ORIGINAL RESEARCH

Does the Census Miss the Native-Born Children of Immigrant Mothers? Evidence from State-Level Undercount by Race and Hispanic Status

Janna E. Johnson[1] 🄳

Abstract

Despite research going back over a century showing the U.S. census counts some groups more accurately than others at the national level, little is known about how undercount varies within the country. I focus on a population easily measured with administrative data yet known to suffer high levels of undercount—native-born young children—to document state-level variation in undercount by race and Hispanic status. Although the race-specific analysis is only possible for the 2000 census, the patterns I show for all children are similar to those in 2010, implying the results from 2000 are likely relevant to today. Undercount levels vary widely across states, with non-Black children having the highest rates in the south and southwest, and Black children in the northeast. Results by Hispanic status show non-Black Hispanic young children are highly undercounted in several states with high Hispanic populations, but not all, and are also highly undercounted in the northeast and New England. In several states with high non-Black Hispanic undercounts, non-Black non-Hispanic children are also undercounted at a high rate. I find a very strong correlation between the fraction of births to foreign-born mothers in the state and the undercount of Black and non-Black children—in fact, it is the strongest correlate with the undercount of native-born Black children of those I investigate. The fraction of foreign-born mothers does not correlate with the undercount of non-Black Hispanic and non-Hispanic young children, although Hispanic status of the parents do. My results suggest a group-specific, local focus for future work is needed to determine the causes of census undercount.

Keywords United States Census · Undercount · Children · Immigrants · Race/ethnicity

✉ Janna E. Johnson
   jannaj@umn.edu

1   University of Minnesota, Minneapolis, MN, USA

Introduction

The 2020 census has received more attention, and been the subject of more controversy, than any other United States census in recent memory. Census-related issues—from the proposed addition of a citizenship question to the challenges of collecting data in the midst of the global COVID-19 pandemic—have been ever-present in the headlines. As of December 2020, the controversy around the census is still ongoing: The Supreme Court is now hearing a case proposing to remove unauthorized immigrants from the apportionment counts, the population measurements that re-allocate seats in the House of Representatives (and presidential electoral votes) across states. The potential implications of this ruling cannot be understated and will likely be felt far beyond the next 10 years.

Behind all of this controversy and attention is concern about the ability of the census to accurately count all residents of the United States, commonly referred to as *undercount*. The circumstances around Census 2020 have aroused great concern among demographic experts and policymakers that this census may suffer from unprecedented levels of inaccuracy. Particularly troubling is a possible increase in the undercount of communities of color and the foreign born. These populations are known to have experienced disproportionately high levels of undercount in previous censuses and are likely among the groups most affected by the challenges presented by the pandemic and the current political climate.

Despite the long-standing knowledge that particular groups are missed more frequently in the census at the national level, very little is known about how undercount rates vary locally by race and Hispanic status. For example, Hispanic children in Ohio and Black children in Pennsylvania may be undercounted at a different rate than Hispanic children in New Mexico and Black children in Mississippi. As decennial census counts determine the allocation of over $1.5 trillion annually in federal funding across states and counties (Sullivan, 2020), differences in undercount across geographic areas may leave some communities under-resourced to serve their neediest members, many of whom are children. Indeed, state and local funding allocations for several programs that directly serve low income children, such as the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), Head Start, and the State Children's Health Insurance Program (SCHIP) are determined by local area census counts of young children (Reamer, 2017). The population aged 0–4 has also historically suffered some of the highest census undercount rates of any age group.

In this paper I estimate how the census undercount of young children varies across states by race and Hispanic status. Research on geographic variation in undercount is scarce as the data necessary to form highly accurate subnational population measures independent of the census is difficult to obtain. Besides being of interest due to their outsize role in determining both undercount levels and funding allocations, young children are among the easiest groups to estimate from administrative sources, mostly due to the high accuracy of U.S. birth records. To further increase accuracy, I narrow my focus to native-born children and estimate undercount by state of birth (rather than state of residence) to

remove any reliance on measures of international and internal migration. I am the first to my knowledge to document the substantial variation in undercount rates by state in the 2000 census, and the first to show state-level undercount patterns by race and Hispanic status for any census. As the identification of place of birth relies on information collected on the long form, discontinued after Census 2000, I cannot repeat my analysis for more recent censuses. However, the patterns shown in my 2000 results are very similar to those shown in the single extant similar study of young children in 2010—evidence that such patterns are likely consistent across time and still informative today. I also am the first to correlate state-level undercount patterns with administrative data from birth certificates, showing that for native-born young children, foreign-born status of the mother is strongly associated with higher undercount at the state level, particularly for Black children. Results for Hispanic children suggest caution is needed when using the Hispanic status of the parents to define that of the child, particularly in light of recent work on the fluidity of race and ethnic status (Liebler et al., 2017).

This paper proceeds as follows. The next section provides a short summary of the existing research on census undercount of young children, framing my work within the literature. The following section describes the calculation of the undercount of young children by state of birth, presents results, and tests the robustness of the results to assumptions about net emigration. The penultimate section correlates birth certificate information on mothers with state-level undercounts by race and Hispanic status. The final section concludes.

Background

Young children are one of the highest undercounted age groups in the United States census and those of other countries (O'Hare, 2017), dating back to at least 1850 (Hacker, 2013) through the present day (O'Hare, 2014b, 2015; Robinson et al., 1993, 2002). Despite the prevalence of the phenomenon, the underlying reasons why young children are more often missed than older individuals remain mysterious. To be undercounted, young children are either left off a household's census form while other members are enumerated, or the entire household is missed but only the undercount of young children is evident due to overcount of older age groups. How these two mechanisms play a role in the undercount of children remains unknown. However, children in households with so-called "hard-to-count" characteristics, such as belonging to a minority racial and ethnic group or residing in rental housing, are more likely to be undercounted than children who do not share these characteristics (Bruce & Robinson, 2003).[1]

The two existing studies most similar to mine are O'Hare (2014c) and O'Hare et al. (2016). They both use the demographic analysis (DA) method to estimate state-level undercounts of young children aged 0–4 in the 2010 census. The DA

---

[1] For an excellent review of recent research on the undercount of children in the United States, see O'Hare et al. (2019).

methodology involves creating a measure of the population independent of census counts, usually using births, deaths, and measures of net migration, to estimate undercount.[2] O'Hare (2014c) and O'Hare et al. (2016) use the Census Bureau's Vintage 2010 state population estimates as a benchmark for comparison to the census counts. For children aged 0–4, these population estimates consist of birth records adjusted for deaths and net internal and international migration. O'Hare (2014c) calculates undercount by state for all young children in 2010, and shows this undercount is strongly correlated with the size of the Black and Hispanic populations in each state, as well as with state-level measures of linguistic isolation, low education, and unemployment. O'Hare et al. (2016) focus on Hispanic children only, and show that the undercount of Hispanic children in the 2010 census is concentrated among a small number of states.[3]

This paper expands on this existing work in several important ways. First, despite using slightly different samples of young children, the pattern of undercount across states is remarkably similar across my results and those of the two studies above. As we use two completely different censuses—2000 and 2010—this is evidence that the factors driving these patterns are not unique to a single census and are likely persistent through time. Second, I estimate state-level undercounts for four different racial and ethnic groups: non-Black, Black, non-Black Hispanic, and non-Black non-Hispanic, while previous work only did so for all children and Hispanic children. State-level undercount patterns are quite different across these four groups. Third, unlike previous work, I am able to form uncertainty measures on my undercount estimates and test how assumptions about net emigration affect my results. Finally, I employ multivariate regression analysis to correlate the characteristics of mothers and fathers of young children from birth certificates with state-of-birth level undercounts, providing a more detailed picture of the relationship between the undercount of young children and parents' characteristics than the methods used in previous work.

Undercount of Native-Born Children by State of Birth

Data and Methods

Demographic analysis has been used by the Census Bureau since 1950 to measure the population coverage of the decennial census (Robinson, 2010). DA uses measures of births, deaths, and migration to form a measure of the national population independent of any census count, called the *demographic estimate* (*DE*). This value is then compared to the *census estimate* (*CE*) from the decennial census to calculate the undercount rate of the census:

---

[2] For more details on the Census Bureau's DA methodology for Census 2000, see Robinson (2010).

[3] These states include California, Texas, Florida, Arizona, and New York.

$$\%U = \frac{\mathrm{DE} - \mathrm{CE}}{\mathrm{DE}} \times 100. \tag{1}$$

The undercount rate ($\%U$) is expressed as a percentage. I follow the convention used by the Census Bureau: positive values imply an undercount (the demographic estimate is larger than the census estimate) and negative an overcount (the census estimate exceeds the demographic estimate).

In DA, the demographic estimate is assumed to be the correct measure of the population. However, some components of the demographic estimate are more accurate than others. In particular, three components of net migration have no administrative records: foreign-born emigration, native-born emigration, and "residual foreign-born migration" (including unauthorized and quasi-legal migrants).[4] Measures of these components are based on residual methods, combining estimates from previous censuses with administrative records like death records and legal immigration records.[5] As the estimates of these components are based largely on assumptions rather than administrative records (such as the accuracy of the enumeration of the previous census), it is hard to evaluate their accuracy. By focusing exclusively on native-born young children, I do not need to incorporate potentially inaccurate estimates of the foreign-born. However, children born in the United States may emigrate prior to the census, and these children should therefore be excluded from the demographic estimate. As the United States does not keep records of citizens who leave the country, measuring emigrants is difficult. Given the lack of reliable estimates of emigration, I make the necessary assumption that the emigration of native-born young children is zero. If this assumption is violated, the demographic estimate is biased upward, increasing the likelihood my results show an undercount of native-born children. Later in the paper, I test the implications of this assumption for my results.

Data

Under the assumption of zero out-migration, I only require data on births and deaths to calculate the demographic estimate of native-born young children aged 0–4. Data on births come from the National Center for Health Statistics' (NCHS) Vital Statistics Natality Birth Data. As I calculate undercount for native-born children aged 0–4 on April 1, 2000, I use births that occurred in the United States between April 1, 1995 and March 31, 2000. Each birth record contains information on the child (such as birthweight, sex, month of birth, location of birth, etc.) and the parents (such as age, race, education, place of residence, etc.). In addition to using these records for birth counts by cohort, race/Hispanic status, and state of birth, I use them

---

[4] According to Robinson (2010), the other components of net migration are legal immigration, temporary migrants, civilian citizen migration, and armed forces overseas. Measures of these five components are based on U.S. administrative records.

[5] For specifics on the methods involved, see Mulder et al. (2002) (foreign-born emigration), Gibbs et al. (2002) (native-born emigration), and Constanzo et al. (2002) (residual foreign-born migration). Jensen (2013) provides a summary of all methods employed in estimating emigration.

to compute the state-of-birth-level information on mothers used later in the paper. Registration of births in the United States is assumed to be complete for my cohorts of interest, as the Census Bureau assumes complete registration beginning with the 1985 birth cohort.

I use death data from the NCHS Vital Statistics Multiple Cause-of-Death Mortality Data. The data contain individual records for all deaths occurring in the United States in each calendar year, and include date of death, place of death, state of birth, age, residence, sex, race, and cause of death. I use all deaths occurring between April 1, 1995 and March 31, 2000 of children in my birth cohorts of interest. However, as death records do not contain month of birth, and only month of death and age, a challenge arises in assigning decedents to specific birth cohorts. I describe the procedure I employ to assign deaths to birth cohorts in Appendix 2.

For my census estimate, I use a different source than the Census Bureau. As the Bureau only reports undercount by race and age, they use the 100% census counts for their census estimate. As I measure undercount for the native born by state of birth, I use the 5% PUMS as my census estimate (Ruggles et al., 2020), as place of birth is only recorded on the census long form in 2000. The long form was administered to approximately 16% of the population. However, one can still form full population counts from the 5% PUMS using the provided weights, which are constructed to match the full population counts.

Race and Hispanic Status

One of the largest challenges in demographic analysis is the assignment of race and Hispanic status. In order for DA to be accurate, the values of these characteristics must be consistently defined across multiple sources: birth certificates, death certificates, and census responses. In the case of young children, census responses are not self-responses, but those of a proxy (usually a parent). Race and Hispanic status are similarly recorded by proxy on death certificates. Birth certificates contain this information for both parents (if it's not missing, as discussed below), which must be used to define the race and Hispanic status of the child. Due to the difficulty of ensuring race is defined consistently across these sources, the Census Bureau historically reported DA estimates for two race categories: Black and non-Black, as they did in 2000.[6] Starting in 2010, an additional category was added: Black alone or in combination (Devine et al., 2010).

As the Bureau did in 2000, I also use two race categories in my analysis. I use the same race assignment rule for Black and non-Black as the Census Bureau did in 2000: the father rule. Research done by the Bureau and others showed that this rule produced an estimate of the Black population that most closely matched census estimates (Passel, 1990; Robinson, 2010). This involves assigning the race of the father to the child, and if the father's race is missing, assigning the mother's race.[7] Census

---

[6] Census 2000 was the first census that allowed respondents to select multiple races, but the National Center for Health Statistics did not produce guidelines allowing for recording multiple races on the birth certificate until 2003 (Devine et al., 2010).

[7] 14.81% (2,908,687/19,639,583) births are missing the father's race. Zero births are missing the mother's race.

2000 was the last year this rule was used in DA. With the increase in multiracial births, the Census Bureau developed a file in the early 2000s which linked children aged 0 to 17 with their parents in the 2000 census. This "Kid Link File" was used to establish patterns of race reporting in multiracial households, and the distribution of responses was applied to aggregated birth records to assign race in 2010 (Sink & Jensen, 2019). As this file was not available in 2000, the father rule was used to assign births to the two race categories.

The assignment of Hispanic status to births in 2000 is more complicated than the assignment of race. The Census Bureau did not report DA results by Hispanic status in 2000; they began doing so in 2010 for ages 0–19 only (Devine et al., 2010). Hispanic status was not reported by all states on the birth and death certificates until 1993 and 1997, respectively, preventing the use of DA for the Hispanic population for older birth cohorts. As I focus on young children aged 0–4 in 2000, Hispanic status is available for these cohorts. However, unlike race, Hispanic status is missing on some birth and death records, and the fraction of records missing this characteristic varies by state. In Appendix 3, I show the extent of this variation and describe how I account for missing values when assigning Hispanic status to births and deaths. As I did for race, I use the father rule to assign Hispanic status to births.[8]

The 2000 census was the first census in which a respondent could identify as belonging to multiple races, which could complicate the assignment of race and Hispanic status. To simplify the issue, I use the variable RACESING, generated by IPUMS-USA, to identify a child as Black or non-Black in census data (Ruggles et al., 2020). This variable was created by the IPUMS team to enable the consistent definition of race across earlier censuses that only allowed one race and census 2000 and later. It assigns each multiple-race person to a single race based on age, sex, Hispanic status, and geographic characteristics. Hispanic status is recorded separately from race, and I use the IPUMS variable HISPAN to assign children as Hispanic or non-Hispanic.

The Undercount Rate and Uncertainty

Once I form the demographic and census estimates, I calculate the undercount rate as in Eq. (1). I estimate undercounts for the native-born population aged 0–4 by race (Black and non-Black), Hispanic status,[9] and state of birth. I do not estimate undercount by sex as undercount rates for young children do not vary by sex (O'Hare, 2014a).

---

[8] An alternative Hispanic status assignment rule based on the Census Bureau's procedure for assigning Hispanic status to births in 2010 is also described in Appendix 3. Use of this alternative assignment rule does not substantially impact results, as shown in Appendix Tables 6 and 16 .

[9] I only calculate undercount by Hispanic status for non-Black children due to the small size of the Hispanic Black population.

Typically, undercount rates are reported without any measures of uncertainty.[10] Birth and death records as well as the 100% census counts are assumed to comprise the full population and therefore have no associated standard error. Uncertainty in measures of migration, while suspected to be substantial, is not incorporated in official undercount estimates reported by the Census Bureau. In my case, as my demographic estimate only contains birth and death records of recent birth cohorts, I assume it has no associated error. However, to calculate undercount by state of birth, I use the 5% PUMS for the census estimate. The 2000 5% PUMS is a stratified subsample of the full census sample that received the census long form (about 16% of the population), and therefore has sampling error (U.S. Census Bureau, 2003). The supplied sampling weights can be used to form standard errors and confidence intervals. If the demographic estimate falls outside of the 95% confidence interval of the census estimate, the undercount rate is significantly different from zero at that level of confidence. If not, the undercount rate is statistically indistinguishable from zero.

Results

Undercount by Race

The undercount rate for all native-born children aged 0–4 in the 2000 U.S. census is 5.0%, significant at the 95% level. Undercount of this group by state of birth and race (Black and non-Black) is shown in Fig. 1, and in tabular form in Appendix Table 5. The maps in Fig. 1 show states with statistically significant undercount rates; rates for all states are shown in Appendix Table 5. For non-Black young children (Fig. 1a), the overall national undercount is 4.72, significant at the 95% level. Significant undercount rates range from 1.61 (Pennsylvania) to 10.10 (Hawaii).[11] Twenty-six of the fifty states have statistically significant levels of undercount for native-born non-Black children. States with the highest levels of undercount tend to be those with high Hispanic populations, such as California (9.74%), Arizona (9.52), Texas (8.36), and Florida (6.11). Two exceptions to this are Tennessee (6.30%) and North Dakota (6.12).[12]

The pattern for undercount of Black native-born young children by state of birth in Fig. 1b is quite different. The national overall undercount is higher than for

---

[10] Two examples of studies reporting uncertainty in undercount rates are Robinson et al. (1993), who formally models the uncertainty in undercounts for the 1990 census, and Devine et al. (2012), who applies various assumptions about the range of the components in the demographic estimate to produce a high, medium, and low estimate of undercount for 2010.

[11] State of birth undercount rates not significant at the 95% level range from − 1.47 (Montana) to 5.27 (Delaware). Undercount rates for the District of Columbia are shown in Appendix Table 5 but not in Fig. 1. For both Black and non-Black, the undercount in DC is very high, likely due to recall bias when individuals fill out their census form. A potential explanation is mis-reporting of state of birth for many children born in a DC hospital, as the parents of these children may mistakenly report the state of birth of these children as their state of residence at the time of their birth (such as Virginia or Maryland).

[12] Oregon, not widely known as a high Hispanic state, also has a relatively high undercount of non-Black children: 5.96%. Oregon's population in 2000 was 8% Hispanic, while the percent Hispanic for Tennessee and North Dakota in 2000 was 2.2 and 1.2%, respectively (Guzmán, 2001).

non-Black native-born young children: 6.45%. States with significant undercount range from Louisiana (4.73%) to Massachusetts (24.56). Several states have undercount estimates less than zero, indicating an overcount, with Washington's (− 14.75) and New Mexico's (− 38.73) statistically significant at the 95% level. Fewer states (16 out of 50) have significant undercounts than for non-Black children, although this is unsurprising due to the small Black population in many states leading to very large confidence intervals on the census estimate.[13] States with the highest undercounts are mostly located in the Northeast, with the exception of Florida and Hawaii.[14]

Undercount by Hispanic Status

Figure 2 shows the undercount of native-born non-Black young children in 2000 by state of birth and Hispanic status. Results are shown in tabular form in Appendix Table 6. Undercount for non-Hispanic non-Black young children is shown in Fig. 2a. Total national undercount for this group is 3.51%, and states with significant undercounts range from Pennsylvania (1.72) to Hawaii (12.52). States with statistically insignificant undercounts have values ranging from − 0.89 (Rhode Island) to 5.79 (Wyoming). As with the undercount for all non-Black children, states with the highest undercounts tend to be in the Southwest.

Undercount for Hispanic children is shown in Fig. 2b. The national total for native-born non-Black Hispanic young children is higher than for non-Hispanic and Black young children: 8.97%. States with significant undercount rates range from Colorado (4.44) to Rhode Island (20.78), and those with non-significant rates from Vermont (− 139) to North Dakota (17.45). Six states have significant overcounts of Hispanic young children. The states with the highest Hispanic undercount rates are in the Northeast and Southwest. Several states with high Hispanic undercount rates also have high non-Hispanic undercount rates, such as Arizona and California.

Comparison to 2010 State-Level Undercount

As I am only able to calculate undercount of native-born children for the 2000 census, a natural concern is whether the patterns across states and racial/ethnic groups evident in the 2000 results are unique to that census, or if they are more consistent through time. Additionally, as my results are only for native-born children by state of birth, can they be used to infer the characteristics of undercount for all children and by state of residence? Insight on these questions may be drawn from comparing

---

[13] For example, Vermont has an overcount of − 109.05%, but the demographic and census estimates of Black children aged 0–4 born in Vermont in 2000 are only 210 and 439, respectively, and the 95% confidence interval on the census estimate is (206, 671).

[14] Hawaii has among the highest undercount rates for both Black and non-Black children. This is likely due to the large proportion of mixed-race individuals in Hawaii leading to difficulties in matching race and ethnic classification across administrative records and census responses. Over 39% of young children born in Hawaii were recorded as belonging to 2 or more races in the 2000 census, compared to 5% nationally.

(A) Non-Black (U.S. undercount: 4.72)

(B) Black (U.S. undercount: 6.45)

Note: Undercount calculated as (DE-CE)/DE *100. Positive values indicate undercount, negative overcount. States with values significantly different from zero at 95 percent confidence shown. Values for all states shown in tabular form in Appendix Table 5.

Fig. 1 Census undercount by race and state of birth, native-born children aged 0–4, 2000 census

my results to those reported by O'Hare (2014c). O'Hare reports undercount by state of residence for all young children in the 2010 census, using the Census Bureau's population estimates as the demographic estimate. His results for the 2010 census and mine for the 2000 census are shown in Fig. 3.[15] There are three main differences

---

[15] Values are shown in tabular form in Appendix Table 7.

(A) Non-Hispanic (U.S. undercount: 3.51)
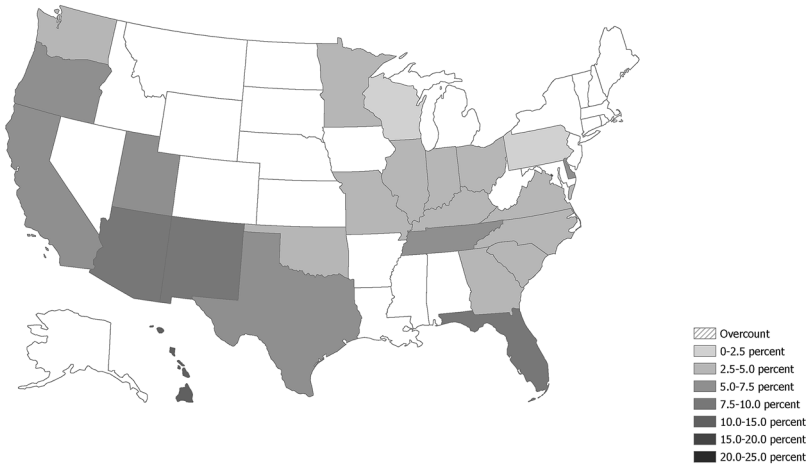
(B) Hispanic (U.S. undercount: 8.97)
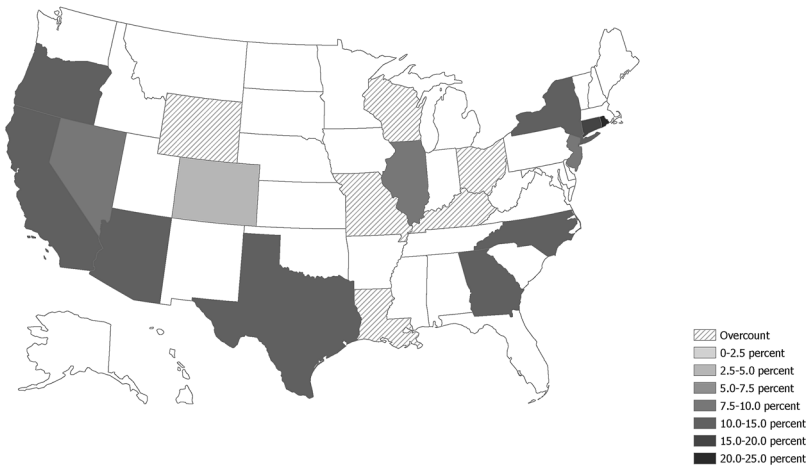
Note: Undercount calculated as (DE-CE)/DE *100. Positive values indicate undercount, negative overcount. States with values significantly different from zero at 95 percent confidence shown. Values for all states shown in tabular form in Appendix Table 6.

Fig. 2 Census undercount by Hispanic status and state of birth, non-Black native-born children aged 0–4, 2000 census
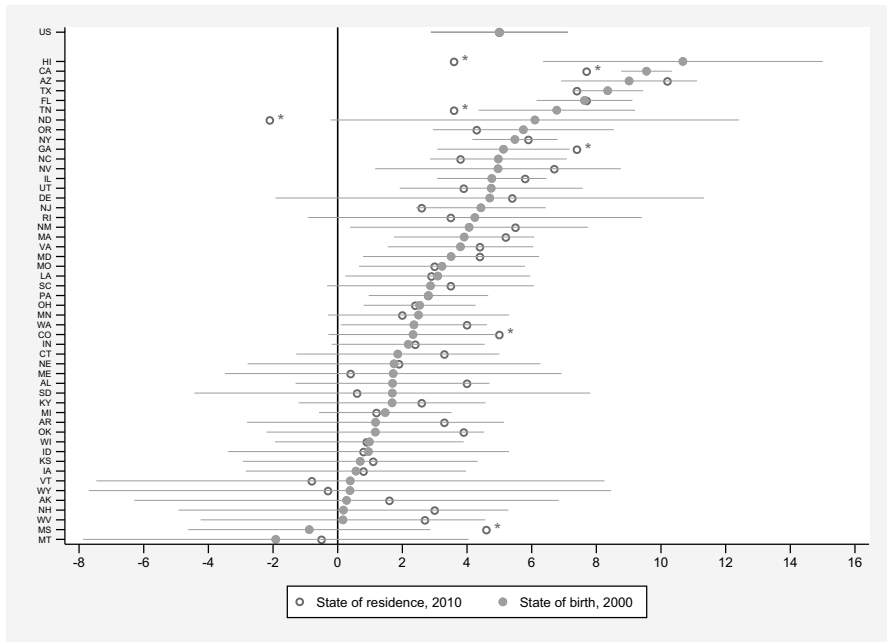
Note: Undercount calculated as (DE-CE)/DE *100. Positive values indicate undercount, negative overcount. U.S. total undercount in both years is 5.0 percent. States sorted in descending order by 2000 undercount values. 2010 values from O'Hare (2014c), who does not report undercount values for the District of Columbia nor statistical significance. 95 percent confidence intervals on 2000 undercounts shown. * indicates 2010 undercount value falls outside of 95% confidence interval of 2000 undercount value. 2010 results include all children aged 0-4 at time of census, 2000 only native-born children aged 0-4 at time of census. Results shown in tabular form in Appendix Table 7.

Fig. 3 Comparison of 2000 and 2010 state-level undercounts, children aged 0–4

between the sample used by me and that used by O'Hare. First, my sample only includes native-born children aged 0–4, while O'Hare's includes all children of these ages.[16] Second, I compute undercount by state of birth, while he calculates it by state of residence. Third, and perhaps most significantly, my results are for the 2000 census while his are for 2010.

There appears to be little difference in undercount of native-born children by state of birth in 2000 and all children by state of residence in 2010. Most states have very similar levels of undercount across the two censuses, and the total undercount rate for the entire country is the same in both years (5.0%). Only seven states have statistically significantly different undercount rates across the two censuses.[17] In terms of numerical difference, 24 states have undercount values less than 1 percentage point

---

[16] As O'Hare is not able to compute undercount by race or ethnic group, my results shown in Fig. 3 and Appendix Table 7 are for all groups combined.

[17] Values are considered statistically significantly different if the estimate reported by O'Hare (who does not report uncertainty values) falls outside of the 95% confidence interval of my 2000 estimate.

apart in 2000 and 2010, and 39 states are less than 2 percentage points apart. The patterns of undercount across states are also quite similar between the two censuses.

Despite the substantial differences between the samples used by me for 2000 and by O'Hare for 2010, both the numerical undercount values and the patterns across states are very similar. Therefore, it appears reasonable to conclude that the patterns and values of undercount for native-born children in 2000 by state of birth are likely evidence of state-level undercount patterns for all young children that are persistent across recent census years.

Emigration

A key assumption in my undercount calculations is that the emigration of native-born young children is zero. Obviously, this is untrue; some children born in the U.S. leave soon after birth. If this emigration is substantial, it could mean my demographic estimates—and therefore my undercount results—are too high. Additionally, if children born in some states emigrate at higher rates than children born in other states, not accounting for this emigration could affect the pattern of undercount across states of birth shown by my results.

I estimate the potential effect of emigration on my results using estimates of emigration rates reported in Van Hook et al. (2006), who estimate emigration circa the year 2000 using the Current Population Survey (CPS). Van Hook and coauthors report estimates for the foreign-born population by age and country of origin. The process for adjusting undercount for emigration is described in Appendix 4. The adjustment is state-of-birth specific, as the fraction of births to foreign-born mothers varies by state, as well as by racial/ethnic group. I use two sets of Van Hook et al. (2006)'s annual emigration rates: the age 25–34 rate (2.7%) and a rate based on national origin: Mexican for non-Black and Hispanic (4.3%), Chinese for non-Hispanic (2.4%), and Caribbean for Black (1.8%). The age 25–34 rate is used as I assume that native-born young children emigrate with their mothers, a large portion of which are likely in this age range. The national origin rates are chosen based on the largest country of origin for foreign-born women aged 15–49 in the 2000 census in the corresponding race/ethnic category. (The birth records only identify Canada, Cuba, and Mexico for foreign-born mothers, with those born in other countries classified as "remainder of world". This is not enough detail to be useful in identifying the place of birth of the mother by race and Hispanic status.) All of these rates are likely upper bound estimates on the emigration of native-born young children of foreign-born mothers, as Van Hook and coauthors report that overall women emigrate at a much lower rate than men (1.7 vs. 4.4%), and all the rates I use are for both genders combined. I only adjust the demographic estimate of native-born children born to foreign-born mothers; I still assume that emigration of native-born children born to native-born mothers is zero.

Selected results from this emigration adjustment are shown in Table 1. Overall U.S. results for the four racial/ethnic groups are shown, as well as the number of states with significant undercounts and the unadjusted and emigration adjusted undercounts for the five states with the highest unadjusted undercount rates in

Table 1 Adjusting state of birth census undercount for emigration, native-born children aged 0–4, 2000

State	Undercount	Undercount adjusted for emigration, age 25–34 rate	Undercount adjusted for emigration, Mexican/Caribbean/ Chinese rate	Fraction births to foreign-born mothers
A. Non-Black				
US total	4.72**	3.14**	2.23**	0.21
Hawaii	10.10**	8.10**	6.94**	0.27
California	9.74**	6.22**	4.12**	0.48
Arizona	9.52**	7.70**	6.65**	0.26
Texas	8.36**	6.24**	5.00**	0.30
Tennessee	6.30**	5.90**	5.68**	0.06
No. of states with significant undercount	27	21	16	
B. Black				
US total	6.45**	5.67**	5.93**	0.11
Massachusetts	24.56**	24.56**	22.82**	0.44
Hawaii	22.01**	22.01**	21.20**	0.19
Rhode Island	20.67**	20.96**	19.26**	0.34
South Dakota	14.12	20.67	13.53	0.17
Florida	12.57**	14.12**	11.47**	0.24
No. of states with significant undercount	17	17	17	
C. Non-Black non-Hispanic				
US total	3.51**	2.40**	2.52**	0.12
Hawaii	12.52**	10.36**	10.60**	0.29
Arizona	8.34**	7.66**	7.74**	0.09
Florida	7.96**	6.41**	6.58**	0.15
New Mexico	7.90**	7.40**	7.45**	0.07
California	7.40**	5.03**	5.29**	0.27
No. of states with significant undercount	25	20	21	

Table 1 (continued)

State	Undercount	Undercount adjusted for emigration, age 25–34 rate	Undercount adjusted for emigration, Mexican/Caribbean/ Chinese rate	Fraction births to foreign-born mothers
D. Non-Black Hispanic				
US total	8.97**	4.74**	2.20**	0.58
Rhode Island	20.78**	16.86**	14.48**	0.65
North Dakota	17.46	16.19	15.45	0.19
Connecticut	15.56**	12.10**	10.03**	0.52
New York	14.26**	10.03**	7.49**	0.61
California	11.92**	7.18**	4.30**	0.65
No. of states with significant undercount	13	9	6	

Five highest undercount states in each category shown for simplicity, excluding District of Columbia. All states plus District of Columbia shown in Appendix Tables 8, 9, 10, and 11. ** indicates 95% confidence interval on undercount estimate does not include zero. Emigration rates are 2.7% for ages 25–34 and 4.3, 1.8, and 2.4% for those originating in Mexico, the Caribbean, and China, respectively. Mexican rate used for non-Black and Hispanic, Caribbean rate for Black, and Chinese rate for Non-Hispanic. These annual net emigration rates are from Van Hook et al. (2006). Undercount accounting for these rates calculated by subtracting emigrants from total demographic estimate. Emigrants calculated using listed rates, compounded annually for each year of life for the fraction of births in each state to foreign-born mothers. For details see Appendix 4. States with significant undercount rates are those for which 95% confidence interval on the undercount estimate does not include zero and the undercount rate is positive. For Black, 2 states have significant overcount (negative undercount) rates using all three demographic estimates, and 1 state has a significant overcount adjusting for emigration using the age 25–34 rate. For non-Hispanic, 1 state has a significant overcount using both the age 25–34 and Chinese rates. Hispanic has the largest number of states with significant overcounts: 7 using the unadjusted estimates, and 11 and 15 using the age 25–34 and Mexican emigration rates, respectively

each category. Results for all states are shown in Appendix Tables 8, 9, 10, and 11. Also shown in Table 1 are the number of states with significant undercount rates before and after adjusting for emigration, as well as the fraction of births in each state to foreign-born mothers. Total U.S. undercount of native-born non-Black children is reduced by approximately 1.5 and 2.5 percentage points when emigration is accounted for using the age 25–34 rate and the Mexican rate, respectively. Of the 27 states with significant unadjusted undercount rates, 21 and 16 remain significant after adjustment using the two different rates, respectively. The reduction for Black children is more modest, from 6.45% to 5.67 and 5.93%, respectively, and all the states with significant unadjusted rates also have significant adjusted undercount rates. For non-Hispanic children, adjusting for emigration reduces undercount by around one percentage point using both rates. Hispanic children are the group with the largest reduction in undercount rates after adjusting for emigration. Adjusting for emigration using the age 25–34 rate cuts the overall national undercount rate by nearly half, and by nearly three-quarters using the Mexican rate. Similarly, the number of states with significant undercount rates falls from 13 to 9 and 6, respectively. In all categories, the states that see the largest reductions in undercount rates are those with the highest proportion of native-born births to foreign-born mothers.

As the emigration rates from Van Hook et al. (2006) are not specifically for children aged 0–4, it is difficult to know whether they accurately estimate the emigration of native-born children. However, recent evidence from Masferrer et al. (2019) uses Mexican census data to estimate approximately 107,000 U.S.-born children aged 0–4 lived in Mexico in 2000. Using the Van Hook et al. rates, I estimate 249,024 native-born young children born to Hispanic mothers emigrated prior to the 2000 census. As the birth records show 72% of births to Hispanic foreign-born mothers are to Mexican-born mothers,[18] over 179,000 of my Hispanic emigrants are Mexican, assuming they emigrate at the same rate. If Masferrer et al. (2019)'s estimates are correct, the method I use over-adjusts for emigration (at least for the Hispanic population), meaning the true undercount value lies between the unadjusted and adjusted values.

In sum, emigration of native-born children of foreign-born mothers may account for high undercount levels of Hispanic, non-Black, and non-Hispanic young children in some states, but rates in large population states with large fractions of births to foreign-born mothers remain high and significant. Emigration of the children of foreign-born Black mothers does not explain the high undercount in states with high foreign-born Black populations.

Correlates with Undercount by State of Birth

In this section, I examine the association between state-level characteristics and the undercount of native-born young children aged 0–4 in the 2000 census. The results of this analysis are not intended to be interpreted as causal evidence of the reasons

---

[18] Mexico is one of the three countries of origin identified for foreign-born mothers on the birth records, hence the ability to estimate the fraction of foreign-born Hispanic mothers born in Mexico.

for the variation in undercount across states. Drawing such conclusions based on correlations without a source of exogenous variation and using relatively sparse data is incorrect at best. Instead, the intent is to provide insights into characteristics associated with the undercount of young children that can potentially provide a foundation for future research into the causes of undercount.

While O'Hare (2014c) examined the correlation between state-level characteristics and the undercount of young children in the 2010 census, my work expands on his in several ways. First, I perform my analysis for each race and Hispanic group separately, while O'Hare only looks at total state undercount. As undercount patterns across states vary substantially by race and Hispanic status, examining the correlates with undercount separately for these groups may provide more insight into potential explanations for the undercount of young children. Second, I correlate state-level undercounts with mothers' characteristics from birth records, while O'Hare (2014c) uses full population characteristics drawn from 2010 census counts and the 2010 American Community Survey (ACS). Birth certificate information on mothers, while imperfect, is likely more reflective of the characteristics of individuals who will respond to the census on behalf of their children than full population census or ACS data. The mothers' characteristics also contain potentially important measures in explaining undercount unexamined by O'Hare, particularly the fraction foreign born. Third, I use OLS regressions to measure the correlation between undercount and the mothers' characteristics, which enables me to consider not only single-variable correlations, as O'Hare does, but also how multiple characteristics combine to explain patterns of undercount across states. These differences enable me to form a more complete, nuanced picture of how the undercount of young children varies by state than previous work.

Data and Methods

To measure the correlation between state-of-birth level undercount and state-level characteristics of mothers, I use the following specification:

$$U_{s,g} = X_{s,g}\beta_g + \varepsilon_{s,g} \tag{2}$$

where $U_{s,g}$ is undercount for native-born children born in state s in racial/ethnic group g (non-Black, Black, non-Black non-Hispanic, and non-Black Hispanic), $X_{s,g}$ is a state-level characteristic of mothers of children born in state s belonging to group g,[19] and $\varepsilon_{s,g}$ is a standard error term. The vector $X_{s,g}$ can consist of up to five variables: the fraction of mothers who are foreign born, who have less than a high school degree, who are unmarried, and the fraction of both mothers and fathers who are Hispanic.[20] These five measures are constructed using the same birth certificate data used to form the demographic estimate of native-born children aged 0–4. As the information reported on birth certificates is limited, these are the only

---

[19] A mother may be of a different racial/ethnic group than her child, as race and Hispanic status is defined using the father rule.

[20] The fraction of fathers who are Hispanic can only be used in the Black/non-Black regressions, as father's Hispanic status determines that of the child.

Table 2 Correlates with state of birth undercount by race and Hispanic status, single variable specifications, native-born children aged 0–4, 2000

	Fraction foreign born mothers	Fraction mothers with less than a high school degree	Fraction unmarried mothers	Fraction Hispanic mothers	Fraction Hispanic fathers
	(1)	(2)	(3)	(4)	(5)
Non-Black	0.178***	0.331***	0.382***	0.147***	0.149***
	(0.018)	(0.030)	(0.084)	(0.012)	(0.012)
R^2	0.670	0.709	0.297	0.767	0.765
Correlation coefficients					
Fraction mothers with less than a high school degree	0.682***				
Fraction unmarried mothers	0.445***	0.524***			
Fraction Hispanic mothers	0.909***	0.856***	0.568***		
Fraction Hispanic fathers	0.906***	0.872***	0.559***	0.996***	
Black	0.275***	− 0.165	− 0.184	0.490***	0.710***
	(0.062)	(0.197)	(0.145)	(0.168)	(0.196)
R^2	0.288	0.014	0.032	0.147	0.211
Correlation coefficients					
Fraction mothers with less than a high school degree	− 0.232				
Fraction unmarried mothers	− 0.316**	0.742***			
Fraction Hispanic mothers	0.802***	− 0.378***	− 0.427***		
Fraction Hispanic fathers	0.877***	− 0.162	− 0.202	0.887***	
Non-Black non-Hispanic	0.112***	0.119	0.114	0.610***	
	(0.041)	(0.092)	(0.085)	(0.097)	
R^2	0.132	0.033	0.035	0.448	
Correlation coefficients					
Fraction mothers with less than a high school degree	− 0.549***				
Fraction unmarried mothers	− 0.454***	0.513***			
Fraction Hispanic mothers	0.748***	− 0.219	− 0.156		

Table 2 (continued)

	Fraction foreign born mothers	Fraction mothers with less than a high school degree	Fraction unmarried mothers	Fraction Hispanic mothers	Fraction Hispanic fathers
	(1)	(2)	(3)	(4)	(5)
Non-Black Hispanic	0.140	0.445***	0.093	0.469***	
	(0.085)	(0.118)	(0.129)	(0.106)	
R^2	0.053	0.226	0.010	0.286	
Correlation coefficients					
Fraction mothers with less than a high school degree	0.039				
Fraction unmarried mothers	− 0.032	− 0.442***	− 0.176		
Fraction Hispanic mothers	0.454***	0.107			
Observations	51	51	51	51	51

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Sample includes 50 states and District of Columbia. Standard errors in parentheses. Estimated using OLS weighted by state population of relevant race/ethic category. Dependent variable in each specification is 2000 undercount by state of birth for children aged 0–4 (expressed as a fraction, not a percent). All specifications include a single independent variable, indicated by the column header. Independent variables calculated from birth certificate data, separately for each race/ethnic category. Race and ethnic status determined using father rule. Cannot include fraction births to Hispanic fathers in non-Hispanic and Hispanic regressions as father's Hispanic status perfectly determines the child's Hispanic status using the father rule

Table 3 Correlates with state of birth undercount by race and Hispanic status, multivariate specifications, native-born children aged 0–4, 2000

	(1)	(2)	(3)	(4)	(5)
A. Non-Black					
Fraction births to foreign-born mothers	0.178***	0.099***	0.096***	0.095**	0.107***
	(0.018)	(0.018)	(0.018)	(0.037)	(0.037)
	[1.00]	[1.87]	[1.91]	[7.47]	[7.87]
Fraction births to mothers with less than a high school degree		0.209***	0.195***	0.193***	0.235***
		(0.033)	(0.035)	(0.052)	(0.059)
		[1.87]	[2.11]	[4.68]	[6.07]
Fraction births to unmarried mothers			0.062	0.062	0.053
			(0.051)	(0.054)	(0.053)
			[1.41]	[1.54]	[1.56]
Fraction births to Hispanic mothers				0.001	0.180
				(0.041)	(0.126)
				[15.82]	[154.27]
Fraction births to Hispanic fathers					− 0.207
					(0.138)
					[177.30]
R^2	0.670	0.821	0.826	0.826	0.835
Adjusted R^2	0.663	0.813	0.815	0.811	0.816
Condition number	2.99	7.31	13.29	21.74	51.76
B. Black					
Fraction births to foreign-born mothers	0.275***	0.276***	0.273***	0.330**	0.289**
	(0.062)	(0.064)	(0.066)	(0.107)	(0.139)
	[1.00]	[1.06]	[1.11]	[2.86]	[4.75]
Fraction births to mothers with less than a high school degree		0.009	0.038	0.009	− 0.006
		(0.174)	(0.255)	(0.260)	(0.264)
		[1.06]	[2.22]	[2.29]	[2.32]
Fraction births to unmarried mothers			− 0.031	− 0.044	− 0.075
			(0.194)	(0.196)	(0.208)
			[2.34]	[2.36]	[2.62]
Fraction births to Hispanic mothers				− 0.191	− 0.337
				(0.282)	(0.422)
				[3.19]	[7.03]
Fraction births to Hispanic fathers					0.283
					(0.605)
					[9.80]
R^2	0.288	0.288	0.289	0.296	0.299
Adjusted R^2	0.274	0.259	0.243	0.235	0.221
Condition number	2.55	10.02	25.35	28.15	30.18

Table 3 (continued)

	(1)	(2)	(3)	(4)	(5)
C. Non-Black Non-Hispanic					
Fraction births to foreign-born mothers	0.112**	0.204***	0.227***	0.024	
	(0.041)	(0.043)	(0.043)	(0.058)	
	[1.00]	[1.43]	[1.52]	[3.91]	
Fraction births to mothers with less than a high school degree		0.356***	0.282**	0.184*	
		(0.092)	(0.094)	(0.082)	
		[1.43]	[1.64]	[1.76]	
Fraction births to unmarried mothers			0.183*	0.114	
			(0.082)	(0.071)	
			[1.44]	[1.51]	
Fraction births to Hispanic mothers				0.641***	
				(0.142)	
				[2.71]	
R^2	0.132	0.340	0.403	0.585	
Adjusted R^2	0.114	0.312	0.364	0.549	
Condition number	2.89	9.89	13.71	17.33	
D. Non-Black Hispanic					
Fraction births to foreign-born mothers	0.140	0.129	0.133	v0.0219	
	(0.085)	(0.075)	(0.069)	(0.060)	
	[1.00]	[1.00]	[1.00]	[1.26]	
Fraction births to mothers with less than a high school degree		0.437***	0.599***	0.587***	
		(0.115)	(0.118)	(0.0920)	
		[1.00]	[1.24]	[1.24]	
Fraction births to unmarried mothers			0.356**	0.434***	
			(0.115)	(0.091)	
			[1.24]	[1.27]	
Fraction births to Hispanic mothers				0.498***	
				(0.088)	
				[1.30]	
R^2	0.053	0.271	0.394	0.642	
Adjusted R^2	0.034	0.241	0.356	0.611	
Condition number	5.62	11.04	16.52	22.01	
Observations	51	51	51	51	

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Sample includes 50 states and District of Columbia. Standard errors in parentheses. Centered variance inflation factors in square brackets. Estimated using OLS weighted by state population of relevant race/ethnic category. Dependent variable is 2000 undercount by state of birth for children aged 0–4 (expressed as a fraction, not a percent), independent variables calculated from birth certificate data, separately for each race/ethnic category. Race and ethnic status determined using father rule. Cannot include fraction births to Hispanic fathers in non-Hispanic and Hispanic regressions as father's Hispanic status perfectly determines the child's Hispanic status using the father rule. Condition numbers calculated including the constant term

characteristics I can examine from this source. The coefficient β_g measures the relationship between the characteristic(s) and undercount for racial/ethnic group g. All variables are measured as fractions. The main specification includes all 50 states and the District of Columbia in the sample, and is weighted by the total state population of group g in the 2000 census. Equation 2 is estimated using OLS for all four ethnic groups g separately.

Results

As an initial analysis, I estimate single variable versions of Eq. 2 for each of the four racial/ethnic groups. Results are shown in Table 2. As all variables are entered as fractions, we can interpret the coefficient for non-Black in column 1 as an increase in foreign-born mothers of 10 percentage points is associated with an increase in undercount of 1.78 percentage points.[21] However, more informative than the coefficients is the R^2 of each regression, which measures the fraction of between-state variation in undercount "explained" by (i.e., associated with) that characteristic. For non-Black children, a very large portion of the between-state variation is correlated with the fraction of foreign-born mothers, mothers with less than a high school degree, and Hispanic mothers and fathers, R^2s of 0.67, 0.71, 0.77, and 0.77, respectively. The fraction of mothers who are unmarried explains a smaller portion with an R^2 of 0.30. However, all of these variables are strongly correlated with one another, as evidenced by the correlation coefficients also shown in Table 2. The fraction Hispanic mothers and fraction Hispanic fathers are extremely strongly correlated (with a ρ of almost 1) and both of these are highly correlated with the fraction of mothers who are foreign-born ($\rho = 0.91$). These high correlations are unsurprising, as the Hispanic population makes up a large proportion of the foreign-born non-Black population. Additionally, higher rates of low education and non-husband/wife households have been shown in previous work to be associated with hard-to-enumerate populations (Bruce & Robinson, 2003).

The pattern is quite different for Black young children. The fraction of mothers who are foreign-born explains the highest proportion of the between-state variation in the undercount of young children for this group, although it is a much lower amount than for non-Black children ($R^2 = 0.29$). The fractions of mothers and fathers who are Hispanic are both positively correlated with undercount, but the fraction of unmarried and low-educated mothers do not significantly explain any of the between-state variation in undercount for Black children.[22] For both non-Black non-Hispanic and non-Black Hispanic children, the fraction Hispanic mothers is most strongly correlated with state-of-birth-level undercount of young children, but in each case this variable measures a slightly different characteristic of the child. For non-Hispanic children, the fraction Hispanic mothers is a measure of the fraction of

---

[21] I report standard errors for the coefficient estimates in Tables 2 and 3 as my undercount estimates use a sample of the full census to measure the population of young children by state of birth. Additionally, it is likely there is measurement error in the information reported on the birth certificates, such as mother's education and marital status.

[22] The strong correlations between Hispanic parentage and foreign-born mothers are likely due to the Black population of Dominican descent, who are Hispanic.

non-Black children with a non-Hispanic father and a Hispanic mother, as Hispanic status is determined using the father rule. (This is also why the fraction Hispanic fathers cannot be included in the Hispanic status regressions, as it is perfectly correlated with Hispanic status using the father rule.) For Hispanic children, this variable measures the fraction of children with two Hispanic parents. Therefore, states with higher fractions of children with a Hispanic mother and a non-Hispanic father tend to have higher undercounts of non-Black non-Hispanic children, and states with a higher proportion of children with two Hispanic parents tend to have higher undercounts of Hispanic young children. The fraction of foreign-born mothers is associated with higher undercount of non-Black non-Hispanic children, but not Hispanic children, while the reverse is true for mothers with less than a high school degree. The fraction of unmarried mothers is not correlated with state-level undercounts of non-Black Hispanic nor non-Hispanic children.

Results of multivariate analyses are shown in Table 3. As several of the independent variables are highly correlated, I report variance inflation factors (VIFs) and condition numbers for all specifications. Many specifications in Table 3 may suffer from large standard errors and unstable coefficient estimates due to multicollinearity issues, so again I focus my discussion on the R^2 of each regression, which is unaffected by multicollinearity. The R^2s show how much of the total between-state variation in undercount is correlated with the variables included in the model. Adjusted R^2s, which account for the increasing number of covariates in each specification, are also shown. For non-Black, including all 5 variables in the model explains over 80% of the between-state variation in the undercount of native-born young children, but the addition of the final 3 variables adds very little explanatory power once the fraction of mothers who are foreign-born and have less than a high school degree are controlled for. Indeed, the VIFs and condition numbers indicate strong multicollinearity when all variables are included.[23] In contrast, the 5 variables together only explain around 30% of the between-state variation in undercount for Black native-born children. Nearly all of this explanatory power is from the fraction of mothers who are foreign-born. For non-Black non-Hispanic children, all variables contribute to the nearly 60% of the explained between-state variation in undercount, although the fraction mothers with less than a high school degree, Hispanic mothers (children with Hispanic mothers and non-Hispanic fathers), and fraction foreign-born mothers are more important in terms of increasing R^2 than fraction unmarried mothers. The fraction of mothers of non-Black Hispanic children who are foreign born doesn't contribute much to the explanatory power of the model of undercount. The total R^2 of 0.64 is mostly due to the fraction of mothers who are Hispanic, have less than a high school degree, and are unmarried.

The appendix contains several robustness analyses for the models in Table 3. To see whether the results are driven by states with very high levels of undercount, the highest undercount states in each of the four groups are excluded from the models in

---

[23] A generally accepted rule of thumb in the literature is a condition number greater than 30 and/or a VIF greater than 10 indicates strong, likely problematic multicollinearity. However, multicollinearity can cause issues even if the VIFs and/or condition numbers are below these thresholds. (Allison, 1999; Kennedy, 2008)

Appendix Table 12. Results are nearly identical, as are those using only states with statistically significant undercount levels (Appendix Table 13). Appendix Table 14 uses the emigration-adjusted undercounts as the dependent variable. As these under-counts are adjusted using the fraction of mothers who are foreign born, the fraction of between-state variation explained by this characteristic falls substantially for the non-Black, non-Black non-Hispanic, and non-Black Hispanic populations. However, the reduction is much smaller for Black children, and the fraction of mothers who are foreign-born still contributes more to the total explanatory power of the model than the other four variables. The R^2s for the full models containing all variables are very similar to those in Table 3. Results using state-level measures of foreign-born, Hispanic, and low-educated women aged 15–49 constructed from 2000 census data, shown in Appendix Table 15, are nearly identical to those using birth certificate data. Results using undercounts using Hispanic status assigned using the middle rule as opposed to the father rule are also very similar (Appendix Table 16).

Conclusion

This paper is the first to calculate census undercount by state of birth, race, and Hispanic status for young children aged 0–4—a population known to suffer from high undercount but easy to measure using administrative records—to gain insight into geographic patterns of undercount by these characteristics. States show a wide range of undercount levels. Non-Black undercount is highest in states in the south and southwest with high Hispanic populations. Undercount for Black native-born children is highest in northeastern states and Florida, states that have a high foreign-born Black population. Non-Black non-Hispanic children have the highest under-count rates among those born in Hawaii, Arizona, and Florida, while the highest rates for non-Black Hispanic children are in the northeastern states of Rhode Island, Connecticut, and New York. Several states, such as California and Texas, have high undercount rates of both Hispanic and non-Hispanic non-Black children. Adjusting for the potential emigration of children of foreign-born mothers may account for the high undercount rates of non-Black, non-Black Hispanic, and non-Black non-Hispanic children born in some states, but does not explain the high undercount of native-born Black children born in states with a high fraction of foreign-born Black mothers.

The strong association between foreign-born mothers and the undercount of Black native-born children is confirmed using OLS regressions of state-of-birth

level undercount rates on mothers' characteristics from birth certificates. The fraction of births to foreign-born mothers has the highest correlation with state-of-birth level undercount of the characteristics examined, although it is highly correlated with fraction of mothers of Black children who are Hispanic (likely due to the large Hispanic Black population from the Caribbean). Fraction of mothers who are Hispanic is also highly correlated with state-of-birth undercounts of non-Black, non-Black Hispanic, and non-Black non-Hispanic children, but fraction foreign-born is not for non-Hispanic and Hispanic children. The fraction Hispanic mothers variable for these two groups actually measures the fraction of mothers of non-Hispanic children and Hispanic children who are Hispanic, as Hispanic status is defined using the father rule. However, using an alternative assignment rule that takes the mother's Hispanic status into account does not affect the results. The results for non-Hispanic children may indicate a disconnect between these assignment rules and the classification of children's Hispanic status by their parents in the census. Those for Hispanic children may indicate a higher likelihood of undercount if both parents are Hispanic.

These correlations should not be considered causal evidence of the underlying reasons for undercount. The relatively sparse information contained on birth certificates combined with no source of exogenous variation makes it highly likely that other, unobserved factors are behind the observed undercount patterns, and the state-level correlations may or may not reflect those at a more local (i.e., city or county) level. Despite these caveats, my results highlight the importance of focusing on the foreign-born population to mitigate coverage issues in the census or other surveys. While much attention has deservedly been paid to increasing the census response of foreign-born non-Black Hispanic individuals, my findings show this should not be done at the expense of other foreign-born populations—such as the foreign-born Black population—as they are also likely to suffer from high levels of undercount. Furthermore, my findings show that to address the undercount of young children, we need to focus on parents, particularly mothers. Determining which mechanism is behind child undercount—missing the child only, but counting the parents, or not counting the entire family—is a priority, as remedies would likely take different forms depending on the underlying cause. More research is needed, perhaps using administrative records like birth certificates matched to individual census responses, to uncover the true causes of undercount and ensure the census counts all residents of the United States.

Appendix 1: Appendix Tables and Figures

See Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, and 16.

Table 4 Birth and death records missing Hispanic status by state, ages 0–4 at 2000 census

Births		Deaths	
State	% missing Hispanic status	State	% missing Hispanic status
Rhode Island	10.86	Oklahoma	30.14
New York	8.18	Rhode Island	23.21
Minnesota	5.93	New York	7.11
Michigan	5.08	North Dakota	5.39
Connecticut	4.44	Maine	3.65
Washington	3.08	Kansas	3.52
New Hampshire	2.67	Illinois	2.98
Vermont	2.50	New Hampshire	2.80
Maine	2.47	Montana	2.55
North Dakota	2.06	Connecticut	2.42
Montana	1.92	Michigan	1.66
Nebraska	1.59	Florida	1.35
Idaho	0.92	Minnesota	1.33
Georgia	0.91	Virginia	1.17
Kansas	0.83	New Jersey	1.11
Oklahoma	0.76	Arizona	1.10
Arizona	0.72	Nevada	1.06
Iowa	0.67	Washington	0.95
Maryland	0.67	Georgia	0.91
Nevada	0.65	District of Columbia	0.78
District of Columbia	0.57	Ohio	0.68
California	0.44	Arkansas	0.65
New Jersey	0.41	Nebraska	0.56
Massachusetts	0.38	Tennessee	0.56
Indiana	0.34	Mississippi	0.48
Pennsylvania	0.29	Kentucky	0.47
Alaska	0.27	Indiana	0.45
Louisiana	0.23	Texas	0.43
Oregon	0.23	Hawaii	0.41
Colorado	0.19	Massachusetts	0.40
Texas	0.16	Louisiana	0.34
Mississippi	0.15	Vermont	0.34
Delaware	0.14	Missouri	0.28
Ohio	0.14	Oregon	0.28
Arkansas	0.12	Pennsylvania	0.27
Hawaii	0.11	Colorado	0.26
South Carolina	0.08	California	0.20
Kentucky	0.07	Iowa	0.20
Utah	0.07	Utah	0.18

Table 4 (continued)

Births		Deaths	
State	% missing Hispanic status	State	% missing Hispanic status
Virginia	0.07	South Carolina	0.15
Florida	0.06	Alabama	0.14
Wyoming	0.06	North Carolina	0.14
Missouri	0.05	Idaho	0.12
West Virginia	0.05	West Virginia	0.09
Illinois	0.04	New Mexico	0.08
Tennessee	0.03	Wisconsin	0.04
Alabama	0.02	Alaska	0.00
North Carolina	0.02	Delaware	0.00
South Dakota	0.02	Maryland	0.00
New Mexico	0.00	South Dakota	0.00
Wisconsin	0.00	Wyoming	0.00
US total	1.23	Total	1.70

Birth records from NCHS Vital Statistics Natality Birth Data, death records from NCHS Vital Statistics Multiple Cause-of-Death Mortality Data. Birth records missing Hispanic status are missing both mother's and father's Hispanic status. Birth record sample includes all births between April 1, 1995 and March 31, 2000, death record sample includes deaths to individuals of those cohorts occurring before April 1, 2000

Table 5 Census undercount by race and state of birth, native-born children aged 0–4, 2000 census

State	Non-Black	Black
Alabama	0.79	3.54
Alaska	0.54	− 3.84
Arizona	9.52**	− 2.57
Arkansas	2.39	− 3.37
California	9.74**	7.45**
Colorado	2.55**	− 1.14
Connecticut	2.25	− 0.57
Delaware	5.27	2.78
District of Columbia	12.63**	20.96**
Florida	6.11**	12.57**
Georgia	4.70**	5.94**
Hawaii	10.10**	22.01**
Idaho	1.11	− 21.17
Illinois	3.82**	8.42**
Indiana	2.50**	− 0.17
Iowa	0.47	2.97
Kansas	0.76	0.06
Kentucky	2.55**	− 6.30

Table 5 (continued)

State	Non-Black	Black
Louisiana	1.95	4.73**
Maine	2.14	− 33.04
Maryland	1.31	8.21**
Massachusetts	1.36	24.56**
Michigan	1.29	2.23
Minnesota	2.77**	− 1.41
Mississippi	1.01	− 2.95
Missouri	2.85**	5.23
Montana	− 1.47	− 54.34
Nebraska	1.83	0.47
Nevada	5.43**	0.34
New Hampshire	0.33	− 10.69
New Jersey	2.53**	12.19**
New Mexico	5.18**	− 38.73**
New York	3.91**	11.10**
North Carolina	3.65**	8.59**
North Dakota	6.12**	5.24
Ohio	2.82**	1.05
Oklahoma	2.15	− 6.55
Oregon	5.96**	− 0.68
Pennsylvania	1.61**	9.18**
Rhode Island	2.68	20.67**
South Carolina	2.97**	2.70
South Dakota	1.48	14.12
Tennessee	6.30**	8.44**
Texas	8.36**	8.27**
Utah	4.88**	− 7.24
Vermont	1.13	− 109.05
Virginia	3.32**	5.26**
Washington	3.40**	− 14.75**
West Virginia	0.76	− 11.88
Wisconsin	1.54	− 4.10
Wyoming	1.32	− 54.83
US total	4.72**	6.45**

Undercount calculated as (DE − CE)/DE *100. Positive values indicate undercount, negative overcount. **indicates DE falls outside 95% confidence interval of CE. Race determined using father rule

Table 6 Census undercount by Hispanic status and state of birth, non-Black native-born children aged 0–4, 2000 census

State	Father rule		Middle rule	
	Non-Hispanic	Hispanic	Non-Hispanic	Hispanic
Alabama	0.90	− 2.52	0.93	− 3.66
Alaska	1.71	− 17.99	0.92	− 4.78
Arizona	8.34**	11.28**	12.56**	4.42**
Arkansas	2.47	1.13	2.65	−1.66
California	7.40	11.92**	6.71**	12.52**
Colorado	1.89	4.44**	2.51	2.68
Connecticut	− 0.31	15.56**	0.67	11.03**
Delaware	6.32**	− 6.02	7.34**	− 20.42
District of Columbia	15.19**	0.75	16.93**	− 9.98
Florida	7.96**	− 0.70	7.39**	1.52
Georgia	3.92**	11.32**	4.30**	8.26**
Hawaii	12.52**	− 7.51	10.26**	9.13
Idaho	0.94	2.20	1.14	0.89
Illinois	2.51**	7.91**	3.64**	4.44**
Indiana	2.82**	− 3.13	2.94**	− 5.27
Iowa	0.26	3.94	0.47	0.46
Kansas	1.71	− 6.40	1.91	− 8.06
Kentucky	3.35**	− 47.27**	3.34**	− 46.13**
Louisiana	2.60	− 16.33**	2.39	− 9.79
Maine	2.34	− 18.36	2.19	− 2.68
Maryland	1.31	1.33	1.27	1.83
Massachusetts	1.61	− 1.03	1.96**	− 4.48
Michigan	1.61	− 3.97	1.59	− 3.57
Minnesota	2.76**	3.02	2.98**	− 1.47
Mississippi	1.24	− 11.55	1.25	− 12.14
Missouri	3.58**	− 20.11**	3.48**	− 16.42**
Montana	− 0.69	− 27.09	− 1.25	− 7.55
Nebraska	1.45	5.35	1.72	2.87
Nevada	3.63	9.16**	5.31**	5.69**
New Hampshire	0.47	− 6.41	0.46	− 5.90
New Jersey	0.82	9.28**	1.16	8.02**
New Mexico	7.90**	2.53	12.44**	− 2.66
New York	0.50	14.26**	2.36**	9.00**
North Carolina	2.92**	10.43**	3.34**	6.68**
North Dakota	5.76	17.45	5.65	20.32
Ohio	3.26**	− 13.24**	3.16**	− 9.20**
Oklahoma	3.11**	− 7.99	3.18**	− 8.76
Oregon	5.01**	10.96**	5.19**	10.08**
Pennsylvania	1.72**	− 0.31	1.81**	− 1.82
Rhode Island	− 0.89	20.78**	0.81	13.24**

Table 6 (continued)

State	Father rule		Middle rule	
	Non-Hispanic	Hispanic	Non-Hispanic	Hispanic
South Carolina	3.68**	− 14.70	3.80**	− 18.23**
South Dakota	2.03	− 23.65	2.03	− 23.31
Tennessee	6.18**	9.40	6.21**	8.65
Texas	5.05**	11.60**	8.19**	8.54**
Utah	5.01**	3.84	4.94**	4.41
Vermont	1.85	− 138.99	1.73	− 95.88**
Virginia	3.47**	1.62	3.97**	− 4.69
Washington	3.86**	0.69	3.96**	0.07
West Virginia	1.01	− 46.55	0.96	− 33.62
Wisconsin	2.48**	− 13.08**	3.11	− 25.98**
Wyoming	5.79	− 39.80**	5.35	− 34.14**
US total	3.51**	8.97**	3.97**	7.37**

Undercount calculated as (DE − CE)/DE *100. Positive values indicate undercount, negative overcount. **indicates DE falls outside 95% confidence interval of CE. Sample includes non-Black only, race for births determined using father rule. Father rule assigns Hispanic status of father to child, and if father's Hispanic status is missing, assigns mother's Hispanic status. For births missing both, Hispanic status assigned using fraction Hispanic by birth cohort and state of birth. Middle rule assigns Hispanic status as follows: (1) if both parents are of same Hispanic status, assign that status to child, (2) if father is Hispanic and mother non-Hispanic, assign 69.8% of these births as Hispanic and the remainder non-Hispanic, (3) if father is non-Hispanic and mother Hispanic,assign 61.4% of these births as Hispanic and the remainder non-Hispanic. If father's Hispanic status missing, assign a status for the father using fraction Hispanic by birth cohort, state of birth, and Hispanic status of mother, followed by applying rules (2) or (3) above. For births missing both mother's and father's Hispanic status, Hispanic status of the child assigned using fraction Hispanic by birth cohort and state of birth

Table 7 Comparison of 2000 and 2010 state-level census undercounts, children aged 0–4

State	2000, state of birth	2010, state of residence	Difference, 2000–2010
Alabama	1.7	4.0	−2.3
Alaska	0.3	1.6	−1.3
Arizona	9.0**	10.2	−1.2
Arkansas	1.2	3.3	−2.1
California	9.6**	7.7	1.9+
Colorado	2.3	5.0	−2.7+
Connecticut	1.9	3.3	−1.4
Delaware	4.7	5.4	−0.7
Florida	7.6**	7.7	−0.1
Georgia	5.1**	7.4	−2.3+
Hawaii	10.7**	3.6	7.1+
Idaho	1.0	0.8	0.2
Illinois	4.8**	5.8	−1.0
Indiana	2.2	2.4	−0.2

Table 7 (continued)

State	2000, state of birth	2010, state of residence	Difference, 2000–2010
Iowa	0.6	0.8	−0.2
Kansas	0.7	1.1	−0.4
Kentucky	1.7	2.6	−0.9
Louisiana	3.1**	2.9	0.2
Maine	1.7	0.4	1.3
Maryland	3.5**	4.4	−0.9
Massachusetts	3.9**	5.2	−1.3
Michigan	1.5	1.2	0.3
Minnesota	2.5	2.0	0.5
Mississippi	−0.9	4.6	−5.5+
Missouri	3.2**	3.0	0.2
Montana	−1.9	−0.5	−1.4
North Carolina	5.0**	3.8	1.2
North Dakota	6.1	−2.1	8.2+
Nebraska	1.7	1.9	−0.2
Nevada	5.0**	6.7	−1.7
New Hampshire	0.2	3.0	−2.8
New Jersey	4.4**.	2.6	1.8
New Mexico	4.1**	5.5	−1.4
New York	5.5**	5.9	−0.4
Ohio	2.5**	2.4	0.1
Oklahoma	1.2	3.9	−2.7
Oregon	5.7**	4.3	1.4
Pennsylvania	2.8**	2.8	0.0
Rhode Island	4.2	3.5	0.7
South Carolina	2.9	3.5	−0.6
South Dakota	1.7	0.6	1.1
Tennessee	6.8**	3.6	3.2+
Texas	8.4**	7.4	1.0
Utah	4.7**	3.9	0.8
Vermont	0.4	−0.8	1.2
Virginia	3.8**	4.4	−0.6
Washington	2.4**	4.0	−1.6
Wisconsin	1.0	0.9	0.1
West Virginia	0.2	2.7	−2.5
Wyoming	0.4	−0.3	0.7
US total	5.0**	5.0	0.0

Undercount calculated as (DE − CE)/DE *100. Positive values indicate undercount, negative overcount. Sample includes all children aged 0–4 at the time of census. 2010 values from O'Hare (2014c), who does not report undercount values for the District of Columbia nor statistical significance. **indicates demographic estimate falls outside 95% confidence interval of census estimate. + indicates 2010 undercount value falls outside of 95% confidence interval of 2000 undercount value. 2010 results include all children, 2000 only native-born children

Table 8 Adjusting state of birth census undercount for emigration, native-born non-Black children aged 0–4, 2000

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Mexican rate	Fraction births to foreign-born mothers
Alabama	0.79	0.40	0.18	0.05
Alaska	0.54	−0.21	−0.63	0.10
Arizona	9.52**	7.70**	6.65**	0.26
Arkansas	2.39	1.91	1.64	0.07
California	9.74**	6.22**	4.12**	0.48
Colorado	2.55**	1.36	0.68	0.17
Connecticut	2.25	0.88	0.10	0.18
Delaware	5.27	4.47	4.02	0.11
District of Columbia	12.63**	10.02**	8.48**	0.38
Florida	6.11**	4.26**	3.19**	0.25
Georgia	4.70**	3.67**	3.09**	0.15
Hawaii	10.10**	8.10**	6.94**	0.27
Idaho	1.11	0.35	−0.08	0.10
Illinois	3.82**	1.91**	0.80	0.26
Indiana	2.50**	2.14**	1.93**	0.05
Iowa	0.47	−0.01	−0.28	0.06
Kansas	0.76	−0.04	−0.49	0.11
Kentucky	2.55**	2.31	2.18	0.03
Louisiana	1.95	1.42	1.12	0.07
Maine	2.14	1.81	1.62	0.04
Maryland	1.31	−0.04	−0.82	0.18
Massachusetts	1.36	0.03	−0.72	0.17
Michigan	1.29	0.60	0.21	0.09
Minnesota	2.77**	2.08	1.68	0.10
Mississippi	1.01	0.71	0.55	0.04
Missouri	2.85**	2.49**	2.29**	0.05
Montana	−1.47	−1.75	−1.90	0.03
Nebraska	1.83	1.20	0.84	0.09
Nevada	5.43**	3.32**	2.10	0.29
New Hampshire	0.33	−0.14	−0.39	0.06
New Jersey	2.53**	0.37	−0.88	0.28
New Mexico	5.18**	4.04**	3.38**	0.15
New York	3.91**	1.57**	0.22	0.30
North Carolina	3.65**	2.78**	2.29**	0.13
North Dakota	6.12**	5.84**	5.68**	0.04
Ohio	2.82**	2.49**	2.31**	0.04
Oklahoma	2.15	1.55	1.22	0.08
Oregon	5.96**	4.73**	4.02**	0.17
Pennsylvania	1.61**	1.03	0.70	0.08
Rhode Island	2.68	1.23	0.40	0.19

Table 8 (continued)

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Mexican rate	Fraction births to foreign-born mothers
South Carolina	2.97**	2.48	2.20	0.07
South Dakota	1.48	1.34	1.26	0.02
Tennessee	6.30**	5.90**	5.68**	0.06
Texas	8.36**	6.24**	5.00**	0.30
Utah	4.88**	4.09**	3.63**	0.11
Vermont	1.13	0.74	0.52	0.05
Virginia	3.32**	2.10**	1.40	0.17
Washington	3.40**	1.96**	1.14	0.19
West Virginia	0.76	0.61	0.53	0.02
Wisconsin	1.54	0.97	0.65	0.08
Wyoming	1.32	0.94	0.72	0.05
US total	4.72**	3.14**	2.23**	0.21

**Indicates 95% confidence interval on undercount estimate does not include zero. Emigration rates are 2.7% for ages 25–34 and 4.3% for those originating in Mexico. These annual net emigration rates are from Van Hook et al. (2006). Undercount accounting for these rates calculated by subtracting emigrants from total demographic estimate. Emigrants calculated using listed rates, compounded annually for each year of life for the fraction of births in each state to foreign-born mothers. For details see Appendix 4

Table 9 Adjusting state of birth census undercount for emigration, native-born Black children aged 0–4, 2000

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Caribbean rate	Fraction births to foreign-born mothers
Alabama	3.54	3.44	3.47	0.01
Alaska	−3.84	− 4.74	− 4.45	0.10
Arizona	−2.57	−3.23	−3.01	0.09
Arkansas	−3.37	−3.44	−3.42	0.01
California	7.45**	6.69**	6.94**	0.10
Colorado	−1.14	−1.88	−1.64	0.10
Connecticut	−0.57	−2.31	−1.74	0.22
Delaware	2.78	2.31	2.47	0.06
District of Columbia	20.96**	20.19**	20.44**	0.13
Florida	12.57**	10.93**	11.47**	0.24
Georgia	5.94**	5.55**	5.68**	0.06
Hawaii	22.01**	20.83**	21.21**	0.19
Idaho	−21.17	−22.31	−21.93	0.12
Illinois	8.42**	8.18**	8.26**	0.03
Indiana	−0.17	−0.31	−0.27	0.02
Iowa	2.97	2.50	2.65	0.06
Kansas	0.06	−0.25	−0.15	0.04
Kentucky	−6.30	−6.50	−6.43	0.03
Louisiana	4.73**	4.62**	4.65**	0.02
Maine	−33.04	−34.54	−34.04	0.18
Maryland	8.21**	7.21**	7.54**	0.15
Massachusetts	24.56**	21.95**	22.82**	0.44
Michigan	2.23	2.08	2.13	0.02
Minnesota	−1.41	−2.71	−2.28	0.19
Mississippi	−2.95	−2.99	−2.98	0.01
Missouri	5.23	5.07	5.12	0.02
Montana	−54.34	−55.03	−54.69	0.05
Nebraska	0.47	0.11	0.23	0.05
Nevada	0.34	−0.31	−0.10	0.09
New Hampshire	−10.69	−12.14	−11.66	0.19
New Jersey	12.19**	10.65**	11.16**	0.23
New Mexico	−38.73**	−39.68**	−39.39**	0.09
New York	11.10**	8.41**	9.30**	0.38
North Carolina	8.59**	8.38**	8.45**	0.03
North Dakota	5.24	3.98	4.36	0.19
Ohio	1.05	0.87	0.93	0.02
Oklahoma	−6.55	−6.89	−6.78	0.04
Oregon	−0.68	−1.33	−1.13	0.09
Pennsylvania	9.18**	8.76**	8.90**	0.06
Rhode Island	20.67**	18.57**	19.26**	0.34

Table 9 (continued)

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Caribbean rate	Fraction births to foreign-born mothers
South Carolina	2.70	2.59	2.63	0.01
South Dakota	14.12	13.23	13.53	0.17
Tennessee	8.44**	8.29**	8.34**	0.02
Texas	8.27**	7.82**	7.97**	0.07
Utah	−7.24	−8.36	−8.00	0.14
Vermont	−109.05	−111.06**	−110.05	0.11
Virginia	5.26**	4.75**	4.92**	0.07
Washington	−14.75**	−15.95**	−15.56**	0.14
West Virginia	−11.88	−12.05	−12.01	0.02
Wisconsin	−4.10	−4.24	−4.19	0.02
Wyoming	−54.83	−55.46	−55.14	0.03
US total	6.45**	5.67**	5.93**	0.11

**Indicates 95% confidence interval on undercount estimate does not include zero. Emigration rates are 2.7% for ages 25–34 and 1.8% for those originating in the Caribbean. These annual net emigration rates are from Van Hook et al. (2006). Undercount accounting for these rates calculated by subtracting emigrants from total demographic estimate. Emigrants calculated using listed rates, compounded annually for each year of life for the fraction of births in each state to foreign-born mothers. For details see Appendix 4

Table 10 Adjusting state of birth census undercount for emigration, native-born non-Hispanic children aged 0–4, 2000

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Chinese rate	Fraction births to foreign-born mothers
Alabama	0.90	0.58	0.61	0.03
Alaska	1.71	1.04	1.11	0.08
Arizona	8.34**	7.66**	7.74**	0.09
Arkansas	2.47	2.21	2.24	0.03
California	7.40**	5.03**	5.29**	0.27
Colorado	1.89	1.17	1.25	0.09
Connecticut	−0.31	−1.83	−1.66	0.16
Delaware	6.32**	5.72**	5.78**	0.06
District of Columbia	15.19**	11.77**	12.15**	0.17
Florida	7.96**	6.41**	6.58**	0.15
Georgia	3.92**	3.09**	3.18**	0.07
Hawaii	12.52**	10.36**	10.61**	0.29
Idaho	0.94	0.55	0.59	0.05
Illinois	2.51**	1.50	1.61**	0.10
Indiana	2.82**	2.55**	2.58**	0.03
Iowa	0.26	−0.09	−0.05	0.04
Kansas	1.71	1.20	1.26	0.06
Kentucky	3.35**	3.13**	3.16**	0.03
Louisiana	2.60	2.07	2.13	0.04
Maine	2.34	1.85	1.90	0.07
Maryland	1.31	−0.30	−0.13	0.14
Massachusetts	1.61	0.12	0.28	0.17
Michigan	1.61	0.64	0.74	0.10
Minnesota	2.76**	1.66	1.78	0.13
Mississippi	1.24	0.95	0.98	0.02
Missouri	3.58**	3.26**	3.30**	0.04
Montana	−0.69	−1.11	−1.06	0.05
Nebraska	1.45	1.01	1.06	0.05
Nevada	3.63	2.61	2.72	0.12
New Hampshire	0.47	−0.14	−0.08	0.08
New Jersey	0.82	−1.23	−1.01	0.20
New Mexico	7.90**	7.40**	7.45**	0.07
New York	0.50	−3.10**	−2.70**	0.30
North Carolina	2.92**	2.38**	2.44**	0.05
North Dakota	5.76	5.37	5.41	0.05
Ohio	3.26**	2.94**	2.97**	0.04
Oklahoma	3.11**	2.71	2.75	0.05
Oregon	5.01**	4.31**	4.39**	0.09
Pennsylvania	1.72**	1.19	1.24	0.06
Rhode Island	−0.89	−2.96	−2.73	0.21

Table 10 (continued)

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Chinese rate	Fraction births to foreign-born mothers
South Carolina	3.68**	3.28**	3.32**	0.03
South Dakota	2.03	1.90	1.91	0.02
Tennessee	6.18**	5.85**	5.89**	0.04
Texas	5.05**	4.06**	4.17**	0.11
Utah	5.01**	4.53**	4.58**	0.07
Vermont	1.85	1.25	1.31	0.07
Virginia	3.47**	2.41**	2.53**	0.11
Washington	3.86**	2.67**	2.80**	0.15
West Virginia	1.01	0.86	0.88	0.02
Wisconsin	2.48**	2.06	2.11	0.05
Wyoming	5.79	5.56	5.58	0.03
US total	3.51**	2.40**	2.52**	0.12

Sample includes non-Black only. **indicates 95% confidence interval on undercount estimate does not include zero. Emigration rates are 2.7% for ages 25–34 and 2.4% for those originating in China. These annual net emigration rates are from Van Hook et al. (2006). Undercount accounting for these rates calculated by subtracting emigrants from total demographic estimate. Emigrants calculated using listed rates, compounded annually for each year of life for the fraction of births in each state to foreign-born mothers. For details see Appendix 4

Table 11 Adjusting state of birth census undercount for emigration, native-born Hispanic children aged 0–4, 2000

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Mexican rate	Fraction births to foreign-born mothers
Alabama	−2.52	−6.85	−9.46	0.62
Alaska	−17.99	−21.46	−23.48	0.37
Arizona	11.28**	7.63**	5.46**	0.53
Arkansas	1.13	−3.49	−6.28	0.66
California	11.92**	7.180**	4.30**	0.65
Colorado	4.44**	1.68	0.05	0.40
Connecticut	15.56**	12.10**	10.03**	0.52
Delaware	−6.02	−11.44	−14.75	0.67
District of Columbia	0.75	−6.87	−11.65	0.93
Florida	−0.70	−6.70**	−10.39**	0.74
Georgia	11.32**	6.34**	3.28	0.78
Hawaii	−7.51	−8.89	−9.68	0.16
Idaho	2.20	−1.56	−3.79	0.48
Illinois	7.91**	2.93	−0.10	0.67
Indiana	−3.13	−6.12	−7.89	0.42
Iowa	3.94	0.20	−2.04	0.52
Kansas	−6.40	−10.53**	−13.00**	0.52
Kentucky	−47.27**	−51.87**	−54.54**	0.46
Louisiana	−16.33**	−20.54**	−23.03**	0.47
Maine	−18.36	−20.60	−21.94	0.23
Maryland	1.33	−4.65	−8.32	0.77
Massachusetts	−1.03	−5.49	−8.16**	0.55
Michigan	−3.97	−6.48	−7.94**	0.34
Minnesota	3.02	−0.58	−2.72	0.53
Mississippi	−11.55	−15.04	−17.06	0.46
Missouri	−20.11**	−23.59**	−25.63**	0.41
Montana	−27.09	−28.06	−28.63	0.09
Nebraska	5.35	1.58	−0.67	0.53
Nevada	9.16**	4.58	1.80	0.66
New Hampshire	−6.41	−9.59	−11.49	0.42
New Jersey	9.28**	4.56**	1.70	0.64
New Mexico	2.53	0.68	−0.39	0.24
New York	14.26**	10.03**	7.49**	0.61
North Carolina	10.43**	5.74	2.86	0.75
North Dakota	17.45	16.19	15.45	0.19
Ohio	−13.24**	−15.49**	−16.78**	0.26
Oklahoma	−7.99	−11.79**	−14.06**	0.48
Oregon	10.96**	6.82**	4.33	0.60
Pennsylvania	−0.31	−3.45	−5.29	0.40
Rhode Island	20.78**	16.86**	14.48**	0.65

Table 11 (continued)

State	Undercount	Undercount accounting for emigration, age 25–34 rate	Undercount accounting for emigration, Mexican rate	Fraction births to foreign-born mothers
South Carolina	−14.70	−19.74**	−22.79**	0.64
South Dakota	−23.65	−25.11	−26.03	0.22
Tennessee	9.40	5.91	3.83	0.56
Texas	11.60**	8.27**	6.29**	0.48
Utah	3.84	0.20	−1.97	0.53
Vermont	−138.99**	−143.59**	−145.16**	0.21
Virginia	1.62	−4.05	−7.52**	0.74
Washington	0.69	−3.74	−6.41**	0.57
West Virginia	−46.55	−50.77	−53.11	0.34
Wisconsin	−13.08**	−16.91**	−19.18**	0.45
Wyoming	−39.80**	−42.22**	−43.62**	0.22
US total	8.97**	4.74**	2.20**	0.58

Sample includes non-Black only. **indicates 95% confidence interval on undercount estimate does not include zero. Emigration rates are 2.7% for ages 25–34 and 4.3% for those originating in Mexico. These annual net emigration rates are from Van Hook et al. (2006). Undercount accounting for these rates calculated by subtracting emigrants from total demographic estimate. Emigrants calculated using listed rates, compounded annually for each year of life for the fraction of births in each state to foreign-born mothers. For details see Appendix 4

Table 12 Correlates with state of birth census undercount by race and Hispanic status, excluding highest undercount states, children aged 0–4, 2000

	(1)	(2)	(3)	(4)	(5)
A. Non-Black					
Fraction births to foreign-born mothers	0.088***	0.084***	0.083***	0.079	0.056
	(0.024)	(0.021)	(0.021)	(0.054)	(0.053)
	[1.00]	[1.00]	[1.02]	[6.26]	[6.54]
Fraction births to mothers with less than a high school degree		0.179***	0.174***	0.172***	0.190***
		(0.049)	(0.052)	(0.059)	(0.057)
		[1.00]	[1.12]	[1.37]	[1.40]
Fraction births to unmarried mothers			0.016	0.012	−0.006
			(0.053)	(0.064)	(0.062)
			[1.13]	[1.63]	[1.67]
Fraction births to Hispanic mothers				0.005	0.322*
				(0.057)	(0.161)
				[7.74]	[68.02]
Fraction births to Hispanic fathers					−0.336**
					(0.161)
					[54.16]
R^2	0.239	0.420	0.421	0.421	0.478
Adjusted R^2	0.222	0.393	0.380	0.365	0.413
Condition number	3.28	9.01	14.38	23.52	46.39
Observations	46	46	46	46	46
B. Black					
Fraction births to foreign-born mothers	0.241***	0.248***	0.245***	0.300***	0.331**
	(0.063)	(0.066)	(0.068)	(0.106)	(0.146)
	[1.00]	[1.06]	[1.09]	[2.67]	[4.91]
Fraction births to mothers with less than a high school degree		0.077	0.121	0.075	0.080
		(0.174)	(0.257)	(0.268)	(0.271)
		[1.06]	[2.27]	[2.42]	[2.43]
Fraction births to unmarried mothers			−0.045	−0.051	−0.025
			(0.195)	(0.197)	(0.215)
			[2.34]	[2.35]	[2.75]
Fraction births to Hispanic mothers				−0.198	−0.103
				(0.293)	(0.425)
				[3.05]	[6.25]
Fraction births to Hispanic fathers					−0.209
					(0.662)
					[9.46]
R-squared	0.244	0.247	0.248	0.256	0.258
Adjusted R-squared	0.227	0.213	0.196	0.185	0.167
Condition number	2.66	11.23	26.82	29.39	31.70
Observations	47	47	47	47	47

Table 12 (continued)

	(1)	(2)	(3)	(4)	(5)
C. Non-Hispanic non-Black					
Fraction births to foreign-born mothers	0.100**	0.192***	0.217***	0.002	
	(0.042)	(0.043)	(0.043)	(0.056)	
	[1.00]	[1.43]	[1.55]	[3.93]	
Fraction births to mothers with less than a high school degree		0.355***	0.290***	0.190**	
		(0.090)	(0.093)	(0.078)	
		[1.43]	[1.63]	[1.75]	
Fraction births to unmarried mothers			0.168*	0.094	
			(0.084)	(0.070)	
			[1.49]	[1.57]	
Fraction births to Hispanic mothers				0.669***	
				(0.135)	
				[2.68]	
R^2	0.111	0.336	0.391	0.608	
Adjusted R^2	0.092	0.308	0.350	0.573	
Condition number	3.12	10.19	15.63	19.18	
Observations	49	49	49	49	
D. Hispanic non-Black					
Fraction births to foreign-born mothers	0.127	0.097	0.102	−0.031	
	(0.086)	(0.066)	(0.067)	(0.058)	
	[1.00]	[1.01]	[1.02]	[1.25]	
Fraction births to mothers with less than a high school degree		0.626***	0.644***	0.624***	
		(0.109)	(0.114)	(0.088)	
		[1.01]	[1.09]	[1.08]	
Fraction births to unmarried mothers			0.089	0.235**	
			(0.143)	(0.114)	
			[1.09]	[1.16]	
Fraction births to Hispanic mothers				0.461***	
				(0.086)	
				[1.34]	
R^2	0.0459	0.455	0.46	0.681	
Adjusted R^2	0.025	0.430	0.422	0.650	
Condition number	5.62	11.42	16.70	21.21	
Observations	47	47	47	47	

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors in parentheses. Centered variance inflation factors in square brackets. Estimated using OLS weighted by state population of relevant race/ethnic category. Dependent variable is 2000 undercount by state of birth, independent variables calculated from birth certificate data, separately for each race/ethnic category. Excluded states are DC, HI, MA, RI (Black), AZ, CA, DC, HI, TX (non-Black), DC and HI (non-Hispanic non-Black), and CT, ND, NY, and RI (Hispanic non-Black). Race and ethnic status determined using father rule. Cannot include fraction births to Hispanic fathers in Hispanic regressions as father's Hispanic status perfectly determines the child's Hispanic status using the father rule. Condition numbers calculated including the constant term

Table 13 Correlates with state of birth census undercount by race and Hispanic status, including only states with statistically significant undercount, children aged 0–4, 2000

	(1)	(2)	(3)	(4)	(5)
A. Non-Black					
Fraction births to foreign-born mothers	0.158***	0.073***	0.070***	0.091*	0.105**
	(0.024)	(0.024)	(0.024)	(0.044)	(0.046)
	[1.00]	[1.94]	[1.98]	[6.38]	[6.88]
Fraction births to mothers with less than a high school degree		0.221***	0.208***	0.240***	0.290***
		(0.043)	(0.046)	(0.074)	(0.086)
		[1.94]	[2.20]	[5.46]	[7.50]
Fraction births to unmarried mothers			0.056	0.067	0.056
			(0.066)	(0.070)	(0.070)
			[1.44]	[1.56]	[1.59]
Fraction births to Hispanic mothers				−0.030	0.133
				(0.053)	(0.154)
				[15.39]	[134.34]
Fraction births to Hispanic fathers					−0.193
					(0.172)
					[162.57]
R^2	0.628	0.823	0.828	0.830	0.840
Adjusted R^2	0.613	0.808	0.806	0.800	0.802
Condition number	3.41	7.14	11.92	20.37	55.21
Observations	27	27	27	27	27
B. Black					
Fraction births to foreign-born mothers	0.186**	0.189**	0.195**	0.221	0.163
	(0.077)	(0.079)	(0.082)	(0.137)	(0.187)
	[1.00]	[1.00]	[2.03]	[2.70]	[4.75]

Table 13 (continued)

	(1)	(2)	(3)	(4)	(5)
Fraction births to mothers with less than a high school degree		0.143	0.024	0.007	0.005
		(0.236)	(0.340)	(0.359)	(0.370)
		[1.00]	[1.99]	[2.07]	[2.07]
Fraction births to unmarried mothers			0.131	0.127	0.076
			(0.265)	(0.274)	(0.302)
			[2.03]	[2.04]	[2.34]
Fraction births to Hispanic mothers				−0.081	−0.276
				(0.349)	(0.552)
				[2.89]	[6.82]
Fraction births to Hispanic fathers					0.371
					(0.796)
					[10.79]
R^2	0.254	0.271	0.283	0.285	0.297
Adjusted R^2	0.210	0.180	0.139	0.081	0.027
Condition number	2.75	9.95	26.52	30.68	33.05
Observations	19	19	19	19	19
C. Non-Hispanic non-Black					
Fraction births to foreign-born mothers	0.171***	0.236***	0.251***	0.135*	
	(0.041)	(0.044)	(0.043)	(0.078)	
	[1.00]	[1.44]	[1.50]	[5.49]	
Fraction births to mothers with less than a high school degree		0.302**	0.234*	0.175	
		(0.113)	(0.113)	(0.113)	
		[1.44]	[1.60]	[1.76]	

Table 13 (continued)

	(1)	(2)	(3)	(4)	(5)
Fraction births to unmarried mothers			0.170*	0.145	
			(0.091)	(0.088)	
			[1.32]	[1.36]	
Fraction births to Hispanic mothers				0.320*	
				(0.182)	
				[4.07]	
R^2	0.427	0.567	0.630	0.679	
Adjusted R^2	0.403	0.528	0.577	0.615	
Condition number	2.90	11.31	13.21	17.35	
Observations	25	25	25	25	
D. Hispanic non-Black					
Fraction births to foreign-born mothers	0.229	0.213	0.060	0.037	
	(0.135)	(0.139)	(0.161)	(0.109)	
	[1.00]	[1.02]	[1.53]	[1.53]	
Fraction births to mothers with with less than a high school degree		0.159	0.646*	0.244	
		(0.211)	(0.355)	(0.255)	
		[1.02]	[3.21]	[3.65]	
Fraction births to unmarried mothers			0.481	0.322	
			(0.290)	(0.198)	
			[3.38]	[3.49]	
Fraction births to Hispanic mothers				0.760***	
				(0.168)	
				[1.25]	
R^2	0.137	0.165	0.288	0.698	

Table 13 (continued)

	(1)	(2)	(3)	(4)	(5)
Adjusted R^2	0.089	0.067	0.154	0.618	
Condition number	6.56	11.08	19.72	28.55	
Observations	20	20	20	20	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sample includes states with significant undercounts for each race/ethic category as indicated in Appendix Tables 5 and 6 . Standard errors in parentheses. Centered variance inflation factors in square brackets. Estimated using OLS weighted by state population of relevant race/ethic category. Dependent variable is 2000 undercount by state of birth, independent variables calculated from birth certificate data, separately for each race/ethnic category. Race and ethnic status determined using father's rule. Cannot include fraction births to Hispanic fathers in Hispanic regressions as father's Hispanic status perfectly determines the child's Hispanic status using the father rule. Condition numbers calculated including the constant term

Table 14 Correlates with state of birth census undercount by race and Hispanic status, emigration-adjusted undercounts, native-born children aged 0–4, 2000

	Age-adjusted undercounts					Country of origin adjusted undercounts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Non-Black										
Fraction births to foreign-born mothers	0.104***	0.023	0.020	0.018	0.031	0.0597***	−0.022	−0.025	−0.029	−0.016
	(0.018)	(0.018)	(0.019)	(0.037)	(0.038)	(0.018)	(0.019)	(0.019)	(0.038)	(0.038)
	[1.00]	[1.87]	[1.91]	[7.47]	[7.87]	[1.00]	[1.87]	[1.91]	[7.47]	[7.87]
Fraction births to mothers with LTHS		0.214***	0.200***	0.198***	0.241***		0.217***	0.203***	0.199***	0.243***
		(0.033)	(0.035)	(0.053)	(0.060)		(0.034)	(0.036)	(0.054)	(0.060)
		[1.87]	[2.11]	[4.68]	[6.07]		[1.87]	[2.11]	[4.68]	[6.07]
Fraction births to unmarried mothers			0.061	0.060	0.051			0.060	0.058	0.049
			(0.051)	(0.054)	(0.054)			(0.052)	(0.055)	(0.055)
			[1.41]	[1.54]	[1.56]			[1.41]	[1.54]	[1.56]
Fraction births to Hispanic mothers				0.003	0.186				0.005	0.193
				(0.042)	(0.128)				(0.042)	(0.129)
				[15.82]	[154.27]				[15.82]	[154.27]
Fraction births to Hispanic fathers					−0.211					−0.217
					(0.140)					(0.141)
					[177.30]					[177.30]
R^2	0.399	0.677	0.686	0.686	0.701	0.177	0.559	0.571	0.571	0.592
Adjusted R^2	0.387	0.663	0.666	0.659	0.668	0.160	0.540	0.543	0.534	0.547
Condition number	2.99	7.31	13.29	21.74	51.76	2.99	7.31	13.29	21.74	51.76
B. Black										
Fraction births to foreign-born mothers	0.206***	0.207***	0.205***	0.262**	0.218	0.229***	0.230***	0.228***	0.284**	0.241*
	(0.062)	(0.065)	(0.067)	(0.108)	(0.140)	(0.062)	(0.064)	(0.067)	(0.108)	(0.140)
	[1.00]	[1.06]	[1.11]	[2.86]	[4.75]	[1.00]	[1.06]	[1.11]	[2.86]	[4.75]

Table 14 (continued)

	Age-adjusted undercounts					Country of origin adjusted undercounts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fraction births to mothers with LTHS		0.011	0.042	0.013	−0.003		0.011	0.041	0.012	−0.004
		(0.176)	(0.257)	(0.263)	(0.267)		(0.175)	(0.257)	(0.262)	(0.266)
		[1.06]	[2.22]	[2.29]	[2.32]		[1.06]	[2.22]	[2.29]	[2.32]
Fraction births to unmarried mothers			−0.033	−0.046	−0.079			−0.032	−0.046	−0.078
			(0.196)	(0.198)	(0.210)			(0.195)	(0.197)	(0.209)
			[2.34]	[2.36]	[2.62]			[2.34]	[2.36]	[2.62]
Fraction births to Hispanic mothers				−0.192	−0.348				−0.191	−0.345
				(0.284)	(0.426)				(0.284)	(0.425)
				[3.19]	[7.03]				[3.19]	[7.03]
Fraction births to Hispanic fathers					0.303					0.297
					(0.610)					(0.608)
					[9.80]					[9.80]
R^2	0.183	0.183	0.184	0.192	0.196	0.218	0.218	0.218	0.226	0.230
Adjusted R^2	0.167	0.149	0.132	0.121	0.107	0.202	0.185	0.168	0.159	0.144
Condition number	2.55	10.02	25.35	28.15	30.18	2.55	10.02	25.35	28.15	30.18
C. Non-Hispanic non-Black										
Fraction births to foreign-born mothers	0.013	0.111**	0.135***	−0.109*		0.024	0.122**	0.145***	−0.095	
	(0.045)	(0.048)	(0.048)	(0.063)		(0.045)	(0.048)	(0.048)	(0.063)	
	[1.00]	[1.43]	[1.52]	[3.91]		[1.00]	[1.43]	[1.52]	[3.91]	
Fraction births to mothers with LTHS		0.381***	0.306***	0.188**			0.379***	0.303***	0.187**	
		(0.102)	(0.106)	(0.090)			(0.101)	(0.105)	(0.090)	
		[1.43]	[1.64]	[1.76]			[1.43]	[1.64]	[1.76]	

Table 14 (continued)

	Age-adjusted undercounts					Country of origin adjusted undercounts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fraction births to unmarried mothers			0.186*	0.104				0.186**	0.105	
			(0.093)	(0.078)				(0.091)	(0.077)	
			[1.44]	[1.51]				[1.44]	[1.51]	
Fraction births to Hispanic mothers				0.768***					0.754***	
				(0.156)					(0.154)	
				[2.71]					[2.71]	
R^2	0.002	0.225	0.286	0.533		0.006	0.230	0.293	0.535	
Adjusted R^2	−0.019	0.193	0.241	0.493		−0.015	0.198	0.247	0.494	
Condition number	2.89	9.89	13.71	17.33		2.89	9.89	13.71	17.33	
D. Hispanic non-Black										
Fraction births to foreign-born mothers	0.065		0.057	−0.099		0.016		0.009	−0.148**	
	(0.087)		(0.071)	(0.063)		(0.089)		(0.072)	(0.064)	
	[1.00]		[1.00]	[1.26]		[1.00]		[1.00]	[1.26]	
Fraction births to mothers with LTHS		0.452***	0.622***	0.610***			0.462***	0.637***	0.625***	
		(0.119)	(0.122)	(0.096)			(0.121)	(0.123)	(0.098)	
		[1.00]	[1.24]	[1.24]			[1.00]	[1.24]	[1.24]	
Fraction births to unmarried mothers			0.374***	0.453***				0.385***	0.464***	
			(0.118)	(0.094)				(0.120)	(0.096)	
			[1.24]	[1.27]				[1.24]	[1.27]	
Fraction births to Hispanic mothers				0.502***					0.504***	
				(0.092)					(0.094)	
				[1.30]					[1.30]	

Table 14 (continued)

	Age-adjusted undercounts					Country of origin adjusted undercounts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
R^2	0.011	0.240	0.374	0.620		0.001	0.233	0.371	0.613	
Adjusted R^2	−0.009	0.209	0.334	0.587		−0.020	0.201	0.331	0.579	
Condition number	5.62	11.04	16.52	22.01		5.62	11.04	16.52	22.01	
Observations	51	51	51	51		51	51	51	51	

$* p < 0.1$, $** p < 0.05$, $*** p < 0.01$. Sample includes 50 states and District of Columbia. Standard errors in parentheses. Centered variance inflation factors in square brackets. Estimated using OLS weighted by state population of relevant race/ethic category. "LTHS" denotes "less than a high school degree." Dependent variable is 2000 undercount by state of birth (expressed as a fraction, not a percent) adjusted for emigration using either the age 25–34 emigration rate (columns 1–5) or the country of origin emigration rate (columns 6–10). This rate is the Mexican rate for non-Black and Hispanic, the Caribbean rate for Black, and the Chinese rate for non-Hispanic. For more details see text and Appendix 4. Independent variables calculated from birth certificate data, separately for each race/ethnic category. Race and ethnic status determined using father rule. Cannot include fraction births to Hispanic fathers in Hispanic regressions as father's Hispanic status perfectly determines the child's Hispanic status using the father rule. Condition numbers calculated including the constant term

Table 15 Correlates with state of birth census undercount by race and Hispanic status, state-level census characteristics, native-born children aged 0–4, 2000

	(1)	(2)	(3)	(4)
A. Non-Black				
Fraction women foreign-born	0.215***	0.129***	0.117***	0.0918**
	(0.025)	(0.022)	(0.029)	(0.040)
	[1.00]	[1.49]	[2.62]	[5.01]
Fraction women with less than a high school degree		0.464***	0.484***	0.406***
		(0.067)	(0.074)	(0.114)
		[1.49]	[1.80]	[4.22]
Fraction women unmarried			0.058	0.041
			(0.089)	(0.091)
			[1.77]	[1.84]
Fraction women Hispanic				0.041
				(0.045)
				[9.16]
R^2	0.604	0.802	0.804	0.808
Adjusted R^2	0.596	0.794	0.792	0.791
Condition number	2.99	14.24	45.69	59.75
B. Black				
Fraction women foreign-born	0.277***	0.276***	0.277***	0.520***
	(0.070)	(0.069)	(0.070)	(0.144)
	[1.00]	[1.00]	[1.00]	[4.49]
Fraction women with less than a high school degree		−0.456	−0.466	−0.652**
		(0.278)	(0.285)	(0.293)
		[1.00]	[1.03]	[1.15]
Fraction women unmarried			0.052	0.132
			(0.237)	(0.235)
			[1.03]	[1.06]
Fraction women Hispanic				−1.265*
				(0.658)
				[4.49]
R^2	0.240	0.280	0.281	0.334
Adjusted R^2	0.224	0.250	0.235	0.276
Condition number	2.54	11.16	28.13	32.43
C. Non-Hispanic non-Black				
Fraction women foreign-born	0.139***	0.216***	0.189***	0.001
	(0.048)	(0.061)	(0.064)	(0.072)
	[1.00]	[1.70]	[1.88]	[3.17]
Fraction women with less than a high school degree		0.416*	0.524**	0.545***
		(0.210)	(0.224)	(0.193)
		[1.70]	[1.95]	[1.95]

Table 15 (continued)

	(1)	(2)	(3)	(4)
Fraction women unmarried			0.202	0.257*
			(0.150)	(0.130)
			[1.70]	[1.72]
Fraction women Hispanic				0.138***
				(0.034)
				[3.17]
R^2	0.144	0.208	0.238	0.444
Adjusted R^2	0.127	0.176	0.189	0.396
Condition number	2.91	17.94	54.57	58.77
D. Hispanic non-Black				
Fraction women foreign-born	0.049	0.090	0.088	0.172***
	(0.083)	(0.074)	(0.069)	(0.059)
	[1.00]	[1.02]	[1.02]	[1.11]
Fraction women LTHS		0.572***	0.721***	0.532***
		(0.141)	(0.145)	(0.125)
		[1.02]	[1.20]	[1.33]
Fraction women unmarried			0.529**	0.527***
			(0.200)	(0.164)
			[1.18]	[1.18]
Fraction women Hispanic				0.310***
				(0.063)
				[1.25]
R^2	0.007	0.260	0.355	0.576
Adjusted R^2	−0.013	0.229	0.314	0.539
Condition number	6.28	10.88	28.91	31.29
Observations	51	51	51	51

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Sample includes 50 states and District of Columbia. Standard errors in parentheses. Centered variance inflation factors in square brackets. Estimated using OLS weighted by state population of relevant race/ethic category. Dependent variable is 2000 undercount by state of birth (expressed as a fraction, not a percent), independent variables calculated from 2000 census data for women aged 15–49, separately for each race/ethnic category. Race and ethnic status of child determined using father rule. Condition numbers calculated including the constant term

Table 16 Correlates with state of birth census undercount by Hispanic status, Hispanic status assigned using middle rule, native-born non-Black children aged 0–4, 2000

	(1)	(2)	(3)	(4)	(5)
A. Non-Hispanic non-Black					
Fraction births to foreign-born mothers	0.120***	0.210***	0.221***	0.062	0.069
	(0.043)	(0.040)	(0.038)	(0.045)	(0.050)
	[1.00]	[1.27]	[1.29]	[2.66]	[3.29]
Fraction births to mothers with less than a high school degree		0.407***	0.308***	0.195**	0.196**
		(0.083)	(0.087)	(0.075)	(0.075)
		[1.27]	[1.56]	[1.72]	[1.72]
Fraction births to unmarried mothers			0.203**	0.090	0.090
			(0.077)	(0.067)	(0.067)
			[1.37]	[1.55]	[1.55]
Fraction births to Hispanic mothers				0.637***	0.667***
				(0.128)	(0.163)
				[2.13]	[3.41]
Fraction births to Hispanic fathers					⁻0.134
					(0.435)
					[3.93]
R^2	0.137	0.423	0.498	0.674	0.674
Adjusted R^2	0.119	0.399	0.466	0.645	0.638
Condition number					
	2.90	9.60	12.80	17.04	18.72
B. Hispanic non-Black					
Fraction births to foreign-born mothers	0.184**	0.167**	0.174**	0.023	⁻0.032
	(0.081)	(0.073)	(0.072)	(0.064)	(0.060)
	[1.00]	[1.00]	[1.01]	[1.27]	[1.36]
Fraction births to mothers with less than a high school degree		0.352***	0.414***	0.402***	0.121
		(0.101)	(0.103)	(0.083)	(0.109)
		[1.00]	[1.11]	[1.11]	[2.40]
Fraction births to unmarried mothers			0.230*	0.337***	0.248***
			(0.120)	(0.098)	(0.092)
			[1.11]	[1.16]	[1.26]
Fraction births to Hispanic mothers				0.630***	0.472***
				(0.121)	(0.117)
				[1.33]	[1.56]
Fraction births to Hispanic fathers					0.530***
					(0.150)
					[2.70]
R^2	0.096	0.278	0.331	0.579	0.671
Adjusted R^2	0.077	0.248	0.288	0.543	0.634
Condition number	5.86	10.04	15.82	29.08	44.76
Observations	51	51	51	51	51

*$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Sample includes 50 states and District of Columbia. Standard errors in parentheses. Centered variance inflation factors in square brackets. Estimated using OLS weighted by state population of relevant race/ethic category. Dependent variable is 2000 undercount by state of birth,

Table 16 (continued)

independent variables calculated from birth certificate data, separately for each race/ethnic category. Race determined using father rule, Hispanic status using middle rule. Middle rule described in notes to Appendix Table 6. Condition numbers calculated including the constant term

Appendix 2: Assigning Deaths to Specific Birth Cohorts

I use death data from the NCHS Vital Statistics Multiple Cause-of-Death Mortality Data. The data contain individual records for all deaths occurring in the United States in each calendar year, and include date of death, place of death, state of birth, age, residence, sex, race, and cause of death. I use all deaths occurring between April 1, 1995 and March 31, 2000 of children in my birth cohorts of interest. As the NCHS Vital Statistics Multiple Cause-of-Death Mortality Data do not contain month of birth, and only month of death and age, a challenge arises in assigning decedents to specific birth cohorts. For example, a child who died at age 2 in June 1999 could either have been born in census year 1996 (and her birthday would fall after her date of death) or in census year 1997 (and her birthday would fall before her date of death). I make several assumptions to assign deaths to specific ages. For deaths of those aged 1–4, I assume the likelihood of birth and death to be independent and uniformly distributed throughout the year, and each month is of equal length.[24] Under these assumptions, I assign the child discussed above a 5/24 probability of being born in census year 1997 (the probability of her being born in April or May 1997 plus the probability of being born in June 1997 and dying after her birthday), and a probability of 19/24 that she was born in census year 1996 (the probability of her birthday falling in July 1996 through March 1997 plus the probability of being born in June 1996 and dying prior to her birthday). I follow the same procedure to assign deaths to each possible cohort by month of death.

The process is simpler for children who die at less than 1 year of age, as age in months is provided for these decedents. For example, a child who dies in June 1999 at less than a year of age is either in the 1998 census cohort if they are 3 or more months of age (born in July 1998–March 1999), or in the 1998 cohort if they are 1 month of age or younger (born in May or June 1999). For those who die at 2 months old, I again assume uniformity and independence of date of birth and death, and therefore assign a probability of 0.5 to belonging to the 1998 or 1999 cohort.

---

[24] These assumptions are of course incorrect, as the seasonality of births and deaths has been well-documented (Udry & Morris, 1967; Rojansky et al., 1992; Bobak & Gjonca, 2001). However, as deaths form a very small proportion of the DA estimate of the population of young children, making these simplifying assumptions are unlikely to affect the overall results.

Appendix 3: Assigning Hispanic Status to Births and Deaths

The assignment of Hispanic status to births and deaths is complicated by the presence of missing Hispanic information, the extent of which varies substantially by state of birth. Appendix Table 4 shows the fraction of birth and death records missing Hispanic status for non-Black births in my sample. Overall, 1.23% of births and 1.70% of deaths to those aged 0–4 at the time of the 2000 census are missing Hispanic status information. Births missing Hispanic status are missing both the mother's and father's Hispanic status. Overall, 11.2 and 1.5% of birth records are missing the father's and mother's Hispanic status, respectively. For births, Rhode Island and New York are missing it on the highest percentage of records: 10.86 and 8.18%, respectively. Two other states are missing Hispanic status on around 5% of birth records, and a further 8 are missing it on between 1 and 5% of records. The remaining 39 states have missing rates of less than 1%. Oklahoma, which did not report Hispanic status on the death certificate at all until 1997, is missing it on over 30% of deaths in my sample, followed by Rhode Island at 23%. New York and North Dakota are next, missing Hispanic status on 7.1 and 5.4% of death records of young children, respectively. However, like birth records, the vast majority of states miss Hispanic status on less than 1% of death records in my sample.

To address the missing Hispanic status problem, I assume that Hispanic status of the child is missing at random, conditional on state and year of birth. For example, if birth records with non-missing Hispanic status born in 1997 in a state are 21% Hispanic, I assign 21% of the children with missing Hispanic status born in the same year and state as Hispanic. I use the same process for death records. This assumption implicitly rules out any correlation between Hispanic status and the likelihood a record is missing Hispanic status, and unfortunately there is no way to test this assumption. However, as the vast majority of states are missing Hispanic status on a very small percentage of birth and death records, violations of this assumption should not make much difference in overall undercount results for these states. However, a few states, like Oklahoma, New York, and Rhode Island, are missing Hispanic status on a high enough percentage of birth and death records that the undercount results by Hispanic status for these states should be viewed with caution.

I use the father rule to assign Hispanic status to births. The father rule for Hispanic status works the same way it did for assigning race: assign the child the Hispanic status of the father, and if the father's Hispanic status is missing, assign that of the mother. Those births missing both mother's and father's Hispanic status are assigned Hispanic or non-Hispanic using the fraction Hispanic by year and state of birth. However, when the Census Bureau began reporting undercount by Hispanic status in 2010, they did not use the father rule to assign Hispanic status to births. Instead, they used the 2000 Kid Link file to estimate how parents reported the Hispanic status of their children when one parent was Hispanic and the other non-Hispanic. According to their calculations, 61.4% of children were reported as Hispanic when the mother was Hispanic and the father non-Hispanic, and 69.8% when the father was Hispanic and the mother non-Hispanic (Devine et al., 2012). As an alternative to the father rule, I use these proportions to assign

Hispanic status to children of one Hispanic and one non-Hispanic parent. Children of two Hispanic or two non-Hispanic parents were assigned the Hispanic status of both parents. Those records missing the Hispanic status of both parents are assigned Hispanic or non-Hispanic using the same procedure as for the father rule, using the fraction Hispanic by year and state of birth.

Appendix 4: Adjusting Undercount for Emigration

This appendix describes the process for adjusting undercount for emigration. The adjustment process assumes native-born young children of native-born mothers do not emigrate, and native-born children of foreign-born mothers emigrate at the same rate as their mothers.

(1) Calculate demographic estimates of native-born children born to foreign-born mothers and native-born children born to native-born mothers, by race/ethnic status (Black, non-Black, non-Black non-Hispanic, non-Black Hispanic), single year of age, and state of birth.

(a) Calculate demographic estimate of all native-born children by single year of age, race/ethnic status, and state of birth
(b) Calculate fraction of native-born children born to foreign-born mothers by single year of age, race/ethnic status, and state of birth using birth records
(c) Multiply the above to get the demographic estimate of native-born children born to foreign-born mothers by single year of age, race/ethnic status, and state of birth. Implicitly assume that the death rate is the same for children of foreign-born and native-born mothers within each state of birth/race-ethnic/age cell, as cannot distinguish deaths by mother's birthplace.
(d) Subtract this number from the original demographic estimate to calculate the demographic estimate of native-born children born to native-born mothers.

(2) Adjust demographic estimate of native-born children born to foreign-born mothers for emigration

(a) Assume native-born children born to foreign-born mothers emigrate at one of two possible two annual rates:

(ii) Age 25–34 rate (2.7%)—this assumes that children emigrate with their mothers, who emigrate at this rate
(iii) A rate based on country of origin of the mother: for non-Black and Hispanic, Mexico (4.3%), for non-Black, China (2.4%), and for Black, the Caribbean (1.8%). These rates are chosen based on the largest country of origin for foreign-born women aged 15–49 in the

2000 census in the corresponding race/ethnic category. (The birth records do not provide enough detail on the country of origin of the mother to be useful.)

(b) Assume in each year of life a child is exposed to one of the above rates, and use the following compounding formula to estimate the demographic estimate of native-born children born to foreign-born mothers adjusted for emigration:

$$\mathrm{DE}^{\mathrm{adj}}_{\mathrm{FB},g} = \sum_{i=0}^{4} \mathrm{DE}_{\mathrm{FB},g,i} \times (1 - \mathrm{rate}_g)^{i+1}, \tag{3}$$

where $\mathrm{DE}_{\mathrm{FB},g,i}$ is the demographic estimate of native-born young children for group g (defined by race/ethnic group) and single year of age i, and rate_g is the chosen annual emigration rate.

(3) Add the demographic estimate of native-born children born to native-born mothers to the emigration-adjusted demographic estimate of native-born children born to foreign-born mothers to get the emigration-adjusted demographic estimate of native-born children

(4) Calculate emigration-adjusted undercount rates using these demographic estimates and the corresponding census estimates using Eq. 1.

References

Allison, P. D. (1999). *Multiple regression: A primer. Pine Forge Press series in research methods and statistics*. Pine Forge Press.

Bobak, M., & Gjonca, A. (2001). The seasonality of live birth is strongly influenced by sociodemographic factors. *Human Reproduction, 16*(7), 1512–7.

Bruce, A., & Robinson, J. G. (2003). *Tract-level planning database with census 2000 census data*. Technical documentation, U.S: Census Bureau.

Constanzo, J. M., Davis, C., Irazi, C., Goodkind, D., & Ramirez, R. (2002). Evaluating components of international migration: The residual foreign born. Working Paper 61, U.S. Census Bureau Population Division.

Devine, J., Bhaskar, R., Desalvo, B., Robinson, J. G., Scopilliti, M., & West, K. K. (2012). The development and sensitivity analysis of the 2010 demographic analysis estimates. Working Paper 93, U.S. Census Bureau Population Division.

Devine, J., Sink, L., Desalvo, B., & Cortes, R. (2010). The use of vital statistics in the 2010 demographic analysis estimates. Working Paper 88, U.S. Census Bureau Population Division.

Gibbs, J. C., Harper, G. S., Rubin, M. J., & Shin, H. B. (2002). Evaluating components of international migration: Native emigrants. Working Paper 63, U.S. Census Bureau Population Division.

Guzmán, B. (2001). The Hispanic population. Census 2000 brief, U.S. Census Bureau.

Hacker, J. D. (2013). New estimates of census coverage in the United States, 1850–1930. *Social Science History, 37*(1), 71–101.

Jensen, E. B. (2013). A review of methods for estimating emigration. Working Paper 101, U.S. Census Bureau Population Division.

Kennedy, P. (2008). *A guide to econometrics*. Blackwell Pub., 6th ed. ed.

Liebler, C. A., Porter, S. R., Fernandez, L. E., Noon, J. M., & Ennis, S. R. (2017). America's churning races: Race and ethnicity response changes between census 2000 and the 2010 census. *Demography, 54*(1), 259–284.

Masferrer, C., Hamilton, E. R., & Denier, N. (2019). Immigrants in their parental homeland: Half a million U.S.-born minors settle throughout Mexico. *Demography*, *56*(4), 1453–1461.

Mulder, T. J., Guzmán, B., & Brittingham, A. (2002). Evaluating components of international migration: Foreign-born emigrants. Working Paper 62, U.S. Census Bureau Population Division.

O'Hare, W., Griffin, D., & Konicki, S. (2019). Investigating the 2010 undercount of young children– summary of recent research. Report from the task force on the undercount of young children, U.S. Census Bureau.

O'Hare, W. P. (2014a). Assessing net coverage for young children in the 2010 U.S. decennial census. *International Journal of Population Research*, *2014*.

O'Hare, W. P. (2014b). Historical examination of net coverage error for children in the U.S. decennial census: 1950 to 2010. Study Series Survey Methodology 2014-03, U.S. Census Bureau Center for Survey Measurement.

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

O'Hare, W. P. (2015). The net undercount of children in the 2010 U.S. decennial census. In M. Hoque, & L. B. Potter (Eds.) *Emerging Techniques in Applied Demography*, chap. 3, (pp. 39–51). Springer.

O'Hare, W. P. (2017). An international perspective on the undercount of young children in the U.S. census. *Statistical Journal of the IAOS*, *33*(2), 289–304.

O'Hare, W. P., Mayol-Garcia, Y., Wildsmith, E., & Torres, A. (2016). The invisible ones: How Latino children are left out of our nation's census count. Report, Child Trends Hispanic Institute and the National Association of Latino Elected Officials.

Passel, J. S. (1990). Demographic analysis: A report on its utility for adjusting the 1990 census. Working paper, The Urban Institute.

Reamer, A. (2017). Counting for Dollars 2020: The Role of the Decennial Cernsus in the Geographic Distribution of Federal Funds. Report, The George Washington University.

Robinson, J. G. (2010). Coverage of population in census 2000 based on demographic analysis: The history behind the numbers. Working Paper 91, U.S. Census Bureau Population Division.

Robinson, J. G., Ahmed, B., Das Gupta, P., & Woodrow, K. A. (1993). Estimation of population coverage in the 1990 united states census based on demographic analysis. *Journal of the American Statistical Association, 88*(423), 1061–79.

Robinson, J. G., West, K. K., & Adlakha, A. (2002). Coverage of the population in Census 2000: Results from demographic analysis. *Population Research and Policy Review, 21,* 19–38.

Rojansky, N., Brzezinski, A., & Schenker, J. G. (1992). Seasonality in human reproduction: An update. *Human Reproduction, 7*(6), 735–45.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2020). IPUMS USA: Version 10.0. https://doi.org/10.18128/D010.V10.0.

Sink, L., & Jensen, E. (2019). *Race of children in population estimates*. Presented at the Annual Meeting of the Southern Demographic Association.

Sullivan, T. A. (2020). *Who's Missing? Undercounting and Underreporting* (pp. 33–47). Springer International Publishing.

Udry, J. R., & Morris, N. M. (1967). Seasonality of coitus and seasonality of birth. *Demography, 4*(2), 673–9.

U.S. Census Bureau (2003). 2000 census of population and housing, public use microdata sample. Technical documentation.

Van Hook, J., Zhang, W., Bean, F. D., & Passel, J. S. (2006). Foreign-born emigration: A new approach and estimates based on matched CPS files. *Demography, 43*(2), 361–382.