bureau released the latest version of the Planning Data Base with data reflecting many of the Hard-to-Count factors (U.S. Census Bureau 2014). Some of the measures are related to housing characteristics and some are related to characteristics of people in households. The variables in the 2014 Planning Data Base also include some of those derived using empirical relationships with the Mail Return Rate (Erdman and Bates 2014).

The correlations between the net undercount rate of the population 0–4 and the twelve Hard-to-Count characteristics are shown in Table 5.6. The measures are listed in order from the most highly correlated to the least highly correlated. Recall that the net undercount rate as measured here is reflected as a negative number so a lower figure reflects a larger net undercount.

Ten of the twelve correlation coefficients in Table 5.6 are in the predicted direction but there are large variations in the size of correlations between the Hard-to-Count measures and the net undercount rates of the population age 0–4.

In general, measures that reflect characteristics of people within a household had higher correlations with the net undercount rate of the population age 0–4 than characteristics of the housing units.

The three measures that are most highly correlated with net undercount among age 0–4 (Linguistic Isolation, Lack of a High School Degree, and Unemployment) are related to characteristics of adults in the household. Since adults in a household typically complete the Census questionnaire, it is not surprising that states where adults are more likely to have problems with filling out for the Census questionnaire have higher net undercount rates for young children.

Percent in Rental Housing and Percent of Housing Units That Are Not Single Detached Units are the housing measures that are most highly correlated (−0.42) with net undercount of the population age 0–4. Percent in Crowded Households also has a moderate correlation (−0.37) with net undercount of 0–4 year olds. Many of these measures are highly correlated with each other so it is difficult to sort out causality.

Four of the twelve correlations are not statistically significantly different from zero. Two measures (Percent of Households Receiving Public Assistance and Percent Vacant Housing Units) actually have a positive correlation with undercount rates of young children at the state level but neither of these correlations is statistically significant. The correlation between net undercount rate for the population age 0–4 and the Percent of the Population That Moved in the Past Year as well as the Percent with No Phone in the Household are in the expected direction, but are not statistically significant.

Some of the weak correlations observed in Table 5.6 may be explained by changes in society since these measures were first identified in the 1990s.

Table 5.6 Correlations between state net undercount rates for population age 0–4 and hard-to-count factors across

	Correlation with state net undercount rate for age 0–4
Percent of households that are linguistically isolated[b]	−0.69[a]
Percent of population age 25+ not high school graduates	−0.58[a]
Percent of people unemployed	−0.57[a]
Percent of occupied housing units that are rental occupied	−0.42[a]
Percent of housing other than single detached units	−0.42[a]
Percent of occupied housing units that are crowded (1.51+ persons per room)	−0.37[a]
Percent of people living below poverty	−0.31[a]
Percent of households that are not husband wife households	−0.27[a]
Percent of occupied housing unit with no phone service	−0.24
Percent of persons age 1+ who moved into unit in last year	−0.18
Percent of households with public assistance income	0.17
Percent vacant housing units	0.15

Source All data are from the 2010 census or the 2010 1-year American Community Survey Estimates

[a]Statistically different from zero at the 0.05 level of significance or higher

[b]Linguistically isolation is measure of households that lack an adult who speaks English well

For example, federal welfare reform that was passed in 1996 changed the major federal program providing cash public assistance as Assistance to Families with Dependent Children was replaced with Temporary Assistance to Needy Families. The changes brought about by the new welfare program (a large decline in the number of families receiving cash public assistance) may mean this measure is no longer a good predictor of Census undercounts. Regarding the lack of a statistically significant correlation between Vacant Housing Units and net undercount rates for age 0–4, it is also important to remember that the 2010 Census took place in the midst of a recession and a "housing crisis." The high level of vacancies which accompanied the housing crisis may have undermined the historic connection between the vacancy rate and Census coverage.

The correlation between the Availability of Phone Service and net undercount rate for population age 0–4 is relatively low at −0.24 and it is not statistically significant. The proliferation of cell phones may have changed the meaning of having a phone at home.

Some of the factors that were not statistically significant at the state level are highly correlated with Mail Return Rates at the Block Group and Tract level. Erdman and Bates (2014) found Percent Moved 2005–2009 as well as Percent in Different House One Year Ago and Percent Vacant units to be important predictors of the Mail Return Rates in analysis related to the 2010 Census.

One other factor that was examined here in addition to traditional Hard-to-Count variables was the Mail Return Rate. The Mail Return Rate is defined by the Census Bureau (2014, p. 36) as:

> The number of mail returns received out of the total number of valid occupied housing units (HUs) in the Mailout/Mailback universe which excludes deleted, vacant, or units identified as undeliverable as addressed.

The correlation between final Mail Return Rates and the net undercount rate for age 0–4 at the state level is +0.50. Recall that net undercount rates are negative numbers so this correlation coefficient indicates the higher the Mail Return Rate the smaller the net undercount for age 0–4. The correlation between the Mail Return Rate and the net undercount rates for age 0–4 is on the same order of magnitude as several of the Hard-to-Count characteristics.

5.5 County Level Undercounts of Young Children

This section examines the net undercount of young children by comparing the Census Bureau's Vintage 2010 Population Estimates for the population age 0–4 to the 2010 Census counts across counties. This analysis focuses on the population age 0–4, because the 2010 DA analysis shows this age group has the largest net undercount of any age group (see Chap. 3).

There are more than 3100 counties in the U.S. but many of them have small populations. In the 2010 Census there were 566 counties with fewer than 500 persons age 0–4 and 1129 with less than one thousand in this age range. Yowell and Devine (2013, Table 7) show the Mean Absolute Percent Error for Population Estimates of the smallest counties is about four times that of the largest counties. Differences between the 2010 Census counts and the Vintage 2010 population estimate for many small counties are fraught with estimation error. Consequently, data for individual small counties are not examined here.

The analysis focuses on groups of counties which is consistent with the advice of Adlakha et al. (2003, p. 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." When counties are grouped together some of the random errors in the estimates for individual counties will cancel each other out. Given the more accurate Population Estimates for large counties (Yowell and Devine 2013), separate analysis is conducted for a subset of large counties.

5.6 The Data

Methods and data used to examine counties are similar to those discussed in the previous section. 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–July 1, 2010." The file is also denoted as "CC-EST2010-ALLDATA." The file was released March 2012 and it is available on the Census Bureau's website.

This file contains yearly estimates for 2000 through 2010, but only the estimates from April 1, 2010, for the population age 0–4 are used in this study. These estimates include the results of special Censuses and local challenges during the previous decade.

The data from the 2010 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 total population and for the population age 0–4 were taken from this file.

The 2010 Census results are compared to Vintage 2010 Population Estimates in the 3141 counties or county equivalents (i.e. parishes or independent cities) for which Vintage 2010 Population Estimates were produced. The District of Columbia is treated as a county in this analysis. A few counties are not included in the analysis because they are too small to provide reliable data. Coverage was measured as the Census minus Population Estimates so a negative number means the Census count was less than the Population Estimate. Percentages are derived by dividing the difference by the Population Estimate.

5.7 County-Level Results

Table 5.7 provides several measures of differences between the Vintage 2010 Population Estimates and the 2010 Census counts for counties for the population age 0–4. Across all counties, the mean numerical difference between Census Counts

Table 5.7 County differences between Vintage 2010 population estimates and 2010 census count for population age 0–4

Mean numerical difference	338
Mean percent difference	1.1
Number of counties with a net overcount	1491
Number of counties with a net undercount	1634
Mean absolute numeric difference	1993
Number of counties with absolute numeric difference larger than 5000	57
Mean absolute percentage difference	7.4
Number of counties with absolute percent difference larger than 10 %	674

and the Vintage 2010 Population Estimates for the population age 0–4 was 338. The average county had an overcount of 1.1 % for the population age 0–4. Since this average county overcount is quite different than the national undercount rate (5.0 % based on Vintage 2010 Population Estimates and 4.6 % based on DA) it indicates that the national rate is driven by high net undercount rates in large counties.

Unlike states where almost all of the states (46 of the 50) experienced a net undercount, counties have a more balanced distribution. There were 1634 counties with a net undercount of the population age 0–4 and 1491 counties with a net overcount of the population age 0–4. For sixteen counties the population estimate and the Census count were exactly the same for the population age 0–4.

Because errors in different directions cancel each other out in calculating the mean it is important to look at absolute differences as well. In absolute terms, the mean difference between the county population estimate for age 0–4 and the Census count for that age group was 1993. In percentage terms the average absolute difference was 7.4 %.

5.7.1 Characteristics Associated with County Net Undercount Rates for Population Age 0–4

There is a clear relationship between county size and collective undercount rates with larger counties having the highest undercount rates and smaller counties having net overcounts. Table 5.8 shows the mean percent difference for the smallest counties (less than 5000 people) is a 5.1 % net overcount compared to a 7.8 % net undercount for the largest counties (those of 500,000 or more people).

The correlation between net undercount rate for the population age 0–4 and size of county population (total population) is modest (-0.28). I suspect the correlation coefficient in confounded by the relatively large errors associated with Population Estimates for smaller counties.

Table 5.8 Difference between 2010 census counts and Vintage 2010 population estimates for population age 0–4 by county size

Total population size of county	Number of counties	Aggregate population estimate (in 1000 s)	Aggregate census count (In 1000 s)	Census − estimate (in 1000 s)	Percent difference[a]
Less than 5000	303	48	50	2	5.1
5000–9999	395	181	183	2	1.3
10,000–24,999	845	886	882	−4	−0.5
25,000–49,999	624	1438	1421	−18	−1.2
50,000–99,999	398	1776	1757	−20	−1.1
100,000–499,999	450	6381	6178	−203	−3.2
500,000+	128	10,553	9729	−823	−7.8
Total	3143	21,263	20,201	−1062	−5

[a]Percent calculated on unrounded numbers

Table 5.9 Net undercount of young children in the ten largest counties in 2010

Rank based on total population	County (major city)		Population estimate	Census	Difference (census − estimate)	Percent difference[a]
1	Los Angeles (Los Angeles)	California	719,712	645,793	−73,919	−10.3
2	Cook (Chicago)	Illinois	385,195	342,493	−42,702	−11.1
3	Harris (Houston)	Texas	365,048	336,314	−28,734	−7.9
4	Maricopa (Phoenix)	Arizona	323,013	282,770	−40,243	−12.5
5	San Diego (San Diego)	California	226,006	203,423	−22,583	−10
6	Orange (Anaheim)	California	214,801	191,691	−23,110	−10.8
7	Kings (New York)	New York	199,891	177,198	−22,693	−11.4
8	Miami − Dade (Miami)	Florida	170,662	149,937	−20,725	−12.1
9	Dallas (Dallas)	Texas	223,980	192,838	−31,142	−13.9
10	Queens (New York)	New York	152,336	132,464	−19,872	−13

[a]Percent calculated on unrounded data

Table 5.8 indicates the 128 counties with half a million or more people had a cumulative net undercount of 823,000 persons and a net undercount rate of 7.8 % for the population age 0–4. Thus these 128 counties account for 77 % of the total national net undercount of slightly over one million people age 0–4, even though only 50 % of the national population age 0–4 live in these counties.

Table 5.9 shows the net undercount rates in the ten largest counties in the nation. Nine of the ten largest counties had a net undercount rate for young children of at least 10 %. Harris County, Texas, is the exception with a net undercount rate of 7.9 %. Undercount estimates for individual counties should be viewed cautiously, but the consistently high net undercount rate for all ten large counties in Table 5.9, plus the evidence in Table 5.8, strongly suggest high net undercount rates for the population age 0–4 for the largest counties in the country.

5.7.2 Race and Hispanic Origin DA

This section looks at the relationship between county racial/Hispanic composition and net undercount of the population age 0–4. The first analysis looks at all counties then the analysis is repeated with only the largest counties (population 250,000 or more) where Population Estimates are likely to be more accurate.

Table 5.10 Correlations between racial/ethnic composition of a county and net undercount rates for age 0–4

	All counties	Counties with populations of 250,000 or more
Percent non-hispanic black alone	−0.21	−0.35
Percent hispanic	−0.12	−0.4
Percent non-hispanic black alone + hispanic	−0.25	−0.59
Percent minority (other than non-hispanic white)	−0.25	−0.59

All Correlations in this table are significant at the 0.01 level (2-tailed)

I use the 2010 Census figures for Black Alone, Black Alone or in Combination, total minority population (i.e. anyone who is not Non-Hispanic White) and Hispanics. The data for these populations are taken from the 2010 Census using American Factfinder.

Table 5.10 shows the correlations between four measures of racial/Hispanic composition in a county and the net undercount rate for the population age 0–4. The racial composition is based on the adult population (age 18+) because adults are usually responsible for filling out the Census questionnaire.

For all counties, all the correlations are in the expected direction (negative correlations for every minority group measured) and all of the correlations are statistically significant, but the correlations are relatively modest in size and range from −0.12 to −0.25. The correlation between net undercount rates and the percent Non-Hispanic Black Alone (−0.21) is higher than the correlation between the net undercount rate and Hispanics (−0.12) which is the opposite order of what was found at the state level. The correlations may be confounded by high estimation errors for many individual counties and a very small minority population in many counties.

For the largest counties (those of 250,000 or more people) the correlations are in the predicted direction and statistically significant but higher in magnitude. As with states, there is a higher correlation when Blacks (Alone or in Combination) and Hispanics are combined (−0.59) into one measure of minority population than for either Blacks (−0.35 for Non-Hispanic Black alone) or Hispanics (−0.40).

5.8 Summary

Forty-six of the fifty states experienced a net undercount for the population age 0–4 and 12 states experienced net undercount rates of 5 % of more. At the state level, the net coverage rates for the population age 0–4 in the 2010 U.S. Census varies from a 10.2 % net undercount in Arizona to 2.1 % overcount in North Dakota.

In general, larger states (in population size) had higher net undercount rates than smaller states. The net undercount rates in states are correlated with the size

of the Black and/or Hispanic population, although the correlation is much higher for Hispanics than for Blacks.

The relationships between traditional Hard-to-Count characteristics and net undercount rates for the population age 0–4 at the state level vary. Looking across states, the characteristics that are most highly correlated with net undercount rates for age 0–4 are personal characteristics (Linguistic Isolation, Lack of a High School Degree, and Unemployment Rate) rather than housing characteristics.

The data examined here indicate that the national net undercount rate for the population age 0–4 varies substantially across counties. About half of the counties experienced a net undercount and half of the counties experienced a net overcount. Larger counties account for the vast majority of the national net undercount for the population age 0–4. In the 128 largest counties based on total 2010 Census population, there was a net undercount of 823,000 persons age 0–4 which amounts to 77 % of the national undercount of persons age 0–4 even though only 50 % of young children living in these counties. All of the ten largest counties have net undercount rates of 7.9 % or higher for the population age 0–4.

References

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.

Bruce, A., Robinson. J. G., & Sanders, M. V. (2001). Hart-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, 2001.

Bruce, A., & Robinson, J. G. (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 Oct 24.

Bruce, A., & Robinson, J. G. (2007). *Tract-Level planning database with census 2000 census data*. U.S. Census Bureau, Washington, DC.

Bruce, A., Robinson. J. G., & Devine J. E. (2012). A planning database to identify areas that are hard-to-enumerate and hard to survey in the United States. In *Proceedings of the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations*, American Statistical Association.

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.

Darga, K. (1999). *Sampling and the census*. Washington DC: American Enterprise Institute Press.

Davis, S. T. (1994). Evaluation of post census county estimates for the 1980s. *Current Population reports*, (Series P-25, No. 963, U.S. Census Bureau, Washington DC).

Erdman, C., & Bates, N. (2014). *The census bureau mail return rate challenge: Crowdsourcing to development hard-to-count scores*. Washington, DC: U.S Census Bureau.

Felton, G. T. (1986). Evaluation of population estimation procedures for counties: 1980. *Current Population Reports*, (P-25, NO. 984, U.S. Census Bureau, Washington, DC).

Gormley, W. T. (2012). *Voices for children: Rhetoric and public policy*. Washington DC: Brookings Institution Press.

Hill, R. B., & Steffes, R. B. (1973). *Estimating the 1970 census undercount for state and local areas*. Washington, DC: National Urban League.

Isaki, C. T., Schultz, L. K., Smith, P. J., & Diffendal, G. (1985). *Small Area Estimation Research for Census Undercount—Progress Repor*t. Paper presented at the International Symposium on Small Area Statistics, U.S. Bureau of the Census, Ottawa, Canada, May.

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.

O'Hare, W. P. (1988). How to evaluate population estimates. *American Demographics*, 10(1).

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 Annual Conference of the Population Association of America, New Orleans, LA.

O'Hare, W. P. (2014a). *Estimating the net undercount of young children in the 2010 U.S. census at the county level*. Poster presented at the 2014 Population Association of America Conference, Boston, MA. April.

O'Hare, W. P. (2014b). State-level 2010 census coverage rates for young children. *Population Research and Policy Review, 33*(6), 797–816.

O'Hare, W. P. (2014c). Assessing net coverage for young children in the 2010 U.S. decennial census. *International Journal of Population Research*, 2014. http://dx.doi.org/10.1155/2014/671715.

Robinson, J. G., Bashir, A., & Fernandez, E. W. (1993). Demographic analysis as an expanded program for early coverage evaluation of the 2000 census. 1993 Annual Research Conference, March 21–24, Arlington VA.

Robinson, J. G., & Kobilaric, E. L. (1995). Identifying differential undercounts at local geographic levels: A targeting database approach. Paper presented at the Annual Meeting of the Population Association of America, San Francisco.

Schwede, L., Terry, R., & Hunter, J. (2015). Ethnographic evaluations on coverage of hard-to-count minority in the US decennial censuses. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hart-to-survey populations* (pp. 293–315). Cambridge, England: Cambridge University Press.

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), December, 1977.

U.S. Census Bureau. (2012a). *Documentation for the revised 2010 demographic analysis middle series estimates*. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. (2012b). *Population estimation methodology*. U.S. Census Bureau, Washington, DC.

U.S. Census Bureau. (2014). *Planning database with 2010 census and 2008-2012 American community survey data: At the tract level*. U.S. Census Bureau, Washington, DC.

Velkoff, V. (2011). *Demographic evaluation of the 2010 census.* Paper presented at the 2011 Population Association of America annual Conference, Washington, DC.

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, 2013.