

## Chapter 2

# Methodology Used to Measure Census Coverage

**Abstract** The two primary methods used to assess the accuracy of the U.S. Census (Demographic Analysis and Dual Systems Estimates) are introduced. A short history of Demographic Analysis (DA) in assessing the U.S. Census is presented. The methodologies for DA and Dual Systems Estimates are provided along with the potential errors and limitations in the DA method. The reasons why DA is the preferred method for assessing census coverage for young children are presented.

**Keywords** Demographic analysis • Dual systems estimates • Post enumeration survey • Census coverage measurement

How do we know who is missed in a Census? Several methods have been used over time and in various countries to answer this question but in the U.S. only the Demographic Analysis method and the Dual Systems Estimates method (sometimes called Post Enumeration Survey) provide quantitative answers to the question posed above (Mulry 2014; Hogan et al. 2013; Bryan 2004; Anderson 2004).

Demographic Analysis or (DA) has been used since the 1950s to provide estimates of net undercounts in the U.S. Census. This method creates a separate independent estimate of the expected population based largely on births and deaths which is compared to the Census counts.

The Dual System Estimates (DSE) method compares Census results to the results of a Post-Enumeration Survey to determine the number and characteristics of people who are omitted or included erroneously (mostly those double-counted).

Nomenclature can be confusing in this arena. The terms Dual Systems Estimates (DSE) and Post-Enumeration Survey (PES) are often used interchangeably. Sometimes the DSE or PES approach is simply called the “survey method.” Moreover, the DSE/PES approach has been given a different name in each of the past three U.S. Censuses. In 2010 it was called Census Coverage Measurement (CCM), in the 2000 Census it was called Accuracy and Coverage Evaluation (A.C.E.) and in the 1990 Census it was called the Post Enumeration Survey (PES).

The analysis presented in this book rests largely on the results of the Census Bureau's Demographic Analysis (DA) method for assessing Census accuracy. I am convinced that DA is a better method for assessing the net undercount of young children because DA rests on highly accurate birth and death records and the least accurate component of DA, net international migration, is a very small component of DA Population Estimates for young children. The simplicity of the DA methodology relative to the DSE methodology can also be seen as an advantage. In addition, the DA data are advantageous because the data are available by single-year of age and consistent DA data are available from 1950 to 2010. The reasons for focusing on the results of Demographic Analysis rather than the U.S. Census Bureau's Dual System Estimates results are explored in more detail later in this Chapter.

## 2.1 Demographic Analysis History

The DA method has been used to assess the accuracy of Census figures for more than a half century and its origins are often traced back to an article by Price (1947). The unexpectedly high number of young men who turned up at the first compulsory selective service registration in October 1940, alerted scholars to the possibility of under-enumeration in the 1940 Decennial Census. The selective service data also provided an independent population estimate for assessing the size of such under-enumeration in the Decennial Census.

In one of the first systematic efforts to use DA to examine U.S. Census results, Coale (1955) found children age 0–4 had a relatively high net undercount rate in the Censuses of 1940 and 1950. Coale (1955, p. 35) used a variant of the Demographic Analysis technique to estimate net undercount rates for several population subgroups age 0–4 in 1950. In 1950, the estimated net undercount from Coale's analysis for age 0–4 ranged from a low of 3.8 % for White females to a high of 11 % for Non-White males. All of the estimates for the net undercount of age 0–4 were higher than the corresponding net undercount estimates for the total population.

Siegel and Zelnik (1966) found a substantial net undercount of children age 0–4 in the 1950 and 1960 Censuses. For the 1960 Census, their preferred composite estimate based on demographic analysis, indicated a net undercount of 2.0 % for White males age 0–4 and 1.2 % of White females age 0–4, 8.4 % for Non-White males and 6.8 % for Non-White females.

Coale and Zelnick (1963) found high net undercount rates for young children in the U.S. Census as far back as 1880, a finding supported by Hacker (2013) who shows that native-born White children age 0–4 had higher than average net undercount rates in each U.S. Census from 1850 to 1930. Over the 1850–1930 period, Hacker estimates the net undercount for native born White males age 0–4 varied from a low of 4.0 % in 1890 to a high of 15.2 % in 1850 and for native born White females the net undercount rates vary from a low of 4.1 % in 1890 to a high of 15.4 % in 1850. In every Census between 1850 and 1930, except 1890, the net undercount of native born Whites age 0–4 was higher than the overall average.

Coale and Rives (1973) also found very high net undercount rates for young Black children in every U.S. Census from 1880 to 1970. Estimates for the Black male population age 0–4 range from 28.5 % in 1890 to 7.4 % in 1960.

In addition to the demographic data presented above, genealogical research also shows a pattern of underreporting young children as far back as the 1850s (Adams and Kasakoff 1991).

## 2.2 Demographic Analysis Method

Since there are already several detailed descriptions of the DA methodology available, I will only review the method briefly here (Robinson 2010; Himes and Clogg 1992; U.S. Census Bureau 2010a).

DA is an example of the cohort-component method of population estimation meaning each component of population change (births, deaths and migration) is estimated for each birth cohort. The cohort-component method of Population Estimates is one of the most widely used techniques in population estimation (Bryan 2004).

The DA method employed for the 2010 Census used one technique to estimate the population under age 75 and another method to estimate the population age 75 and older (West 2012). Since this study focuses on children, only the method used for people age 0–74 is discussed here (people under age 1 are classified as age 0).

The 2010 DA estimates for the population age 0–74 are based on the compilation of historical estimates of the components of population change: Births (B), Deaths (D), and Net International Migration (NIM). The data and methodology for each of these components is described in separate background documents prepared for the development and release of the Census Bureau’s 2010 DA estimates (Robinson 2010; Devine et al. 2010; Bhaskar et al. 2010).

As described by the Census Bureau (2010a) the DA Population Estimates for age 0–74 are derived from the basic demographic accounting Eq. (2.1) applied to each birth cohort:

$$P_{0-74} = B - D + NIM \quad (2.1)$$

$P_{0-74}$	Population for each single year of age from 0 to 74
B	Number of births for each age cohort
D	Number of deaths for each age cohort since birth
NIM	Net International Migration for each age cohort

For example, the estimate for the population age 17 on the April 1, 2010 Census date is based on births from April 1992 through March 1993, reduced by the deaths to that birth cohort in each year between 1992 and 2010, and incremented by Net International Migration (NIM) experienced by the cohort.

Births, deaths and Net International Migration detailed figures are not available for single year of age in the DA estimates released in May 2012 which is the

**Table 2.1** Fundamental data for census Bureau’s DA estimate for the population age 0–4

Births (in 5 years prior to the 2010 Census)	21,120,000
Deaths to those born in 5 years prior to Census	154,000
Net international migration	240,000
DA population estimate for age 0–4	21,206,000
Population age 0–4 counted in 2010 Census	20,201,000

Source U.S. Census Bureau (2010b)

primary source of DA data used here. But the December 2010 DA “Middle Series” estimate for the population age 0 to 4 is comprised of 21,120,000 births, 154,000 deaths, and Net International Migration of 240,000 (see Table 2.1). Births are by far the largest component of the DA Population Estimates for young children. In 2010, births accounted for 99.6 % of the DA population estimate for the population age 0–4 (U.S. Census Bureau 2010b).

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. The National Center for Health Statistics (2014, p. 2) states, “A chief advantage of birth certificate data is that information is collected for essentially every birth occurring in the country each year...” After a thorough review of vital statistics prior to the 2010 Census, the Census Bureau (Devine et al. 2010, p. 5) stated:

- The following assumptions are made regarding the use of vital statistics for DA:
- Birth registration has been 100 % complete since 1985.
  - Infant deaths were underregistered at one-half the rate of the underregistration of births up to and including 1959.
  - The registration of deaths for ages 1 and over has been 100 % complete for the entire DA time series starting in 1935.

Although some of the characteristics gathered on birth certificates may be suspect, the number of births and deaths is widely seen as virtually complete.

In addition to regularly published totals, the Census Bureau receives microdata files from NCHS containing detailed monthly data on each birth and death. These files were used for DA estimates by race. Construction of DA estimates by race is discussed later in this Chapter.

The Census Bureau changed the way it calculated Net International Migration for the 2010 set of DA estimates (Bhaskar et al. 2010). The current method relies heavily on data from the Census Bureau’s American Community Survey (ACS) where the location of the Residence One Year Ago (ROYA) is ascertained for everyone in the survey age 1 or older. The total number of yearly immigrants is derived from this question in each year of the ACS, and then that total number of immigrants is distributed to demographic cells (sex, age and race) based on an accumulation of the same data over the last 5 years of the ACS. Five years of ACS data are used to provide more stable and reliable estimates for small demographic groups. On the other hand, we should note that the five-year average may mask changes over time. Given changing economic conditions, it would not be surprising if the immigration pattern in the 2008–2010 period differed from the pattern

before 2008, however, I suspect such errors would be small, especially for those age 0–4. NIM is available by single year of age for Blacks (Black Alone and Black Alone or in Combination) under age 30 and for Hispanics under age 20.

Statistics on emigration of the foreign-born population from the U.S. are based on a residual method comparing data on the foreign-born population from the 2000 Census to later ACS estimates to develop rates and then applying those rates to observed populations (Demographic Analysis Research Team 2010).

Emigration of U.S. citizens (net native migration) is derived by examining Census data from several other countries (Schachter 2008). This method of estimating out migration is problematic for a couple of reasons. Data are not available for every country and the quality of some foreign Censuses is suspect. However, with few exceptions (see Pitkin and Park 2005) it is widely felt that such emigration has little impact on DA Estimates for young children.

The DA estimates released in May 2012 assume a Net International Migration of only 244,000 out of a population of 21,172,000 for age 0–4 (the 244,000 figure was obtained from Census Bureau staff). Thus Net International Migration accounts for only 1.1 % of the DA estimate for the population age 0–4. Since Net International Migration accounts for such a small part of the DA estimate for the population age 0–4, errors in this component of population change would not have a big impact on the final DA population estimate for the 0–4 age group. In addition, potential errors in the overall estimates of the DA estimates for the population age 0–4 are likely to be small, as discussed below.

In preparing for the December 2010 DA release, the Census Bureau developed five estimation series with differing assumptions to reflect the degree of uncertainty in the estimates. For the population age 0–17, the estimates from the five series presented in December 2010 range from 75,042,000 to 76,222,000 and for the population age 0–4 the estimates ranged from 21,181,000 to 21,265,000. This is a relatively small band of uncertainty compared to the estimated net undercount.

The assumptions about births and deaths were the same for each of the five series. Only the assumptions about Net International Migration varied. In those five series the Net International Migration assumptions for the population age 0 to 4 ranged from 214,000 to 297,000 (U.S. Census Bureau 2010b). The Middle Series estimate of net immigration for age 0–4 was 240,000 for the DA estimates released in December 2010. Thus the high end of the immigration assumption was 57,000 persons higher than the Middle Series and the low end was 26,000 persons lower than the Middle Series.

This provides some guidance about the size of potential errors in immigration estimates and population estimates used in DA for young children. If the Net International Migration component for children age 0–4 in the DA estimate from May 2012 had been 26,000 less, the net undercount of children age 0–4 in the 2010 Census would be 4.5 % instead of the value of 4.6 % reported in the May 2012 DA release. If the Net International Migration component for children age 0–4 had been 57,000 higher the net undercount estimates would have been 4.9 %. In either case, the net undercount of young children would remain much higher than any other age group. If one wanted to look at an extreme case and assume

there was no net immigration of children age 0–4, the DA estimate for the net undercount of the population age 0–4 would 3.6 %, which is still much higher than for any other age group.

For older children, Net International Migration plays a bigger role. For the population age 14–17 the May 2012 DA shows a net overcount of 1.4 %. For the population age 14–17, the Net International Migration assumptions for the five DA series released in December 2010, range from 1.023 million for the low series to 1.424 million in the high series and compose 6.1 and 8.3 % of the DA estimate respectively. The Net International Migration assumption for the December 2010 DA Middle Series was 1.186 million. Thus the high end of the series was 238,000 persons higher than the Middle Series estimate and the low end of the series was 163,000 lower than the Middle Series. If the DA estimate for the population age 14–17 were 238,000 higher than the May 2010 Middle Series, it would result in an overcount estimate of essentially zero. If the DA estimate for the population age 14–17 were 163,000 lower than the May 2012 Middle Series, it would result in an overcount estimate of 2.4 %.

### 2.3 Limitations of the Demographic Analysis Method

There are four major limitations to DA. First, it is only routinely available for the nation as a whole. The population age 0–9 is an exception to this rule. Sub-national analysis can be done for the population under age ten, because the Census Bureau's Population Estimates for age 0–9 are not linked to the previous Census. This issue is explored in Chap. 5.

Second, DA estimates are only available for a few race/ethnic groups. Historically the estimates have only been available for Black and Non-Black groups. This restriction is due to the lack of race specificity and consistency for data collected on the birth and death certificates historically. The only group that has been identified relatively consistently over time is Blacks (African-Americans).

The 2010 DA estimates include data for Hispanics for the first time, but only for the population under age 20. Hispanics under age 20 were included in the DA estimates in 2010 because Hispanics have been consistently identified in birth and death certificates since 1990.

The 2010 DA is the first to produce estimates of net undercount of Black Alone and Black Alone or in Combination. Recent changes in how the Census Bureau collects data on race raises questions about the comparability of the data for Blacks in the 2010 Census relative to earlier Censuses. This issue is explored in Chap. 4.

The third limitation of the DA estimates is that they only provide net undercount/overcount figures. A zero net undercount could be the result of no one being missed (omissions) or double counted (erroneous enumerations) or it could be the result of 10 % of the population being missed and 10 % double counted.

The fourth limitation of the DA methodology is the lack of any measures of uncertainty for the estimates similar to standard errors associated with surveys.

However, as mentioned earlier in this Chapter, in the December 2010 DA release, the Census Bureau released five different estimate series based on five sets of assumptions about births, deaths, and Net International Migration to reflect some of the uncertainty regarding the DA estimates.

Despite these limitations, DA has been used for many decades, the underlying data and methodology are strong, and it has provided useful information for those trying to understand the strengths and weaknesses of the U.S. Decennial Census. According to Robinson (2000, p. 1) “The national DA estimates have become the accepted benchmark for tracking historical trends in net Census undercounts and for assessing coverage differences by age, sex, and race (Black, all other).”

As stated earlier, DA is particularly useful for assessing the accuracy of the Census count of young children for two reasons. One of the major uncertainties in using DA to assess the accuracy of total population counts is the assumptions about Net International Migration that must be made. For most age groups other than young children, Net International Migration is subject to more error because of the greater uncertainty of some specific elements such as undocumented immigrants and uncertainty in the estimation of emigrants (Jensen 2012). According to Bhaskar and colleagues (2010, p. 1), “The largest uncertainty in the Demographic Analysis (DA) estimates comes from the international migration component.” For young children, net international immigration is a very small factor, so any errors in the net immigration estimate will have little impact on the DA estimate for this age group.

The second reason DA is the preferred method for assessing the net undercount of young children is improved quality of vital events data. For people born in the United States in the past couple of decade’s vital event data are deemed to be complete. In the five DA scenarios provided in the 2010 DA estimates released in December 2010, the birth and death assumptions are identical for people under age 18 in all five series, which reflects the high level of credibility given to the vital events data for children.

## 2.4 Dual Systems Estimates Methodology

The other major source of data on undercounts and overcounts in the U.S. Census is the Census Bureau’s Dual Systems Estimates (DSE) method. The DSE approach for 2010 is called Census Coverage Measurement. This is an oversimplification, but basically DSE compares results from a Post-Enumeration Survey (PES) to Census records to determine undercounts and overcounts (Mule 2010). The 2010 Census is the first one where DSE has produced data for the population age 0–4, so there is no historical data on young children from DSE. In the 2000 U.S. Decennial Census, DSE estimates were made for age 0–9 and age 10–17, and in the 1990 Census DSE estimates for children were only available for the entire group of children age 0–17.

Table 2.2 shows differences between net undercount estimates of DA and DSE in the 2010 Census for several age groups. For all adult age groups examined, the



**Table 2.2** Comparison of DA and DSE undercount estimates for several demographic groups: 2010

	DSE	DA	Difference (DA–DSE)
Age 0–4	0.7	4.6	–3.9
Age 5–9	–0.3	2.2	–2.5
Age 10–17	–1.0	–0.5	–0.5
Age 18–29 males	1.2	–0.4	1.6
Age 18–29 females	–0.3	–1.5	1.2
Age 30–49 males	3.6	2.3	1.3
Age 30–49 females	–0.4	–1.9	1.5
Age 50 + males	–0.3	–0.5	0.1
Age 50 + females	–2.4	–2.4	0.1

Source O’Hare et al. (2012)

differences are less than 2 % points. However for the population age 0 to 4, the difference is 3.9 % points.

As noted above, for many age groups the DA method and the DSE method produce similar results. However, in the context of comparing the results of DSE and DA in the 2000 Census, and noting the generally consistent results, the U.S. Census Bureau (2003, p. v) observed,

The primary exception to the consistency of results occurs for children aged 0-9. While the A.C.E. Revision II estimates a small net overcount for children 0-9 (the estimate was not statistically significantly different from zero), Demographic Analysis estimated a net undercount of 2.56 %. 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.

A National Research Council report (2004, p. 254) made the same observation about the inconsistency of DA and DSE estimates for young children and the authors note, “No explanation for this discrepancy has been advanced.”

Table 2.3 shows the results of DA and DSE estimates of Census coverage for children in the 1990, 2000 and 2010 Censuses. The data indicate significant inconsistencies between the results of the two methodologies for young children. In 2010, the DSE estimated a 0.7 % net undercount for age 0–4 compared to 4.6 % for DA. In population terms, the 2010 DA estimates a net undercount of 970,000 people age 0–4, while the DSE estimated a net undercount of only 152,000 people in this age group.

Table 2.3 shows that in 2000 and 2010, the DA and DSE coverage estimates for age 10–17, are relatively consistent, but estimates for age 0–9 are different.

**Table 2.3** Comparison of estimated net percent undercount from DA and DSE for population age 0–17 in 1990, 2000 and 2010

	1990		2000		2010	
	DA	DSE	DA	DSE	DA	DSE
Age 0–17	–1.8	–3.2	–0.7	0.8	1.7	0.3
Age 10–17			–1.8	–1.3	–0.5	–1.0
Age 0–9			–2.6	0.5	–3.4	–0.2
Age 0–4					–4.6	–0.7

Source O’Hare et al. (2012)



The 2000 DSE estimates for age 0–9 was +0.5 % compared to –2.6 % for DA. In 2010 the DSE coverage estimates for age 0–9 was –0.2 % compared to –3.4 % for DA.

It should be noted that in the 1990 Census the results of DSE showed a higher net undercount for all children age 0–17 than DA (3.2 % undercount from the DSE method compared to a 1.8 % undercount for the DA method). However, there was no disaggregation of the DSE data for children into smaller age groups. Given the very different net undercount rates for children in different age groups the implications of the 1990 data are not clear.

O'Hare and his colleagues (2012) provide detailed documentation of the inconsistency between DSE and DA estimates for young children and suggest that uncorrected correlation bias may result in an underestimation of the undercount for young children in the DSE methodology. The U.S Census Bureau (2012b p. 1) describes correlation bias as,

Correlation bias results from the failure of the general independence assumption underlying dual system estimation. This form of bias tends to lead to underestimation of dual system estimates if persons missed in the Census are more likely than those found in the Census to also have been missed in the Census Coverage Measurement survey.

The issue of correlation bias in the DSE approach has been discussed by other researchers (Wachter and Freedman 1999; Shores 2002; Shores and Sands 2003). The National Research Council (2009) created a panel to study the issue of correlation bias and coverage measurement in the 2010 Census, but did not seem to take up the issue of correlation bias for young children in their deliberations.

The existence of correlation bias in the DSE method is already recognized for the adult Black male population. Currently, adjustments in the DSE estimates for adult Black males are made to correct for correlation bias (U.S. Census Bureau 2012b). No similar adjustments are made for young children, in part, because there is not a widely accepted method for doing so.

Another issue with the DSE method is the matching that is required to link records from the Census to the records in the Post Enumeration Survey. To oversimplify the situation, to use the DSE method, one must make a decision about whether a person named Jon Smith in the PES is the same as the person named Johnathan Smith in the Census. Of course there are usually other clues like age, sex and address to use in matching. Matching procedures have improved over time but this is still an area where potential errors may occur. It would be useful to know if matching is more difficult for young children.

The DSE method also depends on respondent recall and that introduces another potential problem. The Post Enumeration Survey is usually conducted 4–6 months after the April 1, Census date. In discussing residential location at the time of the Census, Martin (2007, p. 429) notes, "Respondents interviewed months after April 1 may find it difficult to recall accurately when a move occurred." Recall may be potential problem for other data as well.

In the absence of any other reason for the large difference in net undercount estimates for young children between the DA method and the DSE method, uncorrected correlation bias in the DSE method is the leading explanation for

the observed differences. The Census Bureau Task Force on the Undercount of Young Children, (Griffin 2014, p. i) concluded, “The task force believes that Demographic Analysis (DA) provides the best measure of this undercount in the 2010 Census at 4.6 % nationally.”

The strength of the DA method for assessing net undercounts in young children is widely recognized. In comparing the DA results to DSE results in the 2000 Decennial Census, Zeller (2006, p. 320) 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. “Hogan and colleagues (2013, p. 98) also find, “Given the methodology that underlies DA, its estimates of younger populations tend to be quite accurate.” In comparing the results of the Dual Systems Estimates and DA from the 2000 Census, Shores and Sands (2003, p. 10) conclude, “Demographic Analysis has the advantage that its estimates are constructed from administrative data sources, some of which (e.g. birth and death registration data) are quite accurate.”

In the analysis shown in this publication, I rely almost exclusively on DA estimates. I believe the strengths of DA methodology make it a particularly good technique for estimating the number of young children. Moreover, in the decade prior to the 2010 Census, staff at the Census Bureau investigated a number of issues related to the production of DA estimates (Robinson 2010; Divine et al. 2010; Bhaskar et al. 2010). The increased input, review and examination enhance the likelihood that the 2010 DA estimates are accurate and credible.

In the remainder of this publication, the differences between the Census counts and DA estimates are shown as the Census count *minus* the DA estimate. This is consistent with the convention used by Velkoff (2011) in reporting the first results of the 2010 DA. 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. This may be a point of confusion because some studies have used a net undercount rate which subtracts the Census counts from the DA (or DSE) estimates. In that construction, a negative figure implies an overcount. I chose to use the net Census coverage error construction because I feel having an undercount reflected by a negative number is more intuitive. When figures are stated in the text as an undercount or an overcount, the positive and negative signs are not used.

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

## 2.5 Measuring the Net Undercount of Children by Race

Black is the only race group that has been coded relatively consistently in birth and death certificate data over time, so the only groups for which DA estimates could be produced were Black and Non-Black.

Key to the DA method for Blacks and Non-Blacks is making the vital events data and the Census data consistent. There have always been issues in trying to make information from these two data systems consistent, but the challenge of making accurate DA estimates for Blacks and Non-Blacks has increased in recent years since respondents have been allowed to select more than one race. In discussing the use of vital statistics for DA estimates by race the Census Bureau (Devine et al. 2010, p. 4) concludes, "...developing the estimates for DA race categories comes with a more complex, and substantial set of challenges." See Robinson (2010) for a good general discussion of issues associated with racial classifications in the Census and the vital events registers.

There are multiple problems in trying to make data collected in the Census racial categories comparable to the race data collected on birth and death certificates. For example, the "Some Other Race" category is a response category for the race question in the Census but not in birth or death certificates. Because the birth certificate data do not have a "Some Other Race" category, the Census Bureau constructs a set of modified race categories from the Census responses in which respondents in the Some Other Race category are distributed to Black and Non-Black categories. Thus for making comparisons between DA estimates and the Census counts for Blacks and Non-Blacks, one must use the 2010 U.S. Census modified race tabulations available on the Census Bureau's website. Correctly re-assigning people from the "Some Other Race" category to Black and Non-Black categories is a challenge and provides a potential source of errors.

Another issue is the fact that Census respondents in 2000 and 2010 could mark more than one race. In 1997, the U.S. Office of Management and Budget (1997) updated Statistical Policy Directive 15 requiring federal data collection efforts to allow respondents to mark more than one race. Prior to the 2000 Census, respondents were only allowed to mark one race in the U.S. Decennial Census, which meant the race data from the U.S. Census and from vital events were consistent in this regard.

Another issue is that birth certificate forms only record the race of the mother and father while the race of a child is asked directly in the Decennial Census. Thus, for birth certificate data, the race of the newborn must be inferred from the race of the parent(s). This is further complicated by a significant level of missing data. While data on the race of mother is relatively complete, many birth certificates are missing data on the race of the father. In 2009, 19 % of birth certificate forms did not contain the race of the father (Martin et al. 2011).

When both parents report the same race, that race is assigned to the child. When the two parents report different races on the birth certificate, the Census Bureau assigns newborns to one of thirty-one race categories based on the reported race of their mother and father and on empirical parent-child race relationships seen in the 2000 and 2010 Census data (Ortman et al. 2012).

This issue is further complicated by the fact that it wasn't until 2003 that the federal government issued new standard birth certificate and death certificate forms allowing parents to mark more than one race. However, birth and death certificate data are collected by states and the states only adopted the new forms

slowly over time. Every year after 2003, a new group of states adopted the new birth certificate and death certificate forms. Therefore, each year from 2003 to 2010 the Census Bureau received files on births from NCHS with two kinds of racial categories; one file where respondents were allowed to report multiple race data and one file where they were not. By 2010, 35 states and the District of Columbia were using the new federal birth and death certificate forms.

DA analysis requires that the mixed race data from the birth (and death) certificates be categorized as Black or Non-Black, based on both single-race and multiple-race reported by mother and fathers. For the 2010 DA estimates data from birth certificates were used to categorize people into Black Alone or Black Alone or in Combination categories. NCHS provided the Census Bureau with both the multiple races that are reported and the multiple race response “bridged” to the pre-1997 OMB single race categories. Details about the bridging method are provided by NCHS on their website.

Assignment of race on death certificates is also a potential problem but deaths contribute very little to the DA estimates for young children (Aries 2008).

Given the issues described above, one should view DA estimates for Blacks (Alone or Alone or in Combination) cautiously. Small differences or small changes over time could be due to methodological issues rather than real differences or changes.

The 2010 Census DA estimates were first released in December 2010 but in May 2012 the Census Bureau issued revised Demographic Analysis estimates, for the total population, the Black Alone population, the Black Alone or in Combination population, the Not Black Alone population and the Not Black Alone or in Combination population, but not for the Hispanic population (U.S. Census Bureau 2012a). The estimates for the Black Alone or in Combination populations were only provided for the population below age 30. The May 2012 DA estimates were based on the more recent birth and death data and improvements from ongoing research compared to the DA estimates originally released in December 2010. Since the DA estimate for Hispanics were not updated in the May 2012 release, I use the Middle Series of the December 2010 release for that group in my analysis.

## 2.6 Summary

The main methods for measuring coverage in the U.S. Census are Demographic Analysis (DA) and Dual Systems Estimates (DSE). These two methods produce results that are fairly consistent for all age groups except young children. For the population age 0–4, the DA method estimates a net undercount of 4.6 % compared to 0.7 % for the DSE method (the DSE method is called Census Coverage Measurement in the 2010 Census).

The DA method is widely viewed as the better method for estimating net undercount of young children because it relies heavily on vital events data which are very high quality and the most problematic component of DA, Net International

Migration, is only a very small part of the DA estimates for young children. Moreover, the undercount estimates for young children produced by DSE may suffer from correlation bias which results in an underestimates of the net undercount.

Given the challenges and complications to making the racial categories from the birth certificates consistent with those offered in the Census, net undercount estimates of the Black population should be used cautiously.

## References

- Adams, J. W., & Kasakoff, A. B. (1991). Estimates of U.S. decennial census underenumeration based on genealogies. *Social Science History*, 15(4), 527–543.
- Anderson, B. A. (2004). *Undercount in China's 2000 census in comparative perspective*. PSC Research Report, No. 04-565, Population Studies Center, Ann Arbor, MI: University of Michigan.
- Aries, E., Schauman W. S., Eschbach K., Sorlie P. D., & Backlund E. (2008). *The validity of race and hispanic origin reporting on death certificates in the United States* (Vol. 148, 2). Hyattsville: National Center for Health Statistics, Vital Health Statistics.
- Bhaskar, R., Scopilliti, M., Hollman, F., & Armstrong, D. (2010). Plans for producing estimates of net international migration for the 2010 demographic analysis estimates. Census Bureau Working Paper No. 90.
- Bryan, T. (2004). Population estimates. In J Siegel & D Swanson (Eds.), *The methods and materials of demography* (2nd ed.) Elsevier Academic Press, pp. 523–560.
- 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., & 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.
- Coale, A. J., & Zelnick, M. (1963). *New estimates of fertility and population in the United States*. Princeton NJ: Princeton University Press.
- Demographic Analysis Research Team. (2010). *Estimates of net international migration in demographic analysis*. Population Division, U.S. Census Bureau, presentation at 2010 Demographic Analysis Conference, Washington DC, December 6.
- 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.
- Griffin, D. H. (2014). Final task force report: Task force on the undercount of young children. *Memorandum for Frank A. Vitranio*. Washington, DC:U.S. Census Bureau, February 2.
- Hacker, J. D. (2013). New estimates of census coverage in the United States: 1850–1930. *Social Science History*, 37(1), 71–101.
- Himes, C. L., & Clogg, C. C. (1992). An overview of demographic analysis as a method for evaluating census coverage in the United States. *Population Index*, 58(4), 587–607.
- Hogan, H., Cantwell, P., Devine, J., Mule, V. T., & Velkoff, V. (2013). Quality and the 2010 census. *Population Research and Public Policy*, 32, 637–662.
- Jensen, E. (2012). *International migration and age-specific sex ratios in the 2010 demographic analysis*. Paper presented at the Applied Demography Conference at the University of Texas at San Antonio Texas, January.
- Martin, E. (2007). Strength of attachment: Survey coverage of people with tenuous ties to residences. *Demography*, 44(2), 437–440.
- Martin, J. A., Hamilton, B. E., & Ventura, S. J. (2011). *Births: final data for 2009, National Vital Statistics System* (Vol. 60, no. 1). Hyattsville, MD; National Center for Health Statistics.
- Mule, V. T. Jr. (2010). *U.S. coverage measurement survey plans*. Paper delivered at the Joint Statistical Meetings, Vancouver, Canada.

- Mulry, M. (2014). Measuring undercounts for hard-to-reach groups. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-survey populations*. Cambridge: Cambridge University Press.
- National Center for Health Statistics. (2014). *Assessing the quality of medical and health data from the 2003 Birth certificate revision: results from two states* (Vol. 62, no. 2). National Vital Statistics Reports. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
- National Research Council. (2004). The 2000 census: Counting under adversity, panel to review the 2000 census. In C. F. Citro, D. L. Cork & J. L. Norwood (Eds.), *Committee on national statistics, division of behavioral and social science and education*. The National Academy Press, Washing DC, Page 254.
- National Research Council. (2009). Coverage measurement in the 2010 census. In R. M. Bell & M. Cohen (Eds.), *Panel on correlation bias and coverage measurement in the 2010 decennial census*. Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: National Academy Press.
- O'Hare, W. P., Robinson, J. G., West, K., & Mule, T. (2012). *Comparing demographic analysis and dual-systems estimates results for children*. Paper presented at the Southern Demographic Association Conference, Williamsburg VA, October 11–12.
- Ortman, J. M., Hollman, F. W., & Guarneri, C. E. (2012). *Measuring multiple-race births in the United States*. Presented at the Annual Meeting of the Population Association of America, San Francisco, CA: May 3–5, 2012.
- Pitkin, J., & Park, J. (2005). *The gap between births and census counts of children born in California: Undercount or transnational movement?* Paper presented at the Population Association of America Conference, Philadelphia PA. March.
- Price, D. O. (1947). A check on the underenumeration in the 1940 census. *American Sociological Review*, 12, 44–49.
- Robinson, J. G. (2000). *Accuracy and coverage evaluation: Demographic analysis results*. U.S. Census Bureau, DSSD Census 2000 procedures for operations Memorandum Series B-4, U.S. Census Bureau. Page 1.
- Robinson, J. G. (2010). *Coverage of population in census 2000 based on demographic analysis: The history behind the numbers*. U.S. Census Bureau, Working Paper No. 91.
- Schachter, J. (2008). Estimating native emigration from the United States. Memorandum date December 24, delivered to the U.S. Census Bureau.
- Shores, R. (2002). Accuracy and coverage evaluation revision ii adjustment for correlation bias. DSSD, A.C.E. REVISION II MEMORANDUM SERIES PP-53, U.S. Bureau of the Census, Dec.
- Shores, R., & Sands, R. (2003) Correlation bias estimation in the accuracy and coverage evaluation revision II. In *Proceedings of the Survey Research methods Section, Joint Statistical Meetings*.
- 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*: (1966): 71–85. Washington, D.C.: American Statistical Association.
- U.S. Census Bureau. (2003). *Technical assessment of A.C.E. revision II*. Washington, DC: U.S. Census Bureau.
- 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, U.S. Census Bureau, Washington, DC.
- U.S. Census Bureau. (2010b). Tables released at December 2010 conference.
- U.S. Census Bureau. (2012a). Documentation for the revised 2010 demographic analysis middle series estimates. U.S. Census Bureau, Washington, DC.
- U.S. Census Bureau. (2012b). *DSSD 2010 census coverage measurement memorandum series #2010-G-11*. 2010 Census Coverage Measurement Estimation Reports: Adjustment for Correlation Bias, U.S. Bureau of the Census, Washington, DC.

- U.S. Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity*. Statistical Policy Directive 15, Federal Register Notice, October 30.
- Velkoff, V. (2011). *Demographic evaluation of the 2010 census*. Paper presented at the 2011 PAA annual Conference, Washington, DC.
- Wachter, K.W., & Freedman, D. A. (1999). The fifth cell: correlation bias in U.S. census adjustment. Technical Report Number 570. Berkeley: Department of Statistics, University of California.
- West, K. (2012). *Using medicare enrollment file for the DA 2010 estimates*. Paper presented at the Applied Demography Conference at the University of Texas at San Antonio Texas, January.
- 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.



The Undercount of Young Children in the U.S. Decennial  
Census

O'Hare, W.

2015, IX, 112 p. 11 illus., Softcover

ISBN: 978-3-319-18916-1