Assessing 2010 Census Data with Differential Privacy for Young Children

William P. O'Hare

CNSTAT

December 11-12, 2019

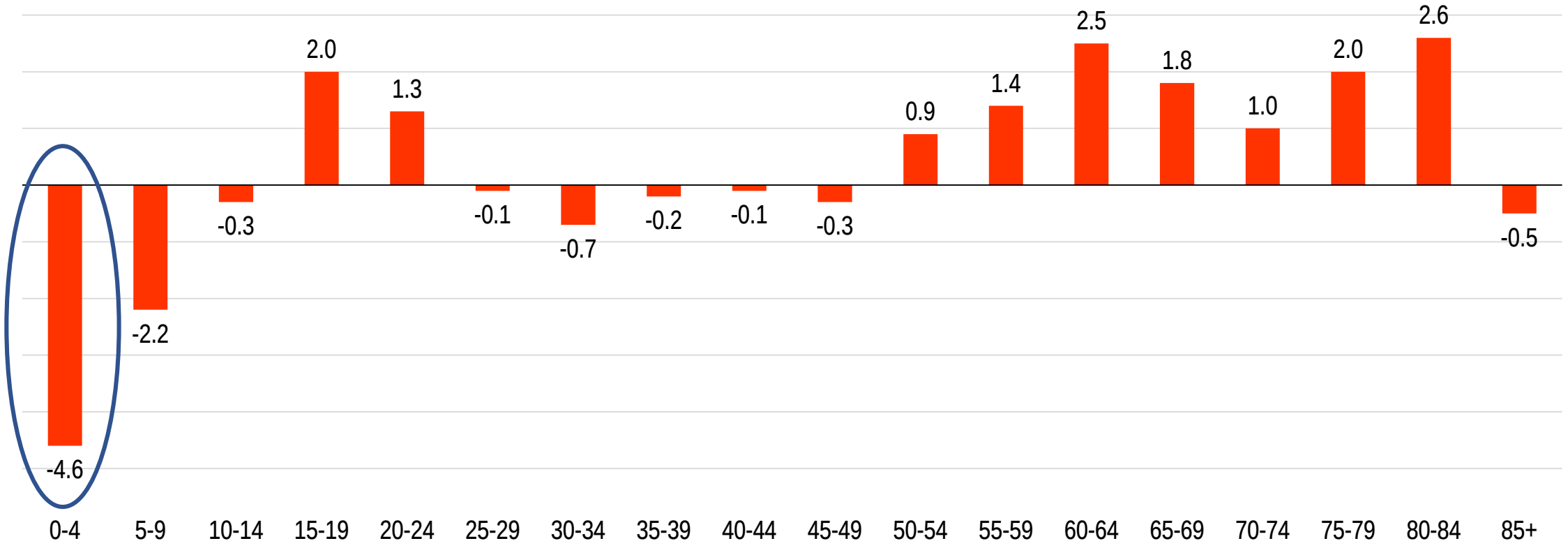
Support was provided from the 2020 Census Project,
a funders collaborative to promote a fair and accurate census.

Outline of Presentation

- Background on the undercount of young children
- Methodology and data sources used in the analysis
- State data
- County data
- Comments and conclusions

Young Children Had a Higher Net Undercount (By Far) Than Any Other Age Group in the 2010 Census

Net Undercount Rates (in Percentages) in 2010 Census, by Age Group

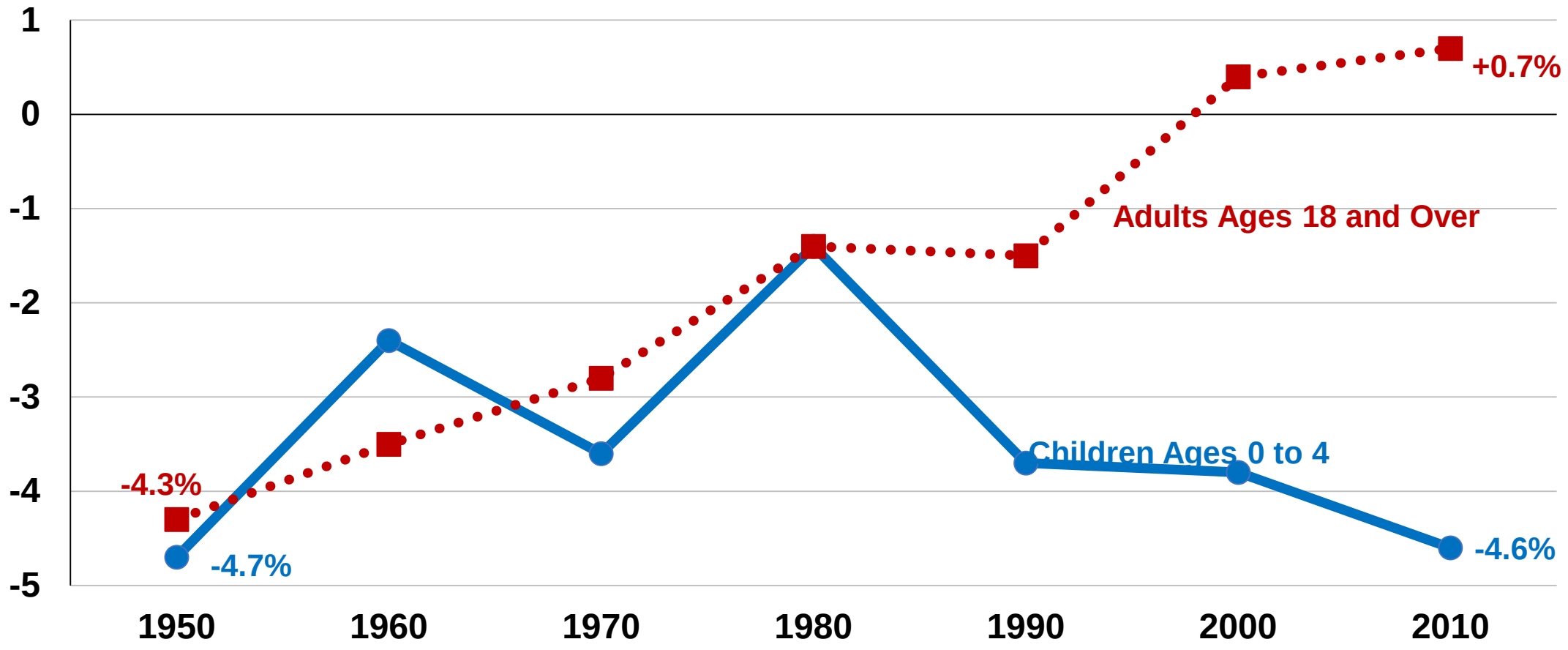


Undercount = 2010 Census Count minus 2012 Demographic Analysis (DA) Estimate

Source: U.S. Census Bureau, May 2012 Demographic Analysis (DA) release.

The Net Undercount of Young Children is Getting Worse While the Coverage of Adults Has Improved

Net Undercount Rates (in Percentages) of Young Children and Adults in the Census: 1950-2010



Source: O'Hare (2015), Chapter 4.

Fundamental Data for Census Bureau DA Estimate for Age 0 to 4 in 2010	
Births in 5 years prior to the 2010 Census	21,120,000
Deaths to those born in 5 years prior to 2010 Census	154,000
Net International Migration Age 0 to 4	240,000
Demographic Analysiis (DA) population estimate for age 0 to 4	21,206,000
2010 Census population count for age 0 to 4	20,201,000
Net Undercount	1,005,000
Source: U.S. Census Bureau	

Data Sources for Analysis

- Summary File data from the 2010 Census (SF)
- Data from the 2010 Census with Differential Privacy Applied (DP)
- Vintage 2010 Population Estimates from the Census Bureau (PEP)
- Census Bureau 2018 Experimental DA net undercount estimates for states and large counties (2018DA)

2010 Population Age 0 to 4 from Selected Sources	
	Population age 0 to 4 (in 1000s)
2010 Demographic Analysis December 2010	21,205
2010 Demographic Analysis May 2012	21,171
Vintage 2010 Census Bureau Population Estimates (PEP)	21,263
2018 Census Bureau Experimental DA	20,978
2010 Census Original Summary File (SF)	20,201
2010 Census File with Differential Privacy (DP)	20,199

Source of data for DP and SF data

IPUMS files

https://www.nhgis.org/differentially-private-2010-census-data

Citation

IPUMS NHGIS, University of Minnesota, www.nhgis.org.

or

Steven Manson, Jonathan Schroeder, David Van Riper, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis, MN: IPUMS. 2019. http://doi.org/10.18128/D050.V14.0

Source of Data for PEP and 2018DA

- PEP
  - U.S. Census Bureau website
  - https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates.html

- 2018 DA
  - Source for 2018 Experimental DA
  - **King, H., Ihrke, D. and Jensen, E., (2018). "Subnational Estimates of Net Coverage Error for the Population Aged 0 to 4 in the 2010 Census**,"" paper present the 2018 Population Association of American Conference, April 25-28 , Denver Colorado, Downloaded May 6, 2018 https://paa.confex.com/paa/2018/meetingapp.cgi/Paper/21374.

Methodology for Subnational Error Estimates

- Numerical Errors = SF or DP – PEP
- Percent Errors = ((SF or DP – PEP)/PEP)
- In this presentation net undercounts are expressed as a negative number

Other Studies That Have Used This Approach

- King, H., Ihrke, D. and Jensen, E., (2018). “Subnational Estimates of Net Coverage Error for the Population Aged 0 to 4 in the 2010 Census,”” paper present the 2018 Population Association of American Conference, April 25-28 , Denver Colorado.

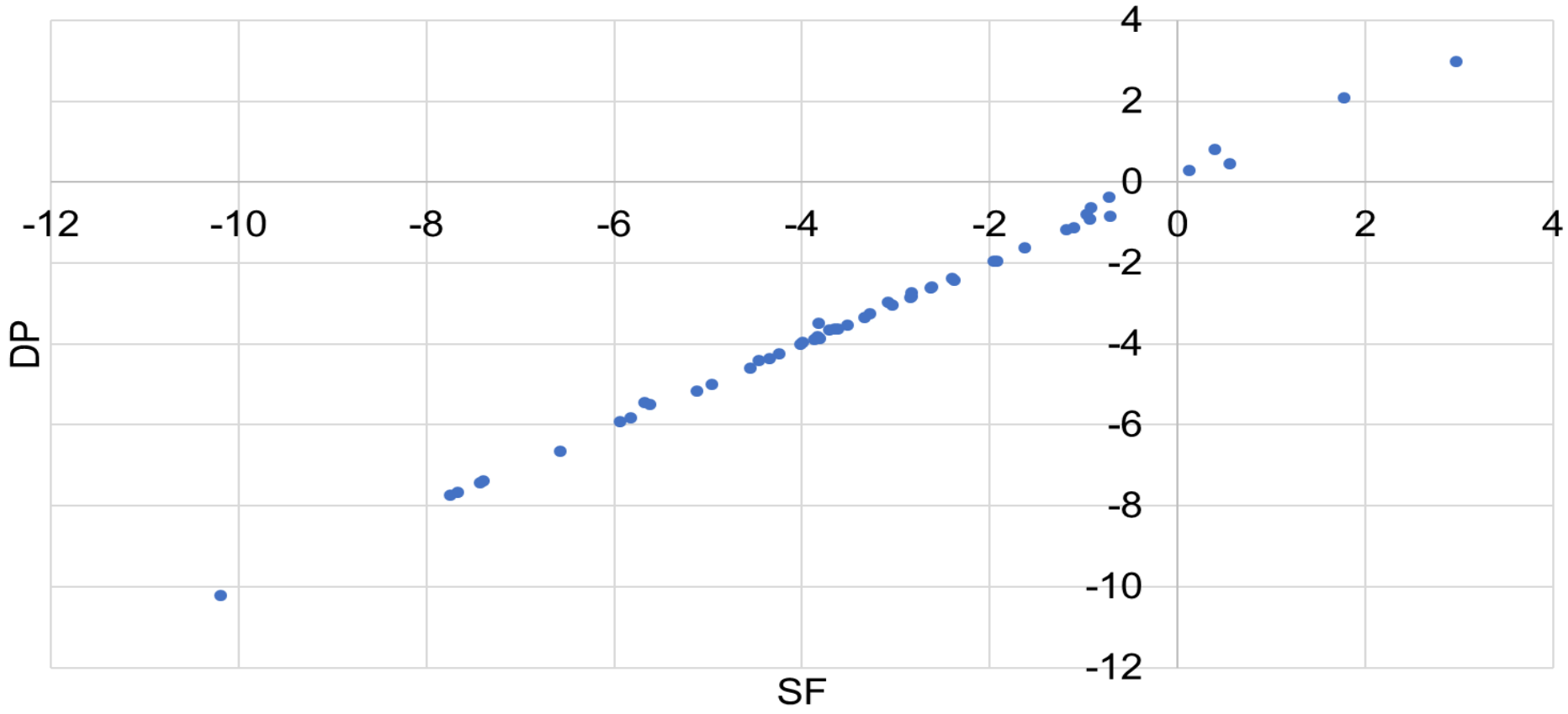
- O’Hare, W.P. (2014). State-Level 2010 Census Coverage Rates for Young Children, *Population Research and Policy Review,* Volume 33, no. 6, pages 797-816.

- O’Hare W.P. (2017). “Geographic Variation in 2010 U.S. Census Coverage Rates for Young Children: A Look at Counties,” *International Journal of Social Science Studies,* Vol. 5, No. 9 Sept. Redframe Publishing.

- Mayol-Garcia, Y., and Robinson, G. (2011).” Census 2010 counts compared to the 2010 population estimates by demographic characteristics.” Poster presented at the Southern Demographic Association Conference, October, Tallahassee FL.
-
- Robinson, G.J., Bashir. A., Das Dupta, P., and Woodward, K.A., (1993). Estimates of Population Coverage in the 1990 United States U.S. Decennial Census Based on Demographic Analysis, *Journal of the American Statistical Association*, Vol. 88, No 423, pp 1061-1071.

Analysis of State-Level Data

Percent Net Undercount for Population Age 0 to 4 for States Based on Comparing SF and DP to PEP



Accuracy Measures for Population Age 0 to 4 for States* Based on SF and DP Compared to PEP		
	SF-PEP	DP-PEP
Number of States with a Net Undercount	46	46
Maximum Value Percent Error	2.1	1.7
Minimum Value Percent Error	-10.2	-10.2
Mean Numerical Error	-21,114	-21,143
Standard Deviation Numerical error	38,189	38,188
Mean Percent Error	-3.4	-3.5
Standard Deviation Percent Error	2.4	2.4
Mean Absolute Numerical Difference	21,176	21,195
Standard Deviation Absolute Numerical Difference	38,154	38,160
Mean Absolute Percent Difference	3.6	3.6
Standard Deviation Absolute Percent Difference	2.2	2.2
* does not include DC		

Accuracy Measures for Population Age 0-4 for States* Based on SF and DP Compared to 2018DA

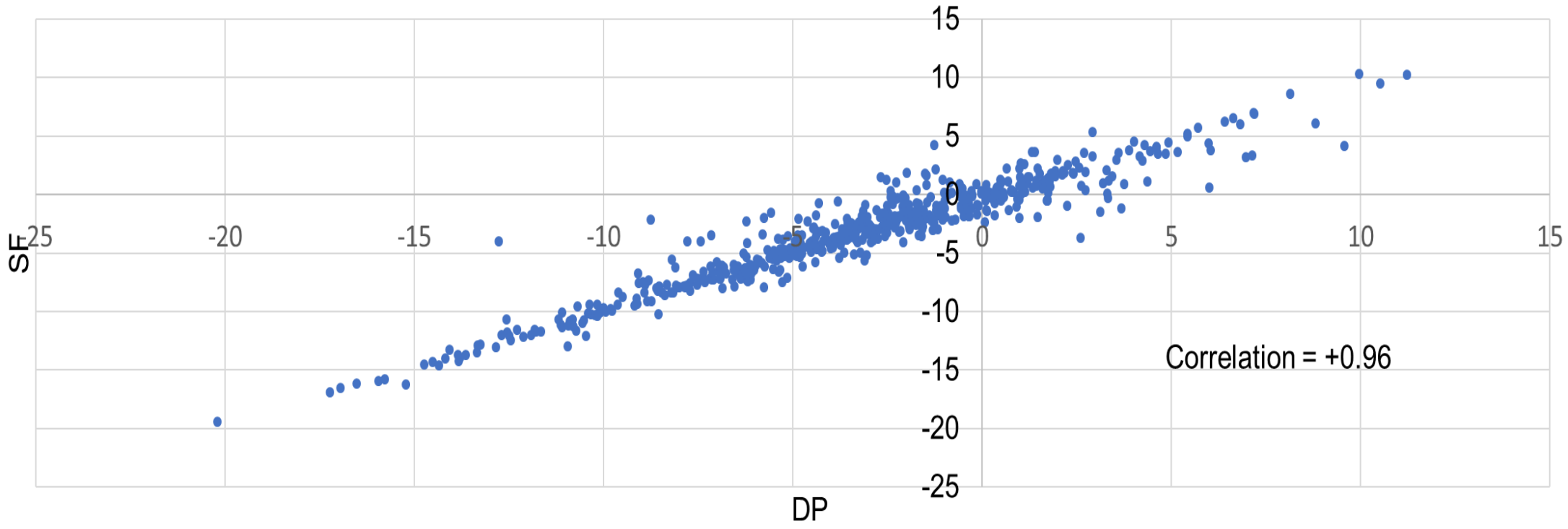
	SF-2018DA	DP-2018DA
Number of States with Net Undercount	49	49
Maximum Value Percent Error	0.4	0.3
Minimum Value Percent Error	-6.2	-6.2
Mean Numerical Difference	16,183	16,212
Mean Percent Difference	-3.2	-3.2
Standard Deviation Percent Difference	1.7	1.7
* does not include DC		

Analysis of Counties

- Large Counties = 100,000 or more total population in 2010 (577 counties and DC)

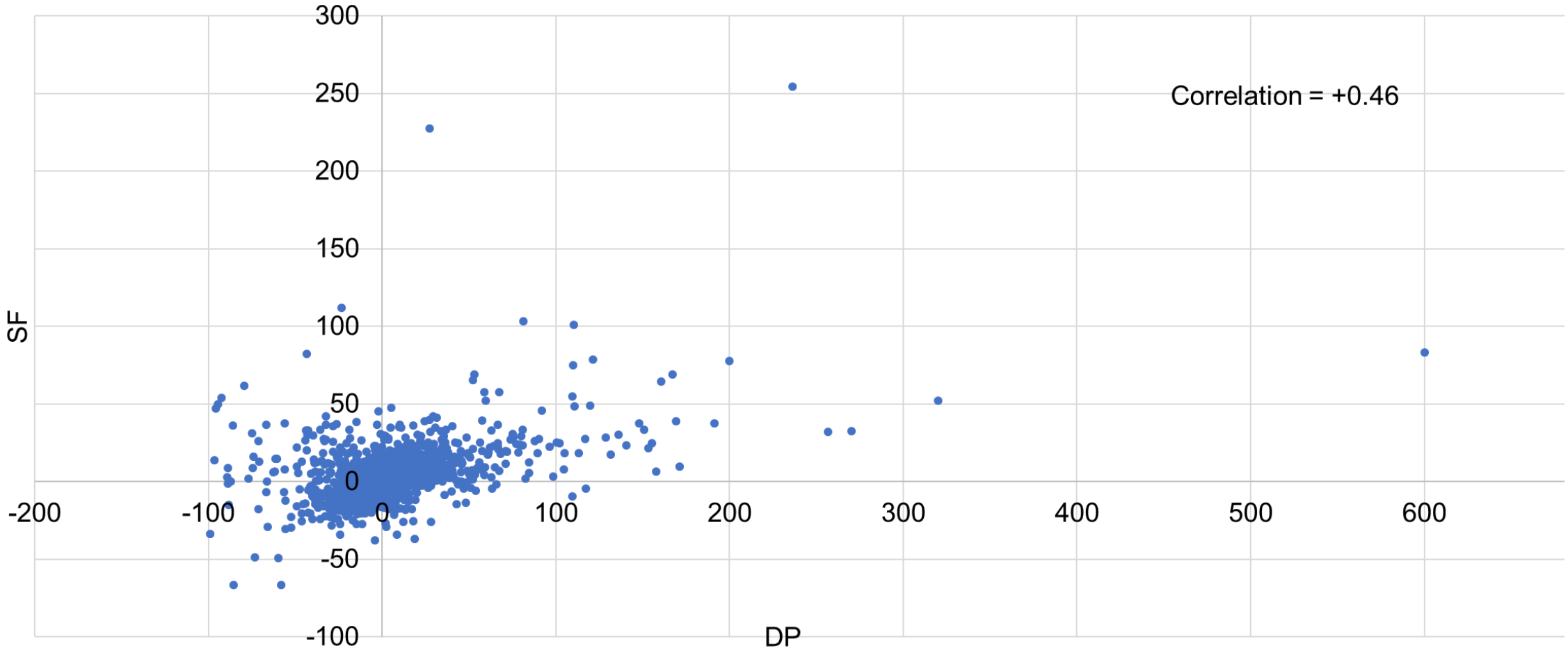
- 2,562 Small Counties = those with less than 100,000 total population in 2010

Percent Errors for Population Age 0 to 4 for Large Counties (100,000+) comparing DP and SF to 2018 DA



Accuracy Measures for Population Age 0 to 4 for 579 Large Counties* (100,000+) Based on SF and DP Compared to 2018DA		
	SF-PEP	DP-PEP
Collective Undercount for age 0 to 4	-4.2	-4.3
Number of Counties with a Net Undercount	448	445
Maximum Percent Error	10.3	11.2
Minimum Percent Error	-16.2	-20.2
Mean Numerical Difference	-1,775	-1,779
Mean Percent Difference	-3.3	-3.4
Standard Deviation Percent Difference	4.5	4.8
Mean Absolute Numerical Difference	1,899	1,926
Mean Absolute Percent Difference	4.3	4.6
Standard Deviation Absolute Percent Difference	3.6	3.6
* Includes the District of Columbia		

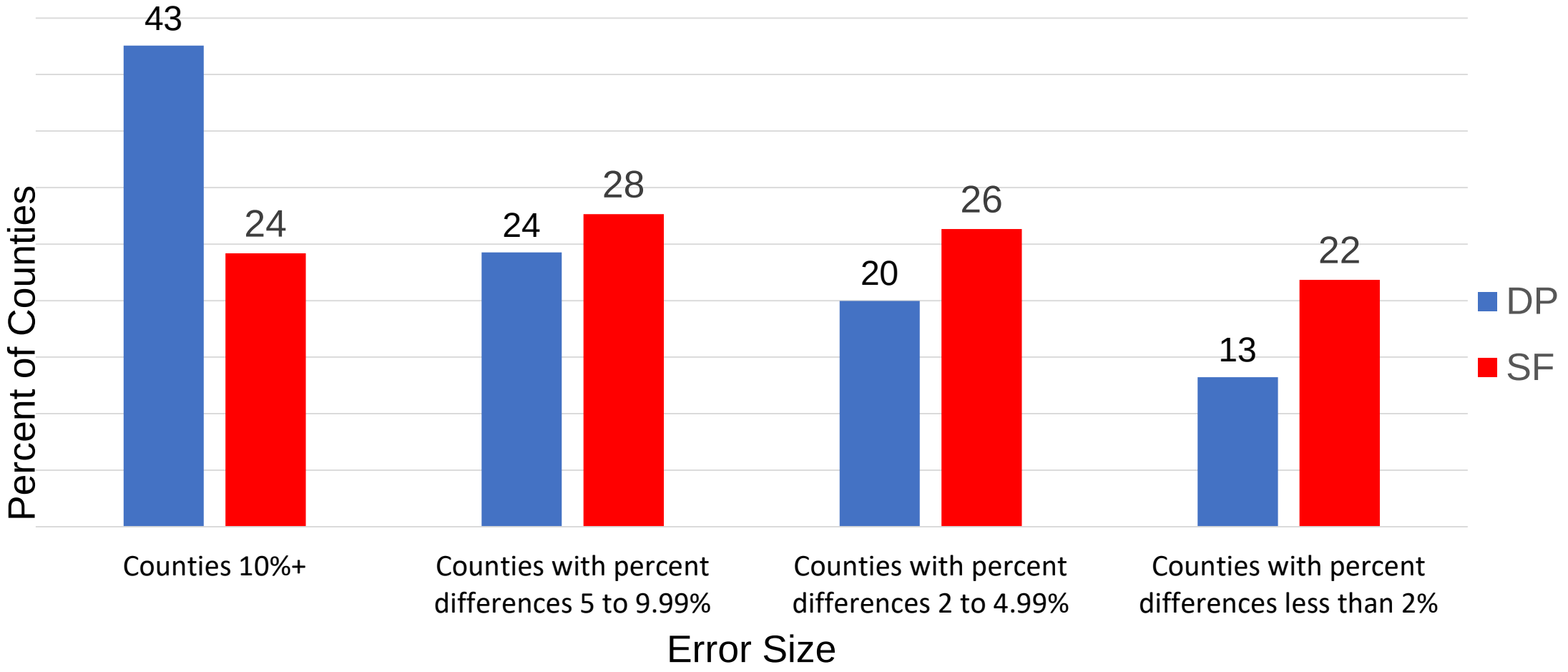
Pecent Error for Age 0 to 4 for 2,562 Counties Under 100,000 Total Population Based on DP and SF Compared to PEP



Accuracy Measures for Population Age 0 to 4 for 2,562 Smaller Counties (less than 100,000 total population) based on SF and DP compared to PEP

	SF-PEP	DP-PEP
Number of Counties with a Net Undercount	1,187	1,298
Collective Coverage for age 0 to 4	-0.8	-0.8
Mean Numerical Difference	-14	-14
Mean Percent Difference	2.1	2.8
Mean Absolute Numerical Difference	86	129
Mean Absolute Percent Difference	8.1	15.2

Distribution of Absolute Percent Errors for DP and SF for Smaller (less than 100,000 population) Counties



Conclusions

- For state level data, DP makes little difference with respect to the net undercount of young children
- For large counties (100,000+), DP makes little difference with respect to the net undercount of young children
- For smaller counties (less than 100,000) DP makes data for young children less accurate but does not impact the average net undercounts

Use Case for Counting and Reporting Young Children Accurately

Use for Planning

"Knowing how many children live in a community is the foundation of many important municipal decisions. Should a community build a new school? A new hospital? Should Head Start for pre-K children be expanded?"

Source: U.S. Census Bureau, "Big Push to Count Every Newborn and Young Child in the Census" November 2019

Use Case – Distribution of Federal Dollars

- 316 Federal Programs used Census-derived data to distributed $1.5 trillion in Fiscal Year 2017
- 215 of the 316 (68%) used local data
- Census errors impact funding for ten years – most of a young child's childhood

Source: Counting for Dollars website

Selected Federal Programs that Use Census Counts for Age 0 to 4 in Distribution of Funds

Program	Fiscal Year 2017 Dollars
Head Start	$9.4 billion
Special Supplemental Program for Women, Infancts, and chidlren (WIC)	$6.5 billion
Child Care and Development Black Grant	$2.8 billion
Maternal and Child Health Serives Block Grant	$545 million

Source: Counting for Dollars https://gwipp.gwu.edu/counting-dollars-2020-role-decennial-census-geographic-distribution-federal-funds#Latest Release

Key Implications

- Data worse for smallest areas but that is where census data is most important.

- With DP, 43% of the 2500+ smaller counties will face the prospect of 10%+ too little funding (or 10%+ too much funding) each year for ten years based on reported population age 0 to 4.

Communicating with the Public

- Need metrics to describe DP impact on accuracy and privacy
  - Planned DP increases privacy protection by 10%???? Or from 1 in a million chance to 1 in a 100 million chance
  - Planned DP decreases accuracy by 10%???

Fundamental Conflict

- Foundations, child advocates, and the Census Bureau are working hard to eliminate or reduce the high net undercount of young children

- DP will infuse the count of young children in smaller counties with 5-10 percent errors.

THANKS!

Bill O’Hare
billohare1@gmail.com