JANUARY 15, 2019

Investigating the 2010 Undercount of Young Children – Net Census Coverage of Very Young Children

Howard R. Hogan

United States™ Census Bureau

Intentionally blank

Table of Contents

List of Tables and Figures

1. Introduction

The fact that net undercoverage of young children has been relatively high in censuses compared with older children and adults is well documented (see Coale 1955, Siegel 1974, U.S. Census 2012a, Hogan et al. 2013, and U.S. Census Bureau 2014). Many hypotheses have been offered including the unstable living arrangements precipitated by having young children (e.g., the need to move in with the child's grandparents), time stress because of the care needs of the child, and the possibility that the respondents, including proxies, may not remember the child (U.S. Census Bureau 2014). The younger the child, the more salient most of the hypotheses for omission seem.

Figure 1 plots the measured 2010 Census net coverage by the age of the child using the 2010 Demographic Analysis estimates. This figure shows that the net coverage error generally falls as the age of the child increases, which is consistent with most of the hypotheses of causality. However, the net coverage error also falls for very young children, especially for those under one year of age. This memorandum examines the net coverage by month of birth for children born in 2009 and 2010 to understand why very young children (under 1 year) have lower levels of net undercoverage error compared with other young children (1 to 4 years).

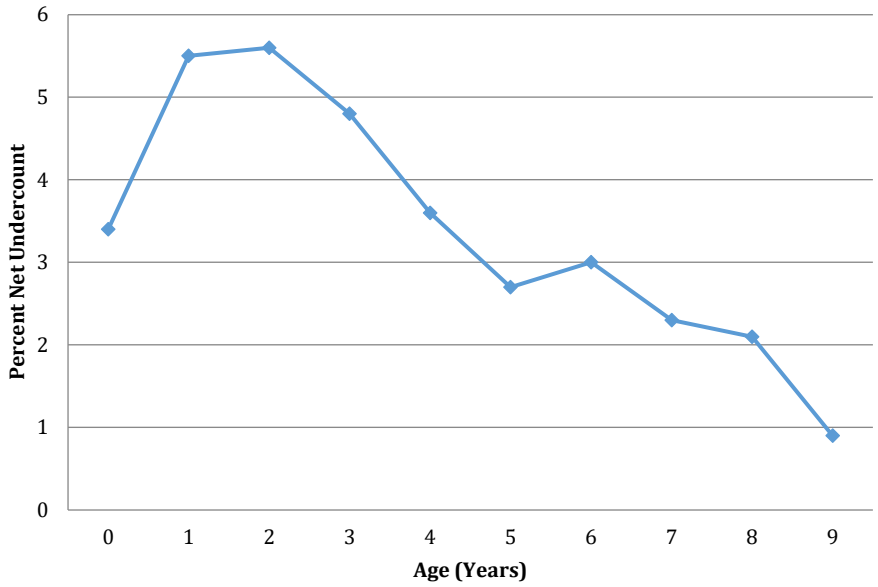

Figure 1. Percent Net Undercount by Year of Age (Ages 0 to 9)
Source: U.S. Census Bureau, 2014

In this report, we compare census counts to counts of births from vital statistics by month. We show that the census counts actually exceeded the vital statistics counts for January and February 2010, implying a net overcount for these months. The 2010 Census records for children age 0 with a year of birth of 2010 and missing a date of birth (DoB) or having an invalid DoB were assigned a DoB for the months of January, February, and March 2010. It is likely that many of these records corresponded to children born after Census Day, children who should not have been included in the 2010 Census count. This helps explain the lower net undercount for children age 0 shown in Figure 1. Without these imputations, children age 0 would have had net coverage errors consistent with, or greater than, those for children ages 1 and 2.

2. Methodology

In this report, we compare counts of births by month from the National Center for Health Statistics to 2010 Census counts to examine the coverage of very young children, those under 1 year. We classify the census counts by whether the DoB was reported or imputed.

2.1 Sources of Data

We obtained census counts from the Census Edited File (CEF). The Decennial Census requests information on the exact DoB for each enumerated person, which enables tabulations of the census population born in 2009 and 2010 by month. Imputation provides month and year of birth values for census enumerations lacking complete reported DoB information. Additionally, some census records correspond to whole-person imputations, instances where all characteristics, including age and DoB, are imputed. The CEF includes detailed information about the source of information for all census records. Imputation flags allow us to determine when age and DoB were reported versus imputed. The specific forms of imputation are also available in the CEF to distinguish between records requiring whole-person imputations and records with some missing data, requiring item imputation of age or DoB.

This research also analyzed results by mode of data collection. Form-type variables on the CEF allowed us to distinguish self-response enumerations from Nonresponse Followup (NRFU) enumerations. We considered all census records from mail returned questionnaires, enumerations of people living in group quarters, and enumerations with updates from the Coverage Followup operation to represent self-response. We categorize the results of Coverage Followup as self-response because the majority of the completed Coverage Followup interviews were for cases that originated from self-responses, rather than NRFU responses (U.S. Census Bureau 2012b). The remaining enumerations came primarily from NRFU and Update/Enumerate. For simplicity, in this document we refer to them as NRFU enumerations.

We obtained the counts of births by calendar month from the National Vital Statistics Reports (National Center for Health Statistics 2011 and 2012).

2.2 Analysis

We compared total births by month from the 2010 Census and vital statistics to reveal important information on the coverage of very young children. Such a comparison ignores early infant deaths as well as net international migration of very young children. The infant mortality rate for the years around 2010 was approximately 6 per 1,000 live births (National Center for Health Statistics 2014). This decrease would be partially offset by a small positive net immigration. After accounting for deaths and net immigration, the Census Bureau's Demographic Analysis estimated the zero-year-old resident population to differ from the registered births by less than 3 per thousand (0.3 percent). For our purposes, this difference was ignored (U.S. Census Bureau 2012c). We calculated differences and coverage ratios.

We used the imputation flags on the CEF to calculate the proportion of the census counts with reported DoB versus imputed DoB. We compared these distributions of reported and imputed DoB by month of birth for the census records.

3. Results

3.1 Month-of-Birth Comparisons of 2010 Census Counts and Vital Statistics Births

Table 1 and Figure 2 present a comparison of the census and vital statistics records measures by month of birth. The ratio of 2010 Census counts to vital statistics births is about 95 percent for most months in 2009, roughly corresponding to a net undercount of 5 percent. The ratio then increases and peaks at around 103 percent for children born in January and February of 2010 and decreases in March and April of 2010. The decennial census measures the population as of Census Day, which is April 1. Only those children born on or before April 1 are eligible to be included in the census.

Table 1. Comparison of 2010 Census Counts by Month of Birth and Vital Statistics on Registered Births, January 2009 - April 1, 2010

Year	Month	2010 Census Count (Census)	Vital Statistics Births (VS)	Difference (VS – Census)	Ratio (100 * Census / VS)
2009	Jan	320,000	338,000	18,000	94.7
2009	Feb	300,000	317,000	17,000	94.6
2009	Mar	328,000	348,000	20,000	94.3
2009	Apr	322,000	337,000	15,000	95.5
2009	May	327,000	345,000	18,000	94.8
2009	Jun	329,000	347,000	18,000	94.8
2009	Jul	346,000	368,000	22,000	94.0
2009	Aug	338,000	360,000	22,000	93.9
2009	Sep	344,000	362,000	18,000	95.0
2009	Oct	330,000	348,000	18,000	94.8
2009	Nov	308,000	320,000	12,000	96.3
2009	Dec	333,000	341,000	8,000	97.7
Subtotal 2009		3,925,000	4,131,000	206,000	95.0
2010	Jan	334,000	323,000	-11,000	103.4
2010	Feb	310,000	302,000	-8,000	102.6
2010	Mar	323,000	339,000	16,000	95.3
2010	Apr[1]	12,000	12,500	500	96.0
Subtotal 2010		979,000	977,000	-2,000	100.2
Subtotal Age 0		3,956,000	4,105,000	149,000	96.4

1. Vital Statistics reduced to 0.039 percent of April estimate to account for April 1 only
Source: National Center for Health Statistics, 2011 and 2012; 2010 Census Edited File – special tabulation

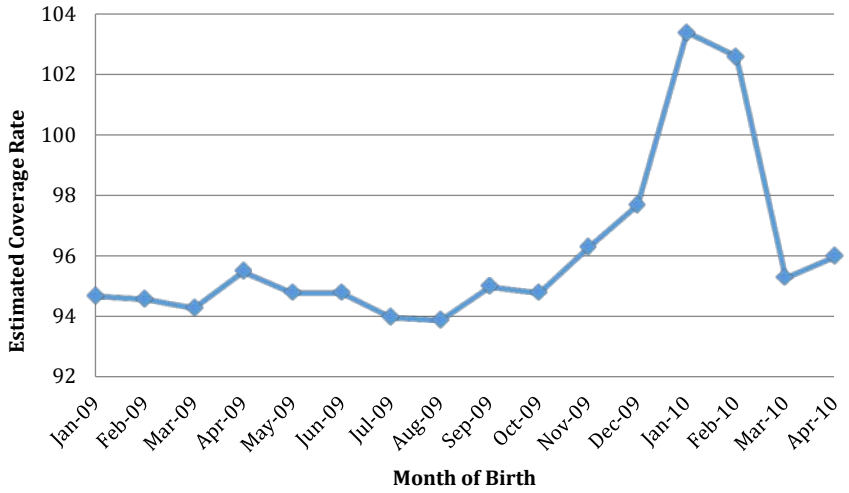

Figure 2. Approximate Coverage Rates by Month of Birth: January 2009 - April 1, 2010
Source: National Center for Health Statistics, 2011 and 2012; 2010 Census Edited File – special tabulation

3.2 Imputations of Age and Date of Birth

How can this trend be explained? Table 2 provides greater detail about the 2010 Census counts of children born in January, February, and March 2010. As a benchmark, Table 2 also includes results for children born in November and December 2009. The data show that of census records tabulated as children born in 2010, only about 83 percent had full DoB reported. The results are very consistent across the three months of January, February, and March. In contrast, about 92 percent of the records for children born in November and December 2009 had full age and DoB reported. While about 6 percent of the children with a final birth month of November or December 2009 required imputation of DoB, that rate rose to 15 percent in January, February, and March 2010. Census records require a minimum amount of data before characteristics, such as DoB, are imputed[1]. When a census record does not include this minimum amount of data, all characteristics are determined via whole-person imputation. Whole-person imputation was very consistent across these five months.

Table 2. Distribution of Age and Date-of-Birth Reporting and Imputation by Month and Year of Birth

Description	November 2009	December 2009	January 2010	February 2010	March 2010
Total Count	308,000	333,000	334,000	310,000	323,000
Full age and DoB reported (%)	91.7	92.0	82.9	83.4	82.2
DoB imputed (%)	6.1	5.8	14.9	14.5	15.5
Whole-person imputation (%)	2.3	2.2	2.1	2.1	2.3

DoB: Date of Birth
Source: 2010 Census Edited File

Table 3 provides greater detail about the form of imputation for the census records with DoB imputed. Close to 10 percent of the census records for children born in 2010 reported an age of 0 and a year of birth of 2010 but had the month imputed, which was imputed as January, February, March, or April (for April, only April 1 is a valid day). For children born in November and December 2009, imputation of month and day of month was an infrequent occurrence (0.4 percent).

Table 3. Distribution of Detailed Type of Imputation of Date of Birth by Month and Year of Birth

Description	November 2009	December 2009	January 2010	February 2010	March 2010
Total Count	308,000	333,000	334,000	310,000	323,000
Full age and DoB reported (%)	91.7	92.0	82.9	83.4	82.2
Reported age and month and year of birth; only day of month imputed (%)	0.5	0.5	0.4	0.5	0.7
Reported age and year of birth; month and day of month imputed (%)	0.4	0.4	9.6	9.2	9.8
Reported age and year of birth; year created from two-digit year, month and day of month imputed (%)	0.2	0.2	0.1	0.1	0.1
Reported age with no DoB information provided; DoB imputed consistent with reported age (%)	2.4	2.3	2.3	2.2	2.3
No age or DoB reported; DoB imputed consistent with allocated age (%)	2.6	2.5	2.5	2.4	2.6
Whole-person imputation (%)	2.3	2.2	2.1	2.1	2.3

DoB: Date of Birth
Source: 2010 Census Edited File

[1] In the 2010 Census if a census record lacked two characteristics (name, relationship, sex, race, Hispanic origin, age or year of birth), then all characteristics were determined by whole-person imputation.

Figure 3 shows the trend of imputed records for children born in 2009 and 2010. It plots the percentage with imputed DoB (blue) and whole-person imputations (red) by final month of birth. Although there seems to be little trend for whole-person imputations, the percentage of records without a reported DoB jumps dramatically in 2010.

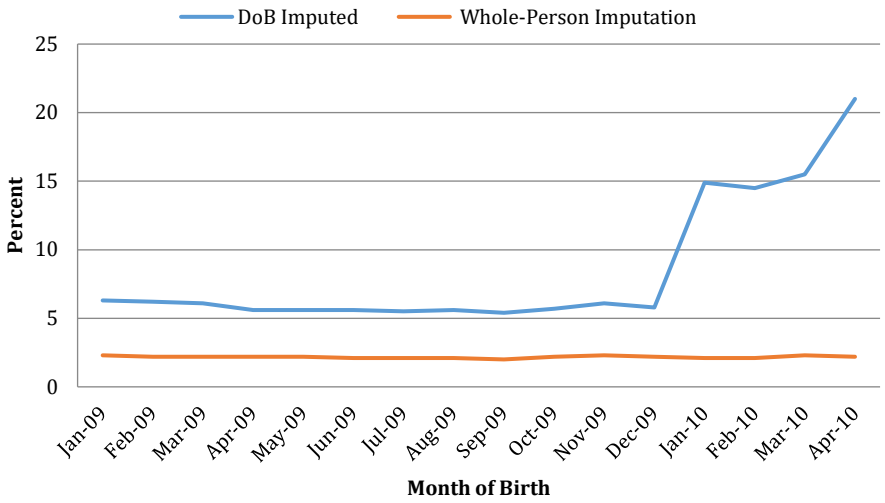

Figure 3. Percent Date of Birth Imputed and Percent Whole-Person Imputation by Month of Birth
DoB: Date of Birth
Source: 2010 Census Edited File

Figure 4 plots the same data in a different way, showing the ratio of 2010 Census records with a reported DoB to those with an imputed DoB (including whole-person imputations). For the 2009 birth year, the ratio is over 10 reported records for every imputed record. For 2010, the ratio falls dramatically to 5 or fewer reported records for every imputed record.

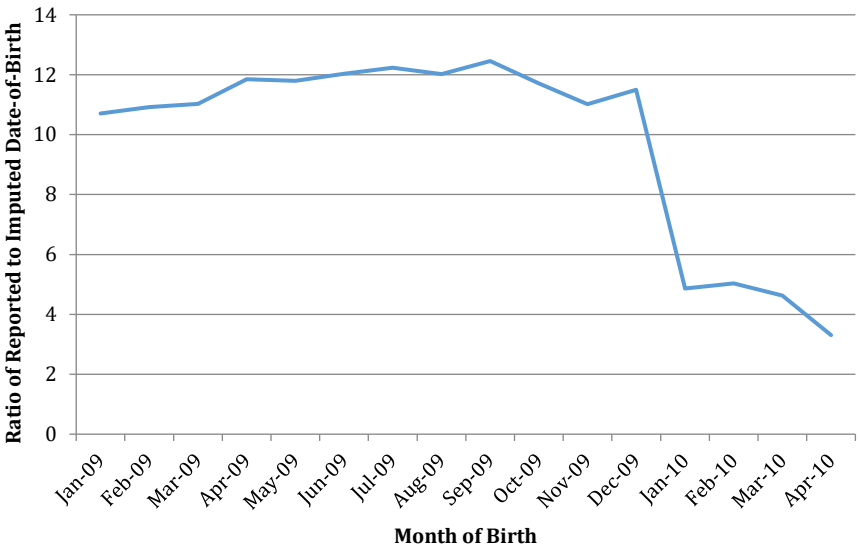

Figure 4. Ratio of Reported to Imputed Date of Birth by Month of Birth 2009-2010
Source: 2010 Census Edited File

The explanation for this is found in Howden (2013):

> *In all, there were 133,543 cases found in the unedited data from the 2010 Census where a date of birth after Census Day was listed…. Of these 133,543 cases, only 8,327 were ultimately dropped from the final Census count, while the remaining 125,216 were included in the final Census count of the total population and edited to reflect a date of birth prior to Census Day.*

While Census Day is April 1, many people responded to the 2010 Census after that date. Respondents were able to return their paper census forms for many months after Census Day. The NRFU operation began in May 2010 and continued through the summer. This lengthy data collection period increased the risk that late respondents could report children who were born after Census Day.

Not all cases with a DoB after Census Day remain age 0 after editing. For example, if DoB is 4/15/2010, reported age is 56, and relationship to head of household is spouse, the record was edited to have a year of birth that matches the reported age of 56. On the other hand, if the DoB is 4/15/2010, reported age is 0, and relationship is child, then that case will most likely have the month edited to be January, February, or March, with the year and age retained. Both self-response and NRFU enumerations with age reported as 0 but without a DoB were presumed to have been born, and allocated a DoB of, April 1 or earlier.

While the overcoverage of children relative to vital statistics for January and February may be because of the inclusion of children born after April 1, 2010, March presents a somewhat different pattern, which is discussed below.

3.3 Imputations of Age and Date of Birth by Mode of Data Collection

If children born after Census Day are included in the 2010 Census with assigned months of birth of January, February, and March 2010, their inclusion masks the true coverage error of children born in these three months. To better understand the errors that might have occurred in 2010, we examined the results by mode of enumeration. Examining the results by mode will allow us to assess if respondents are incorrectly completing forms and including children born after Census Day or if this problem is primarily found in households enumerated during NRFU.

Tables 4 and 5 display the results found in Tables 2 and 3 by data collection mode. On self-response questionnaires (Table 4), over 7 percent of the Census records for children with a year of birth of 2010 reported an age of 0 and a year of birth of 2010 but had an imputed month of birth of January, February, or March. Only 0.2 percent of self-response children born in November or December of 2009 had a year of birth reported as 2009 with month and day imputed. This indicates that substantially more imputation of DoB occurred on self-response questionnaires for children with a final birth month of January, February, or March 2010 compared with children with a final birth month of November and December 2009.

We observe the same pattern for children enumerated during NRFU (Table 5). About 14 percent of census records from NRFU with a birth month of January, February or March 2010 had month and day imputed. The rate was less than 1 percent for children with a birth month of November or

December 2009. While the problem appears greater for people enumerated during NRFU, we also observe potential problems in self-response.

Table 4. Distribution of Age and Date-of-Birth Reporting and Imputation by Month and Year of Birth for Self Response

Description	November 2009	December 2009	January 2010	February 2010	March 2010
Full age and DoB reported	97.1	97.2	90.2	90.6	89.3
Reported age and month/year of birth; only day of month imputed (%)	0.2	0.2	0.2	0.2	0.6
Reported age and year of birth; month and day of month imputed (%)	0.2	0.2	7.2	6.9	7.6
Reported age and year of birth; year created from two-digit year, month and day of month imputed (%)	0.1	0.1	0.1	0.1	0.1
Reported age with no DoB information provided; DoB imputed consistent with reported age (%)	1.5	1.5	1.5	1.5	1.6
No age or DoB reported; DoB imputed consistent with allocated age (%)	0.6	0.5	0.5	0.5	0.6
Whole-person imputation (%)	0.2	0.2	0.2	0.2	0.2

DoB: Date of Birth
Source: 2010 Census Edited File

Table 5. Distribution of Age and Date-of-Birth Reporting and Imputation by Month and Year of Birth for NRFU

Description	November 2009	December 2009	January 2010	February 2010	March 2010
Full age and DoB reported	80.2	80.8	67.6	68.0	68.7
Reported age and month/year of birth; only day of month imputed (%)	1.0	1.0	0.9	1.1	0.9
Reported age and year of birth; month and day of month imputed (%)	0.7	0.6	14.7	14.3	14.0
Reported age and year of birth; year created from two-digit year, month and day of month imputed (%)	0.3	0.2	0.2	0.2	0.2
Reported age with no DoB information provided; DoB imputed consistent with reported age (%)	4.2	4.1	3.9	3.8	3.8
No age or DoB reported; DoB imputed consistent with allocated age (%)	7.0	6.8	6.6	6.6	6.3
Whole-person imputation (%)	6.6	6.5	6.2	6.1	6.1

DoB: Date of Birth
Source: 2010 Census Edited File

Figure 5 summarizes imputation results for children born in 2009 and 2010 by data collection mode. The pattern observed in Figure 3 of high percentages of records with an imputed DoB for children with a year of birth of 2010 holds for both data collection modes. The consistent percentage of whole-person imputations across all months of 2009 and 2010 also holds within each mode. Less than 3 percent of children with a year of birth of 2009 who were enumerated by self-response had their DoB imputed. That rate jumps to almost 10 percent for children with a year of birth of 2010. Similar results are found for children enumerated in NRFU. The children with a year of birth of 2009 had imputation rates of about 13 percent and the children with a year of birth of 2010 had rates over 25 percent.

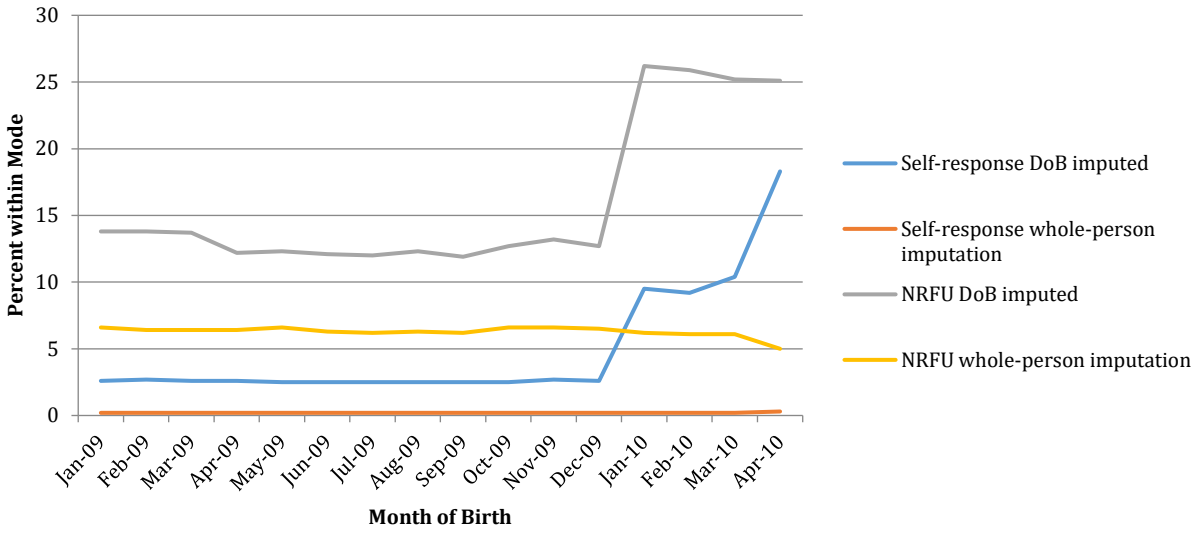

Figure 5. Census Records with an Imputed Month and Day of Birth
Source: 2010 Census Edited File

Figure 6 displays these results in a slightly different way. The stacked bars include the percent of 2010 Census records for each month- and year-of-birth combination with an imputed DoB. The sections of the bar distinguish between mode and type of imputation. The whole-person imputations from self-response and NRFU are very consistent across all months in 2009 and 2010. In combination, they account for about 2 percent of all census records each month. The self-response and NRFU records with an imputed DoB are fairly consistent across the months of 2009 but they both rise dramatically in 2010. We see that both self-response and NRFU contribute to these differences.

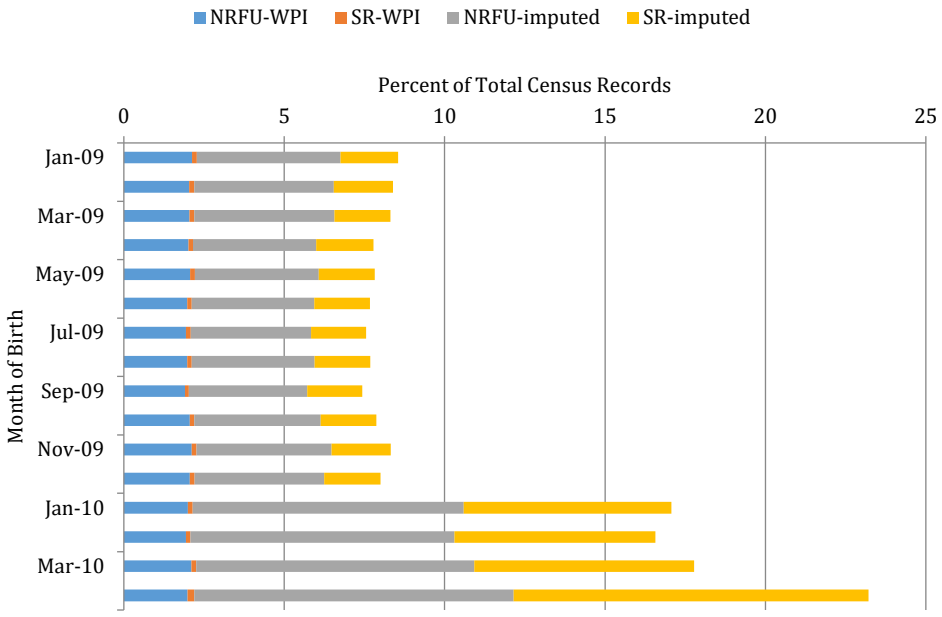

Figure 6. Distribution of Census Records with an Imputed Date of Birth by Type of Imputation and Month of Birth
SR: Self-Response; NRFU: Nonresponse Followup; WPI: Whole-Person Imputation
Source: 2010 Census Edited File

4. Discussion

A focus of this research was to explain why the 2010 Census net undercount for children age 0 was lower than that for children ages 1 to 4. Figure 7 plots first the "implied net coverage rate" including all records with DoB in 2009 and 2010 by month (blue). It shows that net undercount (as approximated here) falls for those born in 2010 turning to a measured overcount. This overcount is likely because of the erroneous inclusion of children born after Census Day for which the census record was edited to reflect a valid DoB in 2010 (i.e., a DoB up to and including April 1).

The plot also shows the implied net coverage error using only those records with reported DoB (red). This measure approximates, or at least better tracks, the gross omission rate. This shows that the coverage worsens for the youngest (less than 3 months old) children. This trend is more consistent with many of the hypothesis offered for the overall coverage pattern for all children.

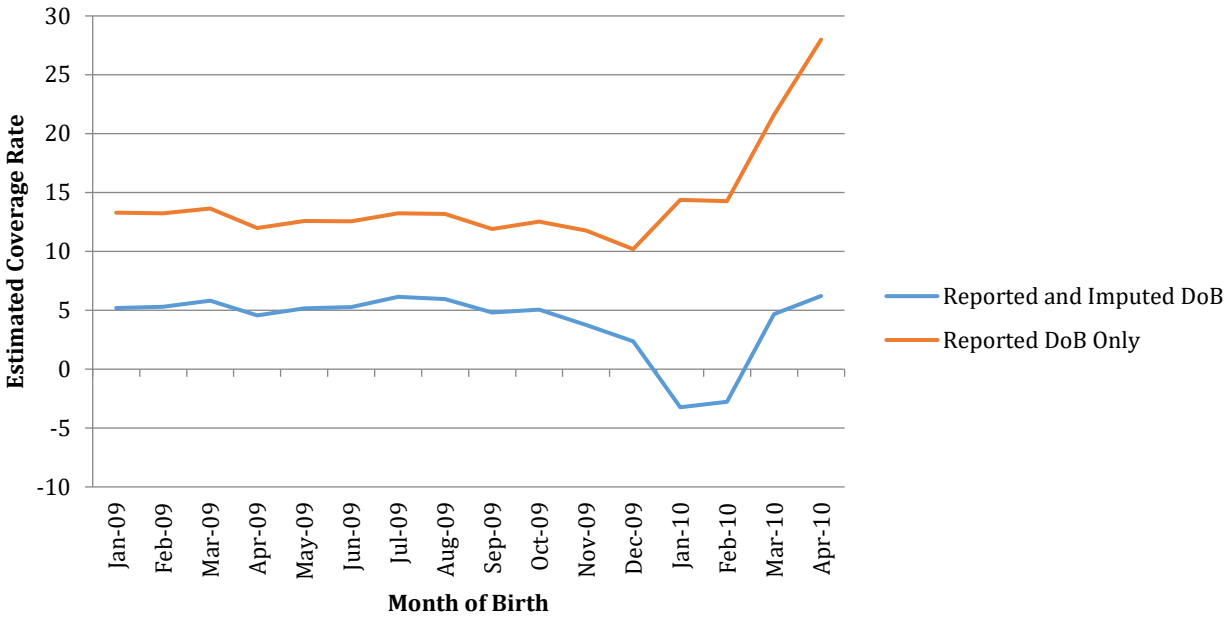

Figure 7. Percent Difference between Vital Statistics and 2010 Census Records with Only Reported Date of Birth and with Reported and Imputed Date of Birth
Source: National Center for Health Statistics, 2011 and 2012; 2010 Census Edited File – special tabulation

Figure 7 presents us with a reasonable hypothesis to explain the relative pattern for March (and April 1). It seems likely that, consistent with many proffered explanations for the undercoverage of young children, those born just before or on Census Day are the most likely to be missed. The omissions rate for these infants is so large that the inclusion of children born after Census Day and attributed to these months still leaves a numeric shortfall.

An implication of this analysis is that if these erroneous inclusions of children with an imputed DoB of 2010 were not included in the census count, all else being equal, the net undercount of these very young children would increase. Using such an editing approach that does not allow these erroneous inclusions would actually decrease the total number of gross errors. However, it could be perceived as worsening the census count, as the net undercount would likely increase.

5. References

Coale, A. J. (1955). "The Population of the United States in 1950 Classified by Age, Sex, and Color--A Revision of Census Figures," Journal of the American Statistical Association, Vol. 50, No. 269 (Mar., 1955), pp. 16-54.

Hogan, H., Cantwell, P. J., Devine, J., Mule, V. T., and Velkoff, V. (2013). Quality and the 2010 Census, Popul Res Policy Rev DOI 10.1007/s11113-013-9278-5.

Howden, L. M. (2013). "Research Note: Babies Born After Census Day: How the Census Bureau Addressed Dates of Birth After Census Day in the 2010 Census," Popul Res Policy Rev (2013) 32:791–801.

National Center for Health Statistics (2011). National Vital Statistics Report, Volume 61, Number 1, "Births: Final Data for 2009."

National Center for Health Statistics (2012) National Vital Statistics Report, Volume 61, Number 1, "Births: Final Data for 2010."

National Center for Health Statistics (2014). QuickStats: Infant Mortality Rates, by Race and Hispanic Ethnicity of Mother — United States, 2000, 2005, and 2010, Morbidity and Mortality Weekly Report (MMWR), Retrieved on May 3, 2017, from: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6301a9.htm.

Siegel, J. S. (1974). "Estimates of Coverage of the Population by Sex, Race, and Age in the 1970 Census," Demography, Vol. 11, No. 1 (Feb., 1974), pp. 1-23.

U.S. Census Bureau (2012a). Alternative Demographic Analysis Estimates of the Undercount by Race, Retrieved from http://www.census.gov/popest/research/da-estimates/Table_3.pdf.

U.S. Census Bureau (2012b). 2010 Census Coverage Followup Assessment Report. 2010 Census Planning Memoranda Series No. 197.

U.S. Census Bureau (2012c). Internal Census file, DA_revised2012_baic.xls.

U. S. Census Bureau (2014). "The Undercount of Young Children," Retrieved on April 18, 2017 from https://www.census.gov/content/dam/Census/library/working-papers/2014/demo/2014-undercount-children.pdf.