NORTH CAROLINA JOURNAL OF LAW & TECHNOLOGY
VOLUME 25, ISSUE 4: MAY 2024

A NOISY DEBATE: SHOULD THE LAW REQUIRE THE U.S. CENSUS BUREAU TO SACRIFICE ACCURACY FOR PRIVACY?

Justin Giles*

Tens of thousands of people use data provided by the U.S. Census Bureau for everything from drawing voting districts, to allocating government funds, to conducting research. But in 2018, the U.S. Census Bureau began using controversial new disclosure avoidance methods. These methods better protected the privacy of the people described by the Census Bureau's data. Unfortunately, they also decreased the data's accuracy, raising concerns such as whether voting districts drawn using 2020 census data will distort the 2024 election.

This Article argues that the Census Bureau's hand was forced. Title 13 of the U.S. Code, as construed by the Supreme Court in Baldrige v. Shapiro, suggests that the Census Bureau must prevent even low privacy risks, requiring sweeping protections like the controversial new disclosure avoidance methods. This is not a good result: The new disclosure avoidance methods may hurt the public via data inaccuracies more than they help the public by protecting privacy. Congress should rewrite the Census Bureau's privacy mandate to allow the Bureau greater flexibility to protect the accuracy of its data.

TABLE OF CONTENTS

* J.D. Candidate, University of North Carolina School of Law, 2025; Master of Public Policy Candidate, Duke University Sanford School of Public Policy, 2025. I would like to thank Rachel Couthino, Yasmin Khodaei, and Sarah Malone for their helpful feedback. I would also like to thank Professor John Quinterno for sparking my interest in this topic. Any errors are my own.

I. INTRODUCTION

Data from the U.S. Census Bureau are a public good. Just ask the voters of Alabama Congressional District 2. In 2021, only one out of seven congressional districts in Alabama—a state where 27% of the population is Black—contained a majority of Black voters.[1]

[1] Kim Chandler, *Black Voting Power Gets Boost in Alabama as New US House Districts Chosen by Federal Judges*, AP NEWS (Oct. 5, 2023, 8:41 PM),

The U.S. District Court for the Northern District of Alabama found these districts diluted the Black vote in a likely violation of the Voting Rights Act.[2] So, in late 2023, Alabama Congressional District 2 was redrawn to give Black voters a majority.[3] This entire process relied on the location data of Black Alabamans provided by the U.S. Census Bureau ("Census Bureau" or "Bureau").[4]

Or, ask Michael Southard, the Economic Development Director of the Choctaw Nation of Oklahoma.[5] Southard uses census data "for everything from making business decisions about where to develop a grocery store or a daycare center to projecting his tribe's future population growth."[6] Losing access to these data could be "devastating" for the local economy.[7]

One might even ask the U.S. Department of Health and Human Services ("DHHS"). Every year, the DHHS allocates federal funds to programs like Medicaid based on state population and income data provided by the Census Bureau.[8] Without these data, over 85 million Americans' basic medical services could go underfunded.[9]

These are only a few examples of the social value that census data produce every day. Thousands of people—from academic researchers to government and private analysts, economists,

---

https://apnews.com/article/redistricting-alabama-voting-rights-act-cd2-9ab2e940b1042f8b28c60b972575ac34 [https://perma.cc/DRT3-UJEN].

[2] Singleton v. Merrill, 582 F. Supp. 3d 924, 935 (N.D. Ala. 2022).

[3] Chandler, *supra* note 1.

[4] *See* Singleton v. Allen, No. 2:21-cv-1291-AMM, 2023 WL 6567895 at *32 (N.D. Ala. 2023).

[5] Matthew Gregg et al., *New 2020 Census Rules Make It Harder to Navigate Native American Data* 1 (Fed. Rsrv. Bank of Minneapolis, Working Paper No. 2023-05, 2023).

[6] *Id*.

[7] *Id*.

[8] ALISON MITCHELL ET AL., CONG. RSCH. SERV., R43357, MEDICAID: AN OVERVIEW 19 (2023).

[9] *November 2023 Medicaid & CHIP Enrollment Data Highlights*, MEDICAID.GOV (Mar. 18, 2024), https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/report-highlights/index.html [https://perma.cc/C2W2-KXC3].

historians, and demographers—rely on accurate census data for their work.[10]

Unfortunately, the accuracy of census data has recently decreased due to the Census Bureau's new disclosure avoidance methods. Disclosure avoidance methods are privacy protections; they prevent users of census data from being able to trace a specific data point back to a specific individual. In 2018, the Census Bureau began discussing two new and controversial disclosure avoidance methods: (1) the addition of random errors to all data from the decennial census,[11] and (2) the use of fully synthetic data to represent the American Community Survey.[12] Data users protested that these new disclosure avoidance methods would produce inferior data.[13] But the Census Bureau moved forward with its plans. All publicly released data from the 2020 census are now "differentially private," meaning that the data contain random errors.[14] The Census Bureau has also used synthetic data in some limited respects, with discussions to broaden its use ongoing.[15]

These changes triggered a wave of criticism from data users. Though the users came from many backgrounds, all shared the same underlying complaint: Less accurate data are less useful. For example, litigants in at least two court cases have alleged that electoral districts drawn with 2020 census data could distort the

---

[10] *See* Steven Ruggles et al., *Implications of Differential Privacy for Census Bureau Data and Scientific Research* 17–18 (Minn. Population Ctr., Working Paper No. 2018–6, 2018) (finding that over 70,000 papers cite census or American Community Survey data).

[11] *See* discussion *infra* Part II.

[12] *See* discussion *infra* Part II.

[13] *See* Mike Schneider, *Census Bureau's Use of 'Synthetic Data' Worries Researchers*, AP NEWS (May 27, 2021, 4:58 PM), https://apnews.com/article/census-2020-technology-data-privacy-business-be938fa5db887a0ae6858dff0be217ef [https://perma.cc/S74L-NSKA] (quoting a data researcher as saying that synthetic data "will not be suitable for research").

[14] U.S. CENSUS BUREAU, DISCLOSURE AVOIDANCE FOR THE 2020 CENSUS: AN INTRODUCTION 6 (2021).

[15] *What Are Synthetic Data?*, U.S. CENSUS BUREAU (May 27, 2021), https://www.census.gov/about/what/synthetic-data.html [https://perma.cc/VTA6-JYCA]; Schneider, *supra* note 13.

outcomes of future elections.[16] Native American nations complain that they lack access to data necessary to conduct their internal affairs due to the Census Bureau's new disclosure avoidance methods.[17] Policy analysts argue that the less accurate data could lead to "substantial misallocations" of government funds.[18]

The resulting debate between the Census Bureau and its critics involves both an issue of law and an issue of policy. First, the Census Bureau argues that it must use the new disclosure avoidance methods because it is statutorily required to protect the privacy of the people who provide its data.[19] Second, the Census Bureau argues that the new disclosure avoidance methods are good policy because they ensure the public continues to trust the Bureau with its data and they do not have too dramatic an effect on the data's accuracy.[20] The Census Bureau's critics dispute both points.[21] This Article takes the first step toward a solution, by arguing that Congress should amend the Census Bureau's privacy mandate to allow the Bureau greater

---

[16] *Cf.* Nairne v. Ardoin, No. CV 22-178-SDD-SDJ, 2024 WL 492688, at *70 (M.D. La. 2024) ("[Plaintiff's witness] hinted that the census data relied upon . . . may be unreliable due to 'differential privacy' protocols employed by the Census Bureau."); Alabama v. U.S. Dep't of Com., 546 F. Supp. 3d 1057, 1066 (M.D. Ala. 2021) ("The crux of Plaintiffs' differential privacy claims is that the Bureau's method will generate intentionally skewed and untrustworthy census data.").

[17] Mike Schneider & Morgan Lee, *Tribal Nations Face Less Accurate, More Limited 2020 Census Data Because of Privacy Methods*, AP NEWS (Sept. 9, 2023, 12:04 AM), https://apnews.com/article/native-americans-census-differential-privacy-tribes-f4fc2869a39a57485220cf2a0ebce18d [https://perma.cc/77F4-T9T9].

[18] Quentin Brummet et al., *The Effect of Differentially Private Noise Injection on Sampling Efficiency and Funding Allocations: Evidence From the 1940 Census*, HARV. DATA SCI. REV., Jun. 2022, at 1, 30.

[19] *E.g.*, Defendants' Response in Opposition to Plaintiffs' Motion for Preliminary Injunction at 2, 8, Alabama v. U.S. Dep't of Com., 546 F. Supp. 3d 1057 (M.D. Ala. 2021) ("[T]he Census Bureau can no longer rely [on pre-2018 disclosure avoidance methods] if it is to meet its obligations to protect respondent confidentiality . . . If the Census Bureau were to continue doing what it did in 2010, it would be violating . . . federal law.").

[20] *E.g.*, *id*. at 2 ("If the Census Bureau were to continue doing what it did in 2010, it would be violating . . . the confidentiality promise that it made to census respondents. And with that bond of trust broken, future census response rates would undoubtedly fall, and the accuracy of future censuses would suffer.").

[21] *See* Ruggles et al., *supra* note 10, at 17–18.

flexibility to weigh privacy against accuracy, and to make policy based on the result.

This Article proceeds in five parts. Part II provides background on the modern privacy threats that prompted the Census Bureau to seek new disclosure avoidance methods. Part III argues that the Supreme Court's treatment of the Census Bureau's privacy mandate has left the Bureau with little choice but to adopt the new disclosure avoidance methods. Nonetheless, Part IV suggests that the new disclosure avoidance methods may be bad policy. Part V examines how congressional action could right the ship.

II. BACKGROUND: MODERN PRIVACY THREATS LEAD TO NEW DISCLOSURE AVOIDANCE METHODS

Every ten years, the Census Bureau attempts to enumerate the entire population of the United States ("U.S."), as required by the U.S. Constitution.[22] This massive collection of data is called the decennial census.[23] It gathers the basic demographic information of the people who respond to the Census Bureau's surveys, including their age, sex, race, and ethnicity.[24]

Most individual-level census data—for example, the information that a specific person is White, aged 47, and female—are never released. However, the Census Bureau releases statistics describing collections of data for geographies as small as a single census block.[25] Census blocks vary in size based on population density.[26] Urban census blocks may be as small as 30,000 square

---

[22] *See* U.S. CONST. art. I, § 2, cl. 3 ("The actual Enumeration [of the U.S. population] shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct.").

[23] *About the Decennial Census of Population and Housing*, U.S. CENSUS BUREAU (Dec. 16, 2021), https://www.census.gov/programs-surveys/decennial-census/about.html [https://perma.cc/48C2-XECF].

[24] *Why We Conduct the Decennial Census of Population and Housing*, U.S. CENSUS BUREAU (Nov. 23, 2021), https://www.census.gov/programs-surveys/decennial-census/about/why.html [https://perma.cc/GQ5W-KBLW].

[25] U.S. DEP'T OF COM., GEOGRAPHIC AREAS REFERENCE MANUAL, ch. 11, at 11-1 (1994).

[26] *Id*. at 19–20.

feet,[27] while rural census blocks may stretch over 250 square miles.[28]

In addition to the decennial census, the Census Bureau collects information from samples of people throughout the United States on a rolling basis.[29] These data make up the American Community Survey ("ACS").[30] The ACS supplements the decennial census with additional information, like education and employment data, as well as more up-to-date information on age, sex, race, and ethnicity.[31]

For nearly two centuries, the Census Bureau has used certain disclosure avoidance methods to protect the public's privacy.[32] These methods have always reduced the accuracy of the data the Bureau releases.[33] The difference between the new methods and the old ones is that the new methods are far more sweeping: They affect the entire dataset, not just particularly sensitive individuals or types of data.[34] The Census Bureau responds that these new methods are necessary in the face of new threats brought on by modern

---

[27] *Id*. at 10.

[28] *Id*. at 20.

[29] *See The Importance of the American Community Survey and the Decennial Census*, U.S. CENSUS BUREAU, https://www.census.gov/programs-surveys/acs/about/acs-and-census.html [https://perma.cc/99ET-L3S6] (last updated Mar. 13, 2024).

[30] *See id*.

[31] *See id.*

[32] *See* Steven Ruggles & Diana L. Magnuson, *"It's None of Their Damn Business": Privacy and Disclosure Control in the U.S. Census, 1790–2020*, 49 POPULATION & DEV. REV. 651, 651 (2023).

[33] *See* discussion of previous disclosure avoidance methods, *infra* Part II.B.

[34] Kriston Capps, *Data Scientists Square Off over Trust and Privacy in 2020 Census*, BLOOMBERG (Aug. 12, 2021, 5:44 PM), https://www.bloomberg.com/news/articles/2021-08-12/data-scientists-ask-can-we-trust-the-2020-census?sref=QFCZ3YPm. [https://perma.cc/43JL-H3AX] (statement of Professor Norm Matloff) ("With the census approach to differential privacy, every piece of data is going to be synthetic, as opposed to data swapping, where only some small fraction of the data is swapped."); Schneider, *supra* note 13 (noting data researcher, Professor Steven Ruggles, who worries that new discoveries in data will be missed when using synthetic data because the models can only capture what is already known, and synthetic data can amplify outliers).

technology, specifically the threats of database reconstruction and reidentification.[35]

A. Reconstruction and Reidentification

Statistics always reveal information about the data they describe. For example, if a dataset of three incomes has an average of $50,000, a minimum of $25,000, and a maximum of $75,000, a person can deduce that the dataset is composed of incomes of $25,000, $50,000, and $75,000.

For a long time, a lack of computing power prevented this sort of deduction from being performed on large datasets like the decennial census. But in 2003, statisticians proved that "[t]oo many statistics published too accurately from a confidential database exposes the entire database with near certainty," a problem that became known as "database reconstruction."[36]

Reconstruction is related to—but not the same as—reidentification. Reconstruction refers to the deduction of microdata, the lowest-level data, from which all descriptive statistics are derived.[37] Reidentification refers to the attachment of that microdata to personally identifiable information ("PII"), such as a name, address, or social security number.[38] Reconstructed census data might look something like this:

---

[35] *See* John M. Abowd, Chief Scientist & Assoc. Dir. for Rsch. & Methodology, U.S. Census Bureau, Presentation at Joint Statistical Meetings in Vancouver, B.C., Canada: Staring-Down the Database Reconstruction Theorem 6 (July 30, 2018) (presentation available online at census.gov) (citation omitted).

[36] *Id*. (citation omitted).

[37] Simson Garfinkel, John M. Abowd, and Christian Martindale, *Understanding Database Reconstruction Attacks on Public Data*, 62 COMMC'N ACM 46, 46 (2019). For an example of database reconstruction, see id. at 48-51.

[38] *See* Kobbi Nissim et al., Differential Privacy: A Primer for a Non-technical Audience 3 (Mar. 3, 2017) (unpublished manuscript) (on file with Privacy Tools for Sharing Research Data project at Harvard University).

Census block ID[39]	Age	Sex	Race	Ethnicity
1000	66	Male	White	Hispanic
1000	41	Female	White	Non-Hispanic
1000	27	Female	Black	Non-Hispanic
1000	19	Male	White	Hispanic
1000	34	Female	Asian	Non-Hispanic

Reconstruction allows for reidentification if an attacker has access to another dataset that both (1) contains PII and (2) shares unique variables, or unique combinations of variables, with the reconstructed data. For example, imagine an attacker reconstructed the above data, which represents every person living within census block 1000. Now, imagine the same attacker has access to the following dataset, which includes PII (names and addresses).

Name	Address	Age
Carl Gauss	123 Normal Road	66
Simone Poisson	456 Discrete Ln	27

If the attacker knows that (1) the addresses of Gauss and Poisson are within Census Block 1000, and (2) that the reconstructed census data contains all residents of Census Block 1000, then the attacker can join the datasets and learn that Gauss is a Hispanic White male and Poisson is a non-Hispanic Black female. This is reidentification: The attacker has connected census data to PII.

Reconstruction risk can be assessed mathematically because it depends only on the characteristics of a dataset and the released statistics describing it.[40] Reidentification risk, on the other hand, is less quantifiable because it depends on the availability of other

[39] The Census Bureau never releases addresses. A census block is therefore the smallest geography available in census data. U.S. DEP'T OF COM., *supra* note 25, at 11-1, 11-10.

[40] For example, for differentially private data, the parameter ϵ (epsilon) is a measure of how confident a reconstruction attacker can be about their reconstruction attempt. Nissim et al., *supra* note 38, at 9–10.

datasets to an attacker.[41] Reidentification is generally understood as more dangerous than reconstruction because it attaches information to a specific person via their PII. In contrast, reconstructed census data only tell an attacker that a block contains people of a certain age, sex, race, and ethnicity—not who those people are.

Recall that the risk of reconstruction depends on the amount of available data. The more statistics are published about a dataset, the easier it is to infer the underlying data.[42] This caught the attention of the Census Bureau, which releases over 150 billion[43] statistics created from the decennial census. These data products include summary statistics, tabulations, and maps describing every part of the country in detail.[44]

Worried that its breadth of releases made its data vulnerable, the Census Bureau began conducting internal experiments to see if it could reconstruct data from the 2010 decennial census.[45] In 2018, the Bureau announced that it had succeeded.[46] The Bureau had reconstructed microdata representing 308,745,538 people, deducing their census block ID, age, sex, race, and ethnicity.[47] It was perfectly

---

[41] *Cf.* Gregory E. Simon et al., *Assessing and Minimizing Re-identification Risk in Research Data Derived from Health Care Records*, 7 eGEMS 1, 8 (2019) (observing that, in the context of health data, "the likelihood of a successful re-identification attack . . . depends on the motivation of and resources available to a potential adversary [which] may not be known at the time of data release").

[42] Abowd, *supra* note 35 ("Too many statistics published too accurately from a confidential database exposes the entire database with near certainty.").

[43] U.S. CENSUS BUREAU, D-FS-GP-EN-0509, COMPARING DIFFERENTIAL PRIVACY WITH OLDER DISCLOSURE AVOIDANCE METHODS 2 (2021).

[44] To get an idea of the availability of census data, go to https://data.census.gov/ and search a location or statistic of interest. *See Explore Census Data*, U.S. CENSUS BUREAU, https://data.census.gov/ [https://perma.cc/NFV6-WTAC] (last visited Apr. 8, 2024).

[45] Mark Hansen, *To Reduce Privacy Risks, the Census Plans to Report Less Accurate Data*, N.Y. TIMES: THE UPSHOT (Dec. 5, 2018), https://www.nytimes.com/2018/12/05/upshot/to-reduce-privacy-risks-the-census-plans-to-report-less-accurate-data.html [https://perma.cc/VS88-6JRN].

[46] Abowd, *supra* note 35, at 11.

[47] John M. Abowd, *Tweetorial: Reconstruction-Abetted Re-Identification Attacks and Other Traditional Vulnerabilities*, BLOGS.CORNELL.EDU: ABOWD, https://blogs.cornell.edu/abowd/special-materials/245-2/ [https://perma.cc/A3LH-9SVA] (last visited Apr. 8, 2024).

accurate for 46% of the population.[48] The Census Bureau then attempted to reidentify respondents by joining the reconstructed data to commercially available datasets containing PII.[49] This allowed the Bureau to correctly reidentify 38% of the population.[50]

There are reasons to believe this experiment was not as successful as it might seem.[51] However, the Census Bureau was alarmed and began seeking a means of preventing reconstruction for the upcoming 2020 census.[52] This led to the adoption of new disclosure avoidance methods, which in turn lead to the accuracy issues now raising concerns for census data users.[53]

B. The Census Bureau's New Disclosure Avoidance Methods

The Census Bureau has historically relied on a variety of disclosure avoidance methods to keep respondents' identities private.[54] For example, respondents' names and addresses are never

---

[48] *Id*. (describing Title 13 as prohibiting "exact attribute disclosure"). For a less technical overview, see Hansen, *supra* note 45.

[49] Michael Hawes, Senior Advisor for Data Access and Privacy Research and Methodology Directorate, U.S. Census Bureau, The Census Bureau's Simulated Reconstruction-Abetted Re-Identification Attack on the 2010 Census 12 (May 7, 2021) (transcript available at U.S. Census Bureau website).

[50] Abowd, *supra* note 47.

[51] *See* discussion *infra* Part IV.A.

[52] *See* Defendants' Response In Opposition To Plaintiffs' Motion For Preliminary Injunction at 9–10, Alabama v. U.S. Dep't of Com., 546 F. Supp. 3d 1057 (M.D. Ala. 2021).

[53] For the most recent—and most sophisticated—Census Bureau evaluation of reconstruction and reidentification risk, see John M. Abowd et al., *The 2010 Census Confidentiality Protections Failed, Here's How and Why* (Nat'l Bureau of Econ. Rsch., Working Paper No. 31995, 2023).

[54] For a history of census privacy concerns and protections, see generally Ruggles & Magnuson, *supra* note 32 (advocating greater permissiveness around release of census data); Margo Anderson, *The Missing History of the Disclosure of Individual Responses in the American Census: What Happened and Why It Matters Now*, 6 GEO. L. TECH. REV. 408 (2022) (detailing the reasons for increases in privacy protections over time); LAURA MCKENNA, U.S. CENSUS BUREAU, DISCLOSURE AVOIDANCE TECHNIQUES USED FOR THE 1970 THROUGH 2010 DECENNIAL CENSUSES OF POPULATION AND HOUSING (2018) (explaining the specific disclosure avoidance methods used by the Census Bureau in recent history).

released.[55] From 1970 to 2010, the Census Bureau also used more advanced methods like swapping (in which the data of two households are interchanged with one another), top/bottom-coding (e.g., income data might include values like "over $500,000" instead of "$612,000"), and category collapsing (in which a large number of categories are consolidated into a few general categories).[56] These methods focused on data that were especially easy to infer, like a lone house within a census block.[57]

Since the 1960s, the Census Bureau has also released public use microdata samples, which contained the individual-level data (e.g., the age and race of specific people) for a small fraction of the U.S. population.[58] Because this microdata did not contain identifying information (names and addresses), the Census Bureau believed that "making records available in this form [did] not violate the provision for confidentiality in the law under which the census was conducted."[59]

The Census Bureau began to reevaluate these methods after its 2018 reconstruction and reidentification experiment.[60] Two new

---

[55] *Title 13, U.S. Code*, U.S. CENSUS BUREAU, https://www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html [https://perma.cc/2SCG-YM4S] (last revised Dec.14, 2023) ("Private information is never published [by the Census Bureau]. It is against the law to disclose or publish any private information that identifies an individual or business such [sic], including names, addresses (including GPS coordinates), Social Security Numbers, and telephone numbers.").

[56] *See* MCKENNA, *supra* note 54, at 3–10.

[57] *See* John M. Abowd & Michael B. Hawes, *Confidentiality Protection in the 2020 US Census of Population and Housing*, 10 ANN. REV. STAT. APPLICATIONS 119, 122 (2023) ("[R]ecords for more vulnerable households were selected with greater probability.").

[58] *See* Steven Ruggles*, Census Data Processing, Part 2*, U.S. CENSUS BUREAU (Aug. 2, 2012) https://www.census.gov/newsroom/blogs/research-matters/2012/08/steven-ruggles-census-data-processing-part-2.html [https://perma.cc/FCT5-NFJ9]; *e.g.,* Press Release, U.S. Census Bureau, Census Bureau Announces Schedule Updates for 2020 Census Data Products (Mar. 27, 2023) (on file with author).

[59] U.S. CENSUS BUREAU, U.S. CENSUSES OF POPULATION AND HOUSING, 1960: 1/1,000,1/10,000, TWO NATIONAL SAMPLES OF THE POPULATION OF THE UNITED STATES 2 (1960).

[60] *See* Abowd, *supra* note 35.

disclosure avoidance methods emerged. First, for the decennial census, the Census Bureau decided that released data would be "differential[ly] priva[te]."[61] To accomplish this, the data would be randomly changed through a method known as noise injection. The changes would be drawn from a distribution centered at zero, which means that on average, the data would remain unchanged, but any particular datapoint may have a different value than it originally did.[62] Second, for the ACS, the Census Bureau began discussing the possibility of releasing only synthetic data—that is, data that are generated by an algorithm to resemble the real data collected.[63]

Both methods operate on the entire dataset, not just particularly sensitive data. Noise injection can result in any datapoint being altered, and synthetic data do not allow any original data to be released. Users of census data therefore worry that these new methods interfere too much with data accuracy.[64] For example, how can a researcher using census data be confident that their results reflect reality as opposed to some artifact of the alterations?

While the Census Bureau maintains that the new methods strike a good balance between data privacy and accuracy, the Bureau has also been transparent that legal concerns influenced its decision.[65] Policy critiques alone are therefore not enough to make the Census Bureau change course. Without Congressional action, the Census Bureau could not be confident a court would find it to be in compliance with its privacy mandate if some of its data were successfully reconstructed.

III. THE LEGAL QUESTION: ARE THE NEW DISCLOSURE AVOIDANCE METHODS REQUIRED BY LAW?

The Census Bureau's privacy mandate includes Title 13 of the U.S. Code ("Title 13"),[66] the Confidential Information Protection

---

[61] *Id*. at 11.

[62] *See* Nissim et al., *supra* note 38.

[63] Schneider, *supra* note 13.

[64] *E.g.*, *id*. (quoting a data researcher as worrying that "new discoveries in data will be missed [when using synthetic data] since the models only capture what is already known" and that "synthetic data can amplify [] outlier[s].").

[65] Abowd et al., *supra* note 53, at 7.

[66] 13 U.S.C. §§ 1–402.

and Statistical Efficiency Act of 2002 ("CIPSEA"),[67] and the Privacy Act of 1974 ("Privacy Act").[68] Title 13 and CIPSEA are both considered below. The Privacy Act's language is duplicative of Title 13's requirements, so it is not discussed here.[69]

The key issue is whether Title 13 and CIPSEA require the Census Bureau to protect against reconstruction, not just reidentification. An interpretation of the law focused solely on reidentification might allow the Census Bureau more flexibility to value data accuracy in addition to privacy. This is because the data published by the Census Bureau directly enables reconstruction, but not reidentification. Reidentification requires an additional step: An attacker must find an external dataset containing PII and connect it to the census data.

A. Title 13 of the U.S. Code

Title 13 bars the Census Bureau from "mak[ing] any publication whereby the data furnished by any particular establishment or individual under this title can be identified."[70] However, it allows the Census Bureau to "furnish copies of tabulations and other statistical materials which do not disclose the information reported by, or on behalf of, any particular respondent."[71] This language is ambiguous, because it does not define "identif[ication]" or "disclos[ure]," leaving courts without guidance on whether Title 13 prohibits the Census Bureau from releasing data that can be reconstructed.[72]

---

[67] Confidential Information Protection and Statistical Efficiency Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (codified at 44 U.S.C. § 101).

[68] Privacy Act of 1974, Pub. L. No. 93-579, 88 Stat. 1896 (codified at 5 U.S.C. § 552a).

[69] The Privacy Act prohibits the Census Bureau, from "disclos[ing] any record [including census data] . . . to any person, or to another agency, except pursuant to a written request by, or with the prior written consent of, the individual to whom the record pertains." 5 U.S.C. § 552a(b). Title 13, Section 8 also prohibits "disclos[ure]." 13 U.S.C. § 8(b).

[70] 13 U.S.C. § 9(a)(2).

[71] *Id.* § 8(b).

[72] Unfortunately, the Code of Federal Regulations provides no additional guidance on Title 13's privacy mandate. *See* 15 C.F.R. §§ 30.1–101.1 (regulating the Census Bureau without describing the privacy protections it must provide). A

This ambiguity was arguably resolved in 1982 by the U.S. Supreme Court in *Baldrige v. Shapiro*.[73] In *Baldrige*, the Court found that Title 13 does not prevent data from being released only where an "individual respondent can be identified."[74] Rather, Title 13 requires *all* "raw data reported by or on behalf of individuals"—even completely anonymous data—"to be held confidential and not available for disclosure."[75] As a result, the Court held that Title 13 barred the Census Bureau from providing a list of vacant addresses to municipal governments even though that list would not reveal any person's identity.[76]

Baldrige's interpretation of Title 13 all but confirms that the statute prohibits the Census Bureau from allowing database reconstruction, not just reidentification. *Baldrige* expressly held that "the data itself"[77] cannot be provided to a third party under Title 13, and the Census Bureau's reconstruction experiment confirmed that "the data itself" can be (partially) inferred from the data products the Bureau releases.[78] It would therefore seem that the Census Bureau needs to use new disclosure avoidance methods—like noise injection or synthetic data—to prevent its releases from allowing members of the public to deduce the underlying microdata.

That said, if courts were to reach this conclusion, it would still be an extension—not merely an application—of *Baldrige*. Unlike the direct provision of data contemplated in *Baldrige*, reconstruction

---

scientific advisor to the Census Bureau looked at Title 13's language and found it "unclear . . . how . . . Title 13 actually guides or constrains specific decisions about disclosure avoidance." GORDON LONG, MITRE CORP., CONSISTENCY OF DATA PRODUCTS AND FORMAL PRIVACY METHODS FOR THE 2020 CENSUS 114 (2022). He argued there is "an urgent need for clarification on the interpretation of Title 13 confidentiality requirements as they apply to census data products, and perhaps even for statutory changes by Congress." *Id*. at 111.

[73] Baldrige v. Shapiro, 455 U.S. 345 (1982).

[74] *Id*. at 355.

[75] *Id*.

[76] The Court found that the list of vacant address could not be provided in response to a Freedom of Information Act request, and that it also could not be provided during discovery. *Id*. at 353–62.

[77] *Id*. at 356.

[78] Abowd, *supra* note 47.

involves an inferential step[79] possible only for an attacker with significant computing power and knowledge of statistics.

However, most courts would probably brush this distinction aside. Despite the inferential step, the risk of identifying specific individuals is probably higher for general reconstructed census data than it was for the list of vacant properties in *Baldrige*. Most judges would probably balk at the idea that Title 13 prohibits the provision of vacant property lists like the one in *Baldridge,* which do not represent any specific people, but allows the inference of accurate microdata, which do represent specific people.

Another argument exists to avoid *Baldrige*'s demanding standard: *Baldrige*'s interpretative reasoning was weak and should therefore be confined to its facts. Variations of this argument might take issue with (1) *Baldrige*'s textual analysis of Title 13, or (2) *Baldrige*'s assessment of Title 13's legislative history.[80] While these arguments are unlikely to be persuasive, they are worth exploring because critics have used them to try to push the Bureau away from its current legal position.[81]

1. Why Criticisms of Baldrige's Textual Analysis Fail

Baldrige's textual analysis rested almost entirely on the Title 13, Section 8 requirement that the Census Bureau publish "statistical materials which do not disclose the information reported by, or on behalf of, any particular respondent."[82] According to the Court, this section's "clear language"[83] demonstrated that Title 13 protected "the data itself," not just the identities of census respondents.[84]

---

[79] Recall that reconstruction is the process of inferring data from, e.g., summary statistics. *See* discussion *supra* Part II.A.

[80] *See Baldrige*, 455 U.S. at 353–62.

[81] *See, e.g.*, Steven Ruggles et al., *Differential Privacy and Census Data: Implications for Social and Economic Research*, 109 AEA PAPERS & PROCS. 403, 404 (2019) (arguing for a definition of disclosure focused on reidentification risk); Ruggles et al., *supra* note 10 (arguing the Census Bureau's history of microdata releases is inconsistent with a legal standard protecting the data itself).

[82] 13 U.S.C. § 8(b); *Baldrige*, 455 U.S. at 356–59.

[83] *Baldrige*, 455 U.S. at 359.

[84] The Court also considered it significant that Section 9 talked about "information" and "data" as opposed to respondents' identities. *See id.* at 356.

Implicit in this reasoning was the idea that "disclose" must be defined as "release."[85] This definition is intuitive, but that does not make it correct. The Office of Budget and Management ("OMB") has provided guidance on statutory concepts of data privacy and identification (though not in the context of Title 13 specifically).[86] OMB's guidance cites a paper that defines "disclosure" as "public identification." [87] Under this definition, *Baldrige* might have come out differently, because releasing a list of vacant properties would not publicly identify any census respondents.

However, courts are unlikely to be persuaded that *Baldrige* wrongly defined "disclose." Title 13 predates OMB's specialized definitions of "disclosure."[88] Even today, "disclose" generally refers to making information known, not to tying that information back to a specific individual's identity.[89] Given especially that textualist analysis is currently in vogue,[90] today's courts would likely agree with *Baldrige* that language prohibiting the Census Bureau from "disclos[ing] the information reported by, or on behalf of, any particular respondent"[91] means exactly what it sounds like: The Census Bureau cannot make accurate census microdata publicly available, even if they are disconnected from individuals' identities.

---

[85] *See* 13 U.S.C. § 8(b); *Baldrige*, 455 U.S. at 356.

[86] Notice of Decision, Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), 72 Fed. Reg. 33362, 33363 (June 15, 2007).

[87] *Id*.; *Report on Statistical Disclosure Limitation Methodology* (Fed. Comm. on Stat. Methodology, Working Paper No. 22, 2005).

[88] Act of Oct. 17, 1976, Pub. L. No. 94–521, 90 Stat 2459 (adding the prohibition on "disclosing" data to Title 13, Section 8 in 1976); *Report on Statistical Disclosure Limitation Methodology*, *supra* note 87, at 2 (defining "disclose").

[89] *E.g.*, *Disclose*, MERRIAM-WEBSTER.COM, https://www.merriam-webster.com/dictionary/disclose [https://perma.cc/TK6D-UVR5] (last visited Mar. 18, 2024) (defining "disclose" as "to make known or public" or "to expose to view").

[90] *See generally* Kevin Tobia, *We're Not All Textualists Now*, 78 N.Y.U. ANN. SURV. AM. L. 243 (2023) (describing, among other things, the "significant sense in which modern American legal interpretive culture is textualist").

[91] 13 U.S.C. § 8(b).

2. Why Criticisms of Baldrige's Analysis of Title 13's Legislative History Fail

To better understand Title 13, *Baldrige* considered past amendments to it. These included Congress's 1929 removal of the Census Director's discretion to release information, Congress's 1960 and 1970 rejections of proposals to allow local officials limited access to census microdata, and Congress's ratcheting up of privacy protections over time.[92] The Court understood these developments as demonstrating a trend of Congress desiring more protection for census data, not less.[93]

However, *Baldrige* ignored the fact that, since 1960, the Census Bureau has released public use microdata samples ("PUMS")[94] containing information like the age and race of specific people.[95] Because the microdata released in PUMS were anonymized and represented only a small fraction of census survey respondents, the Census Bureau previously maintained that "making records available in this form does not violate the provision for confidentiality in the law under which the census was conducted."[96] If Congress intended Title 13 to prevent the release of "the data itself,"[97] it is difficult to imagine why Congress never acted to prevent these releases. The conclusion might be that *Baldrige* was misguided: Releasing anonymous microdata does not violate Title 13.

But this reasoning is also unlikely to persuade courts to depart from *Baldrige*. Although courts do sometimes draw meaning from Congressional inaction, they also recognize that this interpretive method is suspect.[98] Generations of legal scholars have debated the

---

[92] Baldrige v. Shapiro, 455 U.S. 345, 356–58 (1982).

[93] *See id.*

[94] *See Ruggles*, *supra* note 58; discussion *supra* Part II.B.

[95] *See* discussion *supra* Part II.A; *See* Press Release, United States Census Bureau, Census Bureau Announces Schedule Updates for 2020 Census Data Products (Mar. 27, 2023) (on file with author).

[96] U.S. CENSUS BUREAU, *supra* note 59.

[97] *Baldrige*, 455 U.S. at 356.

[98] *See* Brown & Williamson Tobacco Corp. v. Food & Drug Admin., 153 F.3d 155, 170 (4th Cir. 1998) (acknowledging a "general reluctance of courts to rely

merits of reading meaning into Congressional inaction without developing any consensus over a set of interpretative rules.[99] Moreover, courts usually interpret Congressional inaction to resolve a question on which U.S. Supreme Court has not spoken, but here, the Court spoke in *Baldrige*. Therefore, although the PUMS releases of "the data itself" are in tension with *Baldrige'*s holding, they remain unlikely to change the decision of any courts.

In sum, courts presented with the difficult task of interpreting Title 13's stance on database reconstruction would likely turn to *Baldrige* for an easy rule. Under *Baldrige*, the Census Bureau could be found in violation of its privacy mandate were it to allow even a partial reconstruction of its data, because *Baldrige* reads Title 13 as protecting "the data itself."[100] Therefore, the Census Bureau is compelled by *Baldrige* to err on the side of caution by protecting its data not just from reidentification, but from reconstruction as well.

B. The Confidential Information Protection and Statistical Efficiency Act

Baldrige offers the clearest warning that the Census Bureau must prevent database reconstruction or else risk being found in violation of its privacy mandate.[101] However, CIPSEA is another important piece of the Census Bureau's privacy mandate.[102] CIPSEA weighs

---

on congressional inaction as a basis for statutory interpretation" despite deciding to do so in the case at bar).

[99] *E.g.*, Paul Stancil, *Congressional Silence and the Statutory Interpretation Game*, 54 WM. & MARY L. REV. 1251 (2013); William N. Eskridge Jr., *Interpreting Legislative Inaction*, 87 MICH. L. REV. 67 (1988); Lawrence H. Tribe, *Toward a Syntax of the Unsaid: Construing the Sounds of Congressional and Constitutional Silence*, 57 IND. L.J. 515 (1982).

[100] *Baldrige*, 455 U.S. at 356.

[101] *See* discussion s*upra* Part III.A.

[102] The Census Bureau claims that "CIPSEA does not apply to the Census Bureau's procedures for guarding the confidentiality of responses to Census surveys." *Disclosure Avoidance: Latest Frequently Asked Questions*, U.S. CENSUS BUREAU, https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates/2020-das-faqs.html [https://perma.cc/52CG-M2PW] (last updated Jan. 24, 2024). But this is probably an overstatement; it would be more accurate to say that CIPSEA cannot lessen Title 13's privacy protections, but it can increase, supplement, or inform them. *See* 44 U.S.C. § 3564(c) ("[CIPSEA] shall

less definitively toward an outright prohibition on allowing database reconstruction. Still, CIPSEA makes such a finding more likely than not.

CIPSEA states that "[d]ata or information acquired by an agency . . . shall not be disclosed by an agency in identifiable form."[103] Unlike Title 13, CIPSEA expressly defines "identifiable form" as "any representation of information that permits the identity of the respondent to whom the information applies to be reasonably inferred by either direct or indirect means."[104] OMB's guidance on CIPSEA further clarifies that "[i]ndirect identification refers to using information in conjunction with other data elements to reasonably infer the identity of a respondent. For example, data elements such as a combination of gender, race, date of birth, geographic indicators, or other descriptors may be used to identify an individual respondent."[105]

CIPSEA's definition of identifiability reads as if it were written with database reconstruction in mind. "[P]ermit[ing] the identity of [a] respondent . . . to be reasonably inferred"[106] is arguably what database reconstruction does by providing a detailed but anonymous dataset that can be linked to a different dataset containing PII. OMB's description of reidentification facilitated by "a combination of gender, race, date of birth, geographic indicators, or other descriptors"[107] also recalls the sort of information that reconstructed census data would provide to an attacker: anonymized but joined individual characteristics that are unique enough to allow

---

not be construed as authorizing the disclosure for nonstatistical purposes of demographic data or information collected by the Bureau of the Census pursuant to section 9 of title 13."); *see also* KELLY PERCIVAL, FEDERAL LAWS THAT PROTECT CENSUS CONFIDENTIALITY 5 (2019) (describing CIPSEA as controlling the protection of census data).

[103] 44 U.S.C. § 3572(c)(1).

[104] § 3561(7).

[105] Notice of Decision, Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), 72 Fed. Reg. 33362, 33363 (June 15, 2007)

[106] *See* 44 U.S.C. § 3561(7).

[107] Notice of Decision, Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), 72 Fed. Reg. at 33363.

reidentification. Moreover, CIPSEA packs a double punch; it is binding on the Census Bureau, and courts will likely find its definition of identifiability persuasive in interpreting Title 13's prohibition on releasing data whereby a respondent "can be identified."[108]

Still, CIPSEA stops short of *Baldrige*'s standard protecting the "data itself."[109] CIPSEA only protects data that allows the reasonable inference of an individual's identity.[110] Whether CIPSEA prevents the Census Bureau from allowing database reconstruction therefore seems linked to an empirical question: If such reconstruction is possible, how high is the risk of reidentification? If the risk of reidentification is high, then CIPSEA might prohibit data releases that allow for reconstruction. Alternatively, in a world where the risk of reidentification is low, CIPSEA would not require protections against reconstruction like the Census Bureau's new disclosure avoidance methods. Which world do we live in?

Part IV explores several reasons to believe reidentification risk remains low. Still, the Census Bureau's successful reidentification of 38% of census respondents[111] would worry courts. Given both the striking nature of the Census Bureau's 2018 reidentification experiment and the language of CIPSEA recalling the threat of reconstruction, CIPSEA would probably weigh toward courts finding that the Census Bureau must prevent database reconstruction.

C. The Consequences for Violating Title 13's Privacy Mandate

A suit against the Census Bureau for a violation of its privacy mandate would involve administrative law issues beyond the scope of this Article. Suffice it to say that, in a world where the deference[112] courts have traditionally given agencies is rapidly

---

[108] 13 U.S.C. § 9(a)(2).

[109] *See* Baldrige v. Shapiro, 455 U.S. 345, 356 (1982).

[110] *See* 44 U.S.C. § 3561(7).

[111] *See* Abowd, *supra* note 47.

[112] *See* Chevron v. Nat. Res. Def. Council, 468 U.S. 837 (1984) (requiring that courts defer to permissible agency interpretations of their governing statutes).

diminishing,[113] the Census Bureau could not be confident it would emerge from such a suit unscathed.

Courts could go far beyond enjoining the Census Bureau's data releases. Title 13 requires that any Census Bureau staff member who violates its privacy mandate be "fined not more than $5,000 or imprisoned not more than [five] years, or both."[114] Given the severity of these penalties, it is understandable that the Census Bureau would take a risk averse approach to privacy protection.

This Article ultimately argues that this legal regime is too strict. To make the best policy, the Census Bureau must be able to weigh marginal increases in accuracy against decreases in privacy. Such flexibility does not exist under the current, *Baldrige*-driven prohibition on releasing data that is even partially reconstructable.[115] But to understand why this issue is pressing, one must look at where *Baldrige* has taken the Census Bureau. There are good reasons to worry that the new disclosure avoidance methods may not be worth their cost.

---

[113] The U.S. Supreme Court has "ignored *Chevron* far more often than it applied *Chevron* over the past 40 years." Richard Pierce, *Court's New Chevron Analysis Likely to Follow One of These Paths*, BLOOMBERG L. (Feb. 7, 2024, 4:30 AM), https://news.bloomberglaw.com/us-law-week/courts-new-chevron-analysis-likely-to-follow-one-of-these-paths [https://perma.cc/7LNP-W9ZS]. The Court looks poised to overturn Chevron in Loper Bright Enterprises, Inc. v. Raimondo, 45 F.4th 359 (D.C. Cir. 2022). *See* Looper Bright v. Raimondo, SCOTUSBLOG.COM, https://www.scotusblog.com/case-files/cases/loper-bright-enterprises-v-raimondo/ [https://perma.cc/R62R-FJVF] (last visited Mar. 26, 2024). And even if *Chevron* remains the law, it has become subject to multiple exceptions. *E.g.*, W. Va. v. EPA, 597 U.S. 697 (2022) (finding that an EPA interpretation of its governing statute to allow it to pass regulations shifting energy production away from coal involved a "major question," and therefore did not qualify for Chevron deference).

[114] 13 U.S.C. § 214.

[115] To be clear, it was not this Part's position that the Census Bureau's privacy mandate, as it currently stands, *must* or even *should* prohibit data releases that enable database reconstruction. Rather, the argument was one of legal realism: courts likely would find the Census Bureau violated its privacy mandate were the Bureau to release data that could be reconstructed.

IV. THE POLICY QUESTION: ARE THE NEW DISCLOSURE AVOIDANCE METHODS WORTH THEIR COST?

Judging the policy merits of the new disclosure avoidance methods requires weighing the benefits of the methods (the privacy they provide) against their costs (the inaccuracies they introduce into census data).[116] This is a difficult task. But nothing can be said on the policy front without it.

While a full cost-benefit analysis is beyond the scope of this Article, it can take the first step in that direction by making the costs and benefits less abstract. The benefits of the new disclosure avoidance methods include how well they reduce the risk of reidentification. The costs of the new disclosure avoidance methods include how the inaccuracy they create affects various uses of census data. This Part fleshes out what these costs and benefits might look like in further detail. Although the evidence does not lend itself to any definite conclusions, it suggests a real possibility that the costs outweigh the benefits—and that such costs are pressing, involving negative effects on elections, Native American self-governance, public benefit allocation, small-scale data users, and the quality of scientific research.

Of course, this does not change the previous Part's conclusion that the law requires the Census Bureau to use the new disclosure avoidance methods. Consequently, Congress must amend the Census Bureau's privacy mandate.[117] The Census Bureau has the knowledge and resources to perform a far better cost-benefit analysis than this Article can. But first, the Bureau's privacy mandate must give it the flexibility necessary to believe it could make policy in line with the results of such an analysis.

---

[116] *E.g.*, Ori Heffetz, *What Will It Take to Get to Acceptable Privacy-Accuracy Combinations?*, HARV. DATA SCI. REV., Jun. 2022, at 1–9.

[117] For a somewhat similar argument with less legal and technical details, see Matthew Yglesias, *Privacy Concerns are Breaking the Census,* SLOW BORING (Jan. 26, 2022), https://www.slowboring.com/p/census-data [https://perma.cc/CWS3-YEGT].

A. How High is the Risk of Reidentification?

The Census Bureau's 2018 experiment managed to accurately reconstruct 46% of census respondents' data.[118] The reconstructed information included the age, sex, race, and ethnicity of people within specific census blocks.[119] The question then becomes whether these results demonstrate such an unacceptable risk of reidentification that the new disclosure avoidance methods were necessary.

There are two reasons to believe that the 2018 results did not represent an unacceptable risk: (1) Reconstruction is unlikely to occur because it involves too much effort for too little reward; and (2) even if a reidentification attack were successful, the data it would provide to an attacker would not be sensitive enough to harm the identified individuals. Each of these arguments are discussed in turn below.

1. Too Much Effort for Too Little Reward

A full-scale reconstruction is not within the means of many people. Even for those with the requisite statistical skill, mustering the computing power sufficient to handle an entire country's data would be costly.[120] The cost might be less prohibitive for reconstructions of smaller areas,[121] but even reconstructing the data of small areas might be more trouble than it is worth. The sort of information a reconstruction would reveal—age, sex, race, and ethnicity—is already available from other sources such as commercial data brokers and social media platforms.[122] If a costly reconstruction experiment cannot offer better information than is

---

[118] Abowd, *supra* note 47.

[119] *Id.*

[120] *See id.* (suggesting that reconstruction of the entire country's census data could cost "millions of dollars").

[121] *See* Hansen, *supra* note 45 (claiming the New York Times performed a reconstruction experiment on the Manhattan's data—but not elaborating on the success or failure of that experiment).

[122] *See* Geoffrey A. Fowler, *How Politicians Target You: 3,000 Data Points on Every Voter, Including Your Phone Number*, WASH. POST (Oct. 27, 2020), https://www.washingtonpost.com/technology/2020/10/27/political-campaign-data-targeting/ [https://perma.cc/D6AK-AXED].

already available at low costs, it is difficult to imagine why any attacker would reconstruct census data.

In fact, reconstruction might not even offer better information than an attacker could get by randomly guessing the contents of census data. In 2022, Steven Ruggles and David Van Riper published a paper called *The Role of Chance in the Census Bureau Database Reconstruction Experiment* arguing exactly that.[123]

To understand their argument, imagine an attacker who knows the most common characteristics of individuals in every census block (for example, that a given block is mostly occupied by White people). This is a realistic assumption; data of this sort continues to be publicly released and is generally not thought to violate individuals' privacy.[124] Now, imagine the attacker simply assumed that everybody within a block could be described by these characteristics. This attacker would sometimes be correct—in fact, for any single characteristic, they would more often be correct than incorrect. Yet this would probably not be a privacy issue: The attacker would also frequently be incorrect and would not be able to tell which of their guesses happened to be successful.

Ruggles and Van Riper argue that this sort of guessing should be thought of as a "control" in a reconstruction experiment—like a

---

[123] *See* Steven Ruggles & David Van Riper, *The Role of Chance in the Census Bureau Database Reconstruction Experiment*, 41 POPULATION RSCH. & POL'Y REV. 781, 781 (2022). For a slightly less technical summary of Ruggles & Van Riper's argument, see Paul Francis, *The Ruggles/Van Rider Critique of the Census Bureau Reconstruction Attack: An Explainer*, LINKEDIN (Aug. 31, 2021), https://www.linkedin.com/pulse/rugglesvan-rider-critique-census-bureau-attack-paul-francis/?trackingId=RogNV4%2BLIaF3xYs2g5VOLA%3D%3D [https://perma.cc/8FTW-CCQ4].

[124] *DATA GEM: How to Access Data for Your Neighborhood in Just a Few Clicks*, U.S. CENSUS BUREAU (July 21, 2020), https://www.census.gov/data/academy/data-gems/2020/how-to-access-data-for-your-neighborhood.html [https://perma.cc/3E6J-ENDU].

placebo[125] in a drug trial.[126] If a sophisticated reconstruction attack cannot recreate individuals' data significantly better than this sort of guessing could, then the attack should not be thought of as successful—similar to how a drug that performs no better than a placebo should not be thought of as effective.

Ruggles and Van Riper then did what the Census Bureau's 2018 reconstruction experiment did not: Using simulated census data, they created such a control.[127] First, they assumed everybody within a census block belonged to the most common racial and ethnicity groups on that block.[128] Then, they paired these race and ethnicity characteristics with randomly guessed age and sex characteristics.[129] They found that about 41% of these age, sex, race, and ethnicity combinations perfectly matched the actual data in the simulated census blocks.[130] Thus, they argued, the Census Bureau's sophisticated reconstruction attack—which had guessed the age, sex, race, and ethnicity data of 46% of the population—performed little better than semi-random guessing.[131]

To understand how this is possible, consider that people tend to live near other people who are similar to them. While the causes for this phenomenon are complex and hotly debated—ranging from housing discrimination to self-segregation of racial and ethnic groups[132]—the fact that such segregation exists is undeniable: "43% of persons reside on a block with one or more other people who

---

[125] A placebo is a fake drug that the recipient believes to be real. Sick people who believe they have taken a drug will often see an improvement in their symptoms, even if they did not in fact take a drug. Drug trials therefore do not compare the results of people who have received a drug to those who received nothing. Instead, they compare the results of people who receive a drug to people who receive a placebo.

[126] *See* Ruggles & Van Riper, *supra* note 123, at 781.

[127] *Id*. at 783–84.

[128] *Id.* at 783.

[129] These random age and sex characteristics were pulled from the actual distribution of age and sex characteristics across the population. *Id.* at 783-84.

[130] *Id*.

[131] *Id*. at 786.

[132] *See generally* Leah Platt Boustan, *Racial Residential Segregation in American Cities* (Nat'l Bureau of Econ. Rsch., Working Paper No. 19045, 2013) (describing the causes of residential segregation).

share their exact characteristics."[133] The Census Bureau's 2018 reconstruction experiment's 46% "success rate" becomes less worrying in this light; it may have been similar to the success rate of guessing census respondents' data based off no more than the basic demographic statistics for their area.[134] Ruggles and Van Riper argued in favor of this interpretation, pointing out that "[d]atabase reconstruction ought to work best with small blocks where the published tables directly reveal unique combinations of respondent characteristics."[135] Instead, in the Census Bureau's 2018 experiment, "[t]he larger the block, the more exact matches; in fact, large blocks had three times the match rate of small blocks . . . The obvious explanation is that larger blocks have higher odds of including by chance any specific combination of age, sex, race, and ethnicity."[136]

Ruggles and Van Riper's findings throw into question how useful a reconstruction attack on census data would be to an attacker. If the attacker could reach similarly accurate results through random guessing, it is difficult to imagine why they would attempt reconstruction at all. Moreover, unlike the Census Bureau, the attacker would not have the original data on hand to determine which of their guesses were correct.[137]

---

[133] *See* Ruggles & Van Riper, *supra* note 123, at 783.

[134] *See also* Paul Francis, *A Note on the Misinterpretation of the US Census Re-identification Attack*, *in* PRIVACY IN STATISTICAL DATABASES 299–311 (2022).

[135] Ruggles & Van Riper, *supra* note 123, at 782–83.

[136] *Id.*

[137] The Census Bureau recently attempted to estimate how a reconstruction attack might compare to a statistical baseline, implicitly acknowledging that the Ruggles and Van Riper critique has merit. *See* Abowd et al., *supra* note 53, at 9. The Bureau reported that the data of people who possessed different characteristics from others on their block remained highly vulnerable to reconstruction. *See id*. However, this interpretation has faced some criticism for being "misleading": the Bureau did not actually construct a statistical baseline given the high cost involved, its results therefore remain uncertain. *See* Paul Francis, *Thoughts on the recent US Census Bureau Attack Paper*, LINKEDIN (Dec. 26, 2023), https://www.linkedin.com/pulse/thoughts-recent-us-census-bureau-attack-paper-paul-francis-regue/ [https://perma.cc/MUE7-E3C8].

2. Insensitive Data

The sort of data a reidentification attack would expose—age, sex, race, and ethnicity—might not be harmful to the people it described. This is a highly contested point.[138] The Census Bureau argues that reidentification could "make it easier to target individuals—particularly in vulnerable populations like communities of color, same-sex couples, older adults, or parents of very young children—for fraud, enforcement actions, disinformation, or physical or virtual abuse."[139] Others suggest that:

> The U.S. Department of Housing and Urban Development could use census data to find households where people are misusing Section 8 vouchers . . . . Domestic abusers might find the data useful for tracking down their victims. Census data could reveal the sexual orientation of people who don't want to be out.[140]

But these worries may be overblown. Federal agencies are statutorily prohibited from using census microdata[141] to the detriment of the individuals the data describe.[142] Additionally, the data reconstructed in the 2018 experiment—age, sex, race, and ethnicity—were relatively insensitive. These are characteristics that most people display publicly; they could just as easily be captured

---

[138] *See* Abowd et al., *supra* note 53, at 7–8 (pushing back against the idea that census data revealed by reconstruction are insensitive).

[139] *See Disclosure Avoidance: Latest Frequently Asked Questions*, *supra* note 102.

[140] Capps, *supra* note 34.

[141] Importantly, this rule does not prevent group-level statistics from being used to the detriment of respondents. Perhaps the worst example of this comes from World War II, when census statistics were used to target Japanese-Americans to send them to internment camps. *See* Margo J. Anderson, *The Census and the Japanese 'Internment': Apology and Policy in Statistical Practice*, 87 SOC. RES. 789, 789–812 (2020). At first glance, this might seem to support to the idea that census respondents should be worried if their data is not kept private. But remember that even under the new disclosure avoidance methods, group-level statistics remain publicly available with no reconstruction necessary. Abuse of group-level statistics is therefore a separate issue from the privacy concerns surrounding individual-level data that this article discusses.

[142] *See* 13 U.S.C. § 8(c) ("In no case shall [individual-level census data] be used to the detriment of any respondent or other person to whom such information relates."). Before this provision was passed, census statistics were used to identify and prosecute draft dodgers during World War I. *See* Anderson, *supra* note 54.

from a camera in a public location or a social media account as through the census.[143]

The insensitive nature of the data that might be reconstructed also calls into question the Census Bureau's argument that a successful reidentification attack would damage its ability to collect data.[144] Admittedly, these fears are not without basis. "[C]oncerns about privacy and confidentiality are among the reasons most often given by potential respondents for unwillingness to participate in surveys."[145] Yet given how insensitive the data revealed by a successful reidentification attack would be, it is worth second guessing the effect such an attack would have on public trust. No public outcry materialized after the Census Bureau publicized its 2018 reconstruction and reidentification experiment as a success, essentially informing anyone who responded to the census in 2010 that their age, sex, race, and ethnicity data might be tied back to their identity (the Ruggles and Van Riper criticism of this interpretation had not been published yet). And the Census Bureau has not reported that the 2018 experiment caused subsequent issues collecting ACS data or census data in 2020. Perhaps, then, the public is attuned to how sensitive the data released are. Or, perhaps the Census Bureau's data collection issues stem more from general distrust of the government than from fears of reidentification.[146] In either case, the types of data that the Census Bureau reconstructed in 2018—age,

---

[143] *See* Matthew Yglesias, *They Deliberately Put Errors in the Census*, SLOW BORING (Aug. 16, 2021), https://www.slowboring.com/i/39976329/what-is-the-census-protecting-us-against [https://perma.cc/P69E-EMZ5] (arguing that reidentification of census data poses no greater risk than security cameras in public businesses or social media data collection).

[144] *E.g.*, Brief for Defendant at 2, Alabama v. U.S. Dep't of Com., 546 F. Supp. 3d 1057 (M.D. Ala. 2021) ("If the Census Bureau were to continue doing what it did in 2010, it would be violating . . . the confidentiality promise that it made to census respondents. And with that bond of trust broken, future census response rates would undoubtedly fall, and the accuracy of future censuses would suffer.").

[145] *See Disclosure Avoidance: Latest Frequently Asked Questions*, *supra* note 102.

[146] For in-depth treatment of the argument that the Census Bureau's data collection issues are rooted in general distrust of the government, not fears of reidentification, see generally Ruggles & Magnuson, *supra* note 32.

sex, race, and ethnicity—would not seem to be cause for additional concern.

A more sweeping argument in favor of protecting even insensitive data comes from Paul Ohm. Ohm worries that

> databases will grow to connect every individual to at least one closely guarded secret. This might be a secret about a medical condition, family history, or personal preference. It is a secret that, if revealed, would cause more than embarrassment or shame; it would lead to serious, concrete, devastating harm. And these companies are combining their data stores, which will give rise to a single, massive database . . . Once we have created this database, it is unlikely we will ever be able to tear it apart.[147]

Ohm, therefore, sees releases of even innocuous data as problematic; such data are links in the chain that will form the "database of ruin."[148]

Without weighing in on the plausibility of Ohm's nightmare scenario—which, if it will ever come to pass, does not seem to have done so yet—it is worth wondering whether accurate census data might be worth the risk of forming a link in the "database of ruin" chain. As previously discussed, the probability that census data will be reconstructed—and will therefore act as such a link—appears low. What seems to be higher is the probability that inaccurate census data deprive users of valuable information—information that keeps elections fair, allows diseases to be tracked, social issues to be studied, and enables people and businesses to understand the world around them.

B. What Harms Might Inaccurate Census Data Cause?

How do the new disclosure avoidance methods—via the inaccurate census data they create—affect various uses of the data? Of the two new methods, noise injection (or "differential privacy")

---

[147] Paul Ohm, *Don't Build a Database of Ruin*, HARV. BUS. REV. (Aug. 23, 2012), https://hbr.org/2012/08/dont-build-a-database-of-ruin [https://perma.cc/6R28-MECL].

[148] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1745–50 (2010) (arguing that reidentification of non-harmful data could be an intermediary step in linking multiple databases, the end result of which would be harmful).

is the most studied in the context of census data.[149] The results are troubling. Differentially private data appear to create distortions in statistics describing small neighborhoods and minority groups.[150] This raises concerns for thousands[151] of socially valuable uses of the data, which can be organized into five categories: (1) Redistricting; (2) Native American Nations; (3) Public Benefit Provision; (4) Small-Scale Data Uses; and (5) Public Health, Social Science, and Policy Research. These categories are neither mutually exclusive nor exhaustive of the areas where inaccurate census data could cause harm.

The research pertaining to each of these data usage categories remains speculative. Different researchers sometimes come to dramatically different conclusions regarding how large of a problem the Census Bureau's new methods create.[152] One source of the uncertainty is that researchers can only simulate the effect of the new disclosure avoidance methods. They cannot compare the released (less accurate) data to the true data because, of course, releasing the accurate data to them would violate the Bureau's privacy mandate.

1. Redistricting

States must draw voting districts for both U.S. congressional representatives and state legislators such that certain districts are not more powerful than others. Otherwise, states risk being found in violation of the "One person, One vote" rule that the Supreme Court

---

[149] *See infra* Part IV.B.1, 3, 5 (detailing research on the effects of noise injection on census data).

[150] *E.g.*, Christoph F. Kurz et al., *The Effect of Differential Privacy on Medicaid Participation Among Racial and Ethnic Minority Groups*, 57 HEALTH SERVS. RSCH. 207, 211 (2022).

[151] Ruggles et al., *supra* note 10, at 17–18 (finding that over 70,000 papers cite census or American Community Survey data).

[152] *Compare* Christopher T. Kenny et al., *The Use of Differential Privacy for Census Data and Its Impact on Redistricting: The Case of the 2020 U.S. Census*, 7 SCI. ADVANCES, (2021) (finding differential privacy creates concerns for the legal validity of census data used for redistricting), *with* Aloni Cohen et al., *Private Numbers in Public Policy: Census, Differential Privacy, and Redistricting*, HARV. DATA SCI. REV., Jun. 2022 (finding errors associated with differential privacy are comparable to those associated with pre-existing miscounts).

has read into the Constitution.[153] But drawing such districts requires knowing where people live—and knowing where people live requires accurate data.

In 2021, Christopher Kenny simulated the creation of voting districts with noisy, differentially private data—that is, data altered by one of the new disclosure avoidance methods.[154] Kenny found that the errors in the data led "to a likely violation of the 'One Person, One Vote' standard."[155] Deviations in population between voting districts were roughly ten times larger than deviations between districts drawn with accurate census data.[156] Additionally, the discrepancies had "unpredictable" effects on the partisan makeup of districts, making partisan gerrymanders more difficult to detect.[157]

Kenny also found that the new disclosure avoidance methods "systemically" created voting districts with lower levels of racial and ethnic diversity than would be created by accurate data.[158] The effect was large enough to lower the total number of majority-minority districts.[159] This could clash with the Voting Rights Act, which sometimes requires that states draw legislative districts that contain a majority of minority voters to prevent their votes from being diluted.[160]

Disagreement remains over how much concern these results warrant. Other researchers have taken issue with the methodology that led to Kenny's results; these researchers have found only minor errors in redistricting, comparable to errors that already existed due to miscounts by the Census Bureau.[161] Additionally, in 2021, the

---

[153] *See, e.g.*, Reynolds v. Sims, 377 U.S. 533, 558 (1964) ("The Equal Protection Clause requires substantially equal legislative representation for all citizens in a State regardless of where they reside.").

[154] Kenny et al., *supra* note 152, at 1.

[155] *Id.*

[156] *Id*. at 6.

[157] *Id*. at 7–8.

[158] *Id*. at 1.

[159] *Id*.

[160] *See* 52 U.S.C. § 10301.

[161] Cohen et al., *supra* note 152 (finding errors associated with differential privacy are comparable to those associated with pre-existing miscounts). *See also* Tommy Wright & Kyle Irimata, Empirical Study of Two Aspects of the Topdown

Census Bureau decided to inject less noise into the data used for redistricting.[162] This could decrease the issues found by Kenny, though it may not entirely eliminate them.[163]

Still, with the 2024 election quickly approaching, multiple parties have voiced concerns over how the new inaccuracies in census data could distort their votes. In Louisiana, Black voters recently cited the Census Bureau's new methods in their suit arguing that the state was diluting their votes.[164] The Louisiana court, however, found the evidence on this point too thin.[165] Similarly, Alabama voters sought an injunction against the Census Bureau releasing only the inaccurate data.[166] Although they were unsuccessful, emails surfaced during discovery showing internal disagreement at the Census Bureau over the new disclosure avoidance methods.[167] The Chief of the Redistricting & Voting Rights Data Office worried that "in our zeal to protect the data we are harming the very same people we are protecting."[168]

2. Native American Nations

Native American nations govern themselves, despite being small in number and spread out across rural areas. To do so, they require information on their population.

Unfortunately, small populations also mean that Native American data is among the most dramatically affected by the

---

Algorithm Output for Redistricting: Reliability & Variability (Aug. 5, 2021) (unpublished manuscript) (on file with author) (suggesting differentially private census data are suitable for redistricting use).

[162] *Meeting Redistricting Data Requirements: Accuracy Targets*, U.S. CENSUS BUREAU (Sept. 23, 2021), https://content.govdelivery.com/accounts/USCENSUS/bulletins/2cb745b [https://perma.cc/94L2-FTEC].

[163] *Cf.* Kenny et al., *supra* note 152, at 14–18.

[164] Nairne v. Ardoin, No. CV 22-178-SDD-SDJ, 2024 WL 492688, at *28 (M.D. La. 2024).

[165] *Id*.

[166] Alabama v. U.S. Dep't of Com., 546 F. Supp. 3d 1057 (M.D. Ala. 2021).

[167] *Cf.* Plaintiffs' Reply in Support of Combined Motion for a Preliminary Injunction and Petition for a Writ of Mandamus at 1, Alabama v. U.S. Dep't of Com., 546 F. Supp. 3d 1057 (M.D. Ala. 2021).

[168] *Id*. (quoting an email between James Whitehorne, Chief of the Redistricting & Voting Rights Data Office at the U.S. Census Bureau, and John Abowd, Associate Director and Chief Scientist at the U.S. Census Bureau).

Census Bureau's new disclosure avoidance methods. Especially small populations could experience errors in their population estimates by as much as 25%, and even some slightly larger areas (e.g., those containing 1,000 people) could still see errors of around 13%.[169] The results are even more dramatic for data representing subgroups (e.g., measurements of the number of people of a certain age or sex).[170] In fact, many small-scale statistics (such as breakdowns of a particularly small population by age) have entirely ceased to be available.[171] Losing access to these data could be devastating for Native Americans in rural counties where "the Tribe is the only economic engine."[172]

3. Public Benefit Provision

From Medicaid, to Pell Grants, to juvenile delinquency prevention, countless federal programs use census data to allocate funds.[173] Even before the controversy over the Census Bureau's new disclosure avoidance methods, statisticians estimated that small census miscounts could result in tens of billions of misallocated federal funds.[174]

Now, research on differentially private census data confirms that they probably do affect public benefit allocation—and as expected, the effects are most pronounced for small geographic areas and minority groups. For example, the inaccurate data could distort measurements of county-level Medicaid participation rates by over 10%, especially for small counties that are dominated by minority groups.[175] In fact, the only group whose Medicaid participation rates are not distorted are the non-Hispanic White population.[176]

---

[169] Gregg, *supra* note 5, at 5–6.

[170] *Id*. at 6.

[171] *Id*. at 7–11, 20.

[172] *Id*. at 1.

[173] *See* Zachary H. Seeskin & Bruce D. Spencer, *Effects of Census Accuracy on Apportionment of Congress and Allocations of Federal Funds* 3061 (Nw. Univ. Inst. for Pol'y Rsch., Working Paper No. 15-05, 2015) (displaying a sample of federal programs that use census data to allocate funds).

[174] For example, for an average relative root mean squared error of 4%, Seeskin and Spencer estimate a misallocation of $80.8 billion. *See id.*

[175] Kurz et al., *supra* note 150, at 211.

[176] *Id*.

Similarly, a 2022 study found that the Census Bureau's new disclosure avoidance methods had a "relatively modest" effect on the allocation of funds to large groups, but that for smaller groups, the methods "potentially diminish the utility of census data that have been collected at the cost of roughly $10 billion."[177] These sorts of findings raise serious concerns for government programs that use especially fine-grained funding allocations, like Housing and Urban Development Community Block Grants, Rural Business Development Grants, and the Rural Microentrepreneur Assistance Program.[178]

4. Small-Scale Data Users

Small-scale users of census data are numerous and diverse. Local demographic information is important to lawyers trying to prove racial discrimination,[179] journalists and businesses trying to understand local economic trends, cities and municipalities attempting evidence-based policymaking, and even informed citizens looking to better understand their world.[180] But given how the new disclosure avoidance methods affect data for small geographies the most, small-scale data users may be among the most affected parties.

One particularly striking example comes from the city of New Orleans. In the early 2010s, old homes across the city frequently lacked smoke alarms, killing residents who were not alerted to fires in time.[181] So, using ACS data, the city developed a statistical model that could predict whether a house lacked a smoke alarm based on its age, the length of time its occupant had lived in it, and whether

---

[177] Brummet, *supra* note 18, at 30.

[178] *See* Ruobin Gong et al., *Harnessing the Known Unknowns: Differential Privacy and the 2020 Census*, HARV. DATA SCI. REV., Jun. 2022, at 1.

[179] *E.g.*, State v. Johnson, 275 N.C. App. 980 (N.C. Ct. App. 2020) (finding a defendant needed to provide statistics regarding the racial makeup of small-scale police districts to present prima facie evidence of discrimination).

[180] *See* LINDA JACOBSON ET AL., AMERICA'S ESSENTIAL DATA AT RISK: A VISION TO PRESERVE AND ENHANCE THE AMERICAN COMMUNITY SURVEY 7–20 (2023).

[181] *See* Earlene K.P. Dowell, *New Orleans Uses Census Data to Hand Out Free Smoke Alarms Where Needed*, U.S. CENSUS BUREAU (Oct. 8, 2019), https://www..census.gov/library/stories/2019/10/fighting-fires-with-data.html [https://perma.cc/A7BN-U6DV].

the occupant was close to or under the poverty line.[182] Before this model, the Fire Department would install smoke alarms in under a thousand homes each year.[183] But with the help of the data-driven model, the Fire Department reached over ten thousand homes in one year.[184] Thus, the ACS data literally saved lives.

It is difficult to tell whether this use of ACS data would be affected by the new disclosure avoidance methods. But this uncertainty is itself a problem. Small-scale users are often the least sophisticated users of census data. They may not realize they are dealing with modified data, and even if they do, they may become confused on how to properly account for the inaccuracies.

5. Social Science, Public Health, and Policy Research

Steven Ruggles, social scientist and creator of the largest population database in the world, warns that "[i]f public use data become unusable . . . [t]he quantity and quality of research about U.S. policies, the economy, and social structure would decline precipitously."[185] For example, the Census Bureau's new disclosure avoidance methods have a "substantial and concerning impact" on statistics describing migration between U.S. counties.[186] While the effects are the most pronounced for statistics describing small counties and minority groups, the issues persist even in counties of over 100,000 people.[187]

Public health researchers looking at the new disclosure avoidance methods have reached similar conclusions. Differentially private census data could distort measurements of mortality rates.[188] The inaccuracies would "more strongly affect mortality rate

---

[182] LINDA A. JACOBSON & MARK MATHER, UNDERSTANDING AND USING AMERICAN COMMUNITY SURVEY DATA 31 (2020).

[183] Dowell, *supra* note 181.

[184] *Id*.

[185] Ruggles et al., *supra* note 10, at 17.

[186] Richelle L. Winkler et al., *Differential Privacy and the Accuracy of County-Level Net Migration Estimates*, 41 POPULATION RSCH. AND POL'Y REV. 417, 430 (2022).

[187] *Id*. at 430–31.

[188] Alexis R. Santos-Lozada et al., *How Differential Privacy Will Affect Our Understanding of Health Disparities in the United States*, 117 PROC. NAT'L ACAD. SCI. 13405, 13405 (2020).

estimates for non-Hispanic Blacks and Hispanics than estimates for non-Hispanic Whites."[189] Mortality rates specific to COVID-19 could also be affected.[190] The "populations most at risk for . . . distortion, namely, elderly and minority populations, are the very groups COVID-19 harms the most and are in need of the most targeted interventions."[191]

Although researchers who use census data have been among the most vocal critics of the Census Bureau's new methods, their concerns are sometimes treated dismissively.[192] It can be difficult to see the value of research not yet completed, whose findings and public impact remain unknown. The benefits of some data may not accrue for many years—or decades. Statisticians Zachary Seeskin and Bruce Spencer point out how recent policy efforts to increase post-secondary education were based on empirical research published in 2008, which used census data collected between 1940 and 1980 as well as Iowa State Census data collected in 1915.[193] "Not only is it difficult to identify such uses of census data after they have occurred, but it is even more difficult to anticipate them ahead of time."[194]

But the unpredictability of any particular research project should not lead to the devaluation of research as a whole. As Seeskin and Spencer's example demonstrates, a world without the sort of research enabled by accurate census data would be a poorer one.

C. Might Informal Privacy Be Preferable to Formal Privacy?

Taken together, the previous two subsections suggested that the costs of the Census Bureau's new disclosure avoidance methods may outweigh the benefits. Admittedly, this analysis was

---

[189] *Id*.

[190] Mathew E. Hauer & Alexis R. Santos-Lozada, *Differential Privacy in the 2020 Census Will Distort COVID-19 Rates*, 7 SAGE J. 1, 1 (2021).

[191] *Id*. at 4.

[192] *Cf.* Capps, *supra* note 34 ("For particular members of these populations, what is more valuable to them . . . the fact that [a researcher] gets accurate statistics about them, or the fact that they have stronger privacy protections?").

[193] *See* Seeskin & Spencer, *supra* note 173, at 3073.

[194] *Id*.

speculative. What is more certain is that there *are* meaningful costs to reducing data accuracy for the sake of privacy.

But the tradeoff between privacy and accuracy is unavoidable—or at least, it is unavoidable under the current law. *Baldrige* effectively requires the Census Bureau to adopt formal privacy protections. Recall that *Baldrige* most likely requires the Census Bureau to protect not only individuals' identities, but "the data itself"[195]—that is, to prevent not just reidentification, but reconstruction too. Protection against reconstruction is what statisticians call formal privacy.[196] By defining privacy as the inability for an attacker to deduce microdata, this concept makes privacy quantifiable.[197] Under the Census Bureau's new disclosure avoidance methods, statisticians can calculate the degree of confidence an attacker can have in their reconstruction attempt. Formal privacy therefore allows the Census Bureau to meet *Baldrige*'s demands in the modern age—to know that microdata cannot be deduced.

The downside of formal privacy is that it involves an unavoidable trade-off with accuracy. John Abowd, the Chief Scientist for the Census Bureau, presents this as a strength, not a weakness, of formal privacy.[198] One could imagine a graph in which privacy rises as accuracy falls; this graph can actually be created when using a formal definition of privacy.[199]

---

[195] Baldrige v. Shapiro, 455 U.S. 345, 356 (1982).

[196] *See, e.g.*, Cynthia Dwork et al., *Calibrating Noise to Sensitivity in Private Data Analysis*, *in* THEORY OF CRYPTOGRAPHY 265 (2006) (defining privacy loss as the ability of a query to reveal information about a dataset, a concept now known as differential privacy).

[197] *Cf. id.*

[198] John M. Abowd & Ian M. Schmutte, *Revisiting the Economics of Privacy: Population Statistics and Confidentiality Protection as Public Goods* 35 (U.S. Census Bureau, Ctr. for Econ. Stud., Working Paper No. 17-37, 2017).

[199] Indeed, Abowd suggests that the quantifiable relationship between formal privacy and accuracy could allow statisticians to find the optimal bundle of privacy and accuracy one day. *See id.*

And yet, formal privacy tells us little about reidentification risk. Even data that are easy to reconstruct may nonetheless be difficult to connect back to specific individuals.[200]

It is therefore *informal* privacy—protection from reidentification, not protection from reconstruction—that the Census Bureau should prioritize.[201] But as long as Title 13 remains under the control of *Baldrige* and seeks to protect "the data itself," the Census Bureau's ability to value informal privacy is constrained. Were it to abandon formal privacy in favor of informal privacy, the Census Bureau would risk exposing itself to legal penalties.

Congress must therefore act to allow the Census Bureau to strike a better balance between privacy and data accuracy. While this would unquestionably raise the risk of a successful reconstruction attack, the evidence does not suggest that the risk of such an attack is high.[202] Moreover, even if the arguments discussed throughout the previous two Subsections are proven wrong—even if reidentification risk turns out to be high and the harms caused by less accurate data turn out to be rare—the changes advocated in this Article would still improve Title 13. The tradeoff between accuracy and formal privacy is a fundamental mathematical truth; the law ought to take it into account and give the Census Bureau the flexibility necessary to determine the optimal privacy-accuracy bundle as the costs and benefits of accurate data change over time. Additionally, new attention from Congress could bring informal privacy options to the table that reduce reconstruction risk without sacrificing accuracy. The next Part examines what a better legal regime might look like.

---

[200] *See generally* David McClure & Jerome P. Reiter, *Differential Privacy and Statistical Disclosure Risk Measures: An Investigation with Binary Synthetic Data*, 5 TRANSACTIONS DATA PRIV. 535 (2012) (finding that reidentification risk can be low even where little formal privacy exists).

[201] For a (slightly) more technical argument arguing for informal privacy—albeit without considering the role of Congress versus the Census Bureau in making policy—see Francis, *supra* note 137 ("I'm glad to see the Bureau acknowledge that the design they ended up with is not future-proof, and hopeful that this signals a willingness to consider non-formal protections.").

[202] *See* discussion *supra* Part IV.A.

V. AN ALTERNATIVE LEGAL REGIME: INFORMAL PRIVACY

A better legal regime would have to start with an amendment to Title 13's language that clearly eschews *Baldrige*'s requirement of formal privacy. Currently, Title 13, Section 9 prohibits data publications "whereby the data furnished by any particular establishment or individual under this title can be identified."[203] Instead, Congress should consider prohibiting publications whereby census data "can be reasonably reconnected to the identity of the particular establishment or individual who furnished that data." A phrase like "reconnected to the identity" more explicitly invokes reidentification rather than reconstruction. The qualifier "reasonably" could allow the Census Bureau room to tolerate especially low risk forms of reidentification. For example, if a reconstruction attacker were to randomly guess the correct race of a person based off the most common race in their area, then, as Ruggles and Van Riper suggest, this should not be considered reasonable reconnection.

Title 13, Section 8 provides that the Census Bureau "may furnish copies of tabulations and other statistical materials which do not disclose the information reported by, or on behalf of, any particular respondent."[204] The word "disclose"—a highly ambiguous word that played an important role in *Baldrige*'s reasoning—should be removed. The statute might instead allow the Census Bureau to "furnish copies of tabulations and other statistical materials which do not violate Section 9." This would standardize the Census Bureau's privacy requirement across the text of Title 13.

Writing informal privacy into the statute would not require the Census Bureau to provide more accurate data; it would merely give the Bureau the option to do so. Congress should therefore also add a provision that requires the Census Bureau to weigh the benefits of privacy against the costs of accuracy when adopting new disclosure avoidance methods. This would mandate that the Census Bureau walk through an analysis similar to the one in the previous Part; it would put the Bureau under greater pressure to show its work and get it right.

---

[203] 13 U.S.C. § 9(a)(2).
[204] *Id.* § 8(b).

Congress should also think creatively about authorizing specific forms of informal privacy, especially if Congress finds the above statutory amendments too radical. One version of informal privacy already exists under the current law. Federal Statistical Research Data Centers ("FSDRCs") allow researchers gain access to otherwise restricted data. However, FSDRCs are currently inaccessible to the vast majority of data users. As a report from the University of Minnesota explains:

> [At FSDRCs] every stage of the research process is significantly more time-consuming than using public use data, and only the most persistent researchers are successful. In addition, most of the branches charge high fees for anyone unaffiliated with an institution sponsoring an FSRDC. Projects are approved only if they benefit the Census Bureau, which by itself makes most research topics ineligible. Prospective users must prepare detailed proposals, including the precise models they intend to run and the research outputs they hope to remove from the center, which are generally restricted to model coefficients and supporting statistics. Most descriptive statistics are prohibited. Researchers are not allowed to "browse" the data or change the outputs based on their results.[205]

Still, the idea behind FSDRCs—that trustworthy data users can be given full access without jeopardizing privacy—remains sound. Perhaps the Census Bureau could release accurate data to those who complete background checks. Extensive federal government infrastructure already exists to give thousands of individuals low-level security clearance for the purpose of government employment.[206] Such infrastructure could be expanded and repurposed to allow access to accurate census data outside FSDRCs. Similarly, the Census Bureau could license certain institutions to receive accurate data, such as universities, local governments, courts, and newspapers.

These sorts of creative approaches could provide substantial privacy protection while also ensuring census data remain accurate and usable. But because solutions like these provide only informal privacy, it is unlikely that the Census Bureau will adopt them so long as its legal mandate continues to favor formal privacy. Congress must take the next step.

---

[205] *See* Ruggles et al., *supra* note 10, at 17–18.

[206] *Cf.* CONG. RSCH. SERV., RL43216, SECURITY CLEARANCE PROCESS: ANSWERS TO FREQUENTLY ASKED QUESTIONS 6 (2023).

VI. CONCLUSION

The benefits of accurate census data are not always as intuitive as the benefits of privacy, but they are no less real. We need accurate data to conduct fair elections, to enable Native American nations to govern themselves, to properly allocate billions of dollars to programs like Medicaid, to allow local governments and businesses to understand the world around them, and to fuel the valuable work of public health researchers, social scientists, and policy analysts.

However, Title 13 now forces the Census Bureau to reduce the accuracy of the data it publishes. Title 13 does not do this by requiring the Census Bureau to carefully weigh the benefits of privacy against the costs of inaccuracy. Instead, Title 13, as interpreted by *Baldrige*, demands that the Census Bureau protect "the data itself"[207]—that it uses new disclosure avoidance methods to prevent a possible reconstruction attack, even if such an attack is unlikely, or even if would yield no better information than an attacker could get through random guessing.[208] Such a rule does not appear justified.

The debate over the new disclosure avoidance methods remains a "noisy" one—both in the sense that many voices take many positions, and also in the sense that it remains difficult to know the cost of the new methods.[209] For example, different scholars come to different conclusions as to whether the new disclosure avoidance methods threaten the integrity of the redistricting process.[210] Moreover, to its credit, the Census Bureau has already lowered the inaccuracies in certain data releases in response to concerns over redistricting.[211] This Article therefore does not attempt to offer a

---

[207] *See* Baldrige v. Shapiro, 455 U.S. 345, 356 (1982).

[208] *See* Ruggles & Van Riper, *supra* note 123, at 781.

[209] "Noise," as used by statisticians, refers to random variations or inaccuracies that obscure patterns in data.

[210] *Compare* Cohen, *supra* note 152 (arguing that the new disclosure avoidance methods do not create legally significant distortions in redistricting), *with* Kenny, *supra* note 152 (arguing that the new disclosure avoidance methods may prevent redistricting to meet constitutional standards).

[211] *See Meeting Redistricting Data Requirements: Accuracy Targets*, *supra* note 162.

final verdict on the new methods; it only suggests that there remain reasons to worry that their costs outweigh their benefits.[212]

But as a matter of law, the issue is more clear. If the Census Bureau is to have the freedom to make optimal policy—whatever that policy might be—it cannot continue to be haunted by the specter of *Baldrige*. Congress should amend Title 13 to give the Bureau greater flexibility to adopt informal privacy protections. The Census Bureau must be able to weigh privacy against accuracy—and, when appropriate, to choose accuracy.

[212] For a starkly different take on the issue than this Article presents, see generally Abowd et al., *supra* note 53.