Analysis of Unified School Districts and Places with Large Errors for the Population Ages 0 to 4 Caused by Application of Differential Privacy

By

Dr. William P. O'Hare

In my analysis of the DP Demonstration Product released by the Census Bureau on August 25, 2022 (O'Hare 2022) I found many Unified School Districts and Places[1] where DP introduced large errors for the population ages 0 to 4. The population ages 0 to 4 is referred to as young children in this paper. Large errors are defined in this paper as absolute *percent* errors of more than 25 percent or absolute *numeric* errors of more the 25 young children.[2]

This paper focuses on those units that have an absolute *percent* error of 25 percent or more and those units that have absolute *numeric* errors of 25 or more young children (age 0 to 4). In a recent presentation, the Census Bureau (2022, slide 3) describes DP as, "DP inserts small differences into counts of people and households, making it very difficult to identify people." In my opinion, errors of 25 percent or more are not "small".

Numeric errors reflect the difference in 2020 Census results with and without DP. Absolute *percent* difference converts the *numeric* difference to a relative difference by dividing by the 2010 Census Summary File value and multiplying by 100.. It may be a

[1] These include incorporated Places as well and Census Designated Places.

[2] Errors in this application are the difference in the 2010 Census numbers with and without the application of DP.

bit confusing presenting both *numerical* and *percent* errors, so I italicize the terms for help readers more easily distinguish which measure is being discussed.

Table 1 shows the number of Unified School Districts and Places that have large errors. They account for a non-trivial share of all Unified School Districts and Places based on the DP Demonstration Produce released by the Census Bureau in August 2022. . For background, there were about 11,000 Unified School Districts and about 29,000 Places in the 2010 Census which the database reflected in the DP Demonstration Produce released by the Census Bureau in August 2022.

Table 1. Unified School Districts and Places with Large Errors for the Population Age 0 to 4 Due to the Application of Differential Privacy

	Units with Absolute *Numeric* Errors of 25 or more young children		Units with Absolute *Percent* Errors of 25 Percent or More	
	Number	Percent of All Units	Number	Percent of All Units
Unified School Districts	590	5	214	2
Places	564	2	4,209	15

Table 1 shows there are 590 Unified School Districts with absolute *numeric* errors of 25 or more young children and those 590 Unified School Districts are 5 percent of all Unified School Districts. Table 1 shows there 564 Places with absolute *numeric* errors of 25 or more young children and those 564 Places are 2 percent of all Places.

Table 1 shows 214 Unified School Districts with an absolute *percent* error of 25 percent or more young children and they are 2 percent of all United School Districts.

Analysis found 4,209 Places with absolute *percent* errors of 25 percent or more and those 4,209 Places are 15 percent or all places.

If most of the large *percent* errors are based on very small *number*s, they may not too concerning. But the tables below, showing the relationship between large *percent* errors and large *numeric* errors. indicates many units with a high absolute *percent* error also have large absolute *numeric* errors as well.

Table 2 shows there are 590 Unified School Districts with absolute *numeric* errors of 25 or more young children. About 40 percent of these 590 Unified School Districts have absolute *percent* errors of 5 percent or more. In other words, many of the Unified School Districts with large absolute *numeric* errors also have relatively large absolute *percent* errors as well.

Table 2. Distribution of Unified School Districts with Absolute *Numeric* Errors of 25 of More Young Children by Size of Absolute *Percent* Error

Absolute *Percent* Error	Number	Percent of Total
0 to 4.9 Percent	346	59
5 to 9.9 Percent	135	23
10 to 24.9 Percent	97	16
25 Percent or more	12	2
Total	590	100

Table 3 shows the distribution of absolute *numeric* errors for all Unified School Districts with an absolute *percent* error of 25 percent or more. Almost two-thirds (65

percent) of Unified School Districts with absolute *percent error* of 25 percent or more also have absolute *numeric* errors of 5 or more young children. This shows that most of the Unified School Districts with large absolute *percent* errors also have relatively large absolute *numeric* errors.

Table 3. Distribution of Unifed School Districts With Absolute *Percent* Errors of 25 Percent or More by Size of Absolute *Numeric* Error		
Absolute *Numeric* Error	Number	Percent of Total
0 to 4 young children	74	35
5 to 9 young children	59	28
10 to 24 young children	77	36
25 or more young children	4	2
Total	214	101*
* total is not 100% because of rounding.		

Table 4 shows there are 564 Places with absolute p*ercent* errors of 25 Percent or more. More than half (55 percent) of these 564 Places have absolute *numeric e*rrors of 5 or more young children. In other words, many of the Places with large absolute *numeric e*rrors also have relatively large absolute *percent* errors as well.

Table 4 Distribution of Places with Absolute *Numeric* Errors of 25 or More Young Children by Size of Absolute *Percent* Error		
Absolute *Percent* Error	Number	Percent of Total
0 to 4.9 Percent	256	45
5 to 9.9 Percent	96	17
10 to 24.9 Percent	138	24
25 Percent or more	74	13
Total*	564	99

* Total is not 100% because of rounding.

Table 1 shows there are 4,209 Places with an absolute *percent* error of 25 percent or more for young children and they represent 15 percent of all Places. Table 5 also shows the distribution of absolute *numeric* errors for Places with an absolute *percent* error of 25 percent or more. Most (56 percent) had absolute *numeric* errors of 5 or more young children. Table 5 shows there are 74 Places with absolute percent errors of 25 percent or more AND Absolute *numeric* errors of 25 or more young children.

Table 5. Distribution of Places With Absolute *Percent* Errors of 25 Percent or More by Size of Absolute *Numeric* Errors

Absolute *Numeric* Error	Number	Percent of Total
0 to 4 young children	2,267	54
5 to 9 young children	1,222	29
10 to 24 young children	646	15
25 or more young children	74	2
Total	4,209	100

I believe geographic units with large errors will be the biggest problem in terms of the application of DP to the 2020 Census, particularly if such errors are accompanied by changes in funding. I understand DP is a complicated methodology, but it is not clear to me why the Census Bureau cannot truncate the distribution from which it draws the random numbers used in applying DP, to make sure large errors are not injected into the reported data. If the errors injected by DP could be kept to less than 5 percent or less than 5 people, I believe the application of DP in the 2020 Census would be much more acceptable.

References

O'Hare. W. P. (2022). Analysis of Census Bureau's August 2022 Differential Privacy Demonstration Product: Implications for Data on Young Children, (Sept) . https://secureservercdn.net/198.71.233.229/2hj.858.myftpupload.com/wp-content/uploads/2022/09/Implications-of-Differential-Privacy-for-kids-9-21-2022-FINAL-.pdf

U.S. Census Bureau, (2022) "Demographic and Housing Characteristics File (DHC) Update," Presentation at the Census Bureau's National Advisory Committee meeting September 23, 2022.