

By Jade Ford and Molly Danahy¹ | May 4, 2021

UNDERSTANDING DIFFERENTIAL PRIVACY AND ITS IMPACT ON 2020 CENSUS DATA

Introduction

The Census Bureau is required by law to maintain the confidentiality of individual responses to the decennial census.² To meet this statutory mandate, every census cycle the Bureau develops a “disclosure avoidance system” (“DAS”), which it applies to the decennial census data before releasing them to the public. This year, the Bureau plans to use a new methodology for disclosure avoidance: “differential privacy.”

This issue brief examines the basic mechanics of using differential privacy to protect census data, the motivations behind the Bureau’s decision to use it, and its potential impact on the 2020 Census data. Throughout, the brief focuses on the impact on redistricting and voting rights, as well as the potential risks and benefits for communities of color. In order to better understand the true impact of differential privacy, this brief also puts differential privacy in context with other factors that affect the accuracy of the decennial census data—including the substantial differential undercount of communities of color, efforts to add a citizenship question to the 2020 Census, and past privacy measures used by the Bureau.

Given that the Census Bureau has not yet released their final DAS plan for 2020, it is still too early to determine precisely how differential privacy will impact the 2020 Census data, and by extension redistricting and voting rights for the next decade. That said, our preliminary findings suggest there are benefits to using a systematic method like differential privacy when it comes to litigating redistricting and voting rights claims. This type of method should enable experts to measure the impact of the Bureau’s DAS on census population data and estimates of racially polarized voting using confidence intervals. At the same time, however, there are legitimate concerns about the potential impact the Bureau’s new DAS methodology will have on the 2020 Census data, and civil rights groups should continue to analyze its effects.

This issue brief includes (1) an update on the current timeline of the census and important dates concerning data delivery; (2) a brief overview of the Bureau’s decision to use differential privacy as its chosen privacy measure and the basic mechanics of their plan to apply it to 2020 Census

¹ We would like to thank Ruth Greenwood for her leadership on this project and Meredith Manda for her assistance in researching and drafting.

² See 13 U.S.C. § 9(a)(2).

data; (3) a closer look at how the Bureau’s use of differential privacy may benefit or impede redistricting and voting rights; and (4) a brief overview of ongoing litigation related to the 2020 Census timeline and differential privacy.

Part I: Overview of the 2021 Census Delivery Timeline

In a typical census cycle, the Census Bureau would have already published the total state population counts used for reapportionment and the dataset used for state redistricting, otherwise known as the PL 94-171 file. Indeed, the statutory deadline for delivering the apportionment count was December 31, 2020, and March 31, 2021 for the PL 94-171 file.³

However, due to delays in census operations caused by the COVID-19 pandemic, the Census Bureau announced it would release the 2020 Census data on a delayed timeline.⁴ The Bureau delivered the apportionment data to President Biden on April 26, 2021,⁵ and plans to release the PL 94-171 file by September 30, 2021.⁶

In response to pressure from states to release the PL 94-171 file earlier, the Bureau has also announced it will release “legacy format” state redistricting data in mid-to-late August, 2021.⁷ The legacy format summary redistricting data file is unlike the final PL 94-171 file in that it does not have the “tabulations” (i.e., data tables) for state redistricting data. The legacy format file also does not have a user-friendly system for data access. Instead, the Bureau has indicated that states attempting to use the legacy format file before the PL 94-171 file is released would have to use “an outside vendor to process [the] legacy format summary redistricting data files if states do not have the capacity to tabulate the data on their own.”⁸ In its press release, the Bureau warned of the “difficulty of using the data in this format,” and noted that “any state using legacy format summary redistricting data files would have to accept responsibility for how they process these files; whether correctly or incorrectly.”⁹

Based on this current timeline, we can expect states to begin redistricting in mid-to-late August at the earliest, and more likely after September 30, 2021.

³ 13 U.S.C. §§ 141(b), (c).

⁴ See James Whitehorn, *Timeline for Releasing Redistricting Data*, U.S. Census Bureau, <https://www.census.gov/newsroom/blogs/random-samplings/2021/02/timeline-redistricting-data.html>.

⁵ *2020 Census Apportionment Results Delivered to the President*, U.S. Census Bureau (Apr. 26, 2021), <https://www.census.gov/newsroom/press-releases/2021/2020-census-apportionment-results.html>.

⁶ *Important Dates*, U.S. Census Bureau, <https://2020census.gov/en/important-dates.html>.

⁷ *U.S. Census Bureau Statement on Release of Legacy Format Summary Redistricting Data File*, U.S. Census Bureau (Mar. 15, 2021) (summarizing declaration in *Ohio v. Raimondo* case), <https://www.census.gov/newsroom/press-releases/2021/statement-legacy-format-redistricting.html>.

⁸ *Id.*

⁹ *Id.*

Part II: Why the Bureau Decided to Use Differential Privacy and How It Works

The story of the Bureau’s decision to use differential privacy for the 2020 Census begins with the failure of its prior disclosure avoidance system to protect against new threats. For the last two cycles, the Census Bureau relied primarily on “data swapping” for disclosure avoidance. Under this approach, the Bureau first began by identifying “[h]ousehold records with a high risk of disclosure” due to “a unique combination of certain variables.”¹⁰ These “unique records [were then] targeted for data swapping,” where certain attributes are swapped between records, but the overall census block population total remains the same.¹¹ To give a basic example:

Consider a census block with just 20 people in it, including one Filipino American. Without any disclosure avoidance effort, it might be possible to figure out the identity of that individual. With data swapping, the Filipino American’s data might be swapped with that of an Anglo American from a nearby census block—a census block where other Filipino Americans reside. The details for the person would be aggregated with others, and therefore not identifiable, and yet the total population in both census blocks would remain accurate.¹²

The Bureau also used “partially synthetic data” (i.e., data produced by a model that replaces the real data) for group quarters data,¹³ as well as “noise infusion” to “hide very unusual characteristics of a person or household” for certain limited categories.¹⁴

¹⁰ *Id.* See also Laura McKenna, *Disclosure Avoidance Techniques Used for the 1960 Through 2010 Decennial Censuses of Population and Housing Public Use Microdata Samples*, U.S. Census Bureau (April 2019), <https://www2.census.gov/adrm/CED/Papers/FY20/2019-04-McKenna-Six%20Decennial%20Censuses.pdf>.

¹¹ *Id.*

¹² *Differential Privacy for Census Data Explained*, National Conference of State Legislatures (Mar. 15, 2021), <https://www.ncsl.org/research/redistricting/differential-privacy-for-census-data-explained.aspx>.

¹³ Group quarters are places “where people live or stay, in a group living arrangement that is owned or managed by an entity or organization providing housing and/or services for the residents.” *2020 Census Group Quarters Definitions and Code List*, U.S. Census Bureau (2020), <https://2020census.gov/content/dam/2020census/materials/group-quarters/code-types/2020%20Census%20Group%20Quarters%20Type%20Codes.pdf>. Group quarters include college dormitories, nursing homes, and correctional facilities. *Id.*; see also *Group Quarters/Residence Rules*, U.S. Census Bureau (Mar. 20, 2018), <https://www.census.gov/topics/income-poverty/poverty/guidance/group-quarters.html>.

¹⁴ Laura McKenna, *Disclosure Avoidance Techniques Used for the 1960 Through 2010 Decennial Censuses of Population and Housing Public Use Microdata Sample*, U.S. Census Bureau 5 (Apr. 2019), <https://www2.census.gov/adrm/CED/Papers/FY20/2019-04-McKenna-Six%20Decennial%20Censuses.pdf>.

The 2010 DAS failed to provide adequate protections in the face of new privacy threats. With the dramatic rise in the volume of personal data that is available publicly or on commercial databases, census data have become increasingly vulnerable to “reconstruction attacks” and “linkage attacks.” In a reconstruction attack, data analysts use public aggregate statistics about the private dataset to reconstruct records in the dataset (e.g., data on census block, ethnicity, age, race, and sex). And in a linkage attack, data analysts cross-reference datasets to actually re-identify data (i.e., determine the identity of the individual respondent).

After the 2010 Census data was released, data scientists at the Bureau conducted a reconstruction-abetted linkage attack that re-identified data from a staggering 17% of the U.S. population by cross-referencing census data with commercial databases.¹⁵ Importantly, that attack only used a small fraction of the total number of statistics released.¹⁶ Moreover, experts believe the disclosure risks for the 2020 Census are even higher, since commercial datasets today are more detailed and comprehensive than those available in 2010.

The takeaway from these attacks was clear: the Bureau needed to develop a new methodology for disclosure avoidance; it chose differential privacy.

Differential privacy is a mathematical definition of privacy, not a specific algorithm or statistical tool.¹⁷ An algorithm is said to be “differentially private” if, by looking at the output, one cannot

¹⁵ See Simson L. Garfinkel, *Deploying Differential Privacy for the 2020 Census of Population and Housing*, U.S. Census Bureau (July 31, 2019), <https://www.census.gov/content/dam/Census/newsroom/press-kits/2019/jsm/presentation-deploying-differential-privacy-for-the-2020-census-of-pop-and-housing.pdf>. A professor at Columbia was able to reproduce a similar reconstruction attack for a smaller population. Mark Hansen, *To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data*, N.Y. Times (Dec. 5, 2018), <https://www.nytimes.com/2018/12/05/upshot/to-reduce-privacy-risks-the-census-plans-to-report-less-accurate-data.html>. And experts from both inside and outside the Census Bureau have addressed the claim by Dr. Stephen Ruggles that there is no real threat of re-identification. See Amicus Brief of Data Privacy Experts, *Alabama v. U.S. Dep’t of Comm.*, No. 3:21-cv-00211-RAH-KFP, ECF. No. 99-1 at 5-7 (M.D. Al. Apr. 23, 2021); Supplemental Declaration of John M. Abowd, Chief Scientist and Associate Director for Research and Methodology at the U.S. Census Bureau, *Alabama v. U.S. Dep’t of Comm.*, No. 3:21-cv-00211-RAH-KFP, ECF. No. 116-1, at 10-20 (M.D. Al. Apr. 26, 2021).

¹⁶ Declaration of John M. Abowd, Chief Scientist and Associate Director for Research and Methodology at the U.S. Census Bureau, *Alabama v. U.S. Dep’t of Comm.*, No. 3:21-cv-00211-RAH-KFP, Ex. B. at 1 (M.D. Al. Apr. 13, 2021).

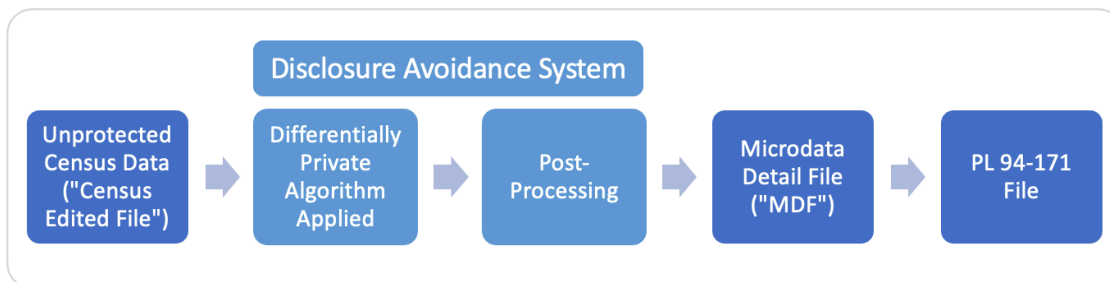
¹⁷ See Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy* 10 (2014), <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>; *Differential Privacy*, Harvard Privacy Tools Project, <https://privacytools.seas.harvard.edu/differential-privacy>; see also Supplemental Declaration of John M. Abowd, Chief Scientist and Associate Director for Research and Methodology at the U.S. Census Bureau, *Alabama v. U.S. Dep’t of Comm.*, No. 3:21-cv-00211-RAH-KFP, ECF. No. 116-1, at 10 (M.D. Al. Apr. 26, 2021).

tell whether an individual's data was included in the original dataset or not.¹⁸ This means that an individual respondent will not suffer any increased risk of harm due to participation.

Differential privacy is ensured by adding carefully calibrated random noise at certain points in the computation. The amount of noise that the Bureau plans to introduce into the census data is governed by a parameter called "epsilon." The Bureau has a total "privacy budget" (i.e., epsilon budget) to allocate across each level of the census hierarchy: nation, state, county, tract, block group, and block.

The Bureau plans to implement a two-step DAS process for the 2020 Census data. First, the Bureau will introduce noise into the census data across each level of the census hierarchy, determined by the allocation of the privacy budget.¹⁹ This is designed to make the data differentially private. Second, the Bureau will engage in post-processing to ensure the data meets certain expectations for what census data "should" look like.²⁰ These post-processing decisions are about the file format and data utility, not protecting privacy. Specifically, post-processing will ensure that there are no (1) partial people; (2) negative people; and (3) inconsistencies between population counts in levels of the hierarchy (e.g., where blocks do not add up to tracts).²¹ Many experts regard post-processing as a "response to a political constraint," since "[m]ost statistically-minded data users could work with the pre-processed data."²² They also worry that post-processing (which is separate from differential privacy) will "introduce a range of oddities into the data" that may be hard to adjust for in data analysis.²³

The 2020 Data Pipeline²⁴



The Census Bureau is still developing its final plan for the 2020 DAS. It has not yet determined how big the privacy budget (i.e., epsilon) will be, where to allocate the privacy budget (i.e., what level of the census hierarchy), and what types of outputs they want to fix in post-

¹⁸ Cynthia Dwork & Aaron Roth, *The Algorithmic Foundations of Differential Privacy* 10 (2014), <https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf>.

¹⁹ danah boyd, *Balancing Data Utility and Confidentiality in the 2020 US Census*, Data & Society at 22 (Apr. 27, 2020), https://datasociety.net/wp-content/uploads/2019/12/Differential-Privacy-04_27_20.pdf

²⁰ *Id.* at 24-25.

²¹ *Id.* at 24.

²² *Id.* at 25.

²³ *Id.* at 25-26.

²⁴ *Id.* at 24.

processing (and how). Until the Bureau releases its final plan, we can only theorize about how differential privacy will impact the 2020 Census data based on demonstration products and what we know about the mechanics of differential privacy versus swapping.

Part III: Potential Benefits and Drawbacks of Using Differential Privacy in the 2020 Census for Redistricting and Voting Rights

This section begins by putting differential privacy and the 2020 DAS into context by laying out the substantial sources of inaccuracy in census data that already exist before any disclosure avoidance mechanism is even applied. Redistricters and experts in litigation already work with this uncertainty in census data, and the noise introduced by differential privacy may amount to a proverbial drop in the bucket when compared with these other sources of error, depending on the Bureau's final plan.

With that context in mind, this section then turns to the key question on the minds of civil rights advocates: how will the Bureau's decision to use differential privacy for the 2020 Census impact redistricting and voting rights? To answer that question, this section lays out what using differential privacy will, and will not, change for apportionment and redistricting. It then examines potential benefits and drawbacks of using differential privacy when compared with swapping, focusing on its impact on the fairness and accuracy of redistricting and the viability of voting rights claims.

A. The Big Picture: Putting Differential Privacy in Context with Other Factors that Affect the Accuracy of Decennial Census Data

While the Bureau was still in the data collection phase for the 2020 Census, two major forces affected the Bureau's enumeration at a scale that was likely many levels of magnitude larger than any potential alterations from differential privacy: (1) the differential undercount of communities of color; (2) the COVID-19 pandemic, and (3) the attempt to add a citizenship question to the census.

1. *The Differential Undercount of Communities of Color*

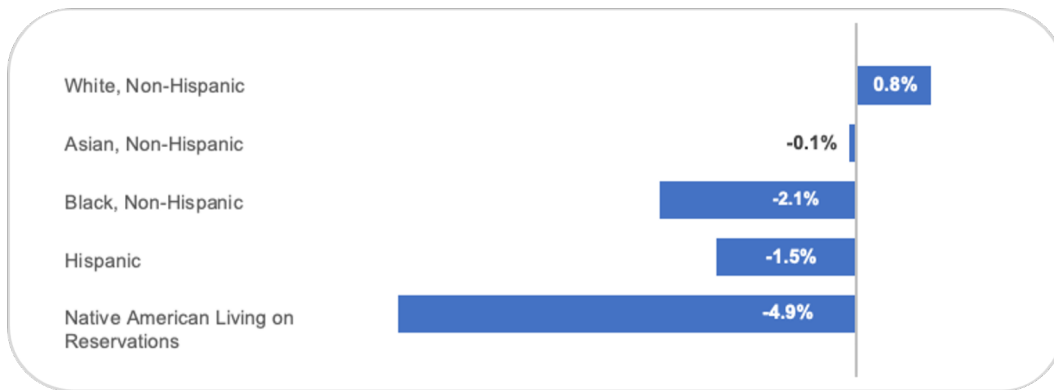
The decennial census has consistently undercounted Black, Hispanic, and Native American communities at higher rates than the white population since at least 1940.²⁵ Earlier reports from the Bureau document the undercount of the Black population as far back as 1870.²⁶ Despite

²⁵ Nathaniel Persily, *The Right to Be Counted Counting on the Census? Race, Group Identity, and the Evasion of Politics*, 53 STAN. L. REV. 1077, 1081 (2001).

²⁶ See Dan Bouk & danah boyd, *Democracy's Data Infrastructure*, Knight First Amendment Inst. at Columbia Univ. (Mar. 18, 2021), <https://knightcolumbia.org/content/democracys-data-infrastructure>.

efforts to mitigate the differential undercount, these problems still persist. The table below shows the differential undercount by race and ethnicity in the 2010 Census.²⁷

Percent Net Undercount/Overcount in the 2010 Census by Race and Ethnicity²⁸



To make these numbers more concrete, just combining the undercount from the Black and Hispanic populations, the 2010 Census failed to count approximately 1.5 million people of color.²⁹ And the differential undercount is expected to be even higher for the 2020 Census due to challenges caused by the COVID-19 pandemic, reduced funding, and the attempt to add a citizenship question.³⁰

This differential undercount reflects biases inherent in the way the census is conducted. For instance, reliance on Internet responses, mail-back forms, and door-to-door enumerator follow-ups “necessarily biases the numbers to the disadvantage of certain groups.”³¹ As a result, communities that face greater socioeconomic barriers, language barriers, and racial discrimination are less likely to be counted by the decennial census. Likewise, communities with higher levels of government distrust, or even fear, experience higher rates of undercounting.³²

²⁷ 2010 Census Coverage Measurement Memo #2010-G-01, *supra* note 2, at 15, 18.

²⁸ Thomas Wolf et al., *Getting the Count Right: Key Context for the 2020 Census*, Brennan Ctr. (Mar. 31, 2020), <https://www.brennancenter.org/sites/default/files/2020-03/CensusPrimer.pdf> (citing 2010 Census Coverage Measurement Estimation Report, 15, table 7, <https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g01.pdf>; Hogan et al., “Quality and the 2010 Census,” 647, table 6, <https://doi.org/10.1007/s11113-013-9278-5>; note this excludes Native Americans living “off Reservation” (1.95%) and Native Hawaiian or other Pacific Islander (-1.34%)).

²⁹ 2010 Census Missed 1.5 Million Minorities, CBS News (May 22, 2012), <https://www.cbsnews.com/news/2010-census-missed-15-million-minorities>.

³⁰ See Diana Elliott et al., *Assessing Miscounts in the 2020 Census*, URBAN INST. 1-2 (June 11, 2019), https://www.urban.org/sites/default/files/publication/100324/assessing_miscounts_in_the_2020_census_1.pdf.

³¹ Persily, *supra* note 25, at 1084.

³² Persily, *supra* note 21, at 1083.

This source of inaccuracy in census population counts—which disproportionately affects communities of color—will very likely surpass any alterations caused by the application of differential privacy or other privacy measures.

2. *The COVID-19 Pandemic*

The COVID-19 pandemic had a significant impact on 2020 Census operations, and by extension the accuracy of the data. Just as the Bureau was preparing for in-person counting, states began instituting lockdowns to try to control the spread of the pandemic. As a result, the Bureau chose to postpone in-person door-knocking until mid-July, and even then the pandemic continued to impede responses.³³ To make matters worse, after initially requesting more time to complete the census due to the pandemic, the Trump administration suddenly reversed its position. In early August, the Bureau announced that it would cut the timeline for collecting and processing the 2020 Census data in half in order to deliver the apportionment count to President Trump by the end of 2020.³⁴ This rush posed a direct threat to the accuracy of the count, particularly for communities of color, since many of the steps the Bureau planned to cut short were designed to ensure a better count of hard-to-count populations.³⁵

Due to litigation by a coalition of civil rights groups and local governments and officials, the timeline was ultimately extended.³⁶ Nevertheless, Census Bureau employees speaking on the condition of anonymity as well as outside experts have warned that the disruptions due to COVID-19 and the changing timelines under the Trump administration affected the accuracy of the 2020 Census data.³⁷

³³ See *2020 Census Operational Adjustments Due to COVID-19*, U.S. Census Bureau (2020), <https://2020census.gov/en/news-events/operational-adjustments-covid-19.html>; see also Hansi Lo Wang, *Here's How the 1st 2020 Census Results Changed Electoral College, House Seats*, NPR (Apr. 26, 2021), <https://www.npr.org/2021/04/26/983082132/census-to-release-1st-results-that-shift-electoral-college-house-seats>.

³⁴ *Statement from U.S. Census Bureau Director Steven Dillingham: Delivering a Complete and Accurate 2020 Census Count*, U.S. Census Bureau (Aug. 3, 2020), <https://www.census.gov/newsroom/press-releases/2020/delivering-complete-accurate-count.html>.

³⁵ See Kelly Percival & Madiba Dennie, *The State of Census Lawsuits on the Eve of Key Data Releases*, Brennan Ctr. (Apr. 23, 2021), <https://www.brennancenter.org/our-work/analysis-opinion/state-census-lawsuits-eve-key-data-releases> (describing challenges to 2020 Census timeline).

³⁶ See *National Urban League v. Ross: Court Case Tracker*, Brennan Ctr. (last updated Apr. 22, 2021), <https://www.brennancenter.org/our-work/court-cases/national-urban-league-v-ross>.

³⁷ See Hansi Lo Wang, *Census Missed Year-End Deadline for Delivering Numbers for House Seats*, NPR (Dec. 30, 2020), <https://www.npr.org/2020/12/30/951566925/census-to-miss-year-end-deadline-for-delivering-numbers-for-house-seats>.

3. *The Attempt to Add a Citizenship Question*

The Trump administration's attempt to add a citizenship question also undermined the accuracy of the 2020 Census data.³⁸ In June 2019, the Supreme Court blocked this effort.³⁹ However, the controversy surrounding the citizenship question and the administration's continued efforts to collect citizenship data through other means likely exacerbated the differential undercount of certain communities—and particularly communities with higher concentrations of noncitizens—by depressing participation.⁴⁰ This chilling effect was also compounded by the July 21, 2020, Presidential Memorandum, which announced that the administration planned to exclude undocumented noncitizens from the apportionment count.⁴¹

Even though the Biden administration has since reversed this policy, the July 21 Presidential Memorandum and the controversy over the citizenship question likely did irreversible harm to participation in the 2020 Census.

These factors—along with many other features of the Bureau's enumeration efforts including imputation methods for non-responsive households—introduce substantial error into census data before any privacy measures are even applied.

B. The Impact of Using Differential Privacy for Redistricting and Voting Rights

Turning to the impact of differential privacy, it is important to first recognize that the Census Bureau is required to alter the census data in *some* way in order to protect the privacy of individual respondents. Specifically, the Bureau is prohibited from “mak[ing] any publication whereby the data furnished by any particular establishment or individual . . . can be identified,”⁴² and is required to keep personally identifiable information from the decennial census confidential for 72 years.⁴³ As such, the question is not *whether* the Census Bureau will alter the PL 94-171 redistricting file before releasing it to the states and the public, but *how* the Bureau will alter the data. Any meaningful debate about differential privacy must therefore compare differential privacy to the Bureau's prior privacy measures or alternatives, rather than evaluate differential privacy as compared with no alterations to the data whatsoever.

³⁸ See Brennan Ctr., *supra* note 23, at 6; see also *New York v. Trump*, 485 F. Supp. 3d 422, 448-49 (S.D.N.Y. 2020).

³⁹ *Department of Commerce v. New York*, 139 S. Ct. 2551, 2576 (2019).

⁴⁰ See Brennan Ctr., *supra* note 23, at 6.

⁴¹ *Excluding Illegal Aliens From the Apportionment Base Following the 2020 Census*, 85 Fed. Reg. 44679 (July 21, 2020), <https://www.federalregister.gov/documents/2020/07/23/2020-16216/excluding-illegal-aliens-from-the-apportionment-base-following-the-2020-census>.

⁴² 13 U.S.C. § 9(a)(2). The Bureau is also prohibited from using census data “for any purpose other than the statistical purposes for which it is supplied.” *Id.* § 9(a)(1).

⁴³ *The 72-Year Rule*, U.S. Census Bureau (citing Pub. L. No. 95-416, 92 Stat. 915), https://www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html.

1. *What Using Differential Privacy Will (and Will Not) Change*

When the Bureau applies the algorithm that is designed to make the census data differentially private, it plans to designate certain statistics as “invariants.”⁴⁴ The use of differential privacy thus will not alter the following statistics at all at the specified levels of geography (and all higher levels): (1) total population at the state level, (2) total housing units at the census block level, and (3) number of group quarters facilities by type at the census block level.⁴⁵ Importantly, this means that using differential privacy will not change the total population count for each state and thus will not affect apportionment.

Instead, the relevant statistics that differential privacy will alter—from the standpoint of civil rights advocates interested in redistricting and voting rights—are total population counts at the census block level, as well as demographic data by census block, including data on race, ethnicity, and age.

2. *Potential Benefits of Using Differential Privacy*

Based on what we know thus far, there are at least two potential benefits of using differential privacy, as opposed to swapping, from the perspective of redistricting and voting rights. The first major benefit is improvement in confidentiality protections, particularly for small census blocks. Confidentiality is critical for participation in the census, and by extension the accuracy of the census. If the Bureau is no longer able to promise that census responses will remain confidential, response rates will likely plummet, particularly among communities that have reason to distrust the government and other actors. Depressed participation would severely undermine the accuracy of the census data, as well as any redistricting plan or voting rights claim that relies on the data. That is precisely why 275 civil rights organizations, civic leaders, and state and local elected officials sought to boost census participation by emphasizing its confidentiality in the lead up to the 2020 Census.⁴⁶

Second, differential privacy may also enable advocates to measure the uncertainty introduced by the Bureau’s chosen DAS. Under the Bureau’s previous privacy technology (swapping), there was no way to determine how the alterations had impacted the data in any systematic way. By contrast, because the Bureau’s plan to apply differential privacy involves adding random noise drawn from a publicly specified probability distribution into the counts followed by certain post-processing steps, it may be possible to create confidence intervals that measure the uncertainty introduced by the privacy protections for certain analyses.

⁴⁴ *Invariants Set for 2020 Census Data Products*, U.S. Census Bureau (Nov. 25, 2020), <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html>. The Bureau also plans to allocate a larger privacy budget (i.e., less will introduce less error) into the AIAN tribal area data in response to some of the concerns raised by Native American communities. *Id.*

⁴⁵ *Id.*

⁴⁶ *Helping Keep the Census Safe*, MALDEF, <https://www.maldef.org/census-pledge/>.

The ability to measure the uncertainty introduced into census data could benefit advocates seeking to bring Voting Rights Act claims, and specifically Section 2 claims. With respect to the requirement that plaintiffs demonstrate the existence of racially polarized voting (“RPV”), differential privacy may enable experts to incorporate confidence intervals to account for the way in which the Bureau’s privacy measure affected the data and the RPV analysis. This would be more accurate than the current approach, which does not (and cannot) account for any changes due to the Bureau’s use of swapping.⁴⁷ Moreover, the inability to measure the uncertainty introduced into the 2000 and 2010 data by swapping in RPV analyses was particularly problematic given that swapping has a homogenizing effect on census blocks.

Relatedly, differential privacy may also enable experts to provide confidence intervals to account for alterations in sub-state population counts due to the Bureau’s privacy protections. These sub-state population counts are relevant to both one person, one vote claims, as well as the majority-minority district analysis in Section 2 claims. While swapping did not alter the population totals for census blocks (and therefore would not have required the use of confidence intervals for those counts), adding in confidence intervals to account for changes from differential privacy may in fact be a more realistic way for courts to understand census data. As detailed above, there are numerous sources of error in the census data, which disproportionately affect communities of color, even before any privacy measure is applied. As such, it may be an improvement for litigants and courts to acknowledge that there is a margin of uncertainty inherent in census population data. If courts are amenable to this approach (which, to be sure, is a big if), differential privacy would at least enable experts to measure the uncertainty caused by the Bureau’s privacy protections for judges adjudicating redistricting and voting rights claims.

3. *Potential Drawbacks of the Bureau’s Proposed 2020 DAS*

There are also a number of potential risks associated with the Bureau’s proposed DAS for 2020, when compared to its past practice of swapping. As a threshold matter, it is important to recognize that these risks are all linked to how the Bureau will implement its 2020 DAS, rather than differential privacy per se. This section therefore examines how differential privacy and post-processing have affected the 2020 DAS and the risks that those changes pose to redistricting and voting rights.

First, there is a risk that second step in the Bureau’s 2020 DAS—post-processing to eliminate outputs that are implausible for census data—will lead to systemic biases in population counts. For instance, the Bureau plans to adjust the differential privacy outputs to ensure that there are

⁴⁷ Additionally, research by Moon Duchin indicates that distortions caused by the use of differential privacy will likely not obscure evidence of RPV. Using data from Irving, TX, Duchin showed that evidence of RPV is clear even when differential privacy is applied, assuming the privacy budget is allocated well for block-level accuracy and as long as you filter out small precincts, which according to Duchin was already required. See Moon Duchin, *Privacy in Census Data: Math Meets Policy*, TUFTS SEMINAR (July 7, 2020), <https://www.youtube.com/watch?v=oiRL5vBWq84>.

no negative integers. While this may increase the perceived legitimacy of the data that is ultimately published and allow existing software packages to process the counts, it may also introduce biases in the data. Indeed, reports analyzing the latest demonstration product published by the Bureau found a small bias that increased the population size of small-population counties at the expense of larger-population counties, typically urban areas.⁴⁸

While this bias may be small compared to the pre-existing differential undercount of communities of color—and while experts in redistricting and voting rights litigation should be able to measure and account for any bias—it is still concerning. Given that current doctrine does not typically account for error in census data in redistricting and voting rights claims, even small biases could exacerbate the under-representation of communities of color. Importantly, these biases may also disproportionately affect the accuracy of census data for small-population areas where even small alterations to the data can have a significant impact. This bias poses a particular risk to rural communities and certain small-population communities of color that live in more remote locations, such as Native American communities.

Second, if not properly accounted for, alterations in census block level population totals caused by differential privacy and post-processing could interfere with accurate redistricting and Section 2 claims. For instance, alterations in voting-age-population percentages by race and ethnicity could lead redistricters and experts to miss or incorrectly identify potential majority-minority districts, as a recent report found using the Bureau’s demonstration data.⁴⁹ And there is always the possibility that the Bureau’s design choices for the algorithm (e.g., the privacy loss budget) or post-processing will introduce so much noise into the data that it becomes unusable even with confidence intervals.

Part IV: Ongoing Lawsuits and How They Might Affect the 2020 Census

Two recent lawsuits, *Alabama v. U.S. Department of Commerce* and *Ohio v. Raimondo* are relevant to the 2020 Census data release timeline and use of differential privacy.

A. Alabama v. U.S. Department of Commerce

In *Alabama v. U.S. Department of Commerce*, the State of Alabama is challenging (1) the Census Bureau’s decision to delay the release of state redistricting data until September 30, 2021, which Alabama alleges violates the March 31 deadline for releasing the state

⁴⁸ See *Preliminary Report: Impact of Differential Privacy & the 2020 Census on Latinos, Asian Americans and Redistricting*, MALDEF & AAJC 1-2, 5-6 (April 2021), <https://www.maldef.org/wp-content/uploads/2021/04/FINAL-MALDEF-AAJC-Differential-Privacy-Preliminary-Report-4.5.2021-1.pdf>.

⁴⁹ *Id.* It is important to note that there are already sources of uncertainty in the underlying data used for majority-minority district analyses, due to the use of sampling to compile data on citizen voting age population in the American Community Survey.

“tabulations of population” used for redistricting under 13 U.S.C. § 141(c); and (2) the Bureau’s decision to use differential privacy as part of its 2020 DAS.

Alabama challenges the use of differential privacy as a violation of the Census Act; one person, one vote; and the Administrative Procedure Act. Throughout, Alabama portrays the Bureau’s decision to use differential privacy as a “decision to provide inaccurate data.”⁵⁰ Alabama also holds out swapping as the gold standard that the Census Bureau should continue to use.⁵¹

On March 26, the court granted Alabama’s request for a three-judge panel. A motion for preliminary injunction is currently pending in the case, and more than a half-dozen amicus briefs have been filed thus far.⁵² The court held an in-person hearing on May 3.⁵³

B. Ohio v. Raimondo

In *Ohio v. Raimondo*, the State of Ohio is challenging the Census Bureau’s delay in releasing the PL 94-171 redistricting data.⁵⁴ Ohio filed a complaint on February 25 challenging the Census Bureau’s decision to delay the release of state redistricting data until September 30, 2021. Ohio alleged that this delay would violate the March 31 deadline for releasing the state “tabulations of population” used for redistricting under 13 U.S.C. § 141(c).⁵⁵ Ohio is seeking injunctive relief or a mandamus compelling the Census Bureau to release the data by March 31, 2021. The district court dismissed the case on standing on March 24,⁵⁶ and Ohio filed a notice of appeal on the same day.⁵⁷

Conclusion

Without the final plan from the Census Bureau, it remains too early to say just how differential privacy will impact redistricting and voting rights. However, what we do know is that the status

⁵⁰ Complaint, *Alabama v. U.S. Dep’t of Comm.*, No. 3:21-cv-00211-RAH-KFP at 45 (M.D. Al. Mar. 10, 2021), https://www.brennancenter.org/sites/default/files/2021-03/Complaint_%202021-03-11_0.pdf.

⁵¹ *Id.* at 13-14.

⁵² As of April 29, 2021, amicus briefs had been filed on behalf of: (1) Jane Bambauer, Professor at University of Arizona, in support of Plaintiffs; (2) a collection of states in support of plaintiffs (UT, AK, AR, FL, KY, LA, ME, MS, MT, NE, NM, OH, OK, SC, TX, and WV); (3) Senate of PA Republican Caucus, Jake Corman, Kim Ward in support of plaintiffs; (4) National Redistricting Foundation in opposition; (5) State Government Leadership Foundation in support of plaintiffs; (6) Historian Margo Anderson in support of plaintiffs; (7) the Electronic Privacy Information Center in opposition; and (8) data privacy experts, including Cynthia Dwork, the co-inventor of differential privacy, in opposition.

⁵³ Order, *Alabama v. U.S. Dep’t of Comm.*, No. 3:21-cv-00211-RAH-KFP, ECF No. 93 (M.D. Al. Apr. 19, 2021).

⁵⁴ *Ohio v. Raimondo*, No. 3:21-CV-064, 2021 WL 1118049, at *8 (S.D. Ohio Mar. 24, 2021).

⁵⁵ Complaint, *Ohio v. Raimondo*, No. 3:21-CV-064 (S.D. Ohio Feb. 25, 2021), https://www.brennancenter.org/sites/default/files/2021-03/Complaint_%202021-02-25.pdf.

⁵⁶ *Ohio v. Raimondo*, No. 3:21-cv-00064-TMR, ECF No. 26 at 18 (S.D. Ohio, Mar. 24, 2021).

⁵⁷ *Ohio v. Raimondo*, No. 3:21-cv-00064-TMR, ECF No. 28 (S.D. Ohio, Mar. 24, 2021).

quo does not provide adequate privacy protections. If left unaddressed, these privacy gaps could jeopardize the confidentiality of the census, which in turn would depress participation, particularly from marginalized communities. We also know that differential privacy itself is not inherently bad or good for fair redistricting and voting rights—it is simply a definition of privacy after all. Instead, how the Bureau chooses to implement differential privacy as part of its 2020 DAS will determine the impact on redistricting and voting rights. It is thus vital for civil rights organizations to continue to analyze the Bureau’s demonstration products and provide input to mitigate the potential risks of this new system, while still capturing its benefits.