Differential Privacy and the 2020 Census
Implications for Regional Agencies
Prepared by the Metropolitan Area Planning Council
December 12, 2019
For more information, contact MAPC Data Services Director Tim Reardon: treardon@mapc.org

Over the past year, the U.S. Census bureau has advanced plans to implement a new method of data privacy protection for the 2020 Decennial Census. This new method, known as "Differential Privacy," adds some amount of statistical "noise" to raw census results before releasing them. While intended to ensure that the Census Bureau complies with Title 13 prohibitions on releasing identifiable data, the use of Differential Privacy is likely to have tremendous impacts on the accuracy of the products that are released, especially tables for small areas or those with multiple cross-tabulations (e.g. age by race.) Not even the total population for any sub-state geography will be held invariant.

The Census Bureau is motivated to adopt this new method of disclosure avoidance based upon internal experiments that combined publicly-released census tables with commercially available "big data" and estimated that 38% of records could be matched. However, after comparing to actual census responses, half of those matches were "false positives" which could be determined only by using confidential census data. Nevertheless, the Bureau is continuing to move forward with the proposed use of Differential Privacy for the 2020 Decennial Census, with plans to use it for the American Community Survey in 2025.

The Census Bureau has released "Demonstration Files" that apply the proposed Differential Privacy (DP) to 2010 decennial census responses, so that data users can assess the differences between the two representations of 2010, and measure the accuracy loss associated with DP. All characteristics ofn an area are affected by Differential Privacy, including the total number of households, total number of residents, and characteristics of residents such as age, race, ethnicity, and household tenure. Researchers across the country have been evaluating these demonstration files and have uncovered many troubling findings. Especially at small area levels, there are incredible changes in population characteristics, as well as glaringly problematic results (one block group in the Boston Region at had an average household size of 22 people after DP was applied.)

The introduction of Differential Privacy will affect program eligibility and resource allocation for all affected geographies. Designations such as Opportunity Zones, Environmental Justice Areas, Qualified Census Tracts, and many others will change as a result of the use of DP. While total apportionment to states will reflect actual enumeration, redistricting will use the noise-infused data, affecting the equal representation of residents. Reduced accuracy of the Decennial Census will negatively affect analysis and modeling done by every MPO, county, and municipality in the country, making it harder to make smart policy decisions and investments. Finally, any aberrant or egregiously inaccurate results observed by the public will undermine confidence in the Census as a reliable data source and fuel skepticism and criticism of the public agencies forced to use this substandard data.

The Census Bureau has solicited comments and feedback on the use of Differential Privacy and specific technical questions about how it will be applied. A National Academy of Sciences workshop was also held on December 11-12 to showcase analysis of the demonstration data. Feedback on the program can be submitted to dcmd.2010.demonstration.data.products@census.gov.