Analysis of Census Bureau's March 2022 Differential Privacy Demonstration Product: Implications for Data on Young Children

By

Dr. William P. O'Hare

Analysis of Census Bureau's March 2022 Differential Privacy Demonstration Product: Implications for Data on Young children

By

Dr. William P. O'Hare

Executive Summary

The U.S. Census Bureau is using a new method called differential privacy (DP) to help protect confidentiality and privacy of respondents in the 2020 Census. This paper provides some information on how the use of DP in 2020 Census is likely to impact the accuracy of data for young children (population ages 0 to 4). The study is based on analysis of the most recent DP Demonstration Product released by the Census Bureau on March 16. 2022. The DP Demonstration Product issued in March 2022 supersedes earlier DP Demonstration Products and focuses on data for the 2020 Census Demographic and Housing Characteristics (DHC) file. This file has most of the tables that were in Summary File 1 in the 2010 Census. The Demonstration Product released in March 2022 has data for population and housing units, but this analysis only examines data from the population file.

This paper presents analysis of the error introduced by DP by comparing the data as reported in the 2010 Census Summary File to the same data after the application of DP. According to the Census Bureau, the demonstration file released by the Census Bureau in March has been optimized for major use cases of the DHC tables.

Analysis presented in this paper found little impact of DP on data about young children for large (highly aggregated) geographic units like states or large counties. However, the story is different for smaller geographic units. Many smaller areas have high levels of error in their data on young children after DP is applied. For example, the count of young children would exhibit absolute error of 5 percent or more in about 27 percent of Unified School Districts after DP is applied. The data also show that 69 percent of Unified School Districts had absolute numeric errors of 5 or more young children after DP is applied.

Errors of the magnitude shown above could have important implications for federal and state funding received by schools and for educational planning. Errors of this magnitude might impact formula funding that is based on Census-derived data and some schools will get less than they deserve.

Bigger absolute error percentages are evident for Hispanic, Black, and Asian young children in Unified School Districts. The mean absolute percent error for Non-Hispanic White young children was 5 percent compared to 27 percent of Hispanic young children, 34 percent for Black young children, and 42 percent for Asian young children. Differential accuracy among race and Hispanic Origin groups raises questions of data equity after DP is applied.

I also examined the accuracy/errors for the single year age 4 child population and found errors for single year of age are particularly large. I found 57 percent of Unified School Districts had absolute percent errors of 5 percent or more for children age 4, and 66 percent had absolute numeric errors of 5 or more children age 4.

Analysis also shows that 39 percent of Places (cities, village, and towns) had absolute percent errors of 5 percent or more for age 0 to 4, and 46 percent of Places had absolute numeric errors of 5 or more young children.

After the injection of DP in the 2010 Census data included in the March 2022 Census Bureau Demonstration Product, there were over 163,000 blocks nationwide that had population ages 0 to 17, but no population ages 18 or over. This result has two important implications, First, blocks with children and no adults is a highly implausible situation and the large number of such blocks may undermine confidence in the overall Census results. Second, these implausible results are likely due to young children being separated from their parents in 2020 Census DHC processing with DP. This separation of children and parent in data processing is an ongoing concern for data on young children and the production of future tables for children. This issue is particularly important in introducing DP into the American Community Survey, which is a key source of child well-being measures (O'Hare 2022b) To understand the well-being of children, it is critical to understand the situation of a child's parents or caretakers. - Moreover, if the same separation of children from their caregivers occurs in the application of DP to the American Community Survey, it will eliminate child poverty data which is based on household income. Child poverty data are the most important type of data on child well-being.

Based on the errors for young child population with the privacy parameters for DP used in the March 2022 DP Demonstration Product, and the lack of clarity about privacy protection from DP, I recommend the Census Bureau take steps to reduce the size of errors injected into the 2020 Census DHC file.

This paper is meant to provide stakeholders and child advocates with some fundamental information about the level of errors DP is likely to inject into the 2020 Census data for the population ages 0 to 4. There are a couple of reasons for sharing this information with child advocates now. The 2020 Census results for some localities may include situations where the number of young children reported looks suspect. It is important to make sure child advocates are aware of the potential impact of DP so they can explain odd child statistics to local leaders.

There is a second reason for sharing this information with state and local child advocates. The U.S. Census Bureau is looking for feedback on the use of DP in the 2020 Census. The Census Bureau is looking for cases where census data are used to make decisions. The Census Bureau is asking data users to examine the DP Demonstration Product to see if the error injected by DP make the data unfit for use. After reading this report, I hope you will convey your thoughts to the Census Bureau.

There is some latitude in how much error the Census Bureau will inject into the DHC files so feedback from census data users is important. If many users feel the current level of accuracy for data on young children in DP Demonstration Product is not accurate enough for some uses, there is a chance the Census Bureau could make the final data more accurate.

Stakeholders, child advocates, and data users should take advantage of this opportunity to communicate their thoughts to the Census Bureau before Census Bureau’s Data Stewardship Advisory Committee makes a final decision on the privacy parameters before the DHC files are released in May of 2023. Comments on the implications of DP in the March 2022 Demonstration File are due Monday, May 16, 2022. **Comments and responses can be sent to 2020DAS@census.gov.**

Analysis of Census Bureau's March 2022 Differential Privacy Demonstration Product: Implications for Data on Young children

By

Dr. William P. O'Hare

Introduction

The U.S. Census Bureau is using a new method called differential privacy (DP) to help protect confidentiality and privacy of Census respondents in releasing data from the 2020 Census.[1] This paper uses several measures to assess the accuracy of census data for young children after DP is applied. Young children are defined in this report as those ages 0 to 4. The analysis is based on the Demonstration Product data released on March 16, 2022, which is the most recent available from the Census Bureau.

In short, DP injects errors in the data provided by respondents to make it more difficult for someone to be identified in the Census records. Adding or subtracting random numbers to the census results makes it more difficult to identify data for specific respondents because the data in the published census results no longer match what respondents submitted. The U.S. Census Bureau (2020e) provides more information

[1] The terminology in this arena can be confusing. Differential Privacy is sometimes called "formal privacy." The system developed for the 2020 Census DHC file has also been called the Top Down Algorithm or TDA. Since the application of differential privacy occurs within the Census Bureau's Disclosure Avoidance Systems (DAS) that term has sometimes been used to describe the use of Differential Privacy. To avoid confusion, I use the term differential privacy (DP) here to distinguish the version of DAS that includes DP from other versions of DAS.

on the use of DP in the 2020 Census along with regular updates of their work (U.S. Census Bureau 2020c). In the fall of 2021, the Census Bureau released a primer on DP. (U.S. Census Bureau 2021d).

For an independent look at differential privacy see Boyd (2019) or Bouk and Boyd (2021). Hotz and Salvo (2020) offer a good review of DP early in the Census Bureau's development. . A good overview of the evolution of the DP issue at the Census Bureau is provided by Boyd and Sarathy (2022). I think it is fair to say that the introduction of DP in the 2020 Census has become a very controversial issue. In their review of the development of the DP issue over the past few years, Boyd and Sarathy (2022, page 1) conclude, "When the U.S. Census Bureau announced its intention to modernize its disclosure avoidance procedures for the 2020 Census, it sparked a controversy that is still underway."

One reason to focus on impact of DP on the population ages 0 to 4 is the high net undercount of that population in the Census. Results of the 2020 Census evaluation, using the Demographic Analysis method, shows a net undercount of 5.4 percent for young children which was much higher than any other age group (U.S. Census Bureau 2022c). Recent trends are also unsettling. From 1950 to 1980, the young children and adults had similar decade-to-decade improvement in terms of census coverage. However, after 1980 the trajectories were quite different. The coverage for adults continued to improve while the coverage of young children decreased dramatically (O'Hare 2022a).

There are a couple of perspectives one could take regarding the high net undercount of young children and DP. On one hand, since the 2020 Census data for young children already has more error than data for other age groups, perhaps the amount of error injected by DP should be limited for this group. It doesn’t seem fair to inject more error into data for groups that already have a lot of error in their census data. On the other hand, one might think that since the 2020 Census data for young children already has a lot of error, the added error from DP will not make much difference.

I focus first on data accuracy for Unified School Districts because schools are the public institution most closely associated with the child population and schools use demographics in a variety of ways. I next look at data for Places. Places include big cities and small villages. They typically have policymaking authority, and they often provide programs for young children such as childcare or preschool programs.

Several issues regarding DP are addressed in the Discussion section included the high error rate for blocks, breaking the relationship between children and parents,questions of equity, and the extent to which DP contributes to the lack of public trust.

Background on Privacy in the Census

In every census, the U.S. Census Bureau faces a trade-off between privacy protection and accuracy. According to the U.S. Census Bureau (2020d),

> “One of the most important roles those national statistical offices (NSOs) play is to carry out a national population and housing census. In so doing, NSOs have two data stewardship mandates that can be in direct opposition. Good data stewardship involves both safeguarding the privacy of the respondents who have entrusted their information to the NSOs as well as disseminating accurate and useful census data to the public.”

The problem that DP is designed to fix is complicated as is the implementation of DP. The passage below from the U.S. General Accountability Office (2020, page 14) is the best short description I have seen on this issue.

> "Differential privacy is a disclosure avoidance technique aimed at limiting statistical disclosure and controlling privacy risk. According to the Bureau, differential privacy provides a way for the Bureau to quantify the level of acceptable privacy risk and mitigate the risk that individuals can be reidentified using the Bureau's data. Reidentification can occur when public data are linked to other external data sources. According to the Bureau, using differential privacy means that publicly available data will include some statistical noise, or data inaccuracies, to protect the privacy of individuals. Differential privacy provides algorithms that allow policy makers to decide the trade-offs between data accuracy and privacy. "

It is important to note that the U.S. Census Bureau has used methods to help avoid disclosure of individual census respondents for many decades. According to U.S. U.S. Census Bureau (2018) some method of disclosure avoidance has been used by the U.S. Census Bureau since 1970. The 2010 Census data include some changes to original responses to help avoid disclosure of information about individual respondents, largely using a method called swapping.

The application of Differential Privacy allows the Census Bureau to control the amount of error injected into the data which is largely controlled by something called "Epsilon." A higher-level epsilon means less error and more risk of violating confidentiality and a lower epsilon means more error and less risk of violating confidentiality. In the latest material from the Census Bureau, Epsilon has been replaced with a term called Rho. It is my understanding Rho works the same way as Epsilon in that a higher value means more accuracy.

Measuring Accuracy

There is no consensus on exactly what measures should be used to assess the accuracy of DP-infused data, and there is no single benchmark to determine if DP-infused figures are “accurate enough for use.” The U.S. Census Bureau (2020a) has suggested several measures of accuracy that could be used to evaluate the DP-infused data.

For simplicity I only look at a few key measures here, but I believe they provide sufficient information to reach some conclusions. The measures used here (mean absolute numeric error, mean absolute percent error, and large errors) are a subset of those discussed by Census Bureau. Like the Census Bureau’s assessment of DP-infused data, I provide data for both absolute numerical errors and absolute percent errors because either can be important and using both perspectives provide a more complete picture of the error profiles for geographic units.

The DP demonstration file released by the Census Bureau on March 16, 2022, provides DP-infused data from the 2010 Census which can be compared to the 2010 Census data without DP to understand the likely impact DP has on data accuracy.

Errors are defined here as the difference between the data as originally reported in the 2010 Census Summary File and the same data after DP has been injected. The data from the Summary File is sometimes referred to as data without the application of DP in this report. Specifically, I subtract the value of the data without DP (Summary File) from corresponding DP-infused data to find the error. For percentages, the difference is divided by the data without DP (i.e., Summary File) value.

I include a measure the Census Bureau calls the Mean Absolute Error (I label this Mean Absolute Numerical Error in the tables to distinguish it from the Mean Absolute Percent Error) and I also include the Mean Absolute Percent Error.

An absolute error reflects the magnitude of the error regardless of direction. A geographic unit with an absolute error of 10 percent could be 10 percent too high or 10 percent too low. Absolute errors are used to make sure positive errors and negative errors do not cancel each other out and make it appear as if there are no errors.

Percent error reflects the size of the error relative to the size of the population. An error of a given magnitude (say 10 young children) may be trivial in large Places but very significant in smaller Places. For example, a numeric error of 10 young children in a school district of 1,000 young children is only a 1 percent error, but a numeric error of 10 young children in a school district of 100 is a 10 percent error.

In addition to measures of average error, I include analysis on the number and percent of geographic units that have relatively large errors. I use two sets of benchmarks to identify large errors: one for absolute numeric errors and one for absolute percent errors.

I believe the number and percent of large errors are likely to be the most important measures of accuracy in the 2020 Census. Large errors are likely to be a statistical problem and a public relationship problem for the Census Bureau, particularly if they are accompanied by large swings in funding that are not connected to real changes in population size. Data from the Census is often used to distribute federal and state dollars based on population (O'Hare 2020a: Reamer 2020). Large errors can result in implausible or impossible results. Such results are likely to cast suspicion on all the data from the Census Bureau and it is likely to undermine the confidence people have in all the census data.

<u>Data Used in This Study</u>

The Demonstration Product released in March 2022 reflects ongoing work at the Census Bureau. Starting in October 2019, the Census Bureau has released several Demonstration Products that reflect the injection of DP into 2010 Census data. The first official data from the 2020 Census with DP infused was the redistricting data file released by the Census Bureau in August 2021.

The data used in my analysis were originally provided by the Census Bureau. The IPUMS- NHGIS unit at the University of Minnesota processed the Census Bureau files and put the data into more user-friendly tables. I analyze the data produced by IPUMS-NHGIS unit which are available at https://nhgis.org/privacy-protected-demonstration-data

According to the IPUMS- NHGIS unit at the University of Minnesota, the privacy loss budget assigned to person-level and housing unit-level counts in the 2022-03-16 vintage file was 20.82 and 22.77, respectively. This contrasts to an Epsilon of 19.6 used for the 2020 Census redistricting files.

Geographic units where there were zero people ages 0 to 4 in either the 2010 data with DP or without DP, were removed from the file for analysis. Observations with zeros for key measures produce very unusual results.   This analysis does not include data for Puerto Rico.

Results for Age 0 to 4 in Four Kinds of Geographic Units

Table 1 provides a few key accuracy measures for the population ages 0 to 4 for four kinds of geographic units. These units were selected because they all have significant policy-making power regarding programs for children.

The results shown in Table 1 indicate that DP is unlikely to have much of an impact on the young child data for states. The mean absolute numeric error for states for the population ages 0 to 4 is about 7 young children and the mean absolute percent error rounds to zero.

Also, DP is unlikely to have much impact on young child county data for most counties. The mean absolute numeric error for counties is about 8 young children and mean absolute percent error is 0.9.

However, of the 3,221 counties examined here 35 percent (1,140) had less than 1,000 children ages 0 to 4. For this subset of counties, DP may distort the data to a considerable degree. For the 302 counties with less than 5,000 people, the mean absolute percent error for ages 0 to 4 was 4.6 percent and the mean numeric error was 5.

Table 1 Key Statistics for Absolute Numeric and Absolute Percent Errors* for Children Ages 0 to 4 for Selected Geographic Units				
	States	Counties or county equivalent	School Districts	Places
Number of Units in the Analysis	50	3,221	10,864	28,729
Mean Size of District (Children ages 0-4 based on Summary File)	39,873	6,342	1,880	546
Mean Absolute Numeric Error**	7	8	12	6
Mean Absolute Percent Error	rounds to zero	0.9	4.3	13.6
Percent of Units with Absolute Numeric Errors of 5 or more young children	58	62	69	46
Percent of Units with Absolut Percent Errors of 5% or more	0	3	27	39
Source: Author's analysis of Demonstration Product data released by the Census Bureau on March 16, 2022 after being prociess by IPUMS NHGIS at the University of Minnesota www.nhgis.org				
Data in this table does not include Puerto Rico or geographic units with zero population age 0 to 4 in 2010 Summary File or DP-infused file.				
* in this paper errors reflect the difference between the 2010 Census data without and with DP injected.				
** The Census Bureau calls this measure Mean Absolute Error. I include the word "Numeric" to distinguish it from Mean Absolute Percent Error.				
DC is not included in the state data but is included in the county data				

The situation is different for Unified School Districts and Places (shown in Table 1), where DP is likely to cause larger distortions (percentagewise) for the young child population. The mean absolute numeric error for Unified School Districts is 12 young children and it is 6 young children for Places. The mean absolute percent error for United School Districts is 4.3 percent and it is 13.6 percent of Places.

Accuracy for Unified school Districts and Places are explored in more detail in the next two sections of this report.

Application of Differential Privacy to School District Data

The analysis first focuses on Unified School Districts since schools are the largest public institution focused on children. The Census Bureau reports there were 61.6 million children ages 3 to 17 enrolled in schools in 2019 (U.S. Census Bureau 2021a).

Schools often provide preschool programs for those under age 5. The Census Bureau shows there were over 5 million children enrolled in preschool in 2019, and more than half of all children age 3 and 4 are in preschool or nursey school (McElrath et al. 2022)

Reamer (2020) shows that $39 billion of federal funds were distributed by the U.S. Department of Education to states and localities in FY 2017 based on census-derived data. Table 2 shows programs run by the U.S. Department of Education that distribute federal funds to state and localities based on census-derived data. In addition, many other government programs also use census-derived data to distribute funds targeted to children.,

Overall, Reamer (2020) identified 316 federal programs that use census-derived data to distribute about $1.5 trillion to states and localities in Fiscal Year 2017. About two-thirds of the 315 programs use substate data which underscores the important of small area census data. . When one is talking about billions of dollars, a small percent error can translate into a large dollar amount. This is even more true when the funding allocation is based only on a particular age group.

Table 2. Federal Programs in the U.S. Department of Education that Distribute Funds to States and Localities based on Census-derived Data	
	Amount Distributed in FY 2017
Adult Education - Basic Grants to States	$581,955,000
Title I Grants to LEAs	$15,459,802,000
Special Education Grants	$12,002,848,000
Career and Technical Education - Basic Grants to States	$1,099,381,000
Vocational Rehabilitation Grants to the States	$3,121,054,000
Rehabilitation Services - Client Assistance Program	$13,000,000
Special Education - Preschool Grants	$368,238,000
Rehabilitation Services - Independent Living Services for Older Individuals Who are Blind	$33,317,000
Special Education-Grants for Infants and Families	$458,556,000
School Safety National Activities	$68,000,000
Supported Employment Services for Individuals with the Most Significant Disabilities	$27,548,000
Program of Protection and Advocacy of Individual Rights	$17,650,000
Twenty-First Century Community Learning Centers	$1,179,756,000
Gaining Early Awareness and Readiness for Undergraduate Programs	$338,831,000
Teacher Quality Partnership Grants	$43,092,000
Rural Education	$175,840,000
English Language Acquisition State Grants	$684,469,000
Supporting Effective Instruction State Grants	$2,055,830,000
Grants for State Assessments and Related Activities	$369,051,000
Teacher Education Assistance for College and Higher Education Grants	$90,955,000
Preschool Development Grants	$250,000,000
Student Support and Academic Enrichment Program	$392,000,000
Total	$38,831,173,000
Source: Counting for Dollars. https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-	

It is also clear that census-related data are often used by states to distribute state government money, but as far as I can tell, there is no systematic data on how much money is distributed by states based on Census data (O'Hare 2020a).

At the Committee on National Statistics DP workshop held in December 2019 there were several presentations reflecting implications of DP-infused data for young children and school districts (Vink 2019; O'Hare 2019; Nagle and Kuhn 2019:). O'Hare (2021) focuses on the accuracy of population ages 0 to 17 for Unified School Districts based on data from the Census Bureau's redistricting file. Note that some of these analyses are now outdated but may be useful for framing issues.

Demographic data are used for several important school district applications. Population projections are often used to plan for expanding (or reducing) school facilities, staff, and other school-related needs. Demographic projections are typically based on Decennial Census data. Current and projected demographic data are often used to construct attendance boundaries to keep classrooms from becoming overcrowded. Such activities often require very small area data such as census blocks. Demographers who work extensively with school districts report that census blocks are a critical geographic unit for their work (Cropper et al. 2021). Constructing attendance boundaries often include sensitivity to racial composition, so small area demographics by race are important.

Many school districts are governed by school boards which are often elected from single member districts. Such districts must meet the usual legal requirements of redistricting such as having districts with equal population size. Such redistricting must also meet the requirements of the Voting Rights Act, which means small area tabulations of population by race and Hispanic origin are important.

Once children get into the K-12 school system, school systems have pretty good data for forecasting the number of children to expect in each grade the following year. From that perspective it is the cohort age 0 to 4 that is the biggest unknown for school systems. Therefore, this is the most important age group for examining the amount of error injected by DP.

Districts where there was a zero for population age 0 to 4 in the DP or SF file were not included in the analysis. Also, recall Puerto Rico is not included.

DP has a bigger impact, percentage wise, in smaller populations and the majority of Unified School Districts are relatively small. Many of the 10,864 Unified School Districts are very small; 729 of the Unified School Districts had total population less than 1,000, and 3,875 districts had total population less than 5,000 in the 2010 Census. The translation of small numeric errors into large percent errors is also more apparent in looking at data for Hispanic, Black, and Asian groups within school districts because those are typically smaller population groups.

Table 3 shows several measures of accuracy/error for 10,864 Unified School Districts in the 2010 Census used in this analysis. The data are provided for all young children (all races) as well as for Non-Hispanic White Alone young children, Hispanic young children, Black Alone young children, and Asian Alone young children. For the remainder of this report when I use the term Black or Asian, it means Black alone or Asian alone. Other race groups were not examined here because the numbers were small, they were often highly clustered, and time was limited.

Data in Table 3 show the vast majority of Unified School Districts have at least one Black child, one Hispanic child, and one Asian child. But many districts have few young children of color. The average number of Hispanic young children in School Districts where there was at least one Hispanic was 524, for Blacks it was 396, and for Asians it was 151. These numbers are well below the overall average of 1,880 young children. The relatively small number of Black, Hispanic, and Asian young children in many districts results in these groups having larger absolute percent errors.

Table 3 shows the mean absolute numeric error for all young children (all races) in Unified School Districts is 12 young children. Data in Table 3 shows for all children, the mean absolute percent error as 4.3. But these measures mask big differences among race and ethnic groups

The mean absolute numeric errors for race and Hispanic Origin groups are smaller than for all children (10 for Non-Hispanic White Alone young children, 7 for Hispanic young children, 5 for Black young children, and 4 for Asian young children), On the other hand, mean absolute percent error was 4.3 percent for all children, 27 percent for Hispanics, 34 percent for Blacks young children, and 42 percent for Asian young children (see Table 3).

Table 3 . Key Error* Statistics for Children Ages 0 to 4 for Unified School Districts					
	All young children	Non-Hispanic White Alone	Hispanic	Black**	Asian**
Number of units in the analysis	10,864	10,838	10,178	7,381	5,932
Mean number of young children in district (in group column heading)	1,880	946	524	396	151
Mean absolute numeric error***	12	10	7	5	4
Mean absolute percent error	4.3	5	27	34	42
Percent of units with errors of 5 or more young children	69%	63%	49%	37%	32%
Percent of units with errors of 5% or more	22%	27%	65%	61%	68%
Source: Author's analysis of Demonstration Product data released by the Census Bureau on March 16, 2022 after being processed by IPUMS NHGIS at the University of Minnesota www.nhgis.org					
Data in this table does not include Puerto Rico or geographic units with zero population age 0 to 4 in 2010 Summary File or DP-infused file.					
* in this paper errors reflect the difference between the 2010 Census data without and with DP injected.					
** The Census Bureau calls this measure Mean Absolute Error. I include the word "Numeric" to distinguish it from Mean Absolute					
DC is not included in the state data but is included in the county data					

Recall that absolute errors reflect the magnitude of the error without regard to the direction of the error. Absolute errors are used so that positive and negative errors do not cancel each other out in constructing an average or mean.

<u>Large Errors in Unified School Districts</u>

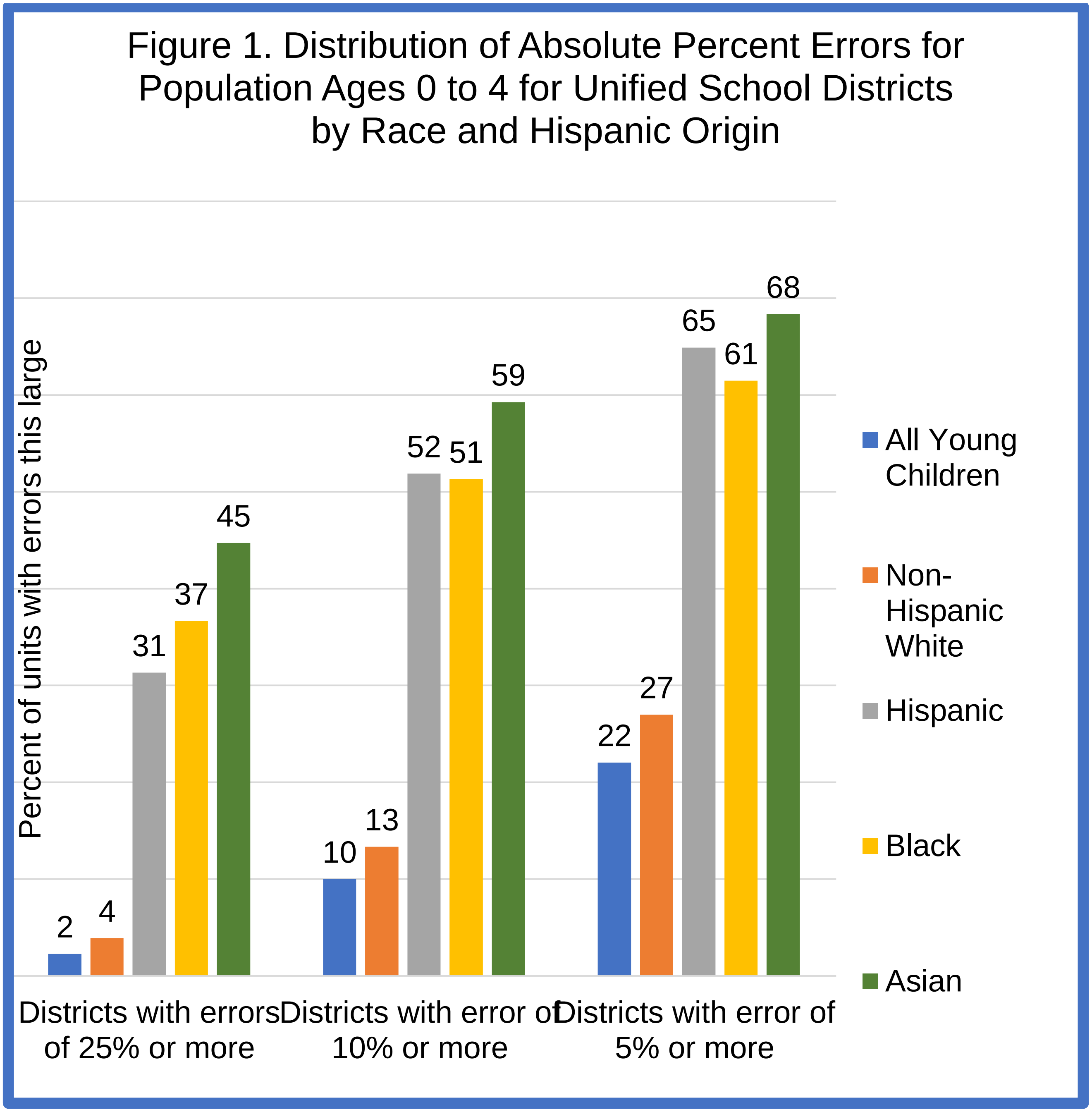
Means or averages are helpful, but they do not reveal the full story. Large errors can be problematic even if the overall mean is relatively low. An examination of the distribution of Unified School Districts by error size can provide more information on the relative accuracy of the DP-infused data.

There is no consensus on what constitutes a large error and definitions probably vary with different applications. I show three benchmarks for large absolute percent errors. The 5 percent or more and 10 percent or more categories are used in several publications. I added the 25 percent plus category to look at the most extreme errors. Errors of 25 percent or more are likely to be very problematic. These thresholds are judgmental, but I think they provide a reasonable range of errors.

To be clear, the districts with more than 25 percent with large errors are also counted in the categories for more than 10 percent error and more than 5 percent error.

Distributions of absolute percent errors are shown in Figure 1 which shows that for all young children, 22 percent of districts had absolute percent errors of 5 percent or more for all children, compared to 27 percent of Non-Hispanic White Alone, 65 percent for Hispanic young children, 61 percent for Black young children, and 68 percent for Asian young children. Since minority groups are smaller in population size, it is not surprising that there are more extreme absolute percent errors.

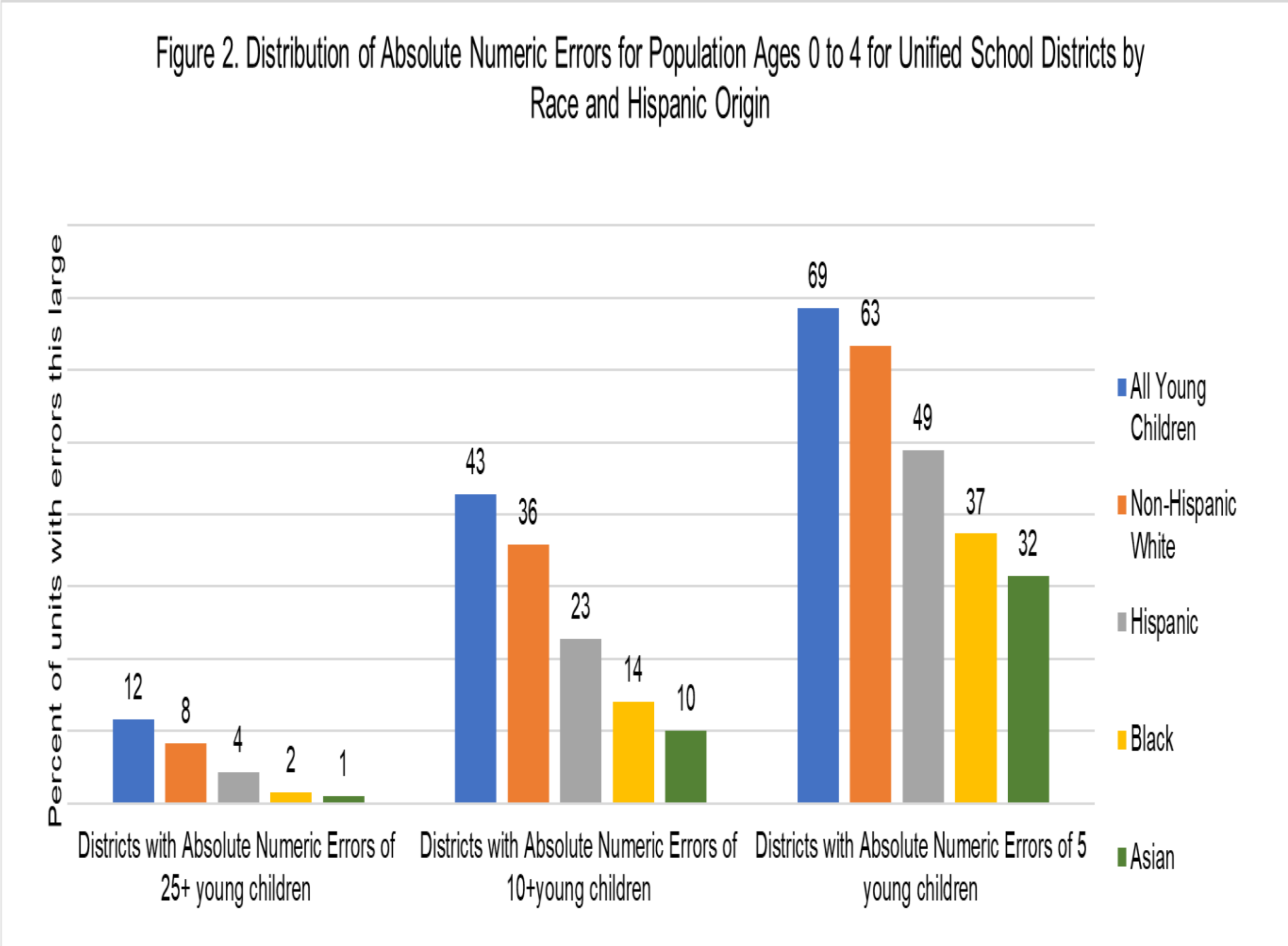
Figure 1 also shows that for young children of color, absolute percent errors of 25 percent or more are not unusual.



I use three benchmarks for large absolute numeric errors. The 5 person and 10 person categories of errors have been used in other publications. I added the 25 persons plus category to look at the most extreme errors. Errors of 25 or more young children are likely to be very problematic.

Figure 2 shows 69 percent of the Unified School Districts had errors of 5 young children or more for young children of all races but the figures for minority groups are smaller: 49 percent for Hispanic young children, 37 percent for Black young children, and 32 percent for Asian young children.

In Figure 2, in each category of absolute numeric errors (5 young children, 10 young children, and 25 young children), there are many fewer districts that have this level of error for Hispanic, Black, and Asian young children than there are districts that have this level of error for all young children or Non-Hispanic White young children.



There are relatively few Unified School Districts with very large absolute numeric errors. Only 12 percent of Unified School Districts have errors of 25 young children or more, compared to 4 percent of Hispanic young children, 2 percent for Black young children, and 1 percent for Asian young children.

The national numbers shown above mask a lot of variation across states. Table 4 shows states ranked on two key measures of accuracy (mean absolute numeric error and mean absolute percent error) for Unified School Districts. The mean absolute numeric error for states ranges from a low of 0 for Hawaii (Hawaii only has one unified school district) to a high of 32 for Montana. The mean absolute percent error ranges from a low of 0 for Hawaii to a high of 6.7 percent in South Dakota.

Table 4 States Ranked by Mean Absolute Numeric Error and Absolute Percent Error for Ages 0 to 4 by Unified School Districts

Rank*		Average of absolute numerical error		Rank*	State	Average of absolute percent error
1	Montana	32		1	South Dakota	6.7
2	Maine	25		2	Nevada	6.4
3	North Dakota	24		3	New York	5.9
4	Washington	20		4	Oklahoma	5.7
5	Nebraska	19		5	New Hampshire	5.4
6	South Dakota	19		6	Iowa	5.3
7	Oklahoma	19		7	Texas	5.2
8	Oregon	18		8	North Dakota	5.1
9	Vermont	17		9	Alaska	5.1
10	Idaho	17		10	Wisconsin	4.8
11	Colorado	17		11	Montana	4.8
12	Texas	16		12	Arkansas	4.7
13	New Mexico	16		13	Colorado	4.6
14	Alaska	16		14	Ohio	4.4
15	Kansas	16		15	Illinois	4.3
16	Missouri	15		16	Oregon	4.3
17	Iowa	15		17	Nebraska	4.3
18	Wyoming	15		18	Michigan	4.3
19	Arkansas	13		19	Kansas	4.2
20	New Hampshire	12		20	Pennsylvania	4.2
21	New York	12		21	Missouri	4.2
22	Michigan	12		22	Minnesota	4.1
23	Minnesota	12		23	Idaho	3.9
24	Illinois	11		24	New Mexico	3.9
25	Wisconsin	11		25	Washington	3.9
26	Ohio	10		26	Connecticut	3.9
27	Nevada	10		27	Arizona	3.9
28	Arizona	9		28	Tennessee	3.7
29	Indiana	8		29	Utah	3.7
30	Mississippi	7		30	Wyoming	3.6
31	California	7		31	Mississippi	3.5
32	Kentucky	6		32	West Virginia	3.4
33	Delaware	6		33	Massachusetts	3.3
34	Pennsylvania	5		34	Indiana	3.1
35	Tennessee	5		35	California	3.0
36	South Carolina	5		36	Georgia	3.0
37	Utah	5		37	Virginia	3.0
38	New Jersey	5		38	Maine	2.9
39	Virginia	4		39	Kentucky	2.9
40	Alabama	4		40	New Jersey	2.8
41	Massachusetts	4		41	Maryland	2.6
42	Georgia	4		42	Vermont	2.4
43	West Virginia	4		43	Alabama	2.4
44	Rhode Island	4		44	Florida	2.4
45	Connecticut	4		45	North Carolina	2.2
46	North Carolina	4		46	Rhode Island	2.1
47	Louisiana	3		47	South Carolina	1.8
48	Florida	2		48	Louisiana	1.8
49	Maryland	2		49	Delaware	1.1
50	Hawaii	0		50	Hawaii	0.0
U.S. Average		12		U.S. Average		4.3

Source: Author's analysis of Demonstration Product released by the Census Bureau on March 16, 2022 afterprocessing by IPUMS NHGIS at the University of Minnesota www.nhgis.org

*Ranks are based on unrounded data.

Analysis for Age 4

In the Demonstration Product released in March 2022, the Census Bureau provided data by single year of age and sex for the population under age 20. I analyze this data for age 4 for Unified School Districts. I selected age 4 because that is often used by school systems to predict the number of kindergarteners to expect in the following school next year. I do not see any reason why the metrics for age 4 would be much different than the metrics for any other single year of age.

Table 5 provides the key metrics for the comparison of age 4 in Unified School Districts in the 2010 Census file with and without DP. Districts with no people age 4 in the DP or SF file were not used in the analysis. The mean absolute numeric error was 11 and the mean absolute percent error was 11 percent for age 4

A big share of Unified School Districts had large errors in both numeric and percentages terms. Two-thirds (66 percent) of Unified School System had absolute numeric errors of 5 or more children and 57 percent of Unified School Districts had absolute percent errors of 5 percent or more for children age 4.

With errors of this magnitude for single year of age, one has to wonder if this data is worth producing, particularly for small districts. It is not clear how users are supposed to manage data with this degree of uncertainty.

Table 5. Unified School District Error* Metrics for Age 4	
Number of Units in Analysis	10,424
Mean number of 4 year old's in Summary File	394
Mean Absolute Numeric Error	11
Mean Absolute Percent Error	11
Percent of units with Absolute Numeric Error 5+ children age 4	66
Percent of units with Absolute Percent Error 5%+	57
Source: Author's analysis of Demonstration Product released by the Census Bureau on March 16, 2022 after processing by IPUMS NHGIS at the, University of Minnesota www.nhgis.org	
* In this paper, errors reflect the difference between the 2010 Census data without and with DP injected.	
Data in this table does not include Puerto Rico or geographic units with zero population age 0 to 4 in 2010 Summery File or DP-Infused file.	

Data for Places

Census Places are geographic units used by the U.S. Census Bureau to publish data. They range from Places with millions of people such as Los Angeles and New York City, to the smallest villages and towns.

Places include both incorporated Places and Census Designated Places (CDPs). There are a little more than 29,000 Places for which the infusion of DP data was produced in the March 16, 2022 (DP Demonstration Product) and most of them (over 19,000) are Incorporated Places rather than Census Designated Places (CDPs). Incorporated Places are legally bounded entities such as cities, boroughs, towns, or villages (names may vary depending on the state). Census Designated Places (CDPs) are statistical entities used in the Census. They are unincorporated communities where there is a concentration of population, housing, and commercial structures and they are identifiable by name. There are nearly 10,000 CDPs for 2010 Census data.

Cities, villages, and towns might want to know about the number of young children in their area for things like planning youth activities, child facilities, and day care centers. The preschool-age population is also useful for forecasting future school enrollments.

The mean absolute numeric error for Places was 6 and the mean absolute percent error was 13.6 percent. The high percent error is not surprising because many of these Places are small. There were 1,422 Places where the number of young children was less than 100, and 9,012 Places where the number of young children was less than 500, based on the 2010 Summary File.

Figure 3 shows the distribution of Places by absolute percent error using the same thresholds used for Unified School Districts. The data in Figure 3 shows that almost half (46 percent) of Places had absolute percent errors of 5 percent or more for the young child population and 15 percent had absolute percent errors of 25 percent or more. Since Places are generally smaller (in population size) than Unified School Districts, it is not surprising that the percentages are larger for Places than for Unified School Districts.

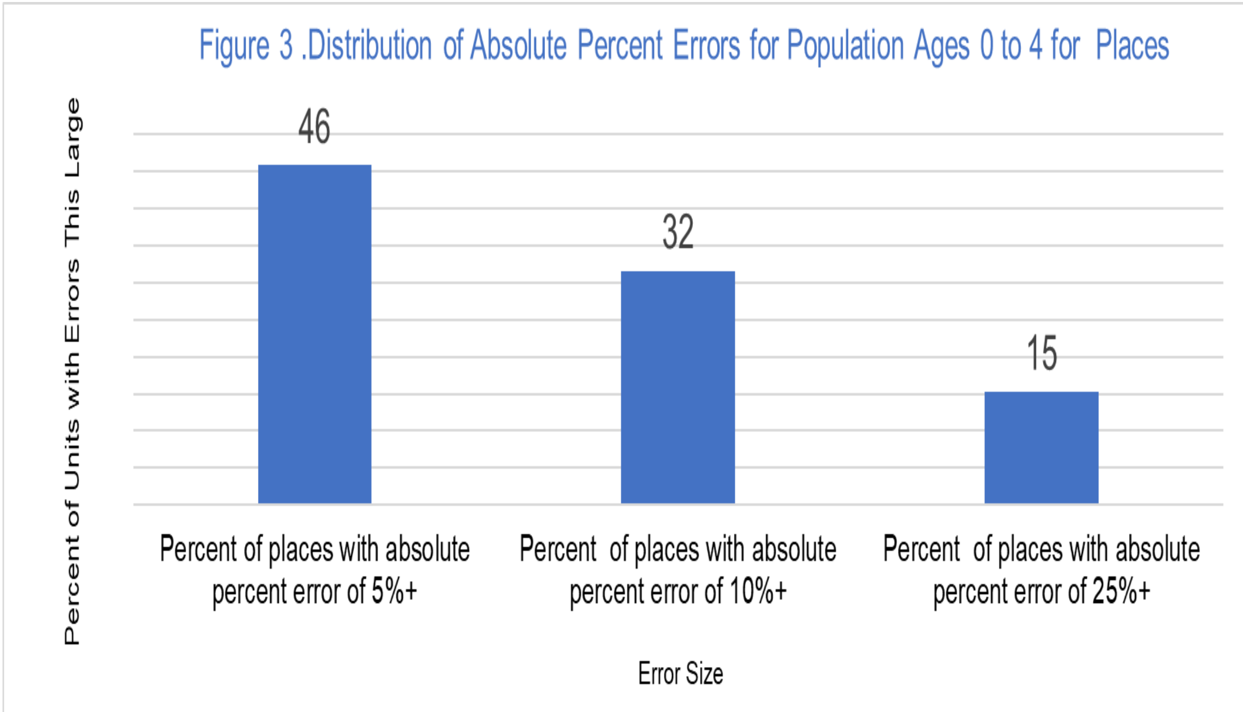


Figure 4 show the distribution of Places by absolute numeric errors using the same categories as Figure 2. Data show 39 percent of the Places had absolute numeric errors of 5 or more young children, and only 2 percent had absolute percent errors of 25 or more young children.

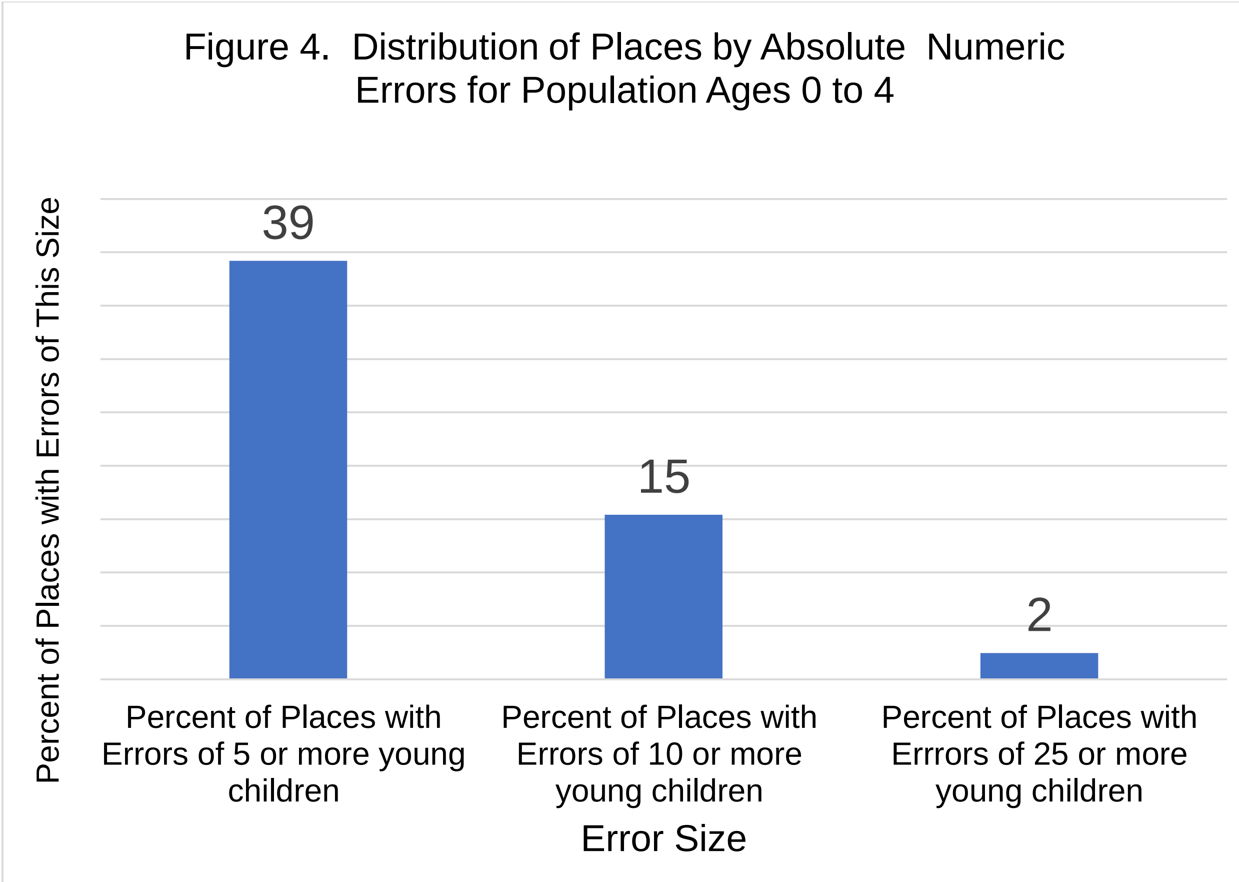


Table 6 shows states ranked on the percent of places in a state with absolute percent errors of 5 percent or more. Data for errors of 10 percent or more and 25 percent or more are also provided in the Table 6.

There is a lot of variation across the states. For example, 68 percent of the Places in Vermont had absolute percent errors of 5 percent or more, compared to 27 percent in New Jersey.

Table 6 States Ranked By Percent of Places is State with Absolute Percent Errors of 5 or more for Population Ages 0 to 4

		Number of Places in State	Percent Distribution Within State		
Rank		State Total	Absolute Percent errors of 5+	Absolute Percent errors of 10+	Absolute Percent errors of 25+
1	Vermont	117	68	50	23
2	New Mexico	414	65	53	30
3	New Hampshire	95	65	42	18
4	Montana	339	64	52	34
5	Alaska	307	62	48	26
6	North Dakota	342	62	50	31
7	Wyoming	182	61	51	26
8	South Dakota	352	61	48	27
9	Nebraska	543	59	44	26
10	Oklahoma	700	58	43	20
11	West Virginia	393	58	42	20
12	Arizona	426	58	46	23
13	Kansas	647	56	42	23
14	Maine	130	53	35	14
15	Iowa	977	51	35	17
16	Rhode Island	34	50	32	18
17	Nevada	122	49	41	26
18	Missouri	996	49	34	17
19	Arkansas	530	49	33	10
20	Virginia	587	48	33	18
21	Pennsylvania	1,741	48	34	15
22	Colorado	429	48	35	19
23	Minnesota	887	47	32	15
24	Kentucky	519	45	28	13
25	New York	1,178	45	28	11
26	North Carolina	731	44	30	13
27	Idaho	217	44	31	12
28	Connecticut	142	44	29	10
29	Washington	608	44	30	14
30	Maryland	502	43	32	21
31	Texas	1,714	43	29	16
32	Oregon	368	43	32	17
33	Wisconsin	760	43	28	13
34	Ohio	1,197	42	28	13
35	Michigan	686	42	28	11
36	Alabama	571	42	28	14
37	Indiana	675	42	27	12
38	South Carolina	392	40	26	12
39	Utah	319	40	27	9
40	Louisiana	469	40	24	9
41	California	1,458	39	28	16
42	Delaware	75	39	21	12
43	Illinois	1,359	39	24	9
44	Georgia	618	38	23	8
45	Massachusetts	242	37	20	6
46	Tennessee	427	35	22	7
47	Mississippi	362	35	20	8
48	Hawaii	150	34	23	7
49	Florida	909	31	18	8
50	New Jersey	536	27	19	9
	U.S Total	28,474	46	32	15

Source: Author's analysis of Demonstration Product released by the Census Bureau on March 16, 2022 after processing by IPUMS NHGIS at the, University of Minnesota www.nhgis.org

* In this paper, errors reflect the difference between the 2010 Census data without and with DP injected.

Data in this table does not include Puerto Rico or geographic units with zero population age 0 to 4 in 2010 Summery File or DP-Infused file.

there was 1 place with no state code

Table 7 shows states ranked on the percent of places in the state with absolute numeric errors of 5 or more young children. Data for 10 percent or more and 25 percent or more are also shown in the table. There is a lot of variation among the states. For example, 74 percent of places in Rhode Island have absoltue numeric errors of 5 or more young children comapred to 15 percent of North Dakota.

Table 7. States Ranked by Percent of of Places with Absolute Numeric Error of 5 or more for Population Ages 0 to 4					
			Percent of Places with Errors This Large		
Rank	Row Labels	Number of Places in the State	errors of 5 or more young children	erros of 10 or more young children	errors of 25 or more young children
1	Rhode Island	34	74	32	3
2	Maine	130	67	35	2
3	Hawaii	150	66	38	8
4	Connecticut	142	64	43	11
5	Massachusetts	242	61	35	7
6	California	1453	60	32	7
7	New Hampshire	95	58	28	5
8	Maryland	501	57	28	7
9	Virginia	587	56	30	9
10	Florida	908	53	28	5
11	New York	1178	53	23	3
12	Arizona	422	51	25	4
13	Washington	605	51	25	5
14	New Jersey	535	49	23	4
15	Nevada	121	48	23	4
16	Texas	1708	46	17	2
17	Vermont	117	45	21	2
18	Utah	317	45	19	3
19	Michigan	686	45	15	2
20	New Mexico	412	44	17	3
21	South Carolina	392	44	20	4
22	Delaware	75	43	20	5
23	Louisiana	470	43	20	4
24	Colorado	426	41	16	2
25	North Carolina	730	40	16	2
26	Oregon	365	40	14	2
27	Pennsylvania	1739	39	16	3
28	Ohio	1196	38	13	3
29	Georgia	617	37	14	2
30	West Virginia	392	36	11	1
31	Tennessee	427	36	14	1
32	Montana	339	33	11	0
33	Alabama	571	33	11	1
34	Oklahoma	698	32	7	0
35	Indiana	674	32	11	1
36	Kentucky	518	32	8	0
37	Wyoming	181	32	10	2
38	Wisconsin	758	31	9	1
39	Illinois	1356	31	9	1
40	Alaska	303	31	10	2
41	Mississippi	362	29	11	0
42	Idaho	216	26	8	0
43	Minnesota	883	26	5	0
44	Arkansas	529	24	5	0
45	Kansas	641	24	4	0
46	Missouri	987	24	5	1
47	South Dakota	350	20	7	1
48	Iowa	968	18	1	0
49	Nebraska	536	15	2	0
50	North Dakota	332	15	3	0
	U.S Total	28,476	39	15	2
Source: Author's analysis of Demonstration Product released by the Census Bureau on March 16, 2022 after processing by IPUMS NHGIS at the, University of Minnesota www.nhgis.org					
* In this paper, errors reflect the difference between the 2010 Census data without and with DP injected.					
Data in this table does not include Puerto Rico or geographic units with zero population age 0 to 4 in 2010 Summery File or DP-Infused file.					
there was 1 place with no state code					

Discussion

It is clear the introduction of DP into the 2020 Census has caused a lot of controversy. I have been following the U.S. Census since 1970, and I do not remember any issue that has caused as much discussion, concern, and debate among data users as the decision to implement DP in the 2020 Census.

Below I review a couple of issues regarding DP that were not addressed in my analysis but may impact stakeholders view of DP

Block-Level Data

Blocks are the smallest geographic unit used in the Census and there are about 8 million blocks in the 2020 Census but only about 6 million are occupied. The average block has a total population of about 41 people and about 3 young children. The small population size of blocks makes them susceptible to large percent errors when random numbers are injected with DP.

The availability of errors at the Census block level makes DP different that normal assessment of Census accuracy. Assessment of Census accuracy using the two standard Census Bureau methods (Demographic Analysis and Post-Enumeration Survey) is only available at the national level as this report is being written (state data will come out soon from PES but there will be no substate data from these methods). But the DP Demonstration Product allows one to look at errors for all levels of Census geography down to the census block level.

There are two broad perspectives on the error DP injects into census blocks. One perspective is that data for census blocks are among the most important data supplied by the Decennial Census, and they need to be as accurate as possible. One of the primary purposes of the Decennial Census is to provide comparable population figures for small areas across the country. To the best of my knowledge, there is no other data source that provides demographic data for all the blocks in the country other than the Decennial Census. Consequently, census accuracy for blocks is especially important. O'Hara (2022) makes a strong case for why block level data are important in terms of creating special or custom districts. The need for such data is often not apparent until well after the Census data has been collected and reported.

Another perspective holds that blocks are typically aggregated into larger units like congressional districts, cities, and counties and in those aggregations the random error injected into blocks cancel each other out and produce relatively accurate data for larger units. From this perspective, errors at the block level are not so important.

Regarding the usability of block level data, the Census Bureau (Devine 2022, slide 17) recently stated, "Block-level data are fit-for-use when aggregated into geographically contiguous larger entities. They are not intended to be fit-for-use as a unit of analysis." It seems likely the high level of inaccuracy for census blocks based on analysis of the DP Demonstration Product influenced the issuance of this statement.

I do not think there is any dispute that the error injected by DP for blocks produces a relatively high absolute percent error and that these errors typically cancel each other out when blocks are aggregated into larger areas. Because the error is random, the amount of error does not become cumulative. It is an open question about how important census block level data are for making decisions.

One problem with use of DP for small areas is the implausible or impossible results. I did not have the computer power to examine blocks for age 0 to 4 in the March 2022 Demonstration Product, but Census Bureau 2022 data show heavy distortions at the block level . For example, more than 163,000 blocks have children (population age 0 to 17) but no adults (population age 18 and over) after DP is applied compared to just 82 such blocks before DP was applied (U.S. Census Bureau 2022_). Many such cases are highly unlikely and raise questions about who these children are living with if there are no adults in their household. The Census Bureau (2022d) offers several other examples of implausible or impossible results in the data after DP is applied

It is not clear to me exactly what statistical problems might be caused by these results, but they undermine the veracity of the census data broadly. A high number of improbable results is identified as a problem of "legitimacy" rather than statistical accuracy by Hogan (2021) and is likely to undermine the confidence the public has in the Census results. When data users see highly implausible results like the large number of blocks with children and no adults, they often wonder what other errors are in the data that are not so apparent.

Despite the statement by the Census Bureau and misgivings among some demographers about the quality of census block data, many data users routinely use the block level data, either because they do not realize the level of potential errors, or because it is the best (or only) data they have at that level.

The data indicate the average percent errors for census blocks is relatively high but does not address how often block-level data are used in decision making. Readers may have their own answer to that question.

Breaking the Link Between Child and Parents

The production of many blocks where there are children, but no adults may be related to the link between children and adults in a household that is broken when 2020 DAS with DP was applied to the DHC file. DP is administered to children and parents independently, so it may eliminate the adults in a household that has children by randomly subtracting data from the number of adults. If the processing retained the link between young children and their parents in a household, it is doubtful that there would be such a high number of blocks with children and no adults.

This statistical disconnection of children and parents is an on-going concern and is likely to have important impacts in later Census products which have more detailed data on young children.[2] For example the connection between children and parents is critical for a lot of data from the American Community Survey. Child poverty is probably the single most important measure of child well-being and determining poverty status requires linking a child to the income of the adults in the households.

The Census Bureau says it will use a different method of DP in the Detailed Demographic and Housing File which will retain the connection between children and

[2] It is my understanding that the use of DP does not necessarily require the disconnect between children and parents in a household. The break between children and parents in the redistricting file and the DHC is a result of the particular DP-related processing chosen by the Census Burau.

parents. Hopefully, that will alleviate our concerns. But data that links children and adults in the Detailed Demographic and Housing file will not be available until late 2023 or 2024. That is getting very close to the date (2025) the Census Bureau said it might start applying DP to the American Community Survey (ACS) Translating the application of DP from the Census to the ACS, is likely to be a difficult process.

Accuracy and Equity

The focus of this report is on census accuracy, but the differential accuracy raises the issue of equity. Equity in terms of data provision has become a more visible aspect of data collection and reporting in the federal government recently (White House Equitable Data working Group 2022). According to the U.S. Census Bureau (2021e, pages 1) " The Census Bureau has an ongoing commitment to producing data that depict an accurate portrait of America, including its underserved communities." Data equity has become a part of broader equity questions. This suggests all results should be examined through the lens of equitable data.

In terms of equity, Figure 1 shows substantial differential accuracy for Unified School Districts by race and ethnicity after DP infused. For Hispanic young children, the mean absolute percent error was 27, for Black young children the mean absolute percent error was 34, and for Asian young children was 42, compared to 5 for Non-Hispanic white children. What does this say about the equity of using the DP method? There is already differential accuracy in census results before DP is applied but it may be the case that DP exacerbates such inequities. Is if fair to inject as much error for groups that already have a lot of error in census data as for those groups that do not have much error? Did the Census Bureau examine equity concerns when they decided to use DP in the 2020 Census?

Selection of a DAS and Public Trust

Disclosure avoidance is not just a statistical issue and examining it only from a statistical perspective may be problematic. Another dimension for assessing alternative DAS methods is the extent to which a given DAS method undermines public trust in the Census data and the Census Bureau itself. There has been a great deal of concern about the erosion of public trust in the Census Bureau recently. According to the National Academy of Sciences, Engineering and Medicine panel assessing the 2020 Census (2022, page 6),

"We are very concerned, based on presentations to the panel and our knowledge of reactions to previous demonstration data, that the Census Bureau's adoption of differential privacy-based disclosure avoidance has increased the level of public mistrust in the 2020 Census and the Census Bureau itself."

In their review of the impact DP has had on the Census Bureau credibility and trust among data users, Boyd and Sarathy (2022, page 1) state, "We argue that rebuilding trust will require more than technical repairs or improved communication: it will require reconstructing what we identify as a "statistical imaginary."

Summary

This report provides information on accuracy of DP-infused data and provides a profile of the likely errors for young children that will be seen in data for in the 2020

Census if the Census Bureau uses the privacy protection parameters reflected in the March 2022 Demonstration Product.

It is important to note that the analysis provided in this paper is just a sample of analyses that could be done. But I believe the data analyzed in this study a relatively good sample of the broader implications of using a DAS method with DP in the DHC with the privacy protection parameters used in this Demonstration Product.

But there are many other data factoids that could have been produced to shed light on the implications of DP. For example, the Census Bureau shows the mean absolute percent error for foster children at the county level is 122 percent and at the Incorporated Place level is 96 percent after DP has been applied. Foster children are very vulnerable population, and this level of error is disturbing.

The question that is not addressed in the previous sections is whether the level of error reflected in this analysis would make 2020 Census for data on young children "unfit for use." Each person will probably have a different answer to how much error in census data for young children is too much error.

Like all disclosure avoidance systems, the use of DP involves a trade-off between privacy protection and census accuracy. There have always been errors in the Census data, but in the 2020 Census, the Census Bureau is trying to decide how much additional error to add to the data in order to enhance privacy protection. By setting privacy parameters, the Census Bureau has control over the level of accuracy and level of privacy protection in the 2020 Census.

Given this balancing act, it would be useful to have more information about two aspects of DP: 1)metrics on privacy protection and 2)information on potential harm of re-identification that DP is designed to protect against. It would be helpful if we could compare the metrics of accuracy to metrics of privacy protection in the March 2022 Demonstration Product. I see many measures of accuracy based on the Demonstration Product. However, I don't see any privacy protection metrics produced by the Census Bureau nor do I see a way to explore the privacy protection aspect with the Demonstration Product. It seems the balance of accuracy and privacy protection is the key reason for using a given disclosure avoidance systems but without metrics for privacy protection I am not sure how to do that.

In assessing the tradeoff between privacy and accuracy it is not clear exactly what harm might be done by a re-identification of someone based on Census data. My recommendation below might be different if lower privacy protection meant hundreds of innocent people would go to jail versus few people getting annoying phone calls. But I have not seen any evidence on this question from the Census Bureau. When I have asked experts about the level of privacy protection afforded by an Epsilon of 19.6 in the redistricting data in terms I can understand it seems like I always get a variation of "it depends." But no metrics.

On the other hand, the problems that are likely to be caused by inaccurate census data on young children are clearer to me. The data in this paper, and many other analyses, provide a rich set of metrics showing the magnitude of error DP injects into Census data and I can envision problems such errors might cause.

When the number of young children in a school district is under-reported by 5 or 10 percent, that could have big implications for their funding and when the number of young children in a community is off by 10 percent or more, that could impact planning in ways that waste taxpayer money and undermine quality education for young children. If the number of young children reported in the Census for a Unified School District is 10 percent too low, it may not automatically translate into 10 percent less money for that jurisdiction. But there is a strong link between underreporting the number of young children and the loss of money in a general sense.

In addition to the money distributed on the basis of census-derived data, Census data are used for many decisions in the public and private sector. The more errors there are in the data, the less likely those decisions will be correct ones.

Given the level of errors in Unified School Districts and Places using the privacy protection level in the most recent DP Demonstration Product, and the lack of clear evidence or measurements about the level or impact of privacy loss, I recommend that the Census Bureau increase the level of accuracy used in the DHC to provide more accuracy small area data for young children.

Author Note

It should be noted that this analysis is not as full and complete as it should be because time did not allow such an analysis. The Census Bureau released the latest DP Demonstration Product on March 16, 2022 but had to re-release it on Apirl 14 because of mistake They request responses by May 16th

Since stakeholders need time to read and absorb this paper it needed to be available well before May 16th. If there had been more time for analysis there is a lot more that could have been done. The data used here could be developed to provide a more granular picture of DP's impact. For example, one could calculate the measures shown here for all counties or all Places within a state, or one could develop the measures for all census tracts within a county.

If more time had been available, it would have been useful to explore data for race and Hispanic groups more thoroughly. Also, it would have been useful to examine accuracy measures for geographic units of different population sizes. If I had more time, I would have used race alone or in combination rather than race alone. There is a good deal more that could be done to provide state-specific data.

It is unfortunate that the time limitations mean the Census Bureau will not receive the quality of feedback they seek.

References

Bouk, D. and Boyd, D. (2021*). Democracy's Data Infrastructure.; The technologies of the U.S. Census*. https://knightcolumbia.org/content/democracys-data-infrastructure

Boyd. D. (2019). "Balancing Data Utility and Confidentiality on the 2020 US_Census," Data and Society, https://datasociety.net/library/balancing-data-utility-and-confidentiality-in-the-2020-us-census/ .

Boyd, D. and Sarathy, J (2022) "Differential Perspectives: Epistemic Disconnects Surrounding the US Census Burau's Use of Differential Privacy," *Harvard Data Science Review* (forthcoming)

Committee on National Statistics (2019). "Workshop on 2020 Census Data Products: Data Needs and Privacy Considerations," presentations are available at https://www.nationalacademies.org/event/12-11-2019/workshop-on-2020-census-data-products-data-needs-and-privacy-considerations .

Cropper, M. McKibben and Stojakovic, Z. (2021). The Importance of Small Area Census Data for School Demographics, Count all Kid website https://ednote.ecs.org/counting-all-kids-how-the-census-impacts-education/

Hogan, H. (2021). "The History of Assessing Census Quality, Presentation at 2021 Population of Association of America Conference, May 5, 2021.

Hotz, J. and Salvo J. (2020). Addressing the Use of Differential Privacy for the 2020 Census: Summary of What We Learned from the CNSTAT Workshop. https://www.apdu.org/2020/02/28/apdu-member-post-assessing-the-use-of-differential-privacy-for-the-2020-census-summary-of-what-we-learned-from-the-cnstat-workshop/ ,

McElrath, K. Bauman, K., and Schmidt, E (2022) Preschool Enrollment in the United states,: 2055 to 2019," U.S. Census Bureau https://www.census.gov/content/dam/Census/newsroom/press-kits/2021/paa/paa-2021-presentation-preschool-enrollment-in-the-united-states.pdf

Nagle, N. and Kuhn, T. (2019). "Implications for School Enrollment Statistics." https://www.nationalacademies.org/event/12-11-2019/workshop-on-2020-census-data-products-data-needs-and-privacy-considerations.

National Academy of Sciences, Engineering and Medicine, (2022). *Understanding the Quality of the 2020 Census, Interim Report* , Washington Dc. The National Academy Press, https://nap.nationalacademies.org/catalog/26529/understanding-the-quality-of-the-2020-census-interim-report

O'Hara, A. (2022) presentation at Analysis of Census Noise Measurements Workshop, April 28-29, Rutgers University.

O'Hare, W.P. (2019). Assessing 2010 Census Data with Differential Privacy for Young Children," https://www.nationalacademies.org/event/12-11-2019/workshop-on-2020-census-data-products-data-needs-and-privacy-considerations .

O'Hare W. P. (2020a). "Many States Use Decennial Census Data to Distribute State Money, The Census Project Website https://thecensusproject.org/2020/01/09/many-states-use-decennial-census-data-to-distribute-state-money/

O'Hare, W.P (2020b). "Implications of Differential Privacy for Reported Data on Young children in the 2020 U.S. Census," Posted on Count All KIDS Website Implications-of-Differential-Privacy-for-kids-11-17-2020-FINAL-00000003.pdf (myftpupload.com) .

O'Hare, W.P. (2021). "Analysis of Census Bureau's August 2021 Differential Privacy Demonstration Product: Implications for Data on Children," Count All Kids website November *https://countallkids.org/resources/analysis-of-census-bureaus-august-2021-differential-privacy-demonstration-product-implications-for-data-on-children/*

O'Hare, W. P. (2022a). "New Census Bureau Data Show Young Children Have a High Net Undercount in the 2020 Census, " Posted on Count All Kids website , March, https://countallkids.org/resources/new-census-bureau-data-show-young-children-have-a-high-net-undercount-in-the-2020-census/

O'Hare , W.P (2022b). U.se of the American Community Survey Data by State Child Advocacy Organizations? Count All Kids website, https://countallkids.org/resources/use-of-the-american-community-survey-data-by-state-child-advocacy-organizations/

Reamer, A. (2020). Counting for Dollars, George Washington University https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds .+

U.S. Census Bureau (2018), "Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing," THE RESEARCH AND METHODOLOGY DIRECTORATE, Mc Kenna, L. U.S. Census Bureau, Washington DC., https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf .

U.S. Census Bureau (2019). "2010 Demonstration Data Products," U.S. Census Bureau, Washington DC., October, https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html .

U.S. Census Bureau (2020a). 2020 Census Disclosure Avoidance Improvement Metrics, U.S Census Bureau, Washington DC., March 18, https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/disclosure-avoidance-system/2020-03-18-2020-census-da-improvement-metrics.pdf?# .

U.S. Census Bureau (2020b), "2020 Census Data Products and the Disclosure Avoidance System", Hawes M. and Garfinkel. S. L., Planned presentation at the Census Scientific Advisory Committee meeting, March 26.

U.S. Census Bureau (2020c) DAS Updates, U.S Census Bureau, Hawes M. June 1 Washington DC., https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/disclosure-avoidance-system/2020-06-01-das-updates.pdf?# .

U.S. Census Bureau (2020d). "Disclosure Avoidance and the Census," Select Topics in International Censuses, U.S. Census Bureau, October 2020. https://www.census.gov/library/working-papers/2020/demo/disclos-avoid-census.html .

U.S. Census Bureau (2020e). "Disclosure Avoidance and the 2020 Census, U.S. Census Bureau," Washington DC., Accessed November 2, https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html .

U.S. Census Bureau (2020f). "Error Discovered in PPM," U.S. Census Bureau, Washington DC. https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html .

U.S. Census Bureau (2020g), "2020 Disclosure Avoidance System Updates," U.S. Census Bureau, Washington DC., https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html .

U.S. Census Bureau (2021a) School Enrollment in the United States: October 2019 - PowerPoint Presentation (census.gov Detailed Tables, School Enrollment in the United States: October 2019 - Detailed Table 1, FEBRUARY 02, 2021.

U.S. Census Bureau (2021b) Developing the DAS: Demonstration Data and Progress Metric, https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-development.html .

U.S. Census Bureau (2021c). "Differential Privacy 101." Webinar May 4, 2021, Michael Hawes. https://www.census.gov/data/academy/webinars/2021/disclosure-avoidance-series/differential-privacy-101.html

U.S. Census Bureau (2021d). " Disclosure Avoidance for the 2020 Census: An Introduction," U.S. Census Bureau, Washington, DC. November https://www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html

U.S. Census Burearu (2021e) “Advancing Equity with Census Bureau Data.” Census Bureau Blog, November 2, 2021, Ron Jarmin , Acting Director Advancing Equity With Census Bureau Data

U.S. Census Bureau (2021f), “Disclosure Avoidance for the 2020 Census: An Introduction,” November 2021, U.S. Census Bureau, Washington DC https://www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html

U.S. Census Bureau (2022a) “Understanding Disclosure Avoidance- Related Variability in the 2020 Census Redistricting data, “ U.S. Census Bureau, Washington DC. January 28. https://www.census.gov/library/fact-sheets/2022/variability.html

U.S. Census Bureau (2022b) “Revised Data Metrics for 2020 Disclosure Avoidance,” U.S. Census Bureau, Washington DC.

U.S. Census Bureau (2022c) Post-Enumeration Survey and Demographic Analysis Help Evaluate 2020 Census Results, March 10 , Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census

U.S. Census Bureau (2022d) Detailed Summary Metrics

U.S. General Accountability Office (2020). “COVID-19 Presents Delays and Risks to Census Counts,” U.S. General Accountability Office, Washington, DC., https://www.gao.gov/products/GAO-20-551R .

Vink, J. (2019). “Elementary School Enrollment,” https://www.nationalacademies.org/event/12-11-2019/workshop-on-2020-census-data-products-data-needs-and-privacy-considerations .

White House Equitable Data Working Group (2022) “A Vision for Equitable Data : Recommendations from the Equitable Data Working Group,” https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf

State of preschool 2021. https://nieer.org/wp-content/uploads/2021/08/YB2020_Full_Report_080521.pdf