National Academy of Sciences: 2020 Census Data Products

Impact of DP on Data about Young Children (ages 0-4)

June 21, 2022

Dr. William P. O'Hare

Consultant to the 2020 Census Count All Kids Campaign

billohare1@gmail.com

COUNT ALL kids CENSUS 2020

Outline of Presentation

- Why focus on Census data for young children?

- Impact of differential privacy on data for young children in the 2020 Census DHC Demonstration files

  My focus is on small geographic units with child policy and programmatic authority

  1) Unified School Districts

  2) Places

- Issues
  - Equity
  - DP processing separating children and parents

Key Terms

- In this presentation, errors reflect the difference between the 2010 Census data with and without DP applied.
- Young children are those ages 0 to 4
- Data from IPUMS at University of Minnesota

Why focus on young children?

- The net undercount rate for young children in the U.S. Census is high and it has been growing

- Many implications of data on young children for public policies and programs
  - Age 0 to 4 are future students
  - Many young children in preschool and childcare
    - 5 million children enrolled in preschool in 2019
    - 59% of children ages 0 to 4 in childcare

Figure 1 Census Net Undercount Rates for Young Children (ages 0 to 4), All Children (ages 0 to 17), and Adults (ages 18 and over): 1950 to 2020

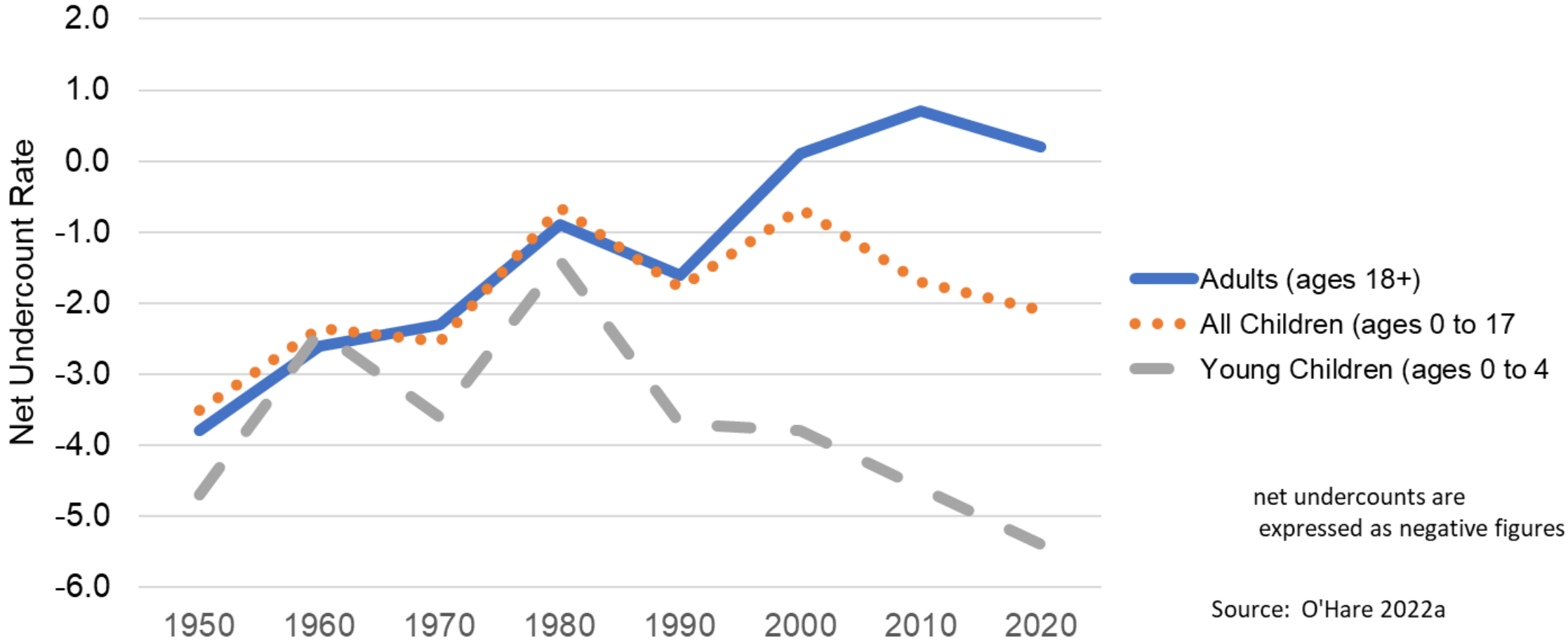
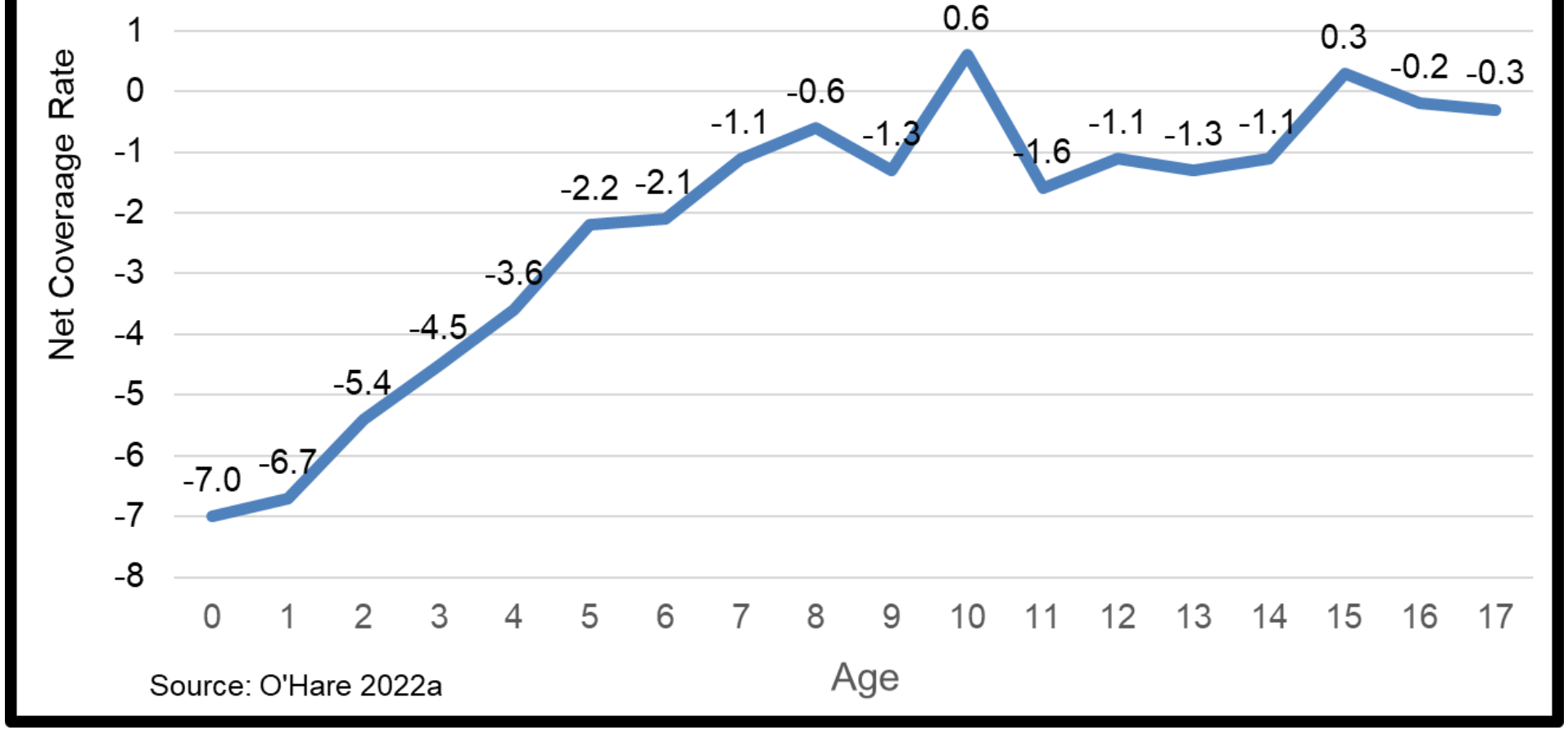


Figure 2. Net Coverage Rates for Children By Single Year of Age in 2020 Census



Source: O'Hare 2022a

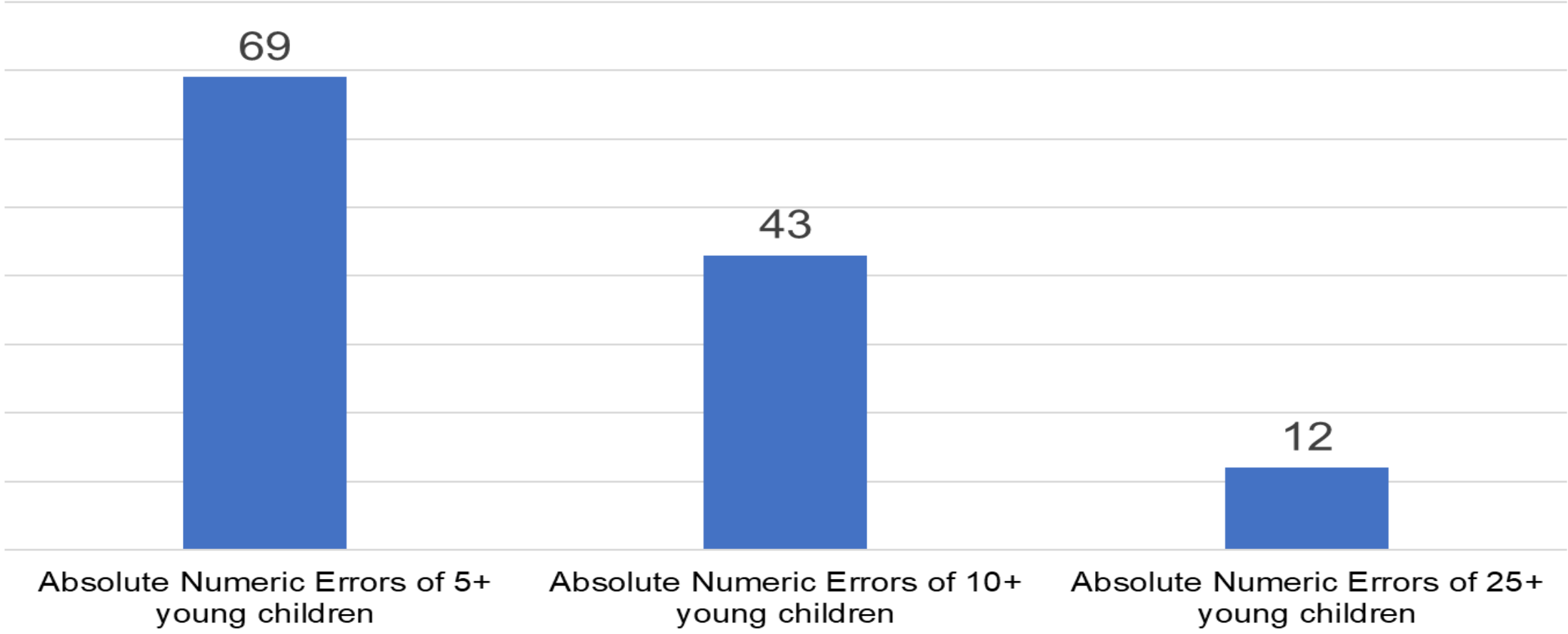
Table 1. Key Statistics for Absolute Numeric and Absolute Percent Errors for Children Ages 0 to 4 for States, Counties, School Districts and Places

	States	Counties*	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: O'Hare 2022b

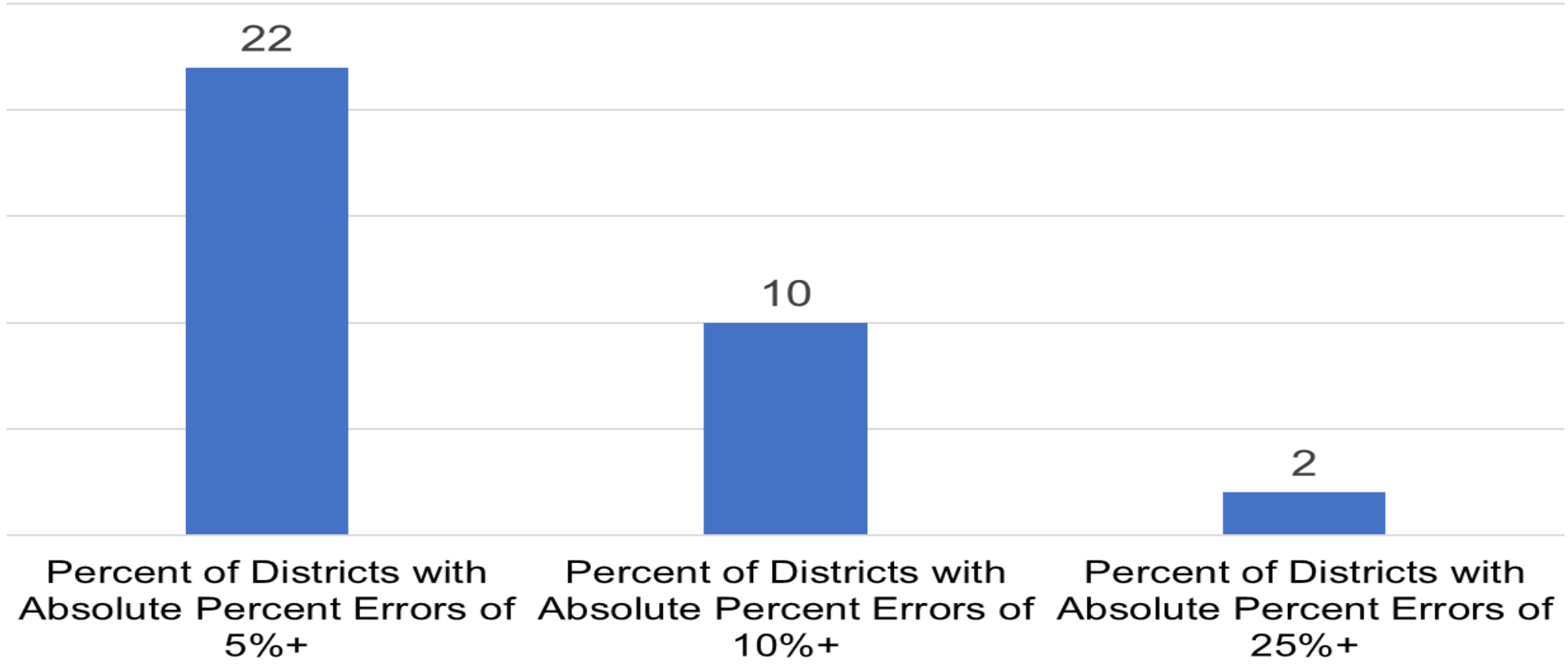
* There are 302 counties with total population less than 5,000.

Figure 3. Percent of School Districts with Given Level of Absolute Numeric Error for Population Ages 0 to 4



Source: O'Hare 2020b

Figure 4. Percent of School Districts with Given Level of Absolute Percent Error for Population Ages 0 to 4



Source: O'Hare 2022b

Figure 5. Distribution of Absolute Numeric Errors for Population Ages 0 to 4 for Unified School Districts by Race and Hispanic Origin

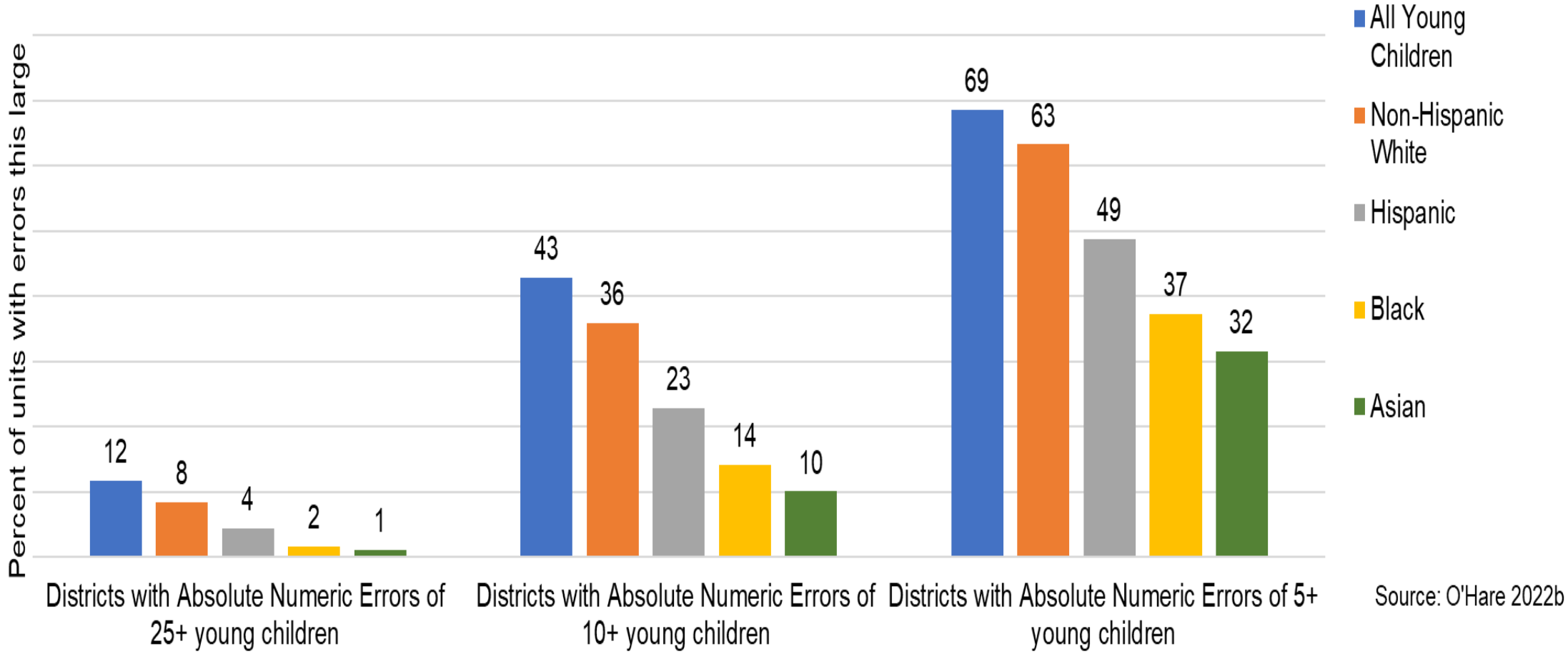


Figure 6. Distribution of Absolute Percent Errors for Population Ages 0 to 4 for Unified School Districts by Race and Hispanic Origin

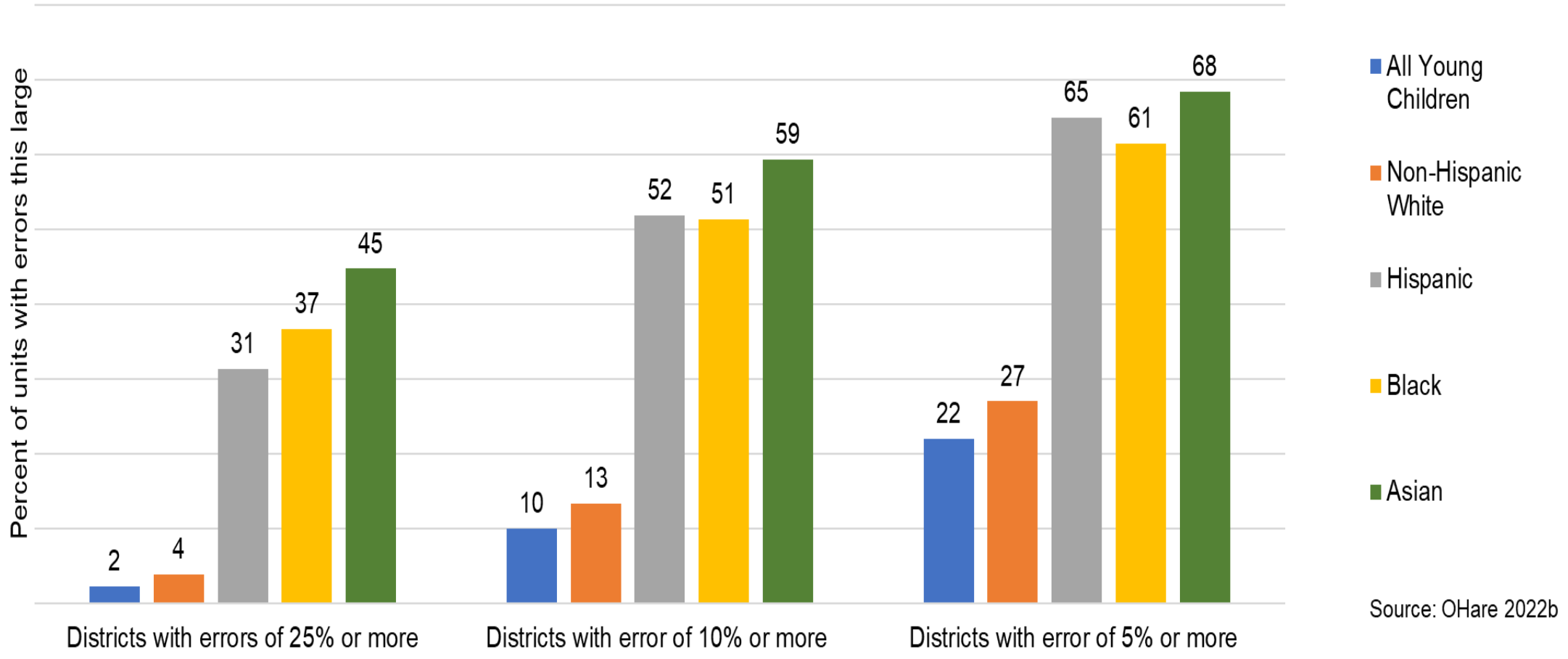


Table 2. Ten States with Highest School District Mean Absolute Percent Errors for Population Ages 0 to 4		
Rank	State	Mean Percent Error for Ages 0 to 4
1	South Dakota	6.7
2	Nevada	6.4
3	New York	5.9
4	Oklahoma	5.7
5	New Hampshire	5.4
6	Iowa	5.3
7	Texas	5.2
8	North Dakota	5.1
9	Alaska	5.1
10	Wisconsin	4.8
Source: O'Hare 2022b		

Table 3. Ten States with the Highest and Lowest Share of School Districts with Errors of 5% or More for Population Ages 0 to 4

Ten States with Highest Percent of School Districts with Errors of 5%+			Ten States with Lowest Percent of School Districts with Errors of 5%+	
State	Share with errors of 5%+		State	Share with errors of 5%+
South Dakota	33		Connecticut	12
Iowa	31		North Carolina	12
Alaska	30		New Jersey	11
Nevada	29		Alabama	11
Arkansas	29		Rhode Island	10
New York	29		Virginia	9
Montana	28		Florida	9
North Dakota	28		Louisiana	6
Colorado	27		Delaware	0
Oklahoma	27		Hawaii	0

Source: O'Hare 2022b

Table 4 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%+ children age 4	57
Source: O'Hare 2022b	

What are 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. Analysis includes incorporated and unincorporated places.

- The vast majority of places are small
  - Over 9,000 have fewer than 500 children ages 0 to 4

Figure 7. Distribution of Places by Absolute Numeric Errors for Population Ages 0 to 4

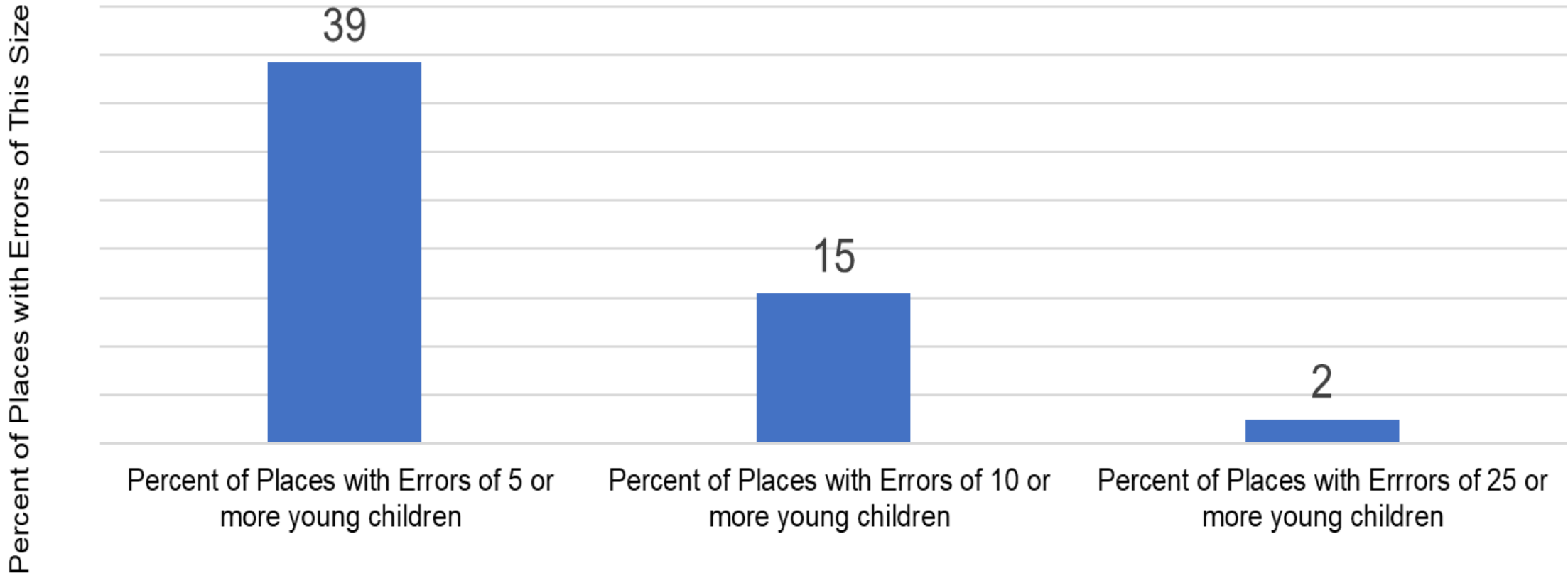
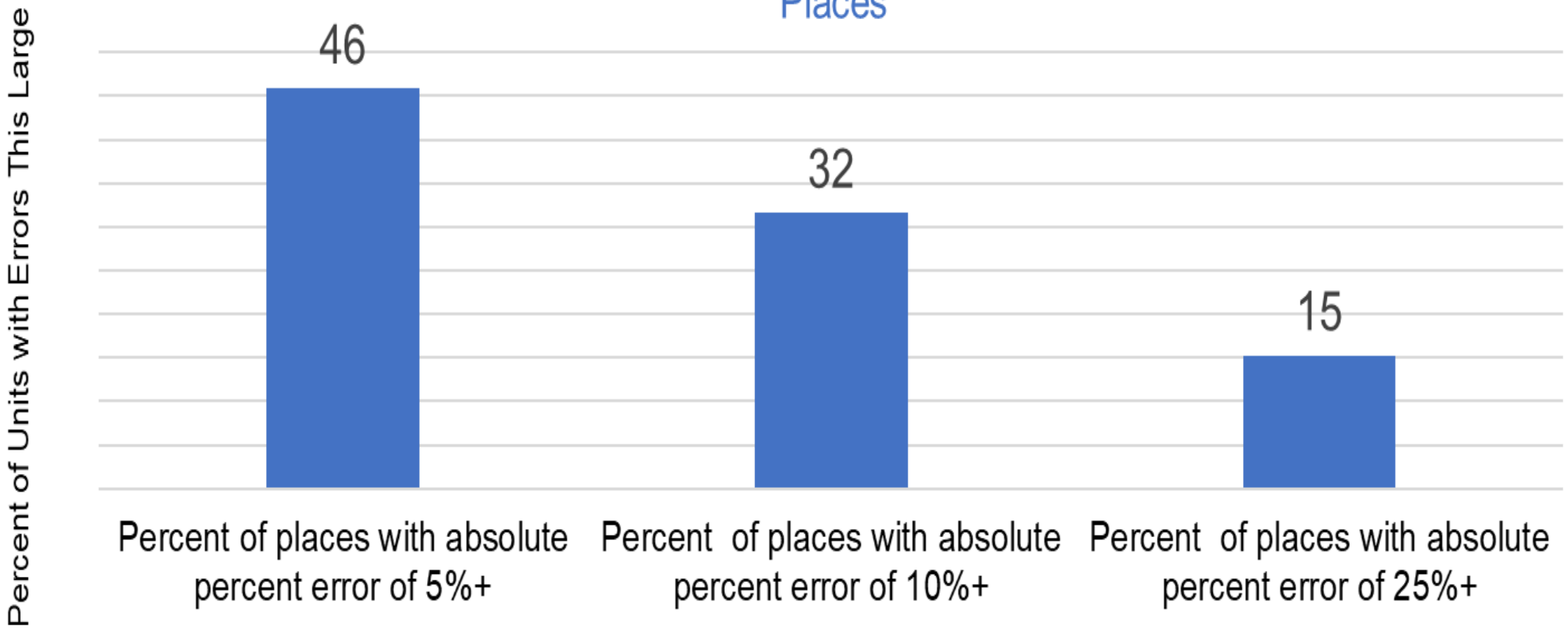


Figure 8. Distribution of Absolute Percent Errors for Population Ages 0 to 4 for Places



Source: O'Hare 2022b

Table 5. Top Ten States By Percent of Places in State with Absolute Percent Errors of 5% for Population Ages 0 to 4

Rank	State	Percent of Places with Absolute Percent Errors of 5%+
1	Montana	32
2	Maine	25
3	North Dakota	24
4	Washington	20
5	Nebraska	19
6	South Dakota	19
7	Oklahoma	19
8	Oregon	18
9	Vermont	17
10	Idaho	17

Source: O'Hare 2022b

A Question of Equity

- Is it fair to inject as much DP error into groups with high census coverage error as groups with low census coverage error?

- Did Census Bureau consider an approach that injected less error (or no error) into groups with high levels of Census coverage errors?

or

- The 2020 census data for young children already has so much error that the added error from differential privacy does not matter much.

Separating Children and Adults in DP Data Processing

- In 2010 Census, 82 blocks with children but no adults before DP applied
- In 2010 Census,163,000 blocks with children but no adults after DP applied (2.7% of all blocks)
- Why? DP used in redistricting and DHC files processes data for children and their parents separately
- Impact of large number of blocks with unrealistic results
  - Decision making??
  - Public perceptions of census data
  - Impact on ACS/CPS/SIPP???????

Key Questions

- How much error does DP inject into population ages 0 to 4 for School Districts and Places? A lot for some School Districts and Places

- How important is that amount of error? Less clear but likely to be a problem for many areas

Conclusions

- Young children have a high and growing census undercount rate
- DP injects a high level of error into the data for the population ages 0 to 4 for Unified School Districts (specially age 4)
- DP injects a high level of error into the data for population ages 0 to 4 for Places
- Questions of DP equity need to be addressed
- The DP algorithm that separates children from parents in processing is worrisome.

References

O'Hare, W. P. (2022a). Analysis of Census Bureau's March 2022 Differential Privacy Demonstration Product: Implications for Data on Young Children, Posted on the Count All Kids website, https://countallkids.org/resources/analysis-of-census-bureaus-march-2022-differential-privacy-demonstration-product-implications-for-data-on-young-children/

O'Hare, W. P. (2022b) 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/

National Academy of Sciences (2022). Understanding the Quality of the 2020 Census: Interim Report, National Academy Press.