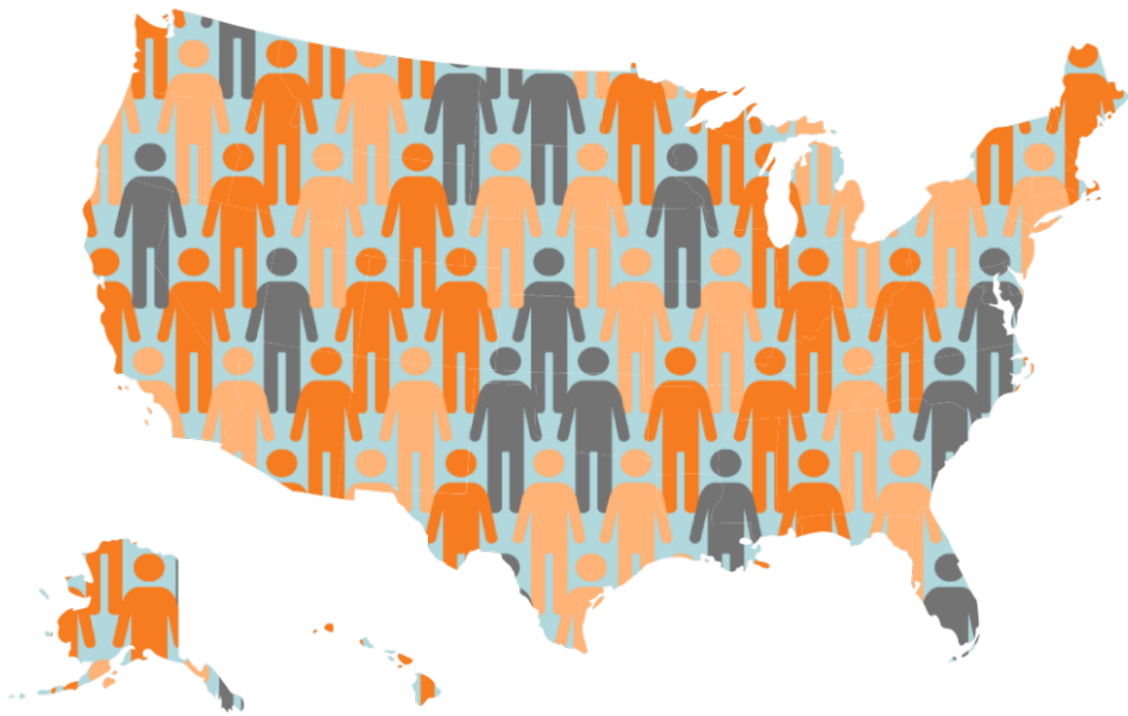
PRELIMINARY REPORT:

Impact of Differential Privacy & the 2020 Census on Latinos, Asian Americans and Redistricting



MALDEF
The Latino Legal Voice for Civil Rights in America.

ASIAN AMERICANS ADVANCING JUSTICE
AAJC

April 2021

About the Organizations

Founded in 1968, **MALDEF (Mexican American Legal Defense and Educational Fund)** is the nation's leading Latino legal civil rights organization. MALDEF's mission is to promote the civil rights of all Latinos living in the United States; in pursuing that mission, we defend the constitutional rights of all Americans. MALDEF currently serves as a member of the Census Bureau's National Advisory Committee on Racial, Ethnic and Other Populations and has served on previous advisory committees to the Bureau. MALDEF has also conducted census outreach and educational campaigns focused on the Latino community for the decennial census each decade. For the 2020 Census, MALDEF led an effort to help support census mobilization efforts and address confidentiality concerns, which were intensified due to actions by the Trump administration, by launching a census confidentiality pledge from a watchdog coalition of over 300 civic leaders, non-profit organizations, elected officials, as well as state and local groups. MALDEF has also litigated several issues arising from attempts to manipulate the outcome of the 2020 Census.

Rooted in the dreams of immigrants and inspired by the promise of opportunity, **Asian Americans Advancing Justice | AAJC (Advancing Justice | AAJC)** advocates for an America in which all Americans can benefit equally from, and contribute to, the American dream. Advancing Justice | AAJC is a national 501 (c)(3) nonprofit founded in 1991 in Washington, D.C. Advancing Justice | AAJC's mission is to advance the civil and human rights of Asian Americans and to build and promote a fair and equitable society for all. Advancing Justice | AAJC has served as a member of numerous advisory committees to the Census Bureau since 2000, including most recently, the National Advisory Committee on Racial, Ethnic and Other Populations. Additionally, Advancing Justice | AAJC currently co-chairs the Leadership Conference on Civil and Human Rights' Census Task Force. Advancing Justice | AAJC has conducted national, state, and local outreach and educational projects focused on the Asian American, Native Hawaiian, and Pacific Islander communities for Census 2000, Census 2010, and Census 2020 and has also litigated issues surrounding Census 2020.

Authors:

Steven A. Ochoa, MALDEF, National Redistricting Coordinator

Terry Ao Minnis, Advancing Justice | AAJC, Senior Director of Census & Voting Program

Executive Summary

The Constitution requires a census every ten years in order to determine the number of seats each state will have in the U.S. House of Representatives. In addition to determining this reapportionment of House seats, Census data are also used for purposes of redistricting, distribution of federal funding, and assisting policy makers, businesses, and other interested stakeholders in assessing and addressing community needs.

The Census Bureau has utilized different methods to meet its federal-law statutory mandate to protect respondents' privacy and confidentiality in published statistics. For the 2020 census, the Census Bureau will use a new methodology for this purpose – "differential privacy." Differential privacy is a mathematical method that uses statistical noise, or false information, to alter data so that the link between the data and specific persons or households cannot be ascertained. How the Census Bureau implements differential privacy could negatively affect civil rights enforcement for the next decade, including with respect to redistricting and voting rights.

This preliminary report – the first such report from civil rights organizations on this topic – compares the demonstration products (which tested different configurations of differential privacy using 2010 Census data) to published 2010 Census data in an effort to assess the impact of differential privacy on a) total population and racial/ethnic populations and b) redistricting.

In assessing impact on population data, our analysis found a small bias of adding population to small population counties at the expense of larger-population counties, and of increasing homogeneity of counties with clear majority race/ethnic populations. These initial findings reinforced our primary concern that data produced using differential privacy would not be of sufficient quality to meet the legal requisites of redistricting – the constitutional requirement of equal population between districts and the necessity of compliance with the federal Voting Rights Act of 1965 (VRA).

Our preliminary analysis shows that changes in total population affected the equal population requirement in redistricting, with 30 congressional plans that previously had deviation ranges under 10 people going over that range in all demonstration products. Similarly, 12 state lower-house plans and 8 state upper-house plans that previously had deviation ranges under 10 percent went over that threshold at some point in the various demonstration products. Recent demonstration products also showed more variation for voting-age population (VAP) percentages than earlier ones, raising a concern that because areas with majority-minority districts (i.e., districts where one racial minority group equals 50% or more of the citizen voting-age population (or, in the case of some jurisdictions, some percentage threshold of the voting age population)) may not be as accurate, redistricters might not draw majority-minority districts where they should and might draw them where they should not. These preliminary findings point to the possible need to drastically alter the way the Census Bureau plans to design and implement a disclosure avoidance system, defined by the government as a process of making changes to data for the purpose of protecting a respondent's privacy and confidentiality in published statistics, on the redistricting data file.

In the limited time remaining to make its final decisions around its disclosure avoidance system, the Census Bureau must meaningfully engage external stakeholders by:

- Providing greater transparency about decisions on differential privacy to date, including providing an explanation for what changes have been made to the system and the impact of those changes; and
- Providing space and opportunity for stakeholders to inform the Bureau on how best to assess options for all final implementation decisions regarding differential privacy.

MALDEF and Advancing Justice | AAJC will continue analyzing the data to further understand the impact of differential privacy on redistricting and provide a final report in the near future.

Introduction

Article 1, Section 2 of the Constitution mandates the count of the nation's entire population once every ten years for purposes of determining the number of seats each state will have in the U.S. House of Representatives. The U.S. Census Bureau is responsible for collecting this data every ten years through the decennial census. Data from the decennial census are also used to redistrict congressional, state legislative, and local jurisdictions (e.g., city council, school board, etc.), as well as to determine how best to provide daily services, products, and support for local communities by lawmakers, business owners, urban planners, and other interested stakeholders. Over $1.5 trillion in federal funding are disbursed to the states every year for hospitals, schools, fire departments, roads, and other services based on census data.[1]

As part of its mandate to collect data, the Census Bureau is bound by Title 13 of the U.S. Code to ensure that no information collected by the Bureau is revealed about any specific individual or household for 72 years.[2] In order to achieve this mandate, the Census Bureau engages in "disclosure avoidance," which is "the process of making changes to data for the purpose of protecting a respondent's privacy and confidentiality in published statistics."[3] For the 2010 census, the Census Bureau relied on "data swapping" as its primary disclosure avoidance method – where

[1] Andrew Reamer, George Washington University Institute of Public Policy, COUNTING FOR DOLLARS 2020 The Role of the Decennial Census in the Geographic Distribution of Federal Funds Brief 7: Comprehensive Accounting of Census-Guided Federal Spending (FY2017) Part A: Nationwide Analysis (2019), https://gwipp.gwu.edu/sites/g/files/zaxdzs2181/f/downloads/Counting%20for%20Dollars%202020%20Brief%207A%20-%20Comprehensive%20Accounting.pdf.

[2] 13 U.S.C. § 9. *See*, U.S. Census Bureau, The "72-Year Rule", https://www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html#:~:text=April%202%2C%202012.,collected%20for%20the%20decennial%20census.

[3] U.S. Census Bureau, Frequently Asked Questions: What is disclosure avoidance?, https://ask.census.gov/prweb/PRServletCustom/app/ECORRAsk_/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pzuiactionzzz=CXtpbn0rTEpMcGRYOG1vS0tqTFAwaENUZWpvM1NNWEMzZ3p5aFpnWUxzVmw0TjJqbndtUHNuQW1pOENMdVRjQndDU05Ndlo5SHpwcm5wVjN0TzFwYUh2UmQ4Zz09*.

block level characteristic data such as race/ethnicity were swapped between nearby residents.[4] This process was highly targeted to those data that had the highest disclosure risk, such as census blocks with very small total populations. Because swapping only affected characteristics rather than total population data with a nearby geographic area, there was no effect on the total population at the block level and only minimal effect on totals for specific characteristics within that geographic area. Thus, the impact on data accuracy was generally not significant.

For the 2020 census, the Census Bureau made the internal decision that data swapping was insufficient to meet its Title 13 obligations due to technological advances: growing computing power, advances in mathematics, and easy access to large, public databases. Furthermore, the Bureau asserted that the amount of data swapping required to maintain its acceptable level of confidentiality for the 2020 census would make its published data unfit for most uses. The Bureau has not provided details to date to support this assertion. Nonetheless, based on this assertion, the Census Bureau embarked on a new disclosure avoidance methodology for the 2020 census – "differential privacy."[5]

Differential privacy is a mathematical method that adds "statistical noise" to published tables in a way that protects each respondent's identity – that is, statistical noise or false information is used to alter data so that the link between the data and a specific person cannot be certain.[6] The Bureau believes that differential privacy would allow it to precisely control the amount of statistical noise added using sophisticated mathematical formulas that permit, in theory, the changing of enough data to protect the confidentiality of census information without harming data quality to the point of dysfunction.

Because the Census Bureau plans to utilize a new disclosure avoidance system that relies on differential privacy together with necessary "post-processing", there is a need to scrutinize the new proposed system to determine whether the resulting data are fit for the various uses of census data. How the Census Bureau sets up its disclosure avoidance system for the 2020 census and how it implements differential privacy could be significant factors in the efficacy of civil rights

---

[4] U.S. Census Bureau, Frequently Asked Questions: How is a differentially private system different from the Census Bureau's traditional disclosure avoidance techniques?, https://ask.census.gov/prweb/PRServletCustom/app/ECORRAsk_/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pzuiactionzzz=CXtpbn0rTEpMcGRYOG1vS0tqTFAwaENUZWpvM1NNWEMzZ3p5aFpnWUxzVmw0TjJqbndtUHNuQW1pOENMdVRjQndDU05Ndlo5SHpwcm5wVjN0TzFwYUh2UmQ4Zz09*#. *See also*, U.S. Census Bureau, Frequently Asked Questions: What is data swapping?, https://ask.census.gov/prweb/PRServletCustom/app/ECORRAsk_/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pzuiactionzzz=CXtpbn0rTEpMcGRYOG1vS0tqTFAwaENUZWpvM1NNWEMzZ3p5aFpnWUxzVmw0TjJqbndtUHNuQW1pOENMdVRjQndDU05Ndlo5SHpwcm5wVjN0TzFwYUh2UmQ4Zz09*#.

[5] U.S. Census Bureau, Frequently Asked Questions: Has the Census Bureau settled on final policy for differential privacy?, https://ask.census.gov/prweb/PRServletCustom/app/ECORRAsk_/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pzuiactionzzz=CXtpbn0rTEpMcGRYOG1vS0tqTFAwaENUZWpvM1NNWEMzZ3p5aFpnWUxzVmw0TjJqbndtUHNuQW1pOENMdVRjQndDU05Ndlo5SHpwcm5wVjN0TzFwYUh2UmQ4Zz09*# .

[6] U.S. Census Bureau, Frequently Asked Questions: What is differential privacy?, https://ask.census.gov/prweb/PRServletCustom/app/ECORRAsk_/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pzuiactionzzz=CXtpbn0rTEpMcGRYOG1vS0tqTFAwaENUZWpvM1NNWEMzZ3p5aFpnWUxzVmw0TjJqbndtUHNuQW1pOENMdVRjQndDU05Ndlo5SHpwcm5wVjN0TzFwYUh2UmQ4Zz09*.

enforcement for the next decade, including with respect to redistricting and voting rights. If the proposed new methodology fails to produce data that are accurate and/or reasonably fit for use, then the new methodology should not be implemented as proposed, and other solutions must be found to address the Bureau's Title 13 mandate.

In order to help answer the question whether the new proposed differential privacy methodology would produce accurate data fit for use, MALDEF and Advancing Justice | AAJC set out to analyze the public demonstration products, produced from the Bureau's testing of its differential privacy methodology, to determine impact on Asian Americans and Latinos in the political process. What is most important to MALDEF's and Advancing Justice | AAJC's work is the fitness of data for redistricting at all levels of government. To that end, the data must be the most accurate possible at the lowest geography (i.e., block-level) possible for a) total population, b) those who are 18 years old and older, and c) across all Census racial /ethnic categories.

This preliminary report -- the first such report from civil rights organizations on this topic -- seeks to answer some questions about how the different demonstration products met accuracy standards for total population among racial/ethnic groups and for purposes of redistricting, with an eye toward the fact that the window is closing for the Census Bureau to make specific decisions about how differential privacy will be implemented (by mid-2021). Our preliminary findings reveal serious concerns about the impact of differential privacy as currently envisioned by the Bureau on our communities' ability to attain our fair share of political power, and surface the need to find solutions to improve the disclosure avoidance system for 2020 census data.

Differential Privacy Research and Decisions to Date

For our purposes, the research on how differential privacy will be implemented for the 2020 census began with the first public release of the Census Bureau's 2010 Demonstration Data Products in October 2019 (Demo 1[7]). Preliminary analysis of that demonstration product raised red flags about how differential privacy would be implemented. A common theme emerged from presenters at the December 2019 Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine (CNSTAT) workshop[8] – the application of differential privacy could result in equality (i.e., the same numbers from the two different datasets) without equity (i.e., different communities were impacted differently by the policy). This is particularly true at smaller geographic levels (i.e., blocks, block groups, tracts), for geographies with zero to small populations, and for smaller racial groups. To date, three other demonstrations products have been made available. This preliminary report primarily focuses on the fourth demonstration product released in November 2020 (Demo 4[9]), but also makes reference to Demo 1, the second

---

[7] U.S. Census Bureau, 2010 Demonstration Data Products (Oct. 2019), https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2010-demonstration-data-products.html ("Demo 1").

[8] Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine, Workshop on 2020 Census Data Products: Data Needs and Privacy Considerations (Dec. 2019), https://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_196518.

[9] National Historical Geographic Information System (NHGIS), IPUMS, Privacy-Protected 2010 Census Demonstration Data (Nov. 2020), https://www.nhgis.org/privacy-protected-demonstration-data#v20201116 ("Demo 4").

demonstration product released in July 2020 (Demo 2[10]), and the third demonstration product released in September 2020 (Demo 3[11]).

In November 2020, the Census Bureau decided that its differential privacy system for the 2020 census would hold invariant (i.e., numbers will not change after application of the new disclosure avoidance system): a) the total population at the state and state-equivalents level, b) the total housing units at the census block level, and c) the number of group quarters facilities by type at the census block level.[12] The Census Bureau will make the remaining decisions about its differential privacy methodology in the first half of 2021, which will play a key role in the level of quality and accuracy achieved for the resulting dataset.

Unfortunately, there has been limited engagement of outside, interested stakeholders. While the Census Bureau created opportunities for a limited group of experts to weigh in on the demonstration products, such as through the December 2019 CNSTAT workshop and through its advisory committees, most interested stakeholders have not been actively engaged by the Census Bureau on differential privacy. Additionally, there has been a lack of transparency or clarity with respect to process, timing, and substance of what is happening with the development of differential privacy, which has made providing feedback and input to the Census Bureau difficult for many stakeholders.

Impact of Differential Privacy on Total Population and Racial and Ethnic Populations

Our first analysis of the different demonstration products focuses on the impact of differential privacy on the total population and population for different racial/ethnic groups. This analysis of county-level data provides an overall baseline on the impact and existing systemic biases (built-in errors that make values wrong for certain population groups at disparate rates) that exist in the way that the Bureau may implement differential privacy and post-processing. We compared the demonstration products (which tested different configurations of differential privacy to 2010 Census data) to data produced by the 2010 Census itself.

Impact on Total Population

Initial analysis of county-level total population yielded a small bias of adding total population to small population counties at the expense of larger-population counties. Original Demo 1 showed incredible swings from large to small counties, with a large bias of total population moving from census blocks designated urban to census blocks designated rural. These total population swings were dramatically improved in later demonstration products, yet smaller counties still feel the

[10] National Historical Geographic Information System (NHGIS), IPUMS, Privacy-Protected 2010 Census Demonstration Data (May 2020), https://www.nhgis.org/privacy-protected-demonstration-data#v20200527 ("Demo 2").
[11] National Historical Geographic Information System (NHGIS), IPUMS, Privacy-Protected 2010 Census Demonstration Data (Sept. 2020), https://www.nhgis.org/privacy-protected-demonstration-data#v20200917 ("Demo 3").
[12] U.S. Census Bureau, Invariants Set for 2020 Census Data Products, https://content.govdelivery.com/accounts/USCENSUS/bulletins/2ae5eda.

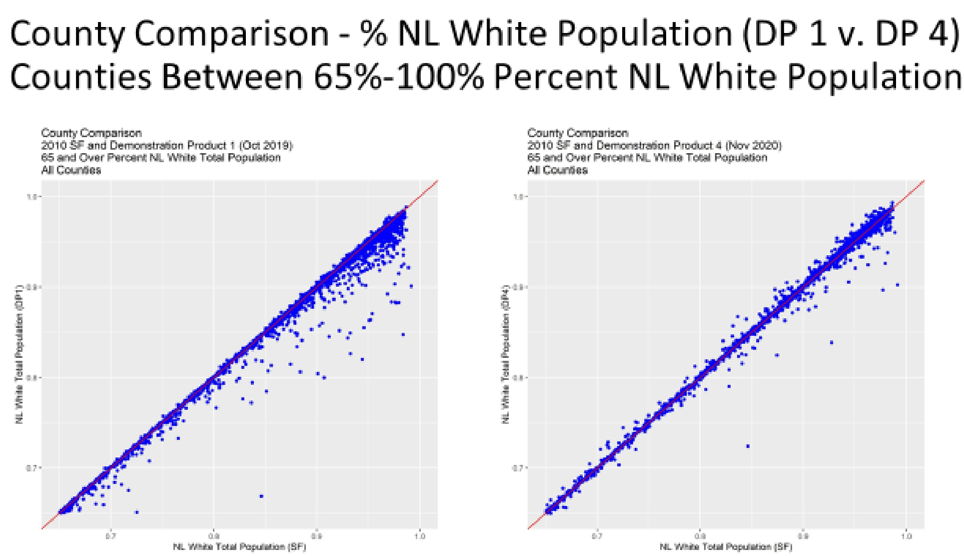
effects of differential privacy in Demo 4 compared to larger counties. Since smaller counties have small total populations to begin with, adding to or subtracting from total population will naturally affect the accuracy of their data more significantly.

Impact on Race and Ethnic Total Populations

While Demo 4 showed great improvement for total-population effects on the 2010 PL 94-171 file (the redistricting file dataset from the 2010 Census) at the county level compared to Demo 1, Demo 4 showed more variation for racial/ethnic population data.

Originally, Demo 1 showed small population counties tending to become more minority. This is to be expected if Demo 1 showed a general shift of urban population to more rural. These smaller, more often rural counties tended to be 80% Non-Latino (NL) White Alone population, on average. However, the opposite pattern emerged by the time Demo 4 was produced. Small counties went from generally gaining Latino, NL Asian, NL Black, and NL American Indian, Alaskan Native (AIAN) population to losing racial and ethnic minority population on average.

This pattern can be demonstrated another way: by looking at high percentage NL White counties. In Demo 1, high NL White counties were almost exclusively losing NL White population. In Demo 4, there is more of a gain in NL White population in these counties, meaning these counties are losing minority populations.



These findings show that differential privacy, as represented by Demo 4, is generally increasing the homogeneity of counties with clear majority race/ethnic populations. This could be alarming if this pattern of increasing homogeneity holds when looking at majority-minority districts drawn with these data products because this would indicate a likely increase in “false positive” majority-minority districts.

Redistricting and Differential Privacy

One of the primary concerns about differential privacy is whether the resulting data produced would be fit for use in redistricting. Data produced by the 2020 census must be of sufficient quality to meet the constitutional requirement of equal population between districts, and allow for full compliance with the federal Voting Rights Act of 1965 (VRA). If data produced are inaccurate and/or of poor quality, especially if there is a systemic bias in the resulting data, the legal requisites of redistricting would not be met.

Constitutional Requirement of Equal Population

The first criterion that must be met in redistricting is the principle of equal population – that is, each district should have the same total population.[13] This is a constitutional requirement. If districts are not close in total population, a redistricting plan is considered malapportioned and therefore unlawful.

Permissible deviations from strict equality may vary by type of district (e.g., congressional, state legislative, local) and overall size of jurisdiction (e.g., state, city). The fairly strict requirement of equal population means that differential privacy could cause proposed electoral districts to appear properly populated that are in fact malapportioned.

Because differential privacy holds statewide total population invariant, ideal district populations for statewide congressional and state legislative districts will not be affected by differential privacy. However individual districts' total populations will be affected because differential privacy would reallocate total populations of census blocks throughout a state.

Differential privacy **will** change total populations for local jurisdictions. This means differential privacy will change a jurisdiction's ideal population size as well as its individual district deviations (how one district differs in population from another) and total plan deviation (how much the entire map differs from every district having the same population). Depending on the quality of data, the changes to total populations at the block level could result in malapportioned plans, especially if there are systemic biases for smaller units of analysis in the application of differential privacy.

With these concerns in mind about the potential impact of differential privacy on the equal population requirement, we asked the following research questions in comparing redistricting plans made with the 2010 PL 94-171 file to the various 2010 demonstration products:

1. How do the Demonstration Products affect statewide and local individual district deviations? Are there patterns for which types of districts lose or gain total population?
2. How do the Demonstration Products affect statewide and local total plan deviations? Would deviation changes be so high that current plans would become malapportioned (e.g., have total plan deviation ranges greater than 10%)?

[13] *Reynolds v. Sims*, 377 U.S. 533 (1964).

Equal Population Analysis: Examining Differential Privacy's Effect on Total Plan Deviations

This preliminary report studied statewide congressional, state upper house, and state lower house districts as of 2018.[14]

Congressional Plan Deviations

Analysis of Demo 4 compared to the 2010 PL 94-171 file showed that differential privacy created malapportioned plans at the congressional level under the commonly-utilized traditional standard of zero deviation. In 2010, seven states only had one statewide congressional district.[15] Almost seventy percent of the remaining states (30 out of 43) with congressional district plans that originally had total plan deviation ranges under 10 people saw their plans become malapportioned under differential privacy. For 29 of those 30 malapportioned plans, their total plan deviation ranges went from under 10 people to an average of 210 people under Demo 4. Additionally, Virginia saw an anomaly in Congressional Districts 2 and 3 – in all four demonstration products, Virginia Congressional District 2 lost 18,000-19,000 total population to Congressional District 3, making the Virginia congressional plan go from a plan deviation range of 1 person to over 5 percent. The current application of differential privacy has a negative effect on fitness for use in the drawing of congressional districts.

State Legislative Plan Deviations

Compared to congressional redistricting, there is more flexibility available to state legislative and local redistricting plans in relation to population deviation (a "safe harbor" of 10 percent). And if we assume a strict 10-percent safe harbor as a malapportionment standard, it stands to reason there will be fewer plans affected because of a more flexible compliance threshold. That said, differential privacy still shows negative consequences on statewide legislative redistricting.

This report analyzed 47 state upper-house/state-senate plans[16] and found eight states had upper-house plan deviation ranges that exceeded 10 percent, through at least one demonstration product, that were previously under 10 percent in the actual 2010 data.[17] In comparison to congressional redistricting, it seems likely that the larger size of the upper-house districts, coupled with the deviation flexibility afforded state redistricting plans, means that state legislative redistricting plans have a greater ability to absorb fluctuations in population resulting from application of differential privacy, which would mean fewer upper-house state legislative plans are driven to malapportionment by the application of differential privacy.

[14] Census block equivalency files of 2018 congressional (116th Congress), state upper house (state senate), and state lower house (state assembly or house) districts were generated with Caliper Corporation data. Block equivalency files were then applied to the 2010 PL 94-171 file and the four released demonstration products to generate district population and demographic data.
[15] Alaska, Delaware, Montana, North Dakota, South Dakota, Vermont, and Wyoming
[16] For this report, Nebraska's Unicameral Legislative plan was included in state lower house analysis; Hawaii and Vermont were excluded since each district elected a varying-number of members.
[17] Idaho, Louisiana, Massachusetts, Mississippi, New York, South Dakota, West Virginia, and Wyoming

However, the same does not hold true with lower-house state legislative plans. Because lower-house state districts are smaller in population and because differential privacy has a bigger impact on smaller jurisdictions, these maps were more affected by the new disclosure avoidance system. Analysis of the demonstration products on 42 lower-house state legislative plans[18] compared to the 2010 PL 94-171 file showed 12 states with plan deviations that exceeded 10 percent under at least one demonstration product that were previously under 10 percent.[19] This distinction between the plans for upper house as compared to lower house is a concern because it points to the likelihood that as redistricting gets more and more local (with resultant smaller districts), differential privacy will become even more of an issue with respect to fitness for use. This is particularly problematic for communities of color because many of our communities are able to draw districts that afford them the opportunity to elect candidates of choice only at the local level.

Compliance with the Voting Rights Act of 1965 (VRA)

The VRA prohibits racial discrimination in voting practices in all federal, state, and local elections.[20] Section 2 of the VRA is a nationwide requirement that often requires the drawing of majority-minority electoral districts in order to avoid minority vote dilution.[21] A majority-minority district is a district where one racial minority group equals 50 percent or more of the citizen voting-age population (or, in the case of some jurisdictions, some percentage threshold of the voting age population). With respect to majority-minority districts, differential privacy could potentially change significantly the voting-age population of minority groups in districts. This could create "false positive" VRA districts that are not actually majority-minority by creating the illusion of a majority-minority district, or "false negative" VRA districts, where the fact that a VRA-mandated district could be drawn appears not to exist. Differential privacy could thus result in a failure to comply with the VRA.

Another potential impact of differential privacy on compliance with the VRA relates to racially-polarized voting (RPV) analysis, which is required for Section 2 vote dilution claims. RPV is a pattern of voting along racial lines where minority voters of the same race consistently support the same candidates and those minority candidates of choice are consistently defeated by candidates preferred by voters of the majority race.

In some instances, voter demographics are gathered from state voter registration files in order to conduct RPV analysis. In other instances, voter demographics can be estimated by the surname of registered voters, such as for Latino and Asian surnames. Where race/ethnicity cannot be generated from the voter file or surname list, census voting-age population data are used to estimate voter demographics. As differential privacy changes block-level voting-age population

[18] Excluded Arizona, North Dakota, and Washington as those plans were analyzed in the state upper house plans; excluded Hawaii, Maryland, New Hampshire, South Dakota, Vermont, and West Virginia since each district elected a varying-number of members. Washington DC City Council included.

[19] Arkansas, Delaware, Idaho, Louisiana, Maine, Massachusetts, Michigan, Mississippi, North Carolina, Tennessee, Texas, and Wyoming

[20] 52 U.S.C. §10301.

[21] *Thornburg v. Gingles*, 478 U.S. 30 (1986).

data, it will change the data used for RPV analysis in these circumstances. This noise could alter polarized voting patterns in ways that reduce apparent polarization, which would impede the ability to prove minority vote dilution and interfere with compliance with the VRA.

In order to assess differential privacy's impact on compliance with the VRA, we asked the following research questions – as compared to the actual 2010 PL 94-171 file:

1. What happens to minority district VAP percentages after differential privacy?
2. Are majority-minority districts more likely or less likely to be affected by differential privacy?
3. What are the potential effects of differential privacy on the ability to draw VRA-protected districts?
4. What are the potential effects of differential privacy on RPV analysis?[22]

Voting Rights Act Analysis: Examining Differential Privacy's Effect on Voting Age Population (VAP) and Potential *Gingles* Prong 1 Effects

With the released demonstration products, the first element of VRA compliance to be analyzed is the ability to draw majority-minority districts. This preliminary report examines all state lower house districts as of 2018.[23] Lower house districts are, on average, smaller than congressional or state upper house districts, allowing for the creation of more majority-minority districts. Therefore, examining current state lower house districts offers a potential glimpse into differential privacy's effects on majority-minority districts.[24]

This preliminary report analyzed state lower house districts' PL 94-171 voting age population (VAP) percentages in three different levels of population concentration by race/ethnicity:[25]

- 0%-44.9% respective race/ethnicity VAP – to examine low concentrations
- 45-65% respective race/ethnicity VAP – to examine current districts that may be just 50% VAP (or 50% CVAP, by inference)[26], to search for "false-positives" or "false negatives"
- 65.1%-100% respective race/ethnicity VAP – to examine high concentrations

[22] This question will be addressed in the final report and is outside the scope of this preliminary report.
[23] There are 4,785 state lower house legislative districts from all 50 states, including Nebraska's unicameral legislature and Washington, DC City Council's 8 districts.
[24] Congressional district and state upper house district data are available upon request.
[25] When calculating VAP percentages, this preliminary report used the U.S. Department of Justice's (DOJ) / Office of Budget and Management's (OMB) guidelines (https://obamawhitehouse.archives.gov/omb/bulletins_b00-02/) for calculating race/ethnicity when examining for ability to draw majority-minority districts:

- LVAP = Latino voting age population = Any Latino person of voting age population
- WVAP = White voting age population = Non-Latino White Alone voting age population
- AVAP = Asian voting age population = Non-Latino Asian Alone voting age population plus Non-Latino Asian & Non-Latino White voting age population
- BVAP = Black voting age population = Non-Latino Black Alone voting age population plus Non-Latino Black & Non-Latino White voting age population
- AIANVAP = American Indian and Alaskan Native voting age population = Non-Latino American Indian and Alaskan Native Alone voting age population plus Non-Latino American Indian and Alaskan Native & Non-Latino White voting age population

[26] Because voting-age population data includes both citizens and non-citizens, a higher percentage of voting-age population is used to infer meeting a 50 percent threshold of citizen population of voting age.

We assessed these issues across all the demonstration products and found that a pattern emerged in general under Demo 4, with high VAP percentage districts (65.1% and over) of any race/ethnicity tending to have higher percentages for that group than the district's original PL 94-171 VAP percentage. This is the converse of what was observed under Demo 1.

- **<u>Latino</u> –** There were 59 districts that had over 65.1% LVAP in the 2010 PL 94-171 file. Demo 1 featured 43 districts having lower LVAP percentages compared to the actual 2010 Census data, and 16 districts having higher LVAP percentages. Conversely, Demo 4 featured 24 districts having lower LVAP percentages and 35 districts having higher LVAP percentages. There was a 32.2% increase in districts gaining more LVAP in Demo 4 compared to Demo 1 (19 districts).
- **<u>Black</u> –** There were 83 districts that had over 65.1% BVAP in the 2010 PL 94-171 file. Demo 1 featured 47 districts having lower BVAP percentages and 33 districts having higher BVAP percentages. Conversely, Demo 4 featured 23 districts having lower BVAP percentages and 59 districts having higher BVAP percentages. There was a 31.3% increase in districts gaining more BVAP in Demo 4 compared to Demo 1 (26 districts).
- **<u>White</u> –** There were 3,586 districts that had over 65.1% WVAP in the 2010 PL 94-171 file. Demo 1 featured 1,929 districts having lower WVAP percentages and 1,555 districts having higher WVAP percentages. Conversely, Demo 4 featured 1,649 districts having lower WVAP percentages and 1,856 districts having higher WVAP percentages. There was an 8.4% increase in districts gaining more WVAP in Demo 4 compared to Demo 1 (301 districts).
- **<u>Asian and American Indian/Alaskan Native</u> –** There were 7 districts that had over 65.1% AVAP and 10 districts that had over 65% AIANVAP in the 2010 PL 94-171 file. There were not enough districts with high concentrations of AVAP or AIANVAP to draw conclusions.

The pattern of increasing VAP percentages also continued in the middle-range VAP percentage districts (45% to 65%), with the exception of middle-range White VAP districts. Middle-range White VAP districts tended to lose more VAP percentage in Demo 4 compared to Demo 1. With middle-range minority VAP districts increasing in percentages, this is to be expected as that increase was more than likely occurring at the expense of White VAP.

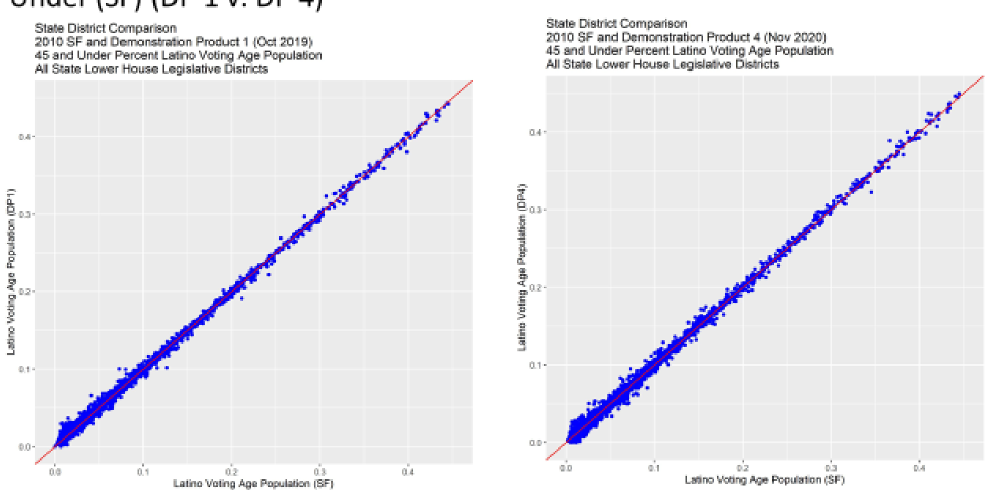
- **<u>Latino</u> –** There were 122 districts that had LVAP between 45% and 65% in the 2010 PL 94-171 file. Demo 1 featured 99 districts having lower LVAP percentages compared to the actual 2010 Census data, and 21 districts having higher LVAP percentages. Conversely, Demo 4 featured 52 districts having lower LVAP percentages and 67 districts having higher LVAP percentages. There was a 37.7% increase in districts that gained more LVAP in Demo 4 compared to Demo 1 (46 districts).
- **<u>Asian</u> –** There were 23 districts that had AVAP between 45% and 65% in the 2010 PL 94-171 file. Demo 1 featured 11 districts having lower AVAP percentages and 12 districts having higher AVAP percentages. Conversely, Demo 4 featured 6 districts having lower AVAP percentages and 17 districts having higher AVAP percentages. There was an increase of 21.7% in districts that gained more AVAP in Demo 4 compared to Demo 1 (5 districts).
- **<u>Black</u> –** There were 300 districts that had BVAP between 45% and 65% in the 2010 PL 94-171 file. Demo 1 featured 191 districts having lower BVAP percentages and 101 districts

having higher BVAP percentages. Conversely, Demo 4 featured 122 districts having lower BVAP percentages and 177 districts having higher BVAP percentages. There was an increase of 25.3% in districts that gained more BVAP in Demo 4 compared to Demo 1 (76 districts).

- **<u>White</u> –** There were 473 districts that had WVAP between 45% and 65% in the 2010 PL 94-171 file. Demo 1 featured 184 districts having lower WVAP percentages and 269 districts having higher WVAP percentages. Conversely, Demo 4 featured 247 districts having lower WVAP percentages and 217 districts having higher WVAP percentages. There was a decrease of 11% in districts that gained more WVAP in Demo 4 compared to Demo 1 (52 districts).
- **<u>American Indian/Alaskan Native</u> –** There were 12 districts that had between 45% and 65% AIANVAP in the 2010 PL 94-171 file. There were not enough districts with middle concentrations of AIANVAP to draw conclusions.

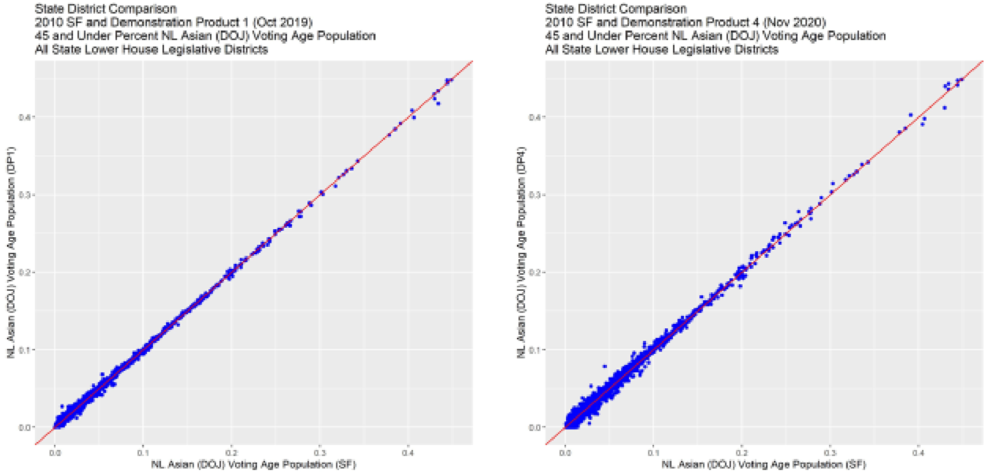
If middle and high-minority VAP districts are gaining more VAP percentage under Demo 4 compared to Demo 1, then the opposite is true of low VAP districts (districts 44.9% VAP and under). Further, Demo 4 also shows greater fluctuations in VAP percentages compared to Demo 1, which could result in an increase in false negative or false positive majority-minority districts under differential privacy. For example, low LVAP districts went from an average loss of -0.22% LVAP and average gain of +0.25% LVAP in Demo 1 to an average loss of -0.27% LVAP and average gain of +0.30% LVAP in Demo 4. Low AVAP districts went from an average loss of -0.14% AVAP and average gain of +0.15% AVAP in Demo 1 to an average loss of -0.24% AVAP and average gain of +0.24% AVAP in Demo 4. Low BVAP districts went from an average loss of -0.15% BVAP and average gain of +0.16% BVAP in Demo 1 to an average loss of -0.28% AVAP and average gain of +0.29% AVAP in Demo 4. Low AIANVAP districts went from an average loss of -0.11% AIANVAP and average gain of +0.11% AIANVAP in Demo 1 to an average loss of -0.19% AIANVAP and average gain of +0.19% AIANVAP in Demo 4. These increases in VAP distribution can be observed in the graphs below.

State Lower House Districts, Latino VAP (LVAP) Comparison – 45% LVAP and Under (SF) (DP 1 v. DP 4)



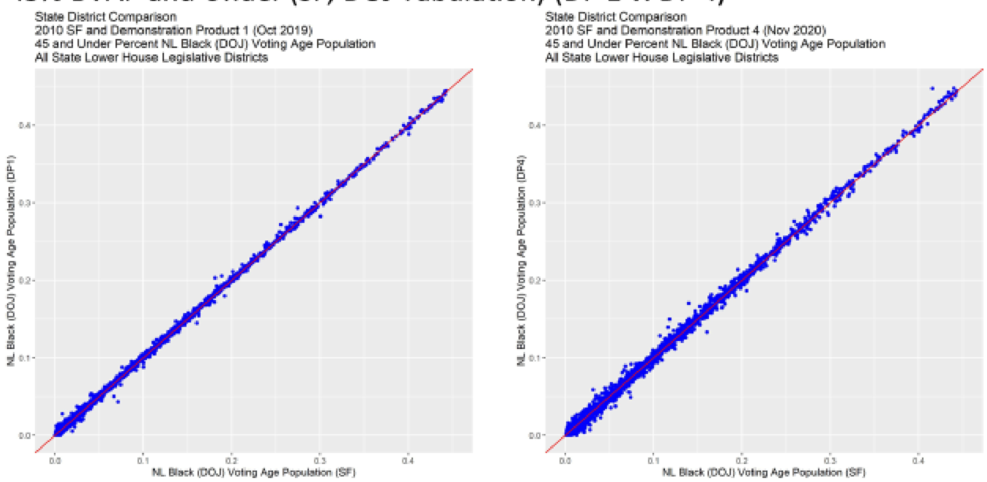
DOJ Latino Tabulation = Latino Population

State Lower House Districts, Non-Latino Asian VAP (AVAP) Comparison – 45% AVAP and Under (SF, DOJ Tabulation) (DP 1 v. DP 4)



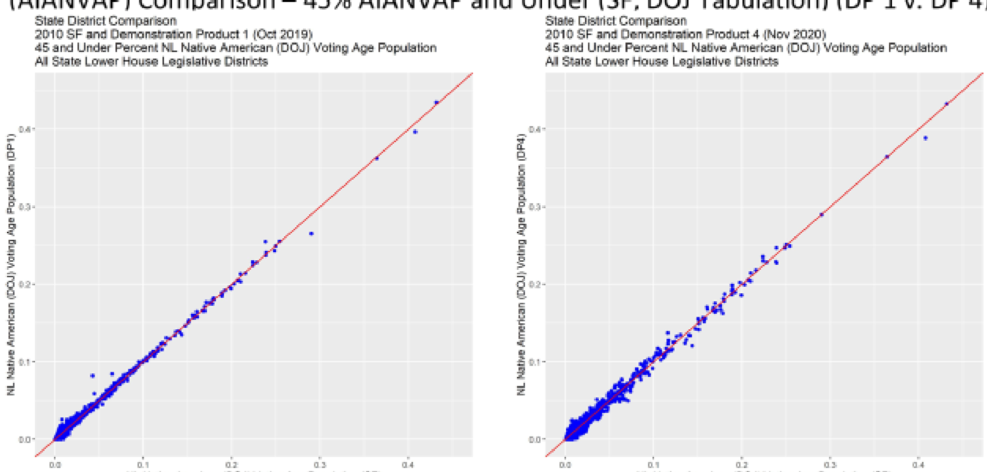
DOJ Asian Tabulation = NL Asian Alone + NL Asian & NL White

State Lower House Districts, Non-Latino Black VAP (BVAP) Comparison – 45% BVAP and Under (SF, DOJ Tabulation) (DP 1 v. DP 4)



DOJ Black Tabulation = NL Black Alone + NL Black & NL White

State Lower House Districts, Non-Latino American Indian/Alaskan Native VAP (AIANVAP) Comparison – 45% AIANVAP and Under (SF, DOJ Tabulation) (DP 1 v. DP 4)



DOJ AIAN Tabulation = NL AIAN Alone + NL AIAN & NL White

Together, these comparisons of Demo 4 and Demo 1 data show shifts where districts with high VAP concentrations of any particular race or ethnic group have an even higher percentage of that group and show greater fluctuations in district VAP changes. It appears that concerns about false positive and false negative majority-minority districts are valid, with the analysis showing that false positive districts are more likely to occur. This research casts some doubt on differential privacy's ability to produce data that is fit for VRA compliance.

Preliminary Conclusions & Next Steps

Based on our preliminary analysis of Census Bureau demonstration products, we currently have grave concerns about the negative effect differential privacy will likely have on the ability to properly redistrict during 2021. Our preliminary analysis shows that changes in total population affected the equal population requirement for redistricting, with 30 congressional plans that previously had deviation ranges under 10 people going over that range in all demonstration products. Similarly, 12 state lower-house plans and 8 state upper-house plans that previously had deviation ranges under 10 percent went over that at some point in the various demonstration products. While recent Demonstration Products (Demo 3, Demo 4) showed improvement in total population matching compared to earlier Demonstration Products (Demo 1, Demo 2), they also showed more variation in changes to race/ethnicity percentages, particularly for smaller jurisdictions.

Further, recent demonstration products showed more variation for population VAP percentages than earlier ones. This raises the concern that because majority-minority VAP districts may not be as accurate, redistricters might not draw majority-minority VAP districts where they should and might draw them where they should not. These preliminary findings point to the possible need to drastically alter the way the Census Bureau plans to design and implement its disclosure avoidance system on the redistricting data file.

An important caveat is that there remain decisions to be made with respect to differential privacy that the Census Bureau asserts would address many of these issues. However, there has been a dearth of transparency, clarity, and engagement with external stakeholders to date on differential privacy. To that end, we support the Census Bureau's effort to slow its current decision-making process on differential privacy in order to take more time to properly assess the use of differential privacy on 2020 census data releases.[27] We also support the Census Bureau's recent announcement to release an additional set of demonstration products for the redistricting data product that will more readily approximate the anticipated privacy/accuracy tradeoff for the forthcoming 2020 Census data products.[28] With the additional time, the Census Bureau must meaningfully engage external stakeholders by:

[27] *See*, U.S. Census Bureau, Fine-Tuning the Disclosure Avoidance System to Ensure Accuracy (Feb. 3, 2021), https://content.govdelivery.com/accounts/USCENSUS/bulletins/2be626f.
[28] *Id*.

- Providing more transparency about what has occurred with differential privacy to date, including providing an explanation for what changes have been made to the system and the impact of those changes; and
- Providing the space and opportunity for stakeholders to inform the Bureau on how best to assess options for "final decisions" regarding differential privacy.

MALDEF and Advancing Justice | AAJC will continue analyzing the data to further understand the impact of differential privacy on redistricting. The final report will include:

- A California-based case study on the impact of differential privacy on local redistricting; and
- Analysis of the impact of differential privacy on estimates of racially polarized voting.