

Comparing the Accuracy of the 2020 Census Counts to Population Estimates Program Blended Base for Age Groups of Children

By

William P. O'Hare

Introduction

Following the 2020 Census, the Census Bureau implemented a new method for developing the population base used for the post-2020 Census estimates (U.S. Census Bureau 2021a). The new method is commonly referred to as the PEP (Population Estimates Program) blended base.

The accuracy of the PEP blended base is important for young children because it may affect the allocation of federal funds, as well as management of various programs serving young children. The PEP blended base will not impact data for political redistricting.

One thing that has hampered evaluation of the PEP blended base estimates is the dearth of age data from the 2020 Census. As of November 2022, the only age group identified so far in the 2020 Census is the population ages 18 and over (one can derive the population ages 0 to 17 by subtracting the ages 18 and over from the total population).

This study develops estimates of the 2020 Census counts for three age groups of children (ages 0 to 4, ages 5 to 13 and ages 14 to 17)¹ based on data the Census

¹ The Census Bureau calls anyone who has not celebrated their first birthday age 0.

Bureau has already released along with some simple algebra. These estimated Census counts for the three age groups are compared to the DA middle series estimates to detect differences. The PEP blended base estimates are also compared to the DA middle series estimates for the same three age groups to assess how well the PEP blended base corrects for the high net undercount of young children in the 2020 Census.

Since the Demographic Analysis (DA) estimates are one of the prime benchmarks used by the Census Bureau (2020 and 2022a) to assess Census accuracy, differences between DA and the Census counts as well as DA and the PEP blended base can be seen as errors. O'Hare et al. (2016) show DA is a particularly useful benchmark for assessing census accuracy for young children. This study compares the size of errors in 2020 Census counts to size of errors in the PEP blended base when compared to the DA middle series.

Analysis shows the PEP blended base does a good job of correcting for the high net undercount of young children (age 0 to 4) in the 2020 Census and the new methodology provides some corrections for the population ages 5 to 13. However, there is little difference between the Census counts and the PEP estimates for ages 14 to 17.

Background

For several decades, the Census Bureau has produced yearly population estimates for states and counties in the decade following each Decennial Census. These post-census population estimates are used for many purposes (U.S. Census

Bureau 2021b) including funding formulas that send more than \$1.6 trillion dollars to states and localities every year (Reamer 2020).

The estimation method used in the past developed a population base and then estimated yearly incremental change from that base. For example, the Census Bureau produced estimates for July 1, 2011, using a base population from April 1, 2010 and the change between April 1, 2010 and July 1, 2011.

In the past, the Decennial Census counts have provided the estimates base. But the 2020 Census detailed data needed for the base was not available in time to use with the 2021 and 2022 population estimates so the Census Bureau staff developed a new blended base methodology. The blended base uses the 2020 Census total population for states and counties blended with distribution of age and sex characteristics from the middle series of the national Demographic Analysis (DA) estimates and the distribution of race/Hispanic characteristics from the Vintage 2020 Population Estimates. State estimates are adjusted to make sure they sum to the national estimate and county estimates are adjusted to make sure they sum to the state estimates.

The Vintage 2021 PEP blended base produces data for April 1, 2020, to match the data from the DA middle series and the 2020 Census counts.

The differences between the PEP blended base and the Census counts have different implications for different population groups. This study focuses on the implications for children.

Estimating the 2020 Census Counts for Children

Evaluation of the PEP blended base is hampered because the Census Bureau has not released data on single year of age or age groups for the population under age 18 from the 2020 Census. But they have released data that allows one to derive Census counts for certain age groups.

The Census Bureau has released the Demographic Analysis (DA) population estimates for single year of age, and they have released the net coverage estimates for those ages. Those can be combined to derive estimated 2020 Census counts for three ages groups: 0 to 4, 5 to 13, and 14 to 17. This paper provides data on how the 2020 Census counts for three age groups of children compared to the data from the PEP blended base for three age groups of children in terms of differences from the DA estimates.

The net coverage error in the 2020 Census is derived by formula (1) below.

$$\text{Net Coverage Error} = 100 \times [(\text{Census Count} - \text{DA Estimate})/\text{DA Estimate}] \quad (1)$$

Since the net coverage error and the DA middle series estimates in equation (1) are known, simple algebra can be used to derive estimated 2020 Census counts shown in formula (2) below.

$$\text{Census Count} = [\text{DA estimate} + ((\text{net coverage error}/100) \times \text{DA estimate})] \quad (2)$$

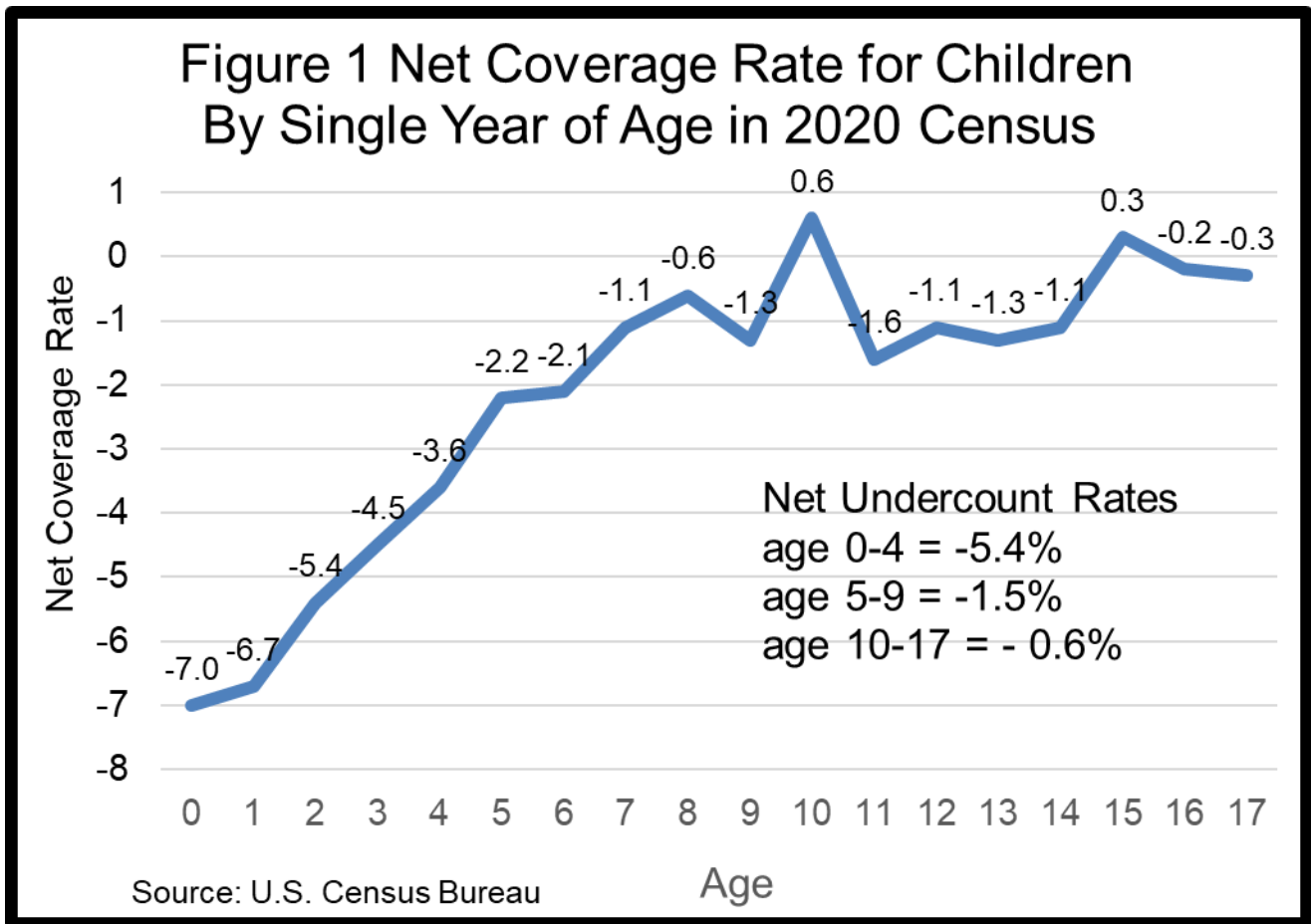
These estimated census counts can then be compared to the DA middle series estimates to measure errors. The PEP blended base estimates are also compared to the DA middle series estimates to see how well the PEP blended base corrects for the

miscounts of children in the 2020 Census. The three age groups reported in the blended base are ages 0 to 4, ages 5 to 13, and ages 14 to 17.

The estimated Census counts shown in this report are not likely to exactly match the 2020 Census results when they are released in the Demographic and Housing Characteristics file in May of 2023 for a couple of reasons. The 2020 Census figures used to calculate the net coverage rates had differential privacy applied to inject noise into the data. In addition, the DA population figures used here are rounded to the nearest thousand and the net coverage rate is rounded to the nearest one hundredth of a percentage point. This rounding impacts the precision of the estimates. In addition, when the single year of age data from the 2020 Census is released in May of 2023, it will have differential privacy applied which will inject noise into the results. Given this situation, the estimated Census counts shown here are not likely to exactly match the reported data in the 2020 Census, but the differences between the figures used here and the data report next May are likely to be small. The differences between the estimated Census counts in this report and the ones reported in the Demographic and Housing Characteristics file released in May of 2023 acknowledged here are unlikely to have much impact on the results of the analysis.

Table 1. Estimated 2020 Census Counts, 2020 DA Middle Series Estimates, and PEP Blended Base Estimates for Three Age Groups of Children					
Age on April 1, 2020	DA Middle Series (rounded to nearest 1,000)*		DA Middle Series Coverage Estimates**	Estimated 2020 Census Count (rounded to nearest 1,000)***	PEP Blended Base April 1, 2020 (rounded to nearest 1,000) ****
0	3,745,000		-7.0	3,483,000	
1	3,788,000		-6.7	3,534,000	
2	3,885,000		-5.4	3,675,000	
3	3,978,000		-4.5	3,799,000	
4	4,062,000		-3.6	3,916,000	
Ages 0 to 4	19,458,000			18,407,000	19,393,000
5	4,090,000		-2.2	4,000,000	
6	4,061,000		-2.1	3,976,000	
7	4,066,000		-1.1	4,021,000	
8	4,086,000		-0.6	4,061,000	
9	4,127,000		-1.3	4,073,000	
10	4,227,000		0.6	4,252,000	
11	4,337,000		-1.6	4,268,000	
12	4,469,000		-1.1	4,420,000	
13	4,447,000		-1.3	4,389,000	
Ages 5 to 13	37,910,000			37,461,000	37,768,000
14	4,347,000		-1.1	4,299,000	
15	4,310,000		0.3	4,323,000	
16	4,347,000		-0.2	4,338,000	
17	4,288,000		-0.3	4,275,000	
Ages 14 to 17	17,292,000			17,236,000	17,225,000
*Source: U.S. Census Bureau, DA release December 2020 https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html					
**Source: National Demographic Analysis Tables: 2020, https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html					
***Source: https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html					
****Source: https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html					

Figure 1 shows the 2020 Census net undercounts for children by single year of age and for three age groups and highlights one of the major problems with the 2020 Census counts: namely the high net undercount of young children. Note the net undercount of young children is much higher than for older children in the 2020 Census. The high net undercount of young children in the U.S. Census is a long-standing problem (O'Hare 2015). Does the PEP blended base correct for the undercounts of children in the 2020 census?



Results

Table 1 shows the derivation of the 2020 Census counts by single years of age for the population ages 0 to 17 based on equation (2) and bundles them into the three age groups for which PEP blended base estimates are provided (ages 0 to 4, ages 5 to 13, and ages 14 to 17). Data for these age groups are also provided for the DA middle series and the PEP blended base estimates in Table 1.

Table 1. Estimated 2020 Census Counts, 2020 DA Middle Series Estimates, and PEP Blended Base Estimates for Three Age Groups of Children						
Age on April 1, 2020	DA Middle Series (rounded to nearest 1,000)*		DA Middle Series Coverage Estimates	Estimated 2020 Census Count (rounded to nearest 1,000)**		PEP Blended Base April 1, 2020 (rounded to nearest 1,000) ***
0	3,745,000		-7.0	3,483,000		
1	3,788,000		-6.7	3,534,000		
2	3,885,000		-5.4	3,675,000		
3	3,978,000		-4.5	3,799,000		
4	4,062,000		-3.6	3,916,000		
Ages 0 to 4	19,458,000			18,407,000		19,393,000
5	4,090,000		-2.2	4,000,000		
6	4,061,000		-2.1	3,976,000		
7	4,066,000		-1.1	4,021,000		
8	4,086,000		-0.6	4,061,000		
9	4,127,000		-1.3	4,073,000		
10	4,227,000		0.6	4,252,000		
11	4,337,000		-1.6	4,268,000		
12	4,469,000		-1.1	4,420,000		
13	4,447,000		-1.3	4,389,000		
Ages 5 to 13	37,910,000			37,461,000		37,768,000
14	4,347,000		-1.1	4,299,000		
15	4,310,000		0.3	4,323,000		
16	4,347,000		-0.2	4,338,000		
17	4,288,000		-0.3	4,275,000		
Ages 14 to 17	17,292,000			17,236,000		17,225,000
*Source: U.S. Census Bureau, DA release December 2020 https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html						
**Source: https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html						
***Source: https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html						

Figure 2 shows the numeric differences between the DA middle series estimates and the Census counts along with the numeric differences between DA middle series estimates and the PEP blended base for three age groups of children.

For age group 0 to 4 and age group 5 to 13, the numeric difference between the DA middle series estimates and the PEP blended base are much smaller than the differences between DA estimate and the Census counts.

For ages 0 to 4, the error comparing the estimated Census counts to the DA middle series is over one million while it is only 65,000 when the PEP blended base is compared to the DA middle series. This means using the PEP blended base rather than the Census counts reduces the error for this population age group by almost one million young children.

For ages 5 to 13, the error comparing the estimated Census counts to the middle series DA is about 449,000 while it is only 142,000 when the PEP blended base is compared to the DA Middle series. This results in an error reduction of about 300,000 children age 5 to 13.

For ages 14 to 17, the errors are small and about the same size for both sources of data (56,000 for the Census count compared to 67,000 for the PEP blended base).

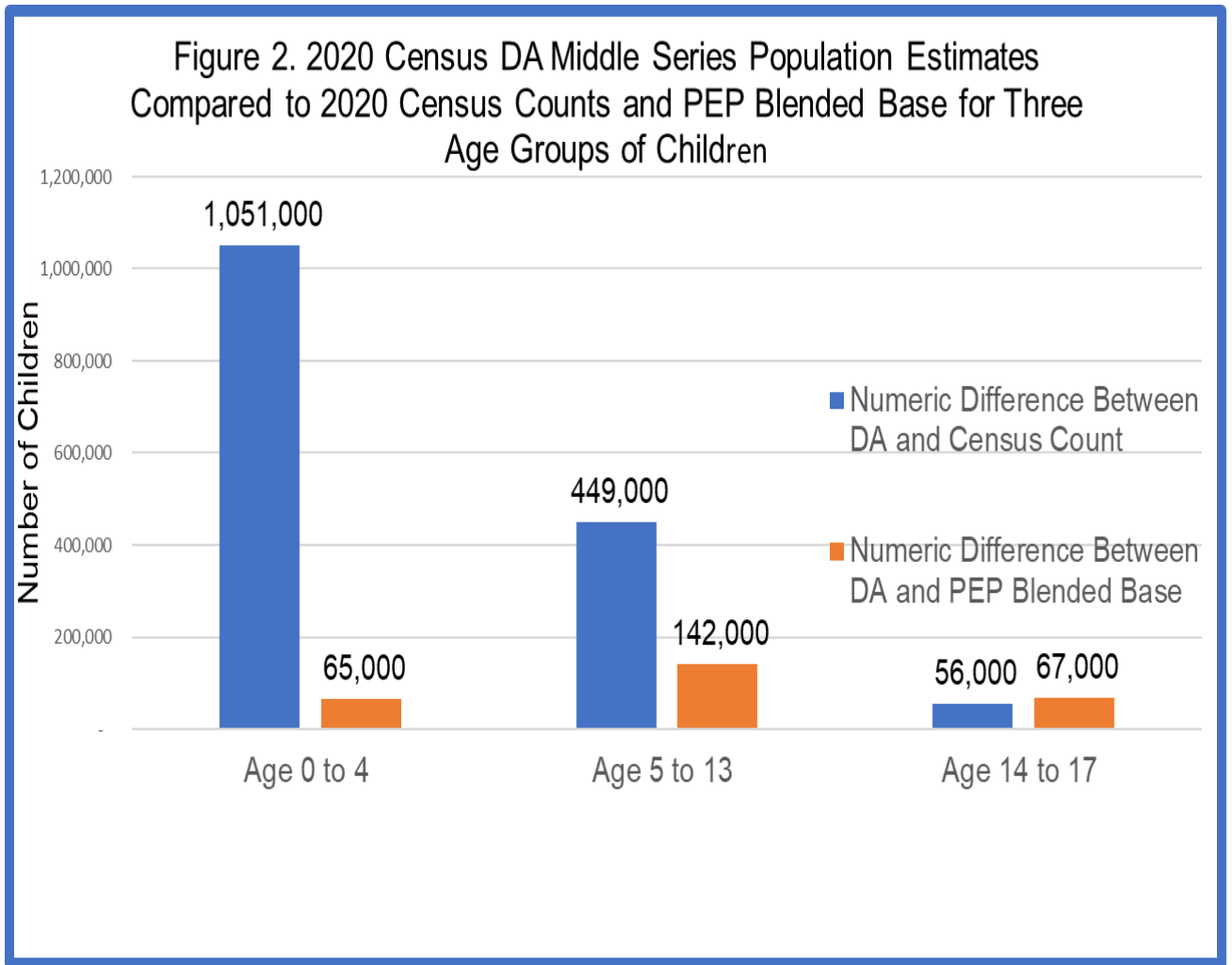
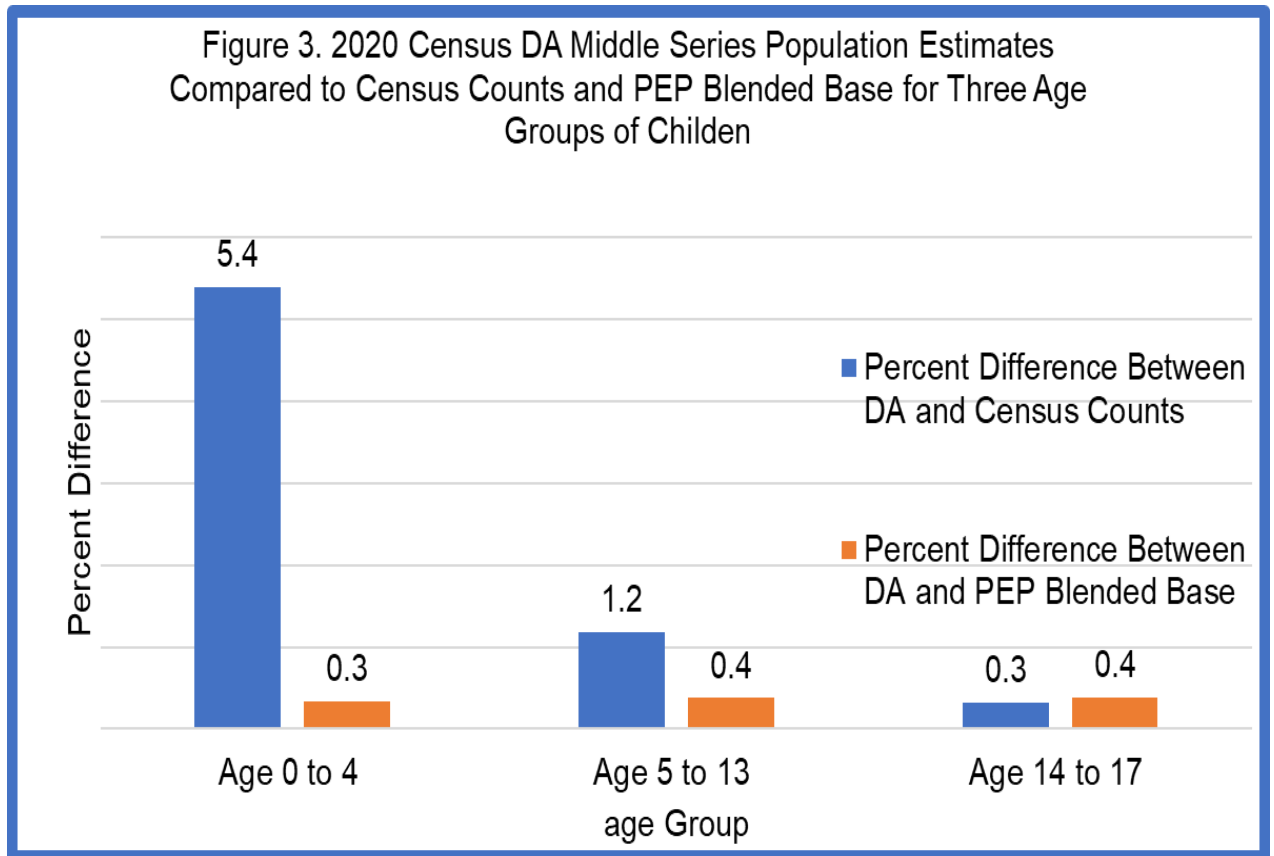


Figure 3 shows the percentage difference between the DA middle series estimates and the Census counts along with the percentage difference between the DA middle series estimates and PEP blended base. The difference between DA estimates and the Census counts is 5.4 percent which is the measured net undercount from the reported by the Census Bureau (2022b). However, the difference between the DA Middle series estimates and the PEP blended base estimates is only 0.3 percent for ages 0 to 4. For the age group 5 to 13, the PEP blended base also shows a much

smaller percent difference from DA middle series than the Census count (1.2 percent compared to 0.4 percent). For age 14 to 17, there is not much difference (0.3 percent compared to 0.4 percent).



Summary

This analysis in this report shows the PEP blended base being used for the 2021 and 2022 post-census population estimates corrects most of the high 2020 Census undercount for young children. In both numeric and percentage terms, the difference between the DA middle series estimate and the PEP blended base are much smaller

than the difference between the DA middle series estimates and the 2020 Census estimated counts for ages 0 to 4.

The PEP blended base is also helpful in correcting the net Census undercount for the population age 5 to 13. For ages 14 to 17, there is little difference between the Census counts and the PEP blended base estimates.

This analysis shows one clear benefit of using the new PEP blended base rather than relying on the 2020 Decennial Census counts for the population base used in the Census Bureau's yearly post-census estimates.

References

O'Hare, W.P. (2015). *The Undercount of Young Children in the U.S. Decennial Census*. Springer Publishers <https://www.springer.com/gp/book/9783319189161#>

O'Hare, W. P., Robinson, J.G., West, K., and Mule, T., (2016). "Comparing the U.S. Decennial Census Coverage Estimates for Children from the Demographic Analysis and Coverage Measurement Surveys," *Population Research and Policy Review*, Vol. 35, Issue 5, pages 685-704. [Comparing the U.S. Decennial Census Coverage Estimates for Children from Demographic Analysis and Coverage Measurement Surveys | SpringerLink](#)

Reamer, A. (2020). "Counting for Dollars: The Role of the Decennial Census in the Geographic Distribution of Federal Funds," *GW Institute of Public Policy*. <https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds>

U.S. Census Bureau (2020). DA release December 2020 <https://www.census.gov/data/tables/2020/demo/popest/2020-demographic-analysis-tables.html>

U.S. Census Bureau (2021a) Hartley, Christine (2021). "Methodology Updates for the Vintage 2021 Estimates," *U.S. Census Bureau*. <https://www.census.gov/data/academy/webinars/2021/methodology-updates-for-the-vintage-2021-estimates.html>

U.S. Census Bureau (2021b). "METHODODOLOGY FOR THE UNITED STATES POPULATION ESTIMATES: VINTAGE 2021 Nation, States, Counties, and Puerto Rico – April 1, 2020 to July 1, 2021, December 2021," *U.S. Census Bureau*. <https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2021/methods-statement-v2021.pdf>

U.S. Census Bureau (2022a). "Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census," *U.S. Census Bureau*. <https://www.census.gov/newsroom/press-releases/2022/2020-census-estimates-of-undercount-and-overcount.html>

U.S. Census Bureau (2022b). "National Census Coverage Estimates for People in the United States by Demographic Characteristics," *U.S. Census Bureau*. <https://www2.census.gov/programs-surveys/decennial/coverage-measurement/pes/national-census-coverage-estimates-by-demographic-characteristics.pdf>