

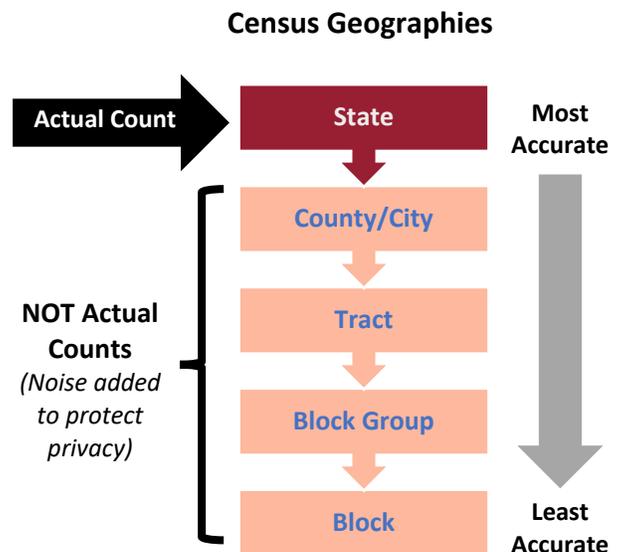
Differential Privacy Potential to Impact 2020 Census Data Accuracy

June 17, 2021

For decades, the U.S. Census Bureau has used different techniques to protect individual privacy through their Disclosure Avoidance System (DAS). Differential Privacy (DP) is the latest technique. This technique was chosen because enhanced privacy protection is needed due to advances in technology that enable the ability to identify individuals when combined with other information. DP will impact data accuracy more than any other technique previously applied to decennial census results. DP will add random data (“noise”) to counts for geographic areas below the state level, to prevent identification of individuals. The amount of noise added to the data will be greater than any prior decennial census technique.

Key Takeaways

- The U.S. Census Bureau is applying DP to the 2020 Census. This will result in different data than the actual counts, especially for smaller census geographies.
- While data privacy is essential, there must be a balance between privacy and accuracy without impacting data usability.
- On June 9, 2021, the U.S. Census Bureau announced final redistricting DP parameters. Much data at the block group level and larger geographies may be accurate enough for use, however, significant accuracy concerns at the Census block level remain.
- The U.S. Census Bureau plans to release the final round of 2010 sample data injected with the final DP parameters for review by stakeholders in September 2021. NVRC and local jurisdictions will complete an analysis to understand the full impact DP will have on the 2020 Census redistricting data.
- The U.S. Census Bureau’s current plan for non-redistricting Census 2020 data product, formally known as the Demographic and Housing Characteristics File, is yet to be determined.



Greatest Impacted Areas: Census Blocks

Census blocks, the smallest census geography, can be as small as a city block or cover a larger area in rural places. The data at this level will be the least accurate out of all geographic levels. Census blocks are used at the local and regional level for small area and neighborhood analysis. The diminished accuracy from implanting DP will impact the ability to perform equity analysis, population estimates and forecast, and neighborhood planning. Blocks can be aggregated together to create data for more meaningful geographic areas and for more accurate, useable data.

Equity Analysis	Forecasting Population and Housing	Planning
Census provides data on race/ethnicity which are used as a baseline to monitor equity. Inaccurate neighborhood or block level data will skew analysis.	Household size and occupancy/vacancy rates from the 2020 Census in most cases will not be usable at block level and will impact local government population estimate and forecasting programs.	Inaccuracies at the lower level geographies will provide difficulties in understanding the population at the block or neighborhood level.

Additional Impacts

Non-Block Level Impacts: Census block groups represent areas with 600 to 3,000 people and will have better accuracy than data at the census block level. Data at the block group level and higher level, larger geographies will be accurate enough at least population and households based on analysis of the April 2021 round. At the County level, persons of one race and the ethnicity data were accurate enough. As of the date of this publication, regional analysis of race/ethnicity for geographies smaller than counties had not been completed. NVRC and local jurisdictions plan to conduct an analysis on the final demonstration data to be released in September 2021 to determine accuracy and usability of various demographics, in particular race/ethnicity data.

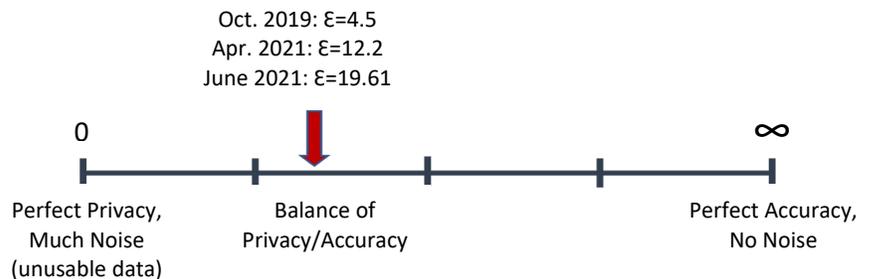
Redistricting Impacts: Redistricting may be impacted for jurisdictions. In working with DOJ and the redistricting community, the Census Bureau developed an accuracy target for geographies of 500 people or more, where the largest race/ethnicity would be +/- 5 percentage points 95% of the time. The race/ethnicity population distribution pattern at the block and block group level is typically used to assess fair representation among districts, but this pattern may be affected by noise which will incur upon the ability to assess racially diverse districts.

Current and Potential Litigation

The State of Alabama filed a lawsuit against the CB regarding the late delivery of redistricting data (anticipated by September 30, 2021) and the implementation of DP due to the resulting inaccurate data products, causing vote dilution in violation of the Fifth Amendment. The following 16 states joined Alabama's challenge: Alaska, Arkansas, Florida, Kentucky, Louisiana, Maine, Mississippi, Montana, Nebraska, New Mexico, Ohio, Oklahoma, South Carolina, Texas, Utah and West Virginia. There is other litigation and potential for more.

Statistical Technique Details

The measure of the balance between privacy versus accuracy is called the privacy-loss budget (PLB) or Epsilon (ϵ). An Epsilon of zero would result in complete privacy with too much noise infused that it would make all the data unusable versus an Epsilon of infinity would result in complete accuracy with no privacy. The CB had multiple rounds released to data users for review. Each round improved the accuracy by increasing the epsilon and/or modifying the data processing algorithm. The June 2021 epsilon and associated algorithm is approved for the 2020 Census redistricting data file. The non-redistricting Census 2020 data product, formally known as the Demographic and Housing Characteristics File, may have a different epsilon, as yet to be determined.



Additional Resources

- [Census Bureau Sets Key Parameters to Protect Privacy in 2020 Census Results, June 9, 2021](#)
- [Census Scientific Advisory Committee Memo, May 25, 2021](#)
- [NPR, June 14, 2021](#)
- [Washington Post, June 3, 2021](#)

...applying differential privacy caused the accuracy of population counts to suffer, particularly in districts that had high racial and ethnic diversity, and cast doubt on the government's ability to rely on this data to enforce the "one person, one vote" principle that ensures every American an equally weighted vote, said Christopher Kenny, a PhD candidate at Harvard's Department of Government."

- Washington Post,
June 3, 2021

The Bureau's implementation of DP has followed an ambitious timeline. But many implications of DP implementation decisions are not yet fully understood...The fitness for use of legal and regulatory uses of the data have not been examined in full, and risks of failing to produce sufficiently accurate data for these uses remain likely but not fully known.

- Census Scientific Advisory Committee
Memo,
May, 25, 2021