Differential Coverage Patterns in the Census by Race: Preparing for 2020 Demographic Analysis by Examining Race Allocation in Births

Heather King
David Ihrke
J. Gregory Robinson (retired)

Population Division
U.S. Census Bureau

Presented at the Annual Meeting of the Population Association of America,
Austin, TX, April 10-13, 2019

ABSTRACT

Demographic Analysis (DA) is one of two methods the U.S. Census Bureau uses to estimate net coverage error in the census. The DA estimates are comprised of data on births, deaths, Medicare data, and international migration that provide independent estimates of the population. Differences between census counts and DA estimates have revealed differential coverage of the population in the census by race. A portion of the differences between DA and census may be driven by (1) race allocation in DA input data, (2) race classification in census data, or (3) changing attitudes on race identification in the census. This paper explores coverage error in the census by accounting for differences that stem from race allocation methods used in the 2010 DA estimates. We will use these findings to refine and improve 2020 DA estimates of coverage differentials by race in the 2020 Census.

Introduction

The Census Bureau has used Demographic Analysis (DA) to evaluate the coverage and quality of each decennial census since 1960 (Robinson, 2011). DA estimates provide independent estimates of the size and composition of the population by using administrative data on births and deaths and estimates of international migration using survey data, foreign censuses, and other sources. Administrative data on Medicare enrollment form the basis of the DA estimates for older cohorts.[1]

Differences between the DA estimates and census counts can reveal patterns of net coverage error in the census. We define net coverage error as the percent difference between decennial census counts and DA estimates.[2] In this paper, we focus on differences in net coverage error by race in the census, referred to as *coverage differentials*, as measured by the 2010 DA estimates. Specifically, we address a discontinuity that begins at age 30 (i.e. 1979 births cohorts[3]) in the 2010 DA estimates of coverage differential by age between Black and non-Black populations in the 2010 Census. The distortion most likely stems from race assignment approaches used in births (the main driver of population change) in the DA estimates and not a change in census coverage.

The 2010 DA estimates provided estimates of Black and non-Black population for ages 30 years and older (and no other race groups) because of limited race detail available on historical birth and death records. Here, we expand the race detail available in the 2010 DA estimates of the Black population for this age group. We produce updated 2010 DA estimates of the population over 30 years old that include separate estimates of Black alone (BA) and Black alone or in combination (BAOIC) populations by adjusting the 2010 DA births series for cohorts born before 1980. The experimental DA estimates presented here have the added benefit of removing the disconnect in coverage differentials seen at age 30 in the 2010 DA program.

---

[1] Births and deaths over the period 1935 to 2010 and estimates of net international migration define the population under 75 years old in the 2010 DA estimates, while adjustments to Medicare rolls form the population 75 years and older.

[2] In this study, we define net coverage error as 100*(Census counts – DA estimates) / DA estimates. For historical context, net coverage error was referred to as "percent net undercount" in publications for previous census evaluations (i.e. 1940 to 2000) and was written 100*(DA estimates - Census counts) / DA estimates.

[3] Since the reference for the census is April 1, the birth cohort for each age spans births occurring from April of one year through March of the next year (i.e. April 2009 through March 2010 for age 0 in 2010; April 1979 through March 1980 for age 30 in 2010).

In the following sections, we discuss issues that drive the discontinuity in coverage differentials by age and the motivation for selecting the method to resolve it. We present a series of charts that compare 2010 DA births (and estimates) with adjusted births (estimates) and close with a discussion on how these results inform our preparation for the 2020 DA program to examine differential coverage patterns by race and age in the 2020 Census.

Background

In 2010 DA, births contributed over 80 percent of gross population change from 1945 to 2010 for the population under 65 (Census Bureau, 2012a). Hence, any evaluation of DA estimates by race largely requires an analysis on births by race. The DA estimate of births uses birth certificate data on live births compiled by the National Center for Health Statistics (NCHS). Race and ethnicity items for the child are not collected. We use information about the race and ethnicity of parents, which is included on the form, to infer the race of the newborn.

The quality, availability, and level of race and ethnicity detail on birth records have improved over time. Race definitions in the census and in vital statistics have also evolved over time. These factors led to different race assignment methods in DA estimates by period and limited the level of race detail produced in the 2010 DA estimates. Accordingly, the 2010 DA procedure used the race of father for births up to 1979, and a more complex method that takes into account race of both parents for births in 1980 and later. This difference in race allocation method drives a discontinuity at age 30 in the 2010 DA estimates of net coverage of the Black population by age and race, which leads to a distortion in coverage patterns of the census by race and age. We highlight and examine this discontinuity in coverage, which we suspect is an issue in the 2010 DA estimates and not an anomaly in census coverage.

In 2010, the DA program provided estimates of two Black populations for ages 0-29: (1) Black alone and (2) Black alone or in combination. The Black alone population comprises those who identify solely as Black or African American. The Black alone or in combination population includes the Black alone population in addition to others who identify as Black or African American ***and*** at least one other race. Because of limited information on individual births for those born prior to 1980 and the

small share of individuals who identify as Black ***and*** another race[4], separate estimates for Black alone and Black alone or in combination were not produced for the 30 years and older population in the 2010 DA. See Figure 1 for the May 2012 Release of 2010 DA estimates of the Black alone (blue line) and Black alone or in combination population (orange line) by age. For expediency, we frequently refer to the Black alone population as "BA" and Black alone or in combination as "BAOIC."

In this work, we expand the 2010 DA estimates for the Black population by producing two separate estimates for the Black alone (BA) and Black alone or in combination (BAOIC) population for ages 30 years and older. The adjusted DA series presented here mitigates the distortion in the age distribution of 2010 DA estimates of net coverage error by age and race for the Black population.

To better understand the effect of race allocation methods on net coverage error in the census based on DA estimates, we compare the levels and distributions of historical births to Black parents prior to 1980 to those in 1980 and later. Specifically, we adjust births to the Black population over the period 1935-1980 to produce separate 2010 DA estimates of Black alone and Black alone or in combination for the population 30 years and older, comparable to the population in ages 0-29. These adjustments should mitigate an expected distortion in net coverage error at age 40 in the 2020 DA estimates (carried over from the distortion at age 30 in the 2010 DA estimates) and improve our 2020 DA estimates of census coverage by race.

Race Assignment to Births

The DA estimate of births uses aggregated birth certificate data from the National Center for Health Statistics (NCHS). As noted previously, these data do not include race and ethnicity detail for the child. The race of the child relies on the reported race of the mother and father. However, there are numerous ways to assign race to births, based on the characteristics of biological parents. The

[4] The Black alone (BA) population comprised about 95 percent of the Black alone or in combination (BAOIC) population in the 2000 Census. Thus, most individuals in the BAOIC group identify as Black or African American and no other races. However, the share of the BA population within BAOIC shrank to about 93 percent in the 2010 Census. The BAOIC population also grew at a faster pace than the BA population from 2000 to 2010 (Rastogi et al, 2011).

Census Bureau has examined three potential assignment rules, based on guidance from NCHS: (1) the "Minority rule", (2) the "Mother rule", and (3) the "Father rule."

Until 1989, NCHS used the Minority rule. The Minority rule assigns the race of a non-white parent in mixed race couples to the birth record. If both parents are non-white, the following rules apply: (1) assign the race of father, (2) assign Hawaiian when either parent is Hawaiian. In 1989, NCHS started using the Mother rule, where all births are assigned the race of their mothers[5]. The Father rule follows in a similar fashion. Passel and research conducted by the Census Bureau (U.S. Census Bureau, 1991) show that the Father rule is the most consistent of the three rules with census race classification.

For births up to 1979, the 2010 DA program used aggregate births by month and year of birth, sex, and paternal race. Based on previous research referenced above, the 2010 DA program used the Father rule to assign race to all births up to 1979.

For births after 1979, the 2010 DA program used individual level data on births and leveraged information from both parents. These more recent data contain information on the date and sex of birth and race and ethnicity of the parents. Race comes from the characteristics of both parents, instead of relying on the race of just one. To this end, the Census Bureau created a file referred to as Kid-link. The Kid-link file is the distribution of race and Hispanic origin of children under the age of 1 in the 2000 and 2010 Census for each maternal and paternal race and Hispanic origin combination.[6] The DA program imposes the Kid-link distribution of race and Hispanic origin seen in census data to each year of births. For example, if X percent of biological children in households

[5] In 1989, the standard certificate of live birth changed to include many characteristics of the mother such as method of delivery and tobacco and alcohol use during pregnancy. Births were then tabulated by maternal race (instead of using the Minority rule) to remain consistent with the additional outcomes of interest (National Center of Health Statistics, 1992).

[6] The original December 2010 release of the 2010 DA estimates used children under 17 to produce the Kid-link file. A revised set of DA estimates released in May 2012 used an updated Kid-link file with children under 1 years of age in households from the 2000 Census and the 2010 Census. For more details on the revisions in the May 2012 release, see https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/da_methodology.pdf

with a Black father and a White spouse/unmarried partner identified as Black in Kid-link, then X percent of births to Black fathers and White mothers are assigned Black.

Net Coverage Error and Coverage Differentials by Race

According to historical DA estimates, the Black population has been undercounted in multiple censuses, compared to non-Black populations. However, this difference in coverage has decreased over time. For example, comparisons with DA estimates showed an estimated 5.5 percent net undercount of the Black population, compared to a 1.6 percent net undercount of the total population in the 1990 Census. In the 2000 Census, the estimated DA undercount for the Black population was reduced to 2.8 percent and 0.1 percent for the total population (Devine et al, 2011). There is no comparable 2010 DA measure of net coverage error for the Black population, which we hope to address here[7].

To illustrate the confounding effect of race allocation methods on estimates of coverage differentials, we first calculate the net coverage error for each race group by age. Net coverage error here is defined as 100*(Census counts−DA estimates) / (DA estimates) and is calculated separately for the Black and non-Black populations by single year of age. A positive net coverage error indicates that the Census count is higher than the DA estimate, which implies a net overcount in the census. A negative net coverage error implies a net undercount.

Coverage differential is the difference in estimated net coverage error between two demographic groups. Figure 2 shows the estimated difference in net coverage error (coverage differential) for non-Black populations compared to Black populations by age in the 2010 DA estimates.

In Figure 2, positive net coverage differentials show census coverage gain among non-Black race groups, relative to the Black population. For example, the net coverage error of the non-Black population at age 40 exceeds that of the Black alone population (blue line) at the same age by

[7] Unlike the DA estimates for ages under 30, separate 2010 DA estimates for ages 30 and over that are comparable to the 2010 Census BA and BAOIC race categories are not available. Therefore, an internal "race consistent" DA estimate across all ages cannot be produced.

about 4 percentage points. A positive coverage differential can indicate either lower net undercount or higher net overcount in the non-Black population in the census.

Figure 2 shows coverage differentials by age when census data are tabulated in the more restrictive race category of Black alone (blue line) and when they are tabulated in the more inclusive race category of Black alone or in combination with other race groups (orange line). This figure shows the impact of how one tallies census race categories on estimates of net coverage error and coverage differentials.

Figure 2 prompts two obvious questions: (1) Why does the pattern in differential coverage in the Black alone or in combination population (orange line) abruptly change at age 30 and (2) why is it so different from that seen in the Black alone group (blue line) for the same ages? Are members of the Black alone or in combination population suddenly "easier to count" starting precisely at age 30, where the coverage differential compared to non-Black populations is abruptly reduced? This is probably not the case, and clearly not consistent with the relatively large coverage differentials seen in previous censuses.

Coverage differentials in the census alone do not explain the discontinuity seen in Figure 2. The population age 30 in the 2010 DA estimates was developed largely from the 1980 birth cohort (with some contribution from international migration). Births from 1980 and onwards in the DA process receive race from the Kid-link method, as opposed to the Father rule. Despite this difference in race assignment, we expect patterns of net coverage error for the Black alone or in combination group to be similar to that seen in the Black alone group and to follow patterns found in previous censuses. See Figure 3 for the 1990, 2000, and 2010 DA estimates of net undercount of the Black population by 5-year age groups. The 2010 DA estimates in Figure 3 are broken out by BA and BAOIC and further illustrate the conflicting coverage patterns seen in Figure 2 in 2010. In particular, the sharp decrease in the net undercount from ages 25-29 to 30-34 seen in the 2010 BAOIC set of DA estimates is inconsistent with the stable coverage patterns at younger ages since 1990.

The absence of separate 2010 DA estimates of the Black alone and Black alone or in combination populations for cohorts born before 1980 exacerbates the discontinuity in coverage differentials by

age seen in Figure 2, which we address here. Because births to parents of different races made up a small share of all births in the years prior to 1980 (Adlakha et al, 2002) and historical data are limited on individual level births, the 2010 DA program produced one series of estimates for the Black population aged 30 years and older.

The Black alone group is by definition a subset of the Black alone or in combination. Other researchers have found that the Black alone and Black in combination groups share similar social, economic, and health characteristics and outcomes, ranging from low birth weight, poverty status, educational attainment, and residential mobility (Ma, 2008; Fryer et al, 2008; 2012; Gabriel, 2016). Since poverty status and housing structure are historically associated with undercount in the census and the BA population makes up a large portion of the BAOIC population, we expect patterns of net undercount for these two groups to mirror one another (U.S. Census Bureau, 2016). Results from the Census Bureau's 2010 Census Coverage Measurement (CCM) program, a set of survey-based estimates of census undercount, further support the assumption of similar patterns of undercount in the census for the Black alone and Black alone or in combination populations.[8] The 2010 CCM results show that the Black alone or in combination and the Non-Hispanic Black alone group both had a 2.1 percent undercount in the 2010 Census (U.S. Census Bureau, 2012b).

Data and Methods

In this paper, we quantify and account for differences between the 2010 DA estimates and the 2010 census by race that stem from using two race assignment methods. To this end, we adjust the 2010 DA estimates by race and age, reconciling the effects of using the Father rule for births up to 1979 and the Kid-link method for births after 1979. To mitigate the discontinuity observed at age 30 in the 2010 DA estimates of net coverage error, we create two separate 2010 DA estimates of the BA population and the BAOIC population for ages 30 and older by adjusting historical births based on the Father rule (1935-1979). We use individual birth data over the period 1968 to 2010, harmonizing race categories to "Black" and "non-Black" for parents, to validate such adjustments.

[8] The Census Coverage Measurement (CCM) program uses a dual-system estimation procedure that samples the Census address frame and compares the sample with the actual decennial enumeration to derive estimates of person and household undercounts by characteristics. For more information about CCM, see https://www.census.gov/coverage_measurement/pdfs/A-19.pdf

We use Black births from the 2010 DA program for ages 0-74 (birth cohorts 1935-2010) and tabulated data on individual births to U.S. residents compiled by the National Center for Health Statistics (NCHS) from 1968 to 2010. The tabulated births from 1968-1984 are based on a 50 percent random sample of all registered births from some states. Beginning in 1985, all registered births are available for tabulation. See Table 1 for the number of states with 100 percent birth data by birth year from 1968 to 1985. Relevant items in the birth data include date and sex of birth, maternal and paternal race and ethnicity, state/county of occurrence birth and state/county residence of mother.

We compare relationships between the 2010 DA Black births derived from Kid-link for 1980-2010 birth cohorts (which largely define the population in ages 0-29 in 2010) to births derived from the race of father over the same period. We use tabulated births from 1968 to 2010 to validate our data and methods. The result is a set of births and DA estimates of the Black alone population and the Black alone or in combination for ages 30 and older. The new births (and new 2010 DA estimates) rely on patterns seen in the 2010 DA birth estimates for ages 0-29 (birth cohorts 1980-2010). The 2010 DA birth cohorts for ages 0-29 serve as benchmarks, because we assume they are based on (1) more recent, higher quality vital statistics data and (2) a more nuanced Kid-link method that leverages information from both parents and household decennial census data.

Using the 2010 DA BA and BAOIC birth series for ages 0-29 as benchmarks, we derive adjustment ratios that convert the single 2010 DA series by age of the Black population into separate estimates of Black alone (BA) and Black alone or in combination (BAOIC) for the population 30 years and older. Birth cohorts over 1935-1979 are the main component of the population 30 years and older population in 2010. We derive birth year specific "conversion factors" that we apply to historical births (1935-1979) based on the Father rule to bring them into better alignment with more recent births based on the Kid-link method (1980-2010).

Two series of ratios by birth year are required: one for BA births and the other for BAOIC births. To derive Black alone births, we model the relationship between year of birth and the ratio of 2010 BA births to Father-rule births over 1980-2010 (ages 0-29 in 2010) and then apply this relationship to births over 1935-1979. We estimate a new 2010 DA series of Black alone or in combination for ages

30 and over in a similar fashion using a model of the ratio of the 2010 DA BAOIC births to Father-rule births by birth year. Allowing the ratios to change over time accounts for changing racial identities and partnering of parents over time. In other words, we indirectly use information about the increasing prevalence of interracial unions in the birth data to parse out the Black alone population from Black alone or in combination.

We fit an exponential model that estimates the relationship between birth year and the ratio of 2010 BA births (Kid-link births) to Father-rule births using birth cohorts from 1980-2010. We use the fitted exponential model to "backcast" the ratios over the period 1935-1979. Thus, we convert the Father-rule births in the historical birth data into comparable estimates of births in the Black alone population. We use a simple linear interpolation method to backcast ratios to derive estimates of Black alone or in combination births, because the relationship between birth year and the ratio of 2010 BAOIC births to Father-rule births is not as well defined. We discuss this issue in greater detail in later sections.

We use backcasted ratios to create a 2010 DA Black alone series for cohorts born between 1935 and 1979 and a birth series for the 2010 DA Black alone or in combination population. Finally, we replace the 2010 DA estimates of births with our new ratio-adjusted births and derive a set of new 2010 DA estimates that include a Black alone and a Black alone or in combination series for all ages 0-74.

The following sections discuss the tabulations of births derived from different race assignment rules, detail on our ratio adjustment strategy, and the effects of the adjusted 2010 DA BA and BAOIC series on estimated net coverage error and coverage differentials. We close with comments on limitations, next steps, and implications for 2020 DA estimates coverage differentials by race.

Results

Figure 4 shows the number of births assigned Black over the period 1968-2010 under the following conditions: (1) Father rule (2) Mother rule (3) *either* parent is black, and (4) *both* parents are black. The 2010 DA estimates of Black alone (BA) births and Black alone or in combination (BAOIC) births are also shown for reference. For births up to 1979, the 2010 DA BA and BAOIC birth series are

identical. Our intent here is to produce separate estimates of BA and BAOIC births over 1935-1979, based on relationships observed between births derived from the race of father and the series of 2010 DA BA and BAOIC births over 1980-2010.

The two approaches assumed to best measure the Black alone population in the census from other research (Adlakha et al, 2002, Devine et al, 2011) – the Mother rule[9] and "Both parents black rule" – yield the smallest numbers of Black births each year, as expected (dark blue line and purple line, respectively, in Figure 4). Births derived from the "Both parents black rule" are the lower bound for all series. The Mother-rule and "Both parents black rule" birth series are almost identical until about 1980, with the differences widening up to 2010. Although births derived from these approaches are similar in levels to the benchmark 2010 DA Black alone series (dashed blue line) across the time series, the Mother rule is slightly more consistent with the benchmark.

The top three series in Figure 4 show births derived from assignment rules assumed to describe the Black alone or in combination population in the census – the Father rule and "Either parent black rule" – and the BAOIC births over 1980-2010 from the 2010 DA estimates. As expected, these three series yield the largest number of births annually (gold line, magenta line, and dashed orange line, respectively). Births from the 2010 DA BAOIC benchmark series over 1980-2010 (dashed orange line) are the upper bound each year, followed by births from the "Either parent black rule" and the Father rule. Father-rule births are more similar in levels to births from the "Either parent black rule" in earlier years from 1968 until about 1990, when the two series start to diverge. From 1980 to 2010, we do not see close agreement between the benchmark series (2010 DA BAOIC births) and births from the "Either parent black rule" assignment rule.

To bring historical births (1935-1979) derived from race of father into alignment with 2010 DA births based on Kid-link (1980-2010), we model the relationship between birth year and the ratio of Kid-link births to race of father births seen in ages 0-29. Figures 5 and 6 show the ratios for Black alone births and Black alone or in combination births, respectively.

[9] The Mother rule generates a smaller number of Black births, compared to the Father rule, because there are more Black fathers annually.

Figure 5 shows the ratio of Black alone births derived from the Kid-link rule for births 1980-2010 to those based on the race of father (dashed blue line – denoted (2010 DA BA) / ROF) by birth year. For reference, we also show the ratio of Mother-rule births to Father-rule births (dark blue line – denoted ROM / ROF) and the ratio of births when both parents are Black to race of father births (purple line – denoted Both / ROF). The ratio of Black alone births from 2010 DA to Father-rule births (2010 DA BA) / ROF) shows an exponential decline from 1980 to 2010, and subsequently, an exponential decline in the agreement between Father-rule births and Kid-link births defined as Black alone. This pattern tracks with that suggested by the race of mother ratio (ROM / ROF), the ratio of births to both parents (Both / ROF), and the increasing frequency of births over time to interracial couples.[10]

To express this exponential decline of agreement, we fit a model of exponential decay relating age in 2010 (and equivalently, birth year) to the ratio of 2010 DA Black alone births derived from Kid-link to Black births based on race of father. We then backcast the ratios for historical births, preserving the exponential relationship seen between birth year and the ratio of Kid-link to Father-rule births in more recent years. The backcasted ratios are shown in Figure 5 (dotted line) for the years 1935-1979 (i.e. ages 30-74 in 2010).

Similar to the procedure for the BA adjustment ratio, we produce a series of ratios over 1980-2010 that define adjustments to race of father births over 1935-1979 to produce estimates of Black alone or in combination births. The ratios of Black alone or in combination births from 2010 DA to Father-rule births – denoted (2010 DA BAOIC) / ROF – are shown in Figure 6 for birth cohorts 1980-2010 (dashed orange line). For reference, the ratios of births using the "Either parents black" rule to births using the race of father (magenta line, denoted Either/ROF) are also shown.

The ratio of 2010 DA BAOIC births to race of father births in Figure 6 shows a linear increase over time from 1980 to 2010. It is not clear that the linear relationship between Kid-link births to Father-rule births seen over the period 1980-2010 holds for historical births, since the Either / ROF ratio

[10] Adlakha determined the share of births to parents of different race was around 1 percent in 1960 and grew to over 5 percent in 2000. (Adlakha, 2002).

shows some gradual increase and change in slope starting for births in 1995. As an alternative, we interpolate the ratios between 1935 and 1979, assuming a ratio of 1 at birth year 1935. A ratio of 1 indicates that the race of father identifies all BAOIC births and vice versa. Assuming a starting ratio of 1 in 1935 is supported by the decreasing gap between all birth series, using different assignment rules shown in Figure 4 from 2010 back to 1968 and the small share of births to parents of different races in historical birth data.

Figure 7 shows the new series of 2010 DA estimates (using ratio-adjusted births for cohorts born between 1935 and 1979) for the Black alone and Black alone or in combination compared to the 2010 DA series. There are small differences between sets of DA estimates by age for the 30 years and older population, but the effect on overall net coverage error and coverage differentials by age is pronounced. Next, we describe the effect of ratio-adjusted births on 2010 DA estimates of overall net coverage error of the Black population, followed by the effect of net coverage error and coverage differentials by age.

Table 3 shows the effect of using ratio-adjusted Black births for ages 30 and older on the 2010 DA estimates of net coverage error of the Black population, compared to previous censuses and DA estimates. The 2010 DA estimates showed an overall 2.3 percent net undercount of the Black alone population (top panel, column 4). There was a 0.7 percent undercount measured for the population under 30 and 3.7 percent undercount for ages 30 and older. The overall net coverage error showed a 1.1 percent undercount for Black alone or in combination, with 1.7 and 0.5 percent undercounts for ages 0-29 and 30 and older, respectively. The overall level of coverage for the Black population in 2010 is quite different depending on the DA set used: the net undercount rate based on the BA estimates (-2.3 percent) is more than double the net undercount rate based on the BAOIC estimates (-1.1 percent). This difference is driven mainly by the huge difference in net coverage error estimates for ages 30 and older (-3.7 percent for the BA set, -0.5 for the BAOIC set).

Table 3 reveals the stabilizing effect of the ratio adjustments to historical births on overall net coverage error and by age group on the 2010 DA estimates. The overall net coverage error of the Black alone population decreased from a 2.3 net undercount to 1.7. The net coverage error for ages

0-29 is unchanged, as expected. The undercount measured from the 2010 DA estimates changed from 3.7 to 2.6.

The effect of the ratio adjustment on 2010 DA estimates of net coverage error of the Black alone or in combination population moves the overall net undercount from 1.1 percent to 2.1 percent (bottom panel, column 4, Table 3). The change is seen in ages 30 years and older, where the undercount measured by the 2010 DA estimates increases from a 0.5 net undercount to a 2.5 percent net undercount.

The changes to 2010 DA measures of net coverage error introduced by the ratio adjustments to historical births provide more reasonable estimates of net coverage error overall and for ages 30 years and older. The ratio adjustments produce net coverage error estimates for the BA and BAOIC populations that are more similar to one another and are more consistent with previous census evaluations. The new estimates (BA or BAOIC) both imply gradual reductions in net undercount from 2000 to 2010, less than the more appreciable coverage change seen from 1990 (-5.5 percent) to 2000 (-2.8 percent). Next, we describe the effect of ratio-adjusted births on 2010 DA estimates of net coverage error by age and coverage differentials by age.

Figure 8 shows net coverage error by 5-year age groups for the Black alone and Black alone or in combination population from the 2010 DA estimates and the new ratio-adjusted 2010 DA estimates. The BA estimates are shaded blue, while the BAOIC estimates are orange. The dashed lines in Figure 8 show the new ratio-adjusted DA estimates, while the 2010 DA estimates are solid. The ratio-adjusted series show appreciable change in age distribution and the lines move closer together for both the Black alone (dashed blue line) and the Black alone or in combination populations (dashed orange line) for ages 30 and older. The patterns of net coverage by age suggested by the new ratio-adjusted estimates are consistent with surrounding age groups and results from previous censuses (1990 and 2000 Censuses in Figure 3).

Figure 9 shows the effect of our ratio adjustments on coverage differentials by age between non-Black and Black populations. The Black alone (blue line) and Black alone or in combination (orange line) have plausible coverage differentials by age. The discontinuity at age 30 has disappeared by

leveraging higher quality data for ages 0-29. Figure 9 is an improvement over the problem posed in Figure 2, where patterns of net coverage error and coverage differential diverged at age 30. Figure 10 places all series together for reference and shows that the largest differences between the BA and BAOIC lines are now the population under age 30. A discontinuity still exists at age 30, but one much less pronounced than that seen in Figure 2 before the birth revisions. This discrepancy for younger ages requires research to assess if differences are real. As noted earlier, CCM estimates find no difference in estimates of net coverage error between the BA and BAOIC populations.

Limitations and Next Steps

Our method to account for different race allocation methods by period has several limitations. For one, we did not adjust any of the other components (deaths, international migration) to derive our two new set of DA estimates for the population 30 years and older. We could attempt a similar adjustment for those components in future work but suspect we would obtain similar results.

Since we aligned historical births by race with patterns seen in more recent births, incorrect assumptions about the Kid-link method will bias our historical birth estimates. In other words, errors in the Kid-link method are imposed on historical births. However, adjustments made to the 2010 DA estimates appear to provide more consistent measures of undercount compared to historical DA and Census Coverage Measurement estimates. Fortunately, the dependence of historical births on more recent births poses not only a risk but an opportunity. Improvements to assumptions about race of births in more recent birth records should lead to improvements to estimates of historical births.

We are also at risk of using an incorrect modeled relationship between birth year and the ratio of Kid-link births to race of father births for cohorts born between 1980 and 2010. This is a particular concern for the Black alone or in combination births. However, using comparisons of simple race rules assumed to best identify the Black alone and Black alone or in combination, respectively, mostly validate our assumptions about how Kid-link births and race of father births change over time. Future work can use annual data on births to parents of different races to further refine relationships between birth year, Kid-link derived BAOIC births, and Father-rule births.

An immediate next step is to apply our new method to the 2000 DA estimates to provide comparable estimates of net coverage error of the Black alone and Black alone or in combination populations, respectively. We would expect the ratio-adjusted 2000 DA estimates of net coverage error to be more similar to one another and reduce the probably overstated improvements in net undercounts seen from 1990 to 2000 (shown in Table 3).

In other 2020 DA research, we are looking to use decade specific Kid-link files for births in the 1980s, 1990s, 2000s, and 2010s. Another avenue of research examines the use of alternative data sources to build the Kid-link file, such as the Current Population Survey[11], to strengthen our understanding of the relationships between household structure and the race of parents and their biological children (Jensen and Eickmeyer, 2019). These improvements may lead to a better picture of the relationship between Kid-link and Father-rule births over time. Finally, we can consider expanding the race detail for older age groups in future DA estimates, since the method presented here does not require detailed race information by race of parents.

Conclusion

The DA estimates provide a reliable benchmark for the decennial census population, because they are (1) independent indicators of the size and composition of the population, (2) derived from a rich constellation of sources that specifically measure the expected ways a population can change, and (3) based on data sources that are highly representative of the population. For these reasons, differences between DA estimates and census counts allow us to study patterns of net coverage error in the census.

The DA estimates are not without their limitations. The data used to build DA estimates generally reflect the population by age and sex with high accuracy. On the other hand, there is substantial uncertainty surrounding the estimates by race that can potentially confound our analysis of

[11] The Current Population Survey (CPS) is a longitudinal survey conducted by the Census Bureau that measures economic and social indicators of households. See https://www.census.gov/programs-surveys/cps/about.html for more details.

coverage of the population in the census by race. In this paper, we have accounted for differences in race allocation methods for two different periods and brought the 2010 DA estimates of net coverage error and coverage differentials into better alignment by age.

The undercount of the Black population is a well-documented phenomenon seen in multiple historical censuses. Differences in race definitions between vital statistics and censuses and changes in self-reporting of race have led to more uncertainty about the extent of this undercount and improvements on undercount since 2000. We have accounted for some of this uncertainty by leveraging relationships between recent and historical births. The adjustments made here should improve our 2020 DA estimates of census coverage by race and enhance efforts to further reduce the undercount of the Black population in future censuses.

References

Adlakha, Arjun L., J. Gregory Robinson, and Amy Symens Smith. (2002). "Alternative Rules for Assigning Race of Birth: Effect on Birth Totals, Implications for Vital Rates and Census Undercount Estimates by Race." Paper presented at the annual meeting of the Southern Demographic Association, Austin, Texas.

Devine, Jason. Larry Sink, Bethany DeSalvo, and Rachel Cortes. (2011). The Use of Vital Statistics in the 2010 Demographic Analysis Estimates. Working Paper Number: TWP0088. https://www.census.gov/population/www/documentation/twps0088/twps0088.pdf

Fryer, Ronald G., Lisa B. Kahn, Steven D. Levitt, and Jorg L. Spengkuch. (2008). The Plight of Mixed Race Adolescents. National Bureau of Economic Research Working Paper 14192. https://www.nber.org/papers/w14192.pdf

Gabriel, Ryan. (2016). "The Residential Stratification of Mixed-Race Couples." PhD Dissertation, Department of Sociology, University of Washington. http://hdl.handle.net/1773/36840

Jensen, Eric and Kasey Eickmeyer. (2019). "Assigning Race and Ethnicity to Birth Records Using a CPS-Based Kid-Link Method." Poster presented at the annual meeting of the Population Association of America, Austin, Texas.

Ma, Sai. (2008). "Paternal Race/Ethnicity and Birth Outcomes." American Journal of Public Health. December 2008, Vol 98, No. 12 98: 2285-2292. https://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.2007.117127

National Center for Health Statistics. (1992). Public Use Data Tape Documentation: 1989 Detail Natality.

Guarneri, Christine E., and Christopher Dick. (2012). "Methods of Assigning Race and Hispanic Origin to Births from Vital Statistics Data." Washington, DC, January: Paper presented at the Federal Committee on Statistical Methodology Annual Meeting.

Rastogi, Sonya. Tallese D .Johnson, Elizabeth M. Hoeffel, and Malcolm P. Drewery. (2011). The Black Population: 2010. 2010 Census Briefs. https://www.census.gov/prod/cen2010/briefs/c2010br-06.pdf

Robinson, J. Gregory. (2011). Coverage of Population in Census 2000 Based on Demographic Analysis: The History Behind the Numbers. U. S. Census Bureau, Population Division Working Paper Number: TWPS0091. https://www.census.gov/population/www/documentation/twps0091/twps0091.pdf

____ "Inconsistencies in Race Classifications of the Demographic Estimates and the Census." (1991). DA Evaluation Project D9, *PERM No. 82,* by J. Gregory Robinson and Susan Lapham.

____ "Documentation for the Revised 2010 Demographic Analysis Middle Series Estimates." (2012a). May 2012 Release of 2010 Demographic Analysis Middle Series Estimates. https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/da_methodology.pdf

____ "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States." (2012b). DSSD 210 Census Coverage Measurement Memorandum Series #2010-G-01, by Thomas Mule.

____ "Administrative Records, Internet, and Hard to Count Population Working Group Final Report." (2016). U.S. Census Bureau National Advisory Committee on Racial, Ethnic, and Other Populations. https://www2.census.gov/cac/nac/reports/2016-07-admin_internet-wg-report.pdf

Tables and Figures

Table 1. Number of States with 100 Percent Registered Births by Year

Birth Year	Number	Percent
1972	6	11.5
1973	9	17.3
1974	16	30.8
1975	23	44.2
1976	29	55.8
1977	34	65.4
1978	37	71.2
1979	43	82.7
1980	45	86.5
1981	45	86.5
1982	46	88.5
1983	47	90.4
1984	47	90.4
1985-2010	52	100.0

Notes

1. Natality data from 1968-1971 are based on a 50 percent sample from all states. Beginning in 1972, all births are available for some states.
2. There are 52 "state" level reporting areas in this study: 49 states (all states except New York), the five boroughs of New York City, New York state outside of NYC, and the District of Columbia.

Sources: National Center for Health Statistics Public Use Natality Detail Documentation, 1968-1985

Table 2. Assumed Relationships between Race Assignment Rules and Demographic Analysis Black Population

Race Assignment Rule	Demographic Analysis Series	Description	Race Categories	
			Parents	Births
Both parents black	Black Alone	Assign black if *both* parents black	Black, Not Black	Black Alone, Not Black Alone
Mother rule		Assign race of mother	Black, Not Black	Black, Not Black
Either parent black	Black Alone or in Combination	Assign black if *either* parent black	Black, Not Black	Black Alone or in Combination, Not Black Alone or in Combination
Father rule		Assign race of father	Black, Not Black	Black, Not Black

Table 3. Demographic Analysis (DA) Estimates of Net Coverage Error of the Black Population

Group	1990 DA[1]	2000 DA[2]	2010 DA	
			2010 DA[3]	Ratio-Adjusted 2010 DA[4]
Black Alone				
All Ages	-5.5	-2.8	-2.3	-1.7
Ages 0-29	-5.5	-1.8	-0.7	-0.7
Ages 30+	-5.6	-3.8	**-3.7**	**-2.6**
Black Alone or in Combination				
All Ages	-5.5	-2.8	-1.1	-2.1
Ages 0-29	-5.5	-1.8	-1.7	-1.7
Ages 30+	-5.6	-3.8	**-0.5**	**-2.5**

Notes

1. 1990 DA estimates are based on Race of Father births (ROF) for ages under 65; Medicare enrollment for ages 65+; Census tabulation is the Black population (No separate estimates for Black alone (BA) or Black alone or in Combination (BAOIC) in 1990 Census)
2. 2000 DA estimates are based on ROF births; Medicare enrollment used for ages 65+; Census tabulations are average of "Black Alone" and "Black Alone or in Combination"
3. 2010 DA May 2012 Release estimates used Black alone (BA) defined births from Kid-link for ages 0-29; ROF births defined 2010 DA BA estimates for ages 30-74; Medicare enrollment used for ages 75+; BA Census tabulations

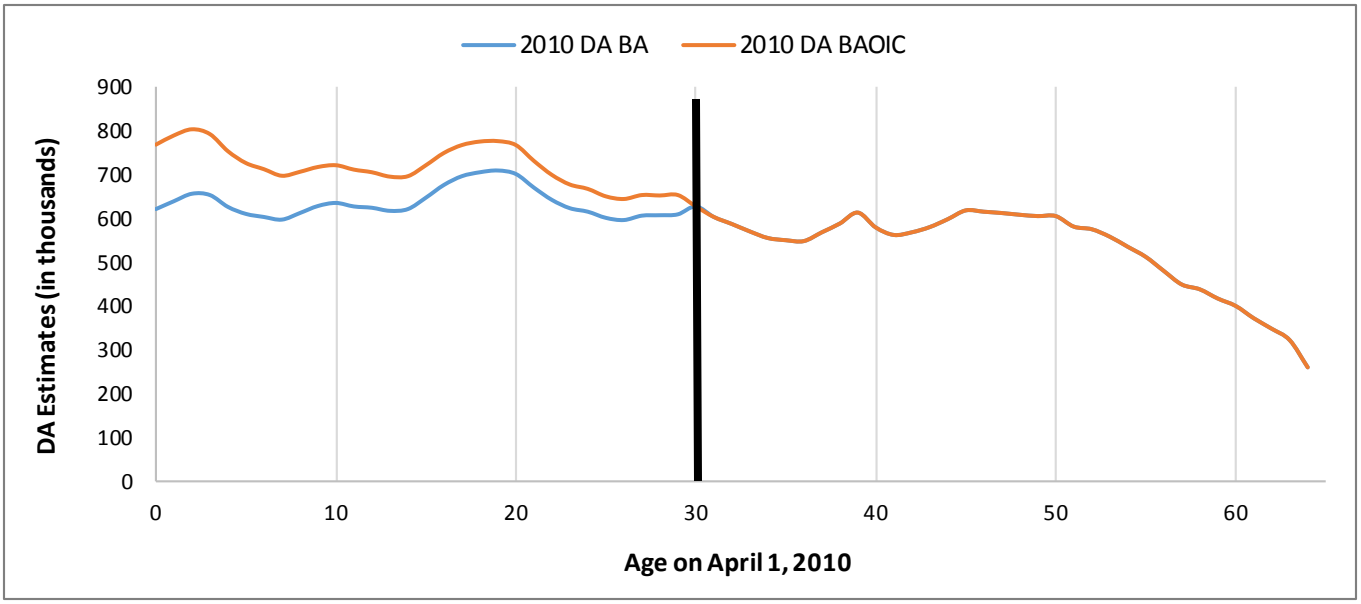
   For Black Alone or in combination (BAOIC), 2010 DA used BAOIC defined births from Kid-link for ages 0-29; ROF births defined 2010 DA BAOIC estimates for ages 30-74; Medicare enrollment for ages 75+; BAOIC Census tabulations

4. Ratio-adjusted 2010 DA estimates use Black alone (BA) defined births from Kid-link for ages 0-29; Ratio-adjusted ROF births define BA estimates for ages 30-74; Medicare enrollment for ages 75+; BA Census tabulations

   Ratio-adjusted 2010 DA estimates use Black alone or in combination (BAOIC) defined births from Kid-link for ages 0-29; Ratio-adjusted ROF births define BAOIC estimates for ages 30-74; Medicare enrollment for ages 75+; BAOIC Census tabulations

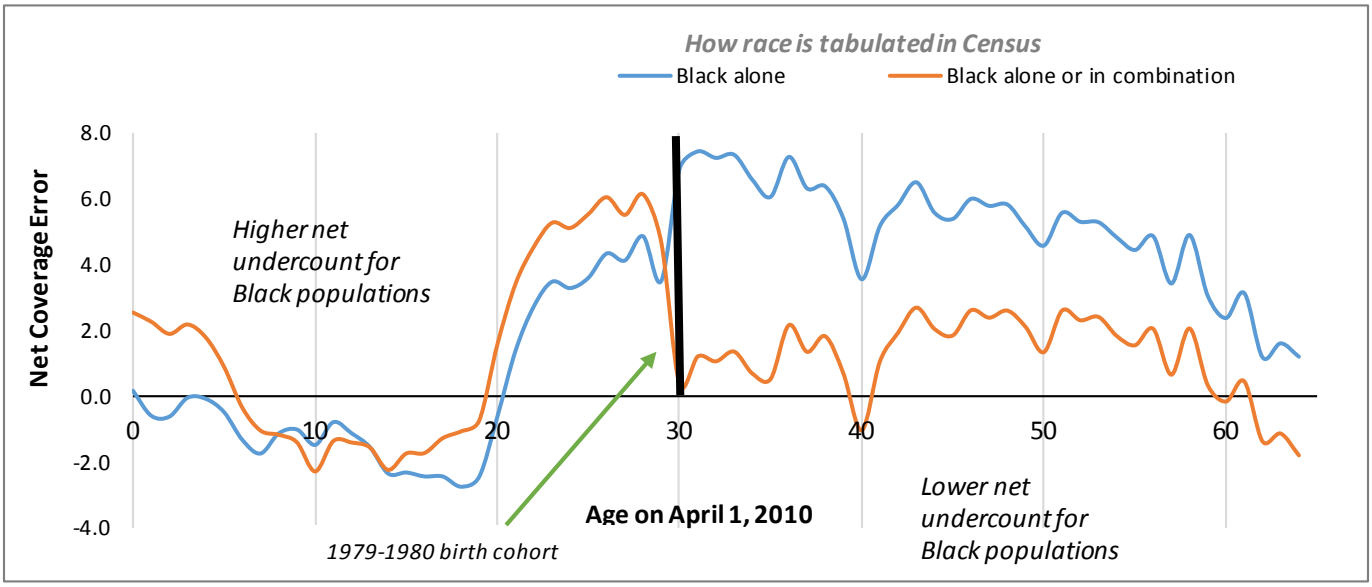
Sources: U.S. Census Bureau: 1990, 2000, and 2010 Census; 1990, 2000, 2010 DA Estimates (May 2012 Release); Ratio-Adjusted 2010 DA Estimates

Figure 1. 2010 Demographic Analysis (DA) Estimates of the Black Population by Age: Black Alone (BA) and Black Alone or in Combination (BAOIC), Ages 0-64



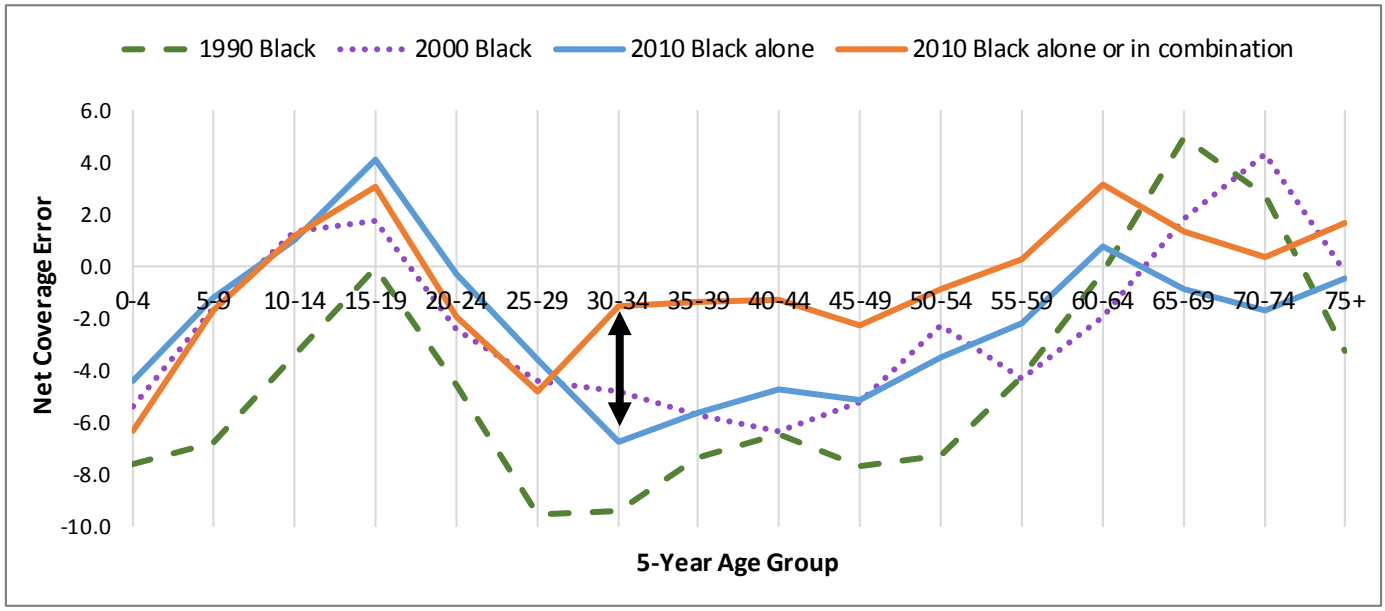
Source: U.S Census Bureau: May 2012 Release of 2010 DA Estimates
Note: Estimates are rounded to nearest thousand.

Figure 2. 2010 Demographic Analysis Estimates of Coverage Differentials between Non-Black and Black Populations by Age, Ages 0-64



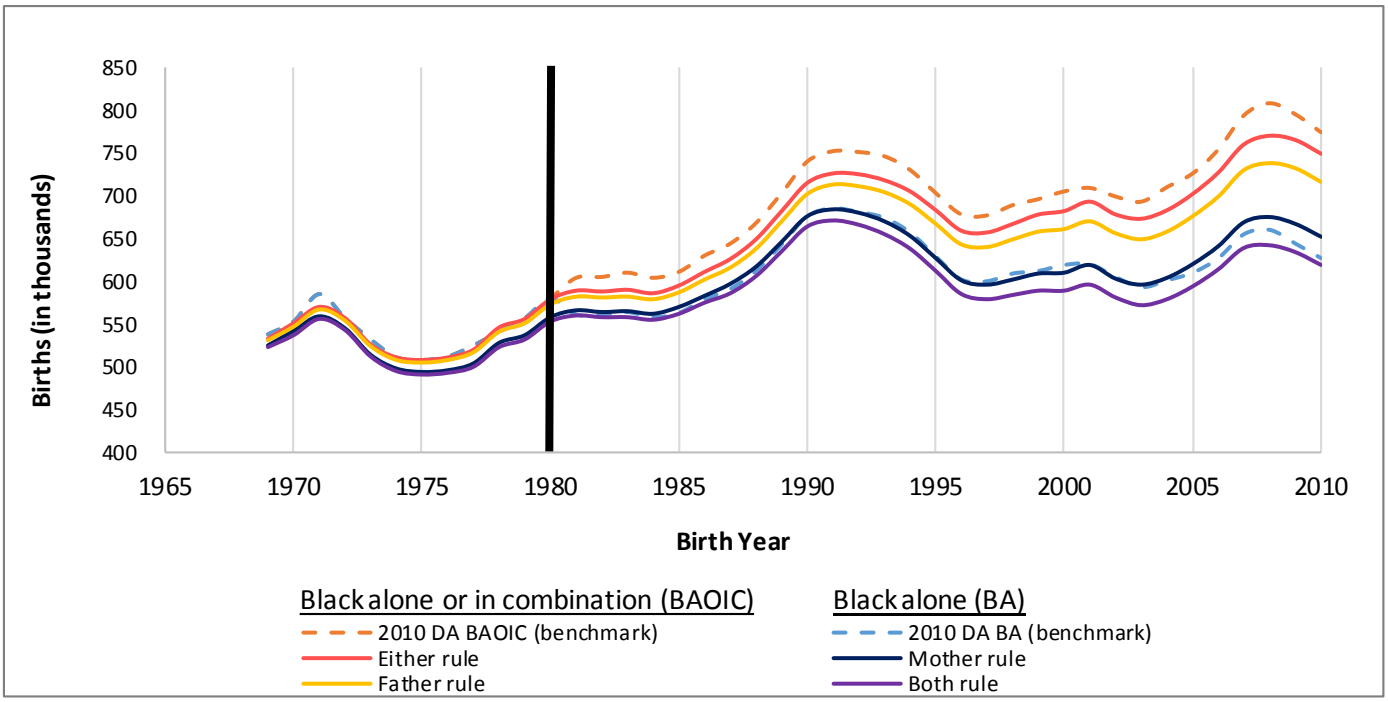
Source: U.S Census Bureau: 2010 Census, May 2012 Release of 2010 DA Estimates

Figure 3. Demographic Analysis Estimates of Net Coverage Error of the Black Population by Age Group in 1990, 2000, and 2010 Census



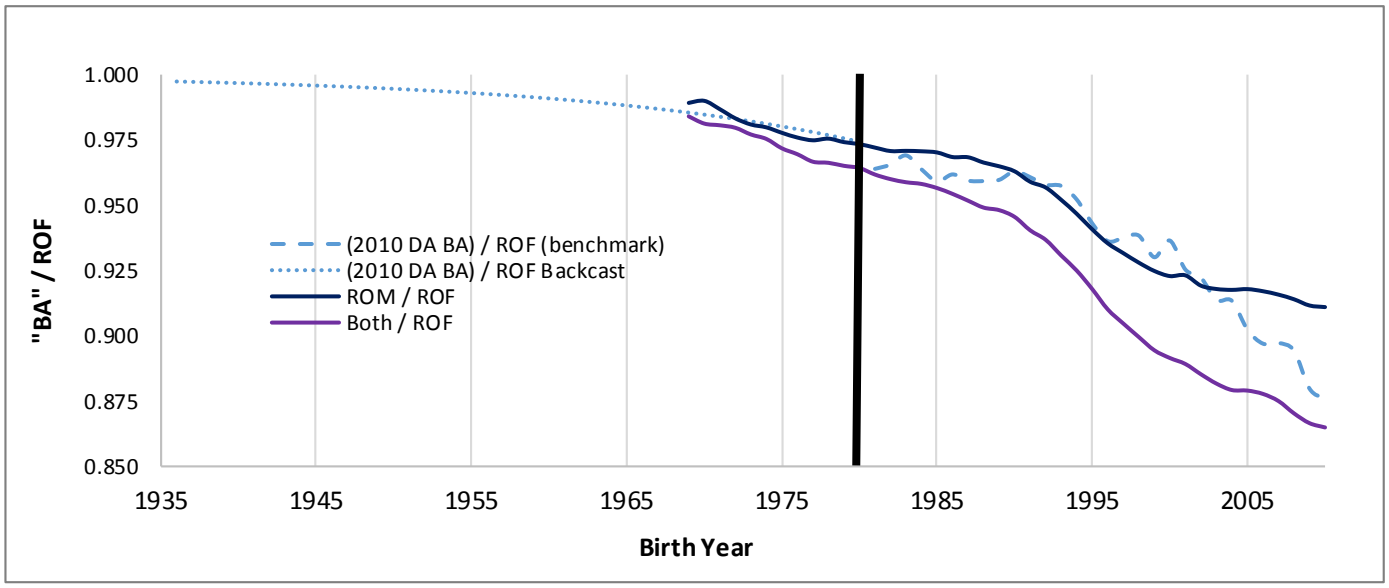
Sources: U.S. Census Bureau: 1990, 2000, and 2010 Census; 1990, 2000, 2010 DA Estimates (May 2012 Release of 2010 DA)
Note: Net coverage error was defined as percent net undercount in the 2000 and 1990 DA publications.

Figure 4. Number of Black Births by Assignment Rule and Birth Year: 1968-2010



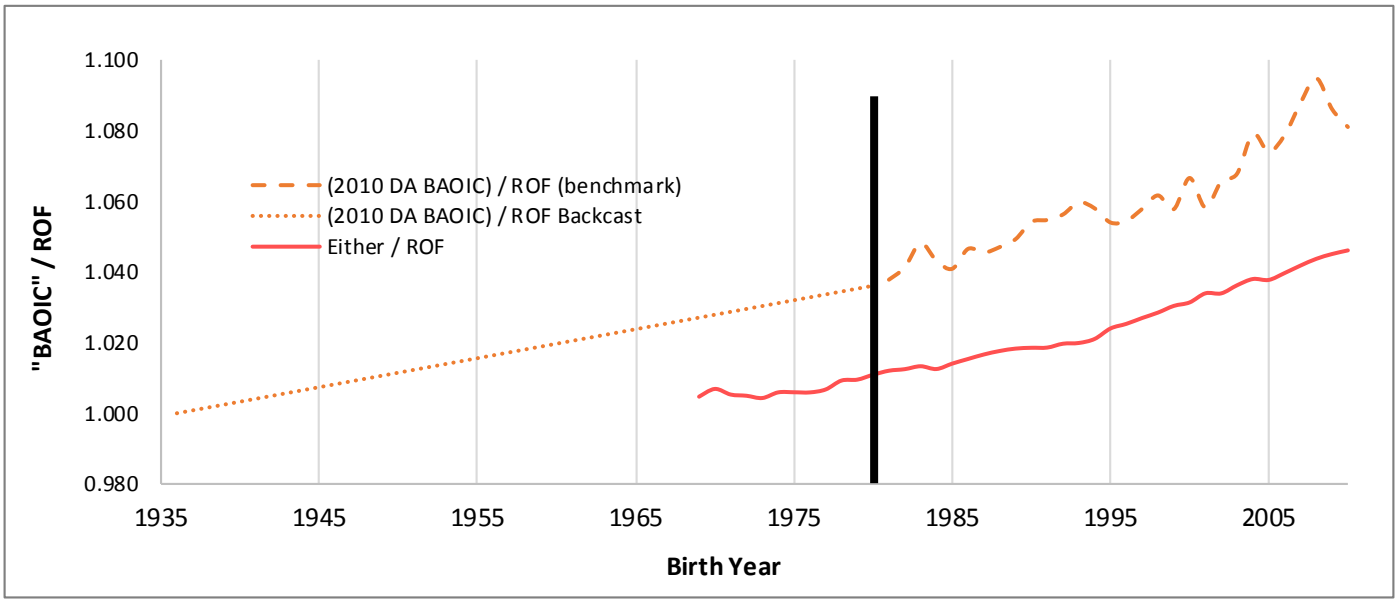
Sources: National Center for Health Statistics: 1968-2010 Natality data; U.S Census Bureau: May 2012 Release of 2010 DA Estimates.
Note: Birth year refers to the ending year of the period from April 1 to March 31 of the following year. Births are rounded to nearest thousand.

Figure 5. Adjustment Ratios to Create Black Alone (BA) Births from Race of Father (ROF) by Birth Year: *Mother-rule births to Father-rule births (ROM / ROF); Both parents black to Father-rule (Both / ROF); 2010 DA Black Alone (BA) births to Father-rule births (2010 DA BA) / ROF*



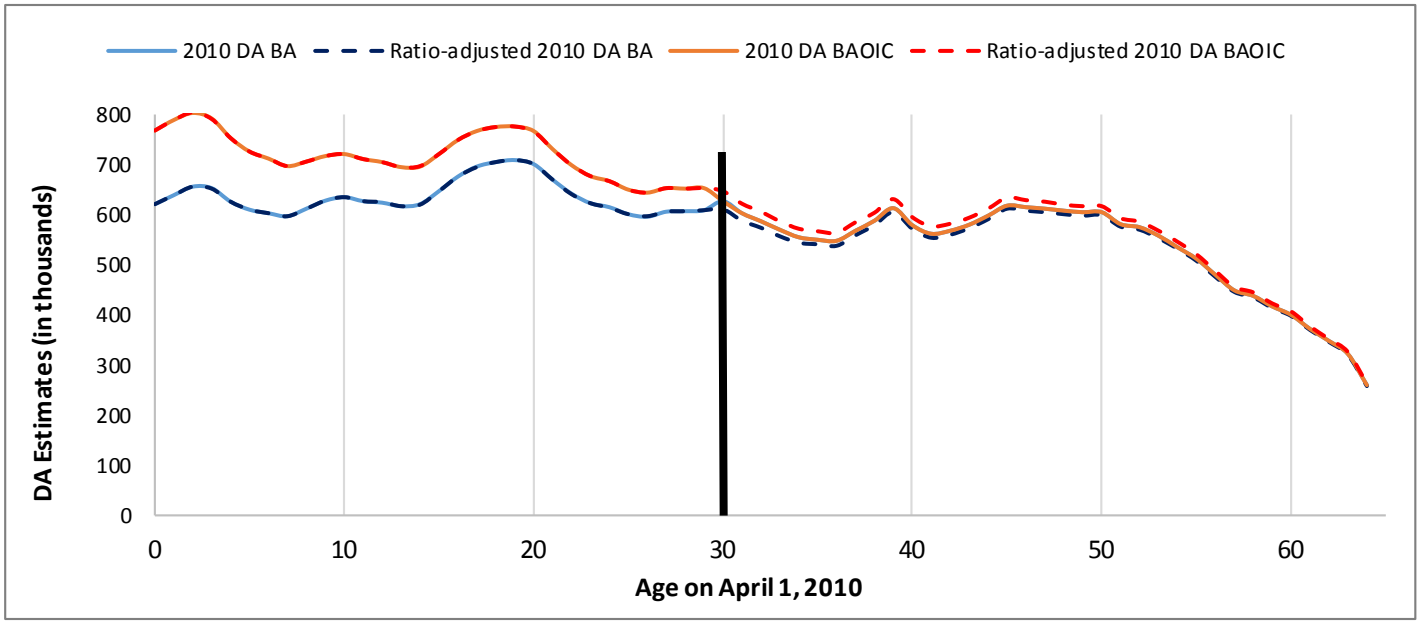
Sources: National Center for Health Statistics: 1968-2010 Natality data; U.S Census Bureau: May 2012 Release of 2010 DA Estimates.
Note: Birth year refers to the ending year of the period from April 1 to March 31 of the following year.

Figure 6. Adjustment Ratios to Create Black Alone or in Combination (BAOIC) Births from Race of Father (ROF) by Birth Year: *Either parent black births to Father-rule births (Either / ROF); 2010 DA Black Alone or in Combination (BAOIC) births to Father-rule births (2010 DA BAOIC) / ROF*



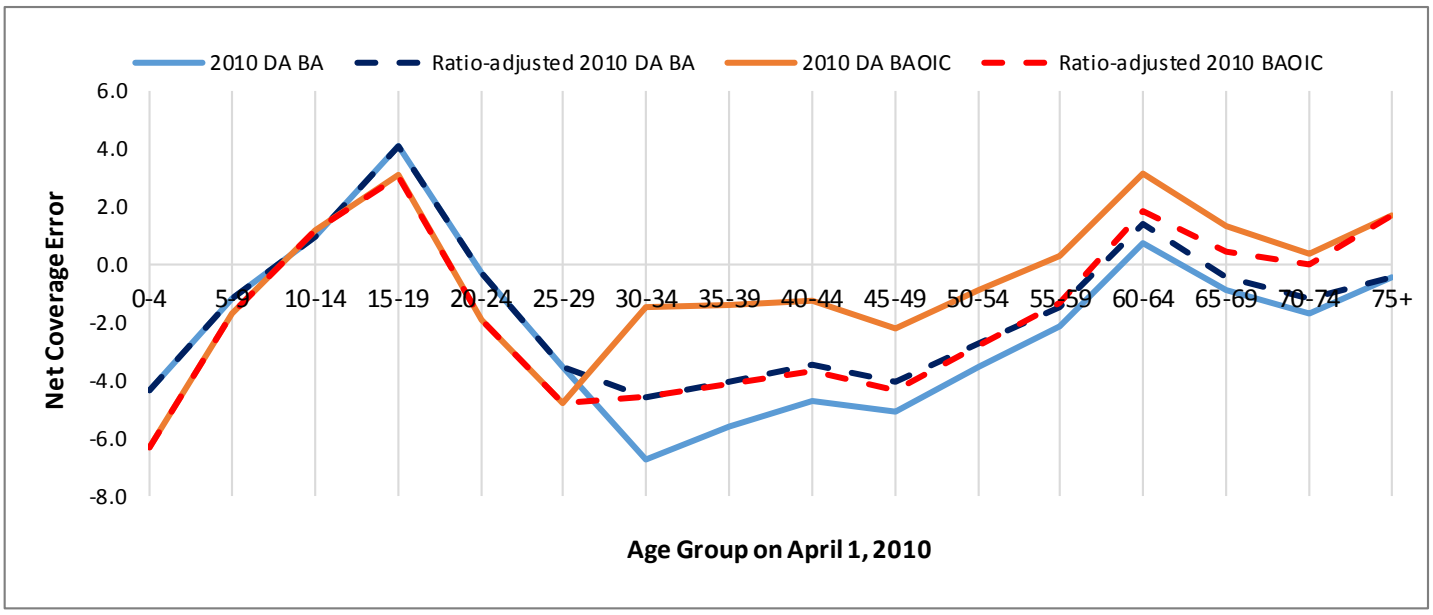
Sources: National Center for Health Statistics: 1968-2010 Natality data; U.S Census Bureau: May 2012 Release of 2010 DA Estimates.
Note: Birth year refers to the ending year of the period from April 1 to March 31 of the following year.

Figure 7. 2010 Demographic Analysis (DA) Estimates of the Black Population by Age: Black Alone (BA) and Black Alone or in Combination (BAOIC), Ages 0-64



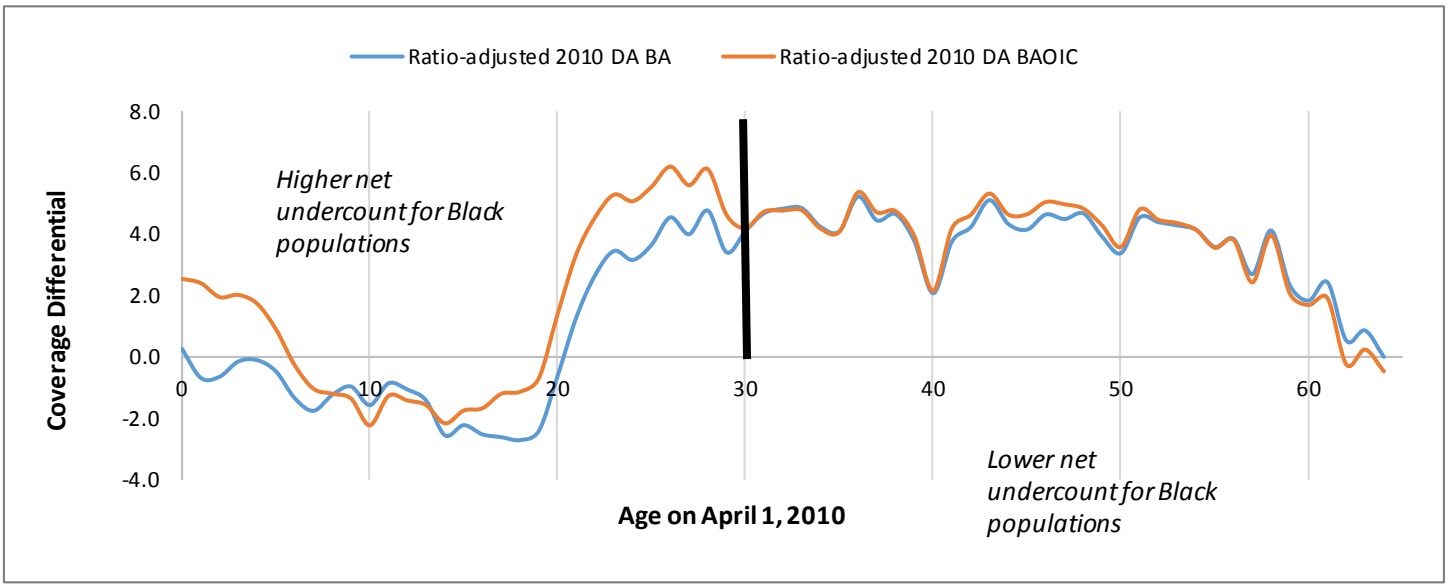
Sources: U.S Census Bureau: May 2012 Release of 2010 DA Estimates, Ratio-Adjusted 2010 DA Estimates
Note: Estimates are rounded to the nearest thousand.

Figure 8. 2010 Demographic Analysis Estimates of Net Coverage Error of the Black Population by Age Group: *2010 Demographic Analysis (DA) Black Alone (BA), 2010 DA Black Alone or in Combination (BAOIC); Black Alone and Black Alone or in Combination from Ratio-Adjusted 2010 DA Black*



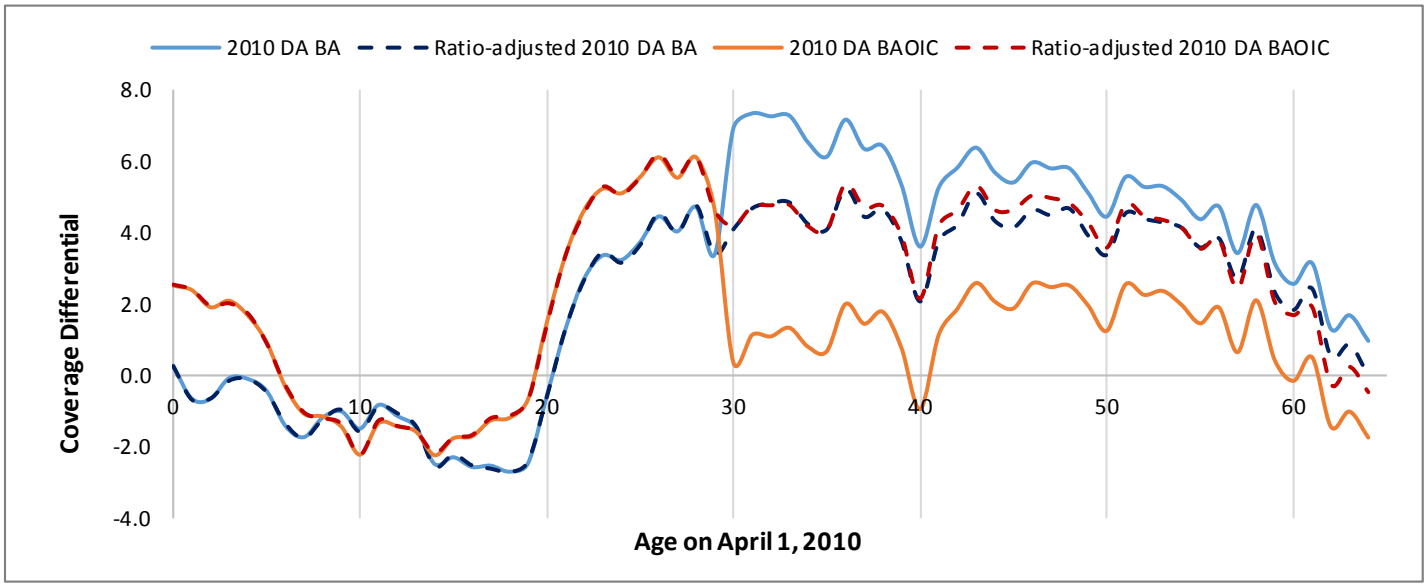
Sources: U.S Census Bureau: 2010 Census, May 2012 Release of 2010 DA Estimates, Ratio-Adjusted 2010 DA Estimates

Figure 9. 2010 Demographic Analysis Estimates of Coverage Differentials between Non-Black and Black Populations by Age, Ages 0-64: *Black Alone (BA) from Ratio-Adjusted 2010 Demographic Analysis (DA) Black, Black Alone or in Combination (BAOIC) from Ratio-Adjusted 2010 DA Black*



Sources: U.S Census Bureau: 2010 Census, Ratio-Adjusted 2010 DA Estimates

Figure 10. 2010 Demographic Analysis Estimates of Coverage Differentials between Non-Black and Black Populations by Age, Ages 0-64: *2010 Demographic Analysis (DA) Black Alone (BA), 2010 DA Black Alone or in Combination (BAOIC); Black Alone and Black Alone or in Combination from Ratio-Adjusted 2010 DA Black*



Sources: U.S Census Bureau: 2010 Census, May 2012 Release of 2010 DA Estimates, Ratio-Adjusted 2010 DA Estimates