



How differential privacy will affect our understanding of health disparities in the United States

Alexis R. Santos-Lozada^{a,1}, Jeffrey T. Howard^b, and Ashton M. Verdery^c

^aDepartment of Human Development and Family Studies, The Pennsylvania State University, University Park, PA 16802; ^bDepartment of Public Health, University of Texas at San Antonio, San Antonio, TX 78249; and ^cDepartment of Sociology and Criminology, The Pennsylvania State University, University Park, PA 16802

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved April 20, 2020 (received for review February 27, 2020)

The application of a currently proposed differential privacy algorithm to the 2020 United States Census data and additional data products may affect the usefulness of these data, the accuracy of estimates and rates derived from them, and critical knowledge about social phenomena such as health disparities. We test the ramifications of applying differential privacy to released data by studying estimates of US mortality rates for the overall population and three major racial/ethnic groups. We ask how changes in the denominators of these vital rates due to the implementation of differential privacy can lead to biased estimates. We situate where these changes are most likely to matter by disaggregating biases by population size, degree of urbanization, and adjacency to a metropolitan area. Our results suggest that differential privacy will more strongly affect mortality rate estimates for non-Hispanic blacks and Hispanics than estimates for non-Hispanic whites. We also find significant changes in estimated mortality rates for less populous areas, with more pronounced changes when stratified by race/ethnicity. We find larger changes in estimated mortality rates for areas with lower levels of urbanization or adjacency to metropolitan areas, with these changes being greater for non-Hispanic blacks and Hispanics. These findings highlight the consequences of implementing differential privacy, as proposed, for research examining population composition, particularly mortality disparities across racial/ethnic groups and along the urban/rural continuum. Overall, they demonstrate the challenges in using the data products derived from the proposed disclosure avoidance methods, while highlighting critical instances where scientific understandings may be negatively impacted.

census | differential privacy | disclosure avoidance | mortality | demography

In September 2018, the US Census Bureau announced that they would implement differential privacy (DP) on data products derived from 2020 census data (1). DP works by infusing noise into data through implementing a top-down algorithm, infusing noise to the nation, then to the states, and on down to blocks (2). The implementation of this method “marks a sea change for the way that official statistics are produced and published” (3). Numerous organizations, data users, researchers, and demographers have expressed concern about the accuracy of the data produced under the DP algorithm and the usefulness of these releases for creating public policy, monitoring population structures and distribution, and expanding scientific understandings of ongoing demographic changes in this country. Recently, the US Census Bureau released 2010 demonstration data products produced using the new disclosure avoidance system (DAS) to allow the data-user community to study the utility of these census tabulations and discuss the trade-offs between accuracy and privacy (4). The Committee on National Statistics conducted a workshop on 2020 census data products in December 2019, where many researchers and data users expressed concerns about the accuracy and usefulness of the noise infused data (4). Given that DP is untested in the decennial census environment, continued use of traditional techniques could be preferable until

more is known about the impact of DP on important uses of census data, particularly in light of concerns discussed at the workshop on December 2019 (4).

Why are there concerns about privacy? Title 13 of the United States Code imposes heavy obligations on the US Census Bureau not to release data that could be successfully reidentified (5). There is concern that modern algorithms, like variations of machine learning (6), may make it possible that standard census data products, like decennial tabulations, could violate this statute because they can reidentify individuals; DP is an attempt to make such efforts futile. However, the US Census Bureau conducted internal reconstruction efforts that have resulted in varying levels of reidentification “success,” depending on the specifications of the reconstruction algorithm (7, 8). Many have noted that the implementation of DP sacrifices the accuracy of upcoming census data releases to protect the privacy of respondents despite “reconstruction” efforts failing between 29% and 62% of the time, with the former lowest fail rate only attained allowing for a 1-y difference in age between the reconstructed files and the population records (7, 9). In prior reconstruction efforts, only 17% of the population was confirmed with confidential data, not using datasets available commercially (10). Despite such failure rates for past reconstruction efforts, census analysts insist that the implementation of a new disclosure methodology due to the

Significance

The decennial census is the primary source of statistical information about the US population, and the quality of its data is of great interest to numerous stakeholders. One particular concern is the accuracy of census data products in light of proposed disclosure avoidance methodologies, especially if these methods disproportionately affect scientific understandings of contemporary social phenomena. Using the 2010 decennial counts produced with proposed differential privacy and traditional techniques, we evaluate how the implementation of differential privacy can affect understandings of mortality rates by obscuring accurate denominators. We find that the implementation of differential privacy will produce dramatic changes in population counts for racial/ethnic minorities in small areas and less urban settings, significantly altering knowledge about health disparities in mortality.

Author contributions: A.R.S.-L. designed research; A.R.S.-L. and J.T.H. performed research; A.R.S.-L. and A.M.V. analyzed data; and A.R.S.-L., J.T.H., and A.M.V. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: All data and code used to produce this analysis is available through an online repository, and access will be granted upon request to the corresponding author following NCHS guidelines for data disclosure. We have mentioned this within the manuscript text.

¹To whom correspondence may be addressed. Email: ar39@psu.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2003714117/-DCSupplemental>.

First published May 28, 2020.

threat of database reconstruction is warranted (5, 11). Although there remains the theoretical prospect for reconstruction, a historical overview indicates that “there is not a single documented case of anyone outside the Census Bureau revealing the responses of a particular identified person in public use decennial census or ACS data” (12). In studies of simulated external attacks, only a small fraction of possible reidentifications turned out to be correct, which has led some to conclude that reidentification risks are small (5, 13). A blog entry by Dr. Ron Jarmin, Deputy Director and COO at the Census Bureau, stated that a study conducted by census researchers concluded that the accuracy of the reconstructed data were limited, and that confirmation of reidentified responses required access to data only available at the Census Bureau (14). Furthermore, although proponents of DP implementation have noted there is a need to better understand the willingness to pay for privacy and statistical accuracy (15) and to develop principles that will guide the evaluation of the protection mechanisms (16), they have documented numerous issues encountered while implementing DP (3). It has been recommended that the Census Bureau proceed with caution and in consultation with key users of census data (5). Better understanding the implications of DP in census products and the perceptions of the data-user community are thus vital. The US Census Bureau has compiled these concerns and continues to listen to the data-user community to identify where the new DAS needs to be improved (17).

The decennial census is the principal source of information about the population of the United States. It operates as a de jure census, where, every 10 y, each resident is counted according to where they usually live on April 1 (18). The original and primary purpose of the census is for political apportionment, which is followed by legislative redistricting. Nevertheless, the decennial census is used much more broadly today. Population counts are used by every branch of the federal government, and census data influence the distribution of federal funds, grants, and other forms of support for local governments while documenting population characteristics such as sex, age, race/ethnicity, and other demographic factors (18). Many state and local stakeholders use census data to assign funds for schools, hospitals, roads, public works, and other forms of government spending (19–21). A Census Bureau report found that 132 programs used census data to distribute more than \$675 billion in funds during fiscal year 2015 (22). Likewise, the decennial census is a key source of information used by businesses for market research and to plan the locations of stores and factories (20). Problems with census data quality can be expected to have ripple effects on job creation and economic activity in a variety of sectors.

Instances in which the decennial census is crucial include, but are not limited to, the study of population size, distribution, and change and measures of population composition such as age/sex composition (18). Census data are also an all-important component of the estimation of population-level indicators for fertility, health, migration, and mortality (23–26) and consumer demographics (20). For example, applied demographers rely on this data source to inform disaster planning and assessment of vulnerable populations (27), determine the number and characteristics of older adults (28), project school enrollment (29), perform health needs assessments (30), and estimate demand for services like transportation (31). Other areas where the accuracy of these data are crucial include, but are not limited to, electoral demography (32–34) and informing business decisions (35). These data are also crucial for expert witness analysis on cases of gerrymandering (36–38). In addition, census data constitute the most reliable source of information used to validate estimation techniques and the coverage of administrative records (39), which are increasingly being used for research and policy decision-making (40–42). Last, census data are vital for planning and implementing the next round of census data collection, and

for the public, research community, and representatives to evaluate those plans.

In this article, we use empirical data to assess one of the most pressing effects of the implementation of DP for county-level population counts, overall and stratified by racial/ethnic groups, and to examine what DP-induced variability in these counts might mean for mortality rate estimates and our understanding of racial/ethnic disparities. Mortality rates are estimated using data from two sources: the numerator comes from vital records and the denominator comes from official population counts (43); they are most often calculated for population subgroups (e.g., by age, sex, race, or interactions of these and other criteria). As such, DP may affect these estimates by changing the denominators’ fidelity to underlying population size. We calculate apparent population changes due to differences in disclosure avoidance methodologies and the magnitude of these changes with respect to county-level population size. Drawing on these calculations, we first document how county-level mortality rate estimates could be affected if DP is implemented, under current parameters, in comparison to estimates produced using the official 2010 population counts. We ask how might county-level mortality rate estimates change for the overall population and three major racial/ethnic groups if differential privacy is implemented using the parameters underlying the demonstration products shared by the US Census Bureau in October 2019. We focus on mortality rate estimates because these are essential population-level metrics for which data are collected and disseminated at the national level, but this analytical approach can be applied to any health condition that affects the population of the United States. Mortality rates are a critical indicator of population health, and recent scholarship has made very broad inferences about societal health and well-being on their basis, for instance, by tying the burgeoning phenomenon of “deaths of despair” to all manner of social challenges and changes (44–46). An advantage of focusing on mortality rate estimates is that death records are published for the whole nation and mortality records satisfy basic criteria of coverage that make them a reliable data source to conduct this analysis (47). Finally, we assess the association between changes in mortality rate estimates and population size and urbanization or adjacency to metropolitan areas (*Approach, Methods, Data, and Measures*).

The results of these analyses indicate substantial variation due to the noise introduced into the population counts included in the 2010 demonstration products in comparison to the 2010 official counts. We then consider how these changes in the denominators impact the accuracy of a population-level health metric by examining changes in the mortality rate estimates due to the change in the denominators. In this test, we examine the magnitude and direction of the changes in mortality rate estimates that DP would introduce. We quantify this change as rate ratios, measuring the deviation from the 2010 crude mortality rate estimates when DP-adjusted data are used as a denominator. This test reveals that the accuracy of mortality rate estimates is significantly affected in areas with smaller populations and lower levels of urbanization, as well as those non-adjacent to a metropolitan area. The tabular analysis and spatial representation show a combination of increases and decreases in both population counts and mortality rates for the four groups of interest, with differences in both direction and magnitude. The results of these tests indicate that implementation of DP would mostly affect understandings of mortality differences among non-Hispanic blacks and Hispanics.

Results

County-Level Population Change. In Table 1, we present a descriptive analysis of the change in population counts for the overall population and three major racial/ethnic groups when comparing the data produced under the traditional and currently

Table 1. Change in 2010 population counts for counties due to the implementation of differential privacy ($N = 3,131$)

Finding	Overall population			Non-Hispanic whites		
	Minimum	Average	Maximum	Minimum	Average	Maximum
Count	-811	-0.39	4,217	-305	-0.30	211
Percent (%)	-2.50	0.44	29.24	-48.84	-0.08	6.72
	Counties			Counties		
	Reduce	No change	Increase	Reduce	No change	Increase
Counties	1,198	17	1,916	1,537	26	1,568
Percent (%)	38.26	0.54	61.19	49.09	0.83	50.08
	Non-Hispanic blacks			Hispanics		
	Counties			Counties		
	Reduce	No change	Increase	Reduce	No change	Increase
Counties	1,686	58	1,359	1,332	15	1,784
Percent (%)	54.33	1.86	43.80	42.54	0.48	56.98

Change was estimated by subtracting the official 2010 population count from the DP count. For the percent calculations, we used the 2010 population counts as a denominator. In the case of non-Hispanic blacks, 28 counties do not have enough information to calculate percent change. For this section, the number of counties (N) is 3,103.

proposed disclosure avoidance methods. Positive changes mean DP leads to overreporting the population, while negative changes mean that it underreports the population. The range of change for the overall population is from -811 to 4,217 people, which roughly translates in a minimum reduction of 2.50% and a maximum increase of 29.24%. On average, the overall population change is zero, as would be expected given the goals of DP. The range for population and percent change is smaller for the non-Hispanic white population in comparison to every other racial/ethnic group. Conversely, both changes in counts and percentages have a wider range for the non-Hispanic black and the Hispanic population, with artificial increases due to DP being over 1,000%. Table 1 also presents the corresponding figures for number of counties where population change is recorded. Change is observed for the overall population and every racial/ethnic group. The group with the most reductions in county population due to DP are non-Hispanic blacks, with a difference of 149 counties in comparison to the corresponding count to non-Hispanic whites (second highest). The group with the most increases in county-level population counts is Hispanics, with 216 counties more than the second highest value (non-Hispanic whites). In *SI Appendix, Fig. S1*, we show maps for percent change attributable to differential privacy for the overall population (*SI Appendix, Fig. S1A*) and the three major racial/ethnic groups (*SI Appendix, Fig. S1 B-D*) of the United States.

In Fig. 1, we show the association between population size under the traditional disclosure protection techniques and percent change due to the implementation of DP for the overall population (Fig. 1A) and the three major racial/ethnic groups (Fig. 1 B-D). Results show that there are higher levels of change for less numerous populations, with these differences being more pronounced for non-Hispanic blacks and Hispanics. In the case of the overall population, population counts show a small increase in less populous areas. For non-Hispanic whites, we observe increases and decreases in smaller areas, with these changes being smaller in more populous areas. On the contrary, for non-Hispanic blacks, we find higher levels of variation both in frequency and magnitude for less populous areas, and we observe this phenomenon even in areas with larger populations. For Hispanics, the implementation of DP results in higher levels of population change, with the more pronounced

effects leaning toward an increase in the Hispanic population in less populous areas.

County-Level Mortality Rates. In Fig. 2, we show a comparison between mortality rate estimates calculated using the official 2010 US Census counts and counts resulting from the implementation of differential privacy as denominators. For the overall population and non-Hispanic whites, we observe clustering of values on top of the line of equality (Fig. 2, blue 45° lines). For non-Hispanic blacks and Hispanics, there are greater deviations from the line of equality than for non-Hispanic whites, but deviations are most pronounced for non-Hispanic blacks. In this instance, using the denominators from the DP dataset produces lower mortality rate estimates than those produced using the 2010 official counts as denominators. This finding reflects the higher level of variation in mortality rates for non-Hispanic blacks and Hispanics than for non-Hispanic whites or for the overall population.

At the county level, differences in mortality rate estimates are better captured by the calculation of mortality rate ratios (MRRs). The MRRs indicate whether the mortality rates calculated using the demonstration product counts result in artificial increases or reductions in comparison to rates produced using 2010 census counts as denominators. We approach these differences using a threshold of ± 0.25 for each population of interest: increase ($MRR > 100.25$), similar ($99.75 < MRR < 100.25$), or decrease ($MRR < 99.75$) in comparison to the mortality rate estimates calculated using the DP counts. A tabulation of these three mutually exclusive categories to assess the direction and magnitude of the change in mortality rate estimates is included in *SI Appendix, Table S1*. We observe changes across every group of interest. However, there is a substantial difference between the estimates by population subgroups. For example, a higher percentage of counties fall within what could be considered similar levels for the overall population (60.14%) and non-Hispanic whites (77.43%). In contrast, the MRRs for non-Hispanic blacks and Hispanics are lower than 40%. Fig. 3 shows the spatial distribution of MRRs for the threshold described earlier for the overall population (Fig. 3A) and the three major racial/ethnic groups (Fig. 3 B-D). We assessed a wider threshold of $\pm 1.00\%$ and the results are consistent (*SI Appendix, Fig. S3*).

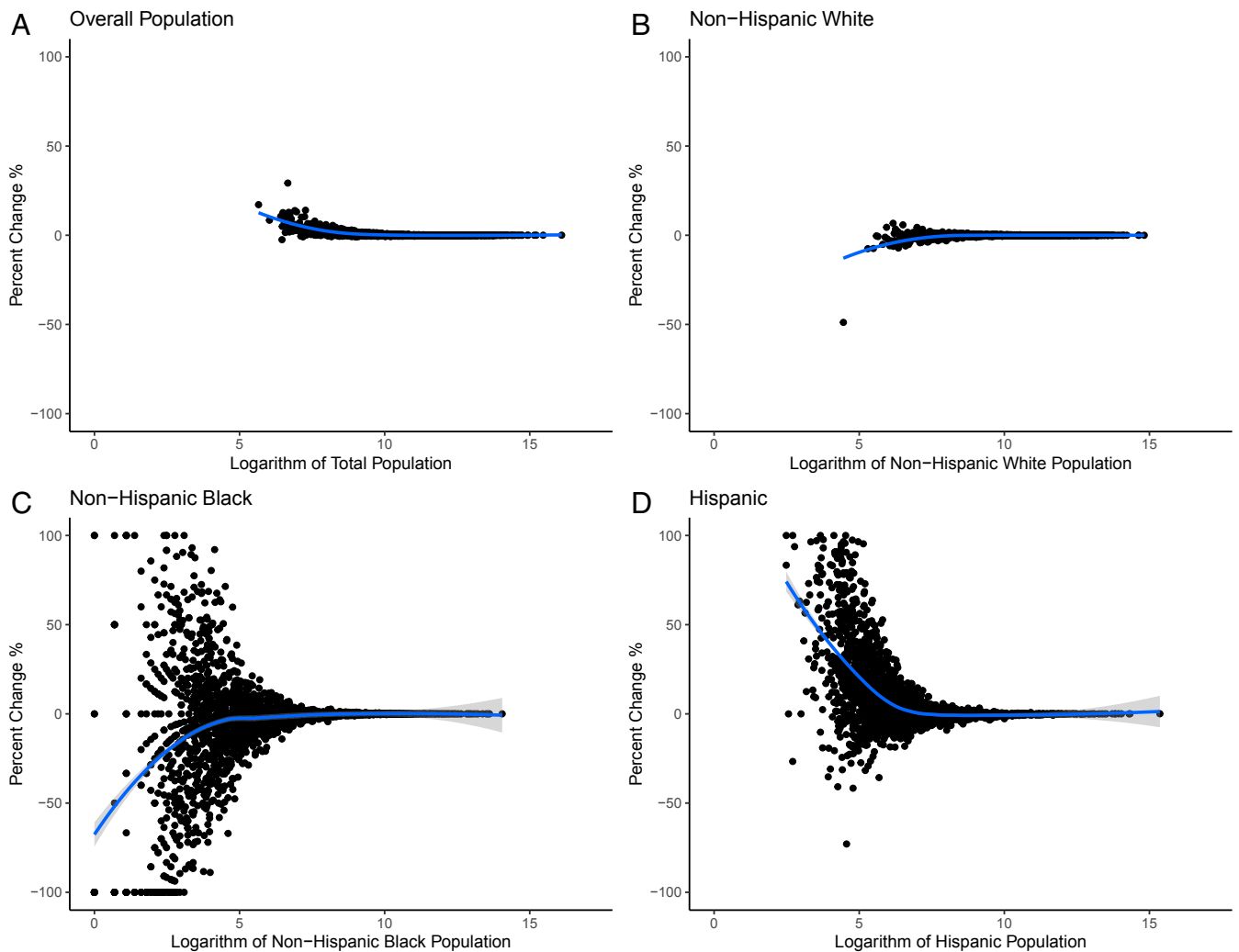


Fig. 1. Percent of population change at the county level due to the implementation of DP and 2010 population size (the logarithm scale of 2010 originally published counts): overall population (A), non-Hispanic whites (B), non-Hispanic blacks (C), and Hispanics (D). Comparisons are between originally published 2010 data and the 2010 data with noise infused by DP.

In Fig. 4, we demonstrate the associations between population size and MRR for the groups of interest. In Fig. 4 A and B, we present the relation between population size and the MRR for the overall population and non-Hispanic whites. There is no clear pattern observed regarding population size and MRR for the total or the non-Hispanic white population. On the contrary, there is a noticeable pattern for non-Hispanic blacks and Hispanics. First, both groups show variation in areas with smaller populations. The MRR for non-Hispanic blacks shows a combination of overestimation and underestimation of mortality rates using alternative denominators. In the case of mortality rates for the Hispanic population, there is a combination of over- and underestimation of the mortality rates, but the underestimation of mortality rates using the DP denominator is more pronounced, with some MRR values exceeding 200. These results mean that, in some cases, the mortality rate estimates calculated using 2010 census counts were twice as high as those produced using the DP denominators. We see similar patterns for non-Hispanic whites living in less populated areas.

We expand the analyses presented in Fig. 4 by exploring whether the MRRs differed by degree of urbanization or adjacency to metropolitan areas. The results are particularly informative regarding the intersection of racial/ethnic and

geographic differences (SI Appendix, Fig. S2). For example, the changes in mortality rate estimates for the overall population and non-Hispanic whites are concentrated in the nonmetro areas with an urban population of 20,000 or more but not adjacent to a metro area. On the contrary, we observe higher levels of variation and the detection of more outliers for non-Hispanic blacks and Hispanics across almost every degree of urbanization or adjacency to a metropolitan area. The cases of the Rural-Urban Continuum codes (RUCC) 6 and higher, which have lower levels of urbanization, are worthy of attention, as there are values where the mortality rates derived using DP denominators are underestimating mortality for non-Hispanic blacks and Hispanics. Such levels of overestimation and underestimation are not observed for the overall population or non-Hispanic whites, with very few exceptions.

Discussion

The implementation of differential privacy to the 2020 Census, as proposed, may affect analyses of the demographic landscape of the United States, and this will have numerous implications for the study of demographic change and health disparities within the nation. First, the change to this disclosure avoidance system will endanger what is known about the demographic

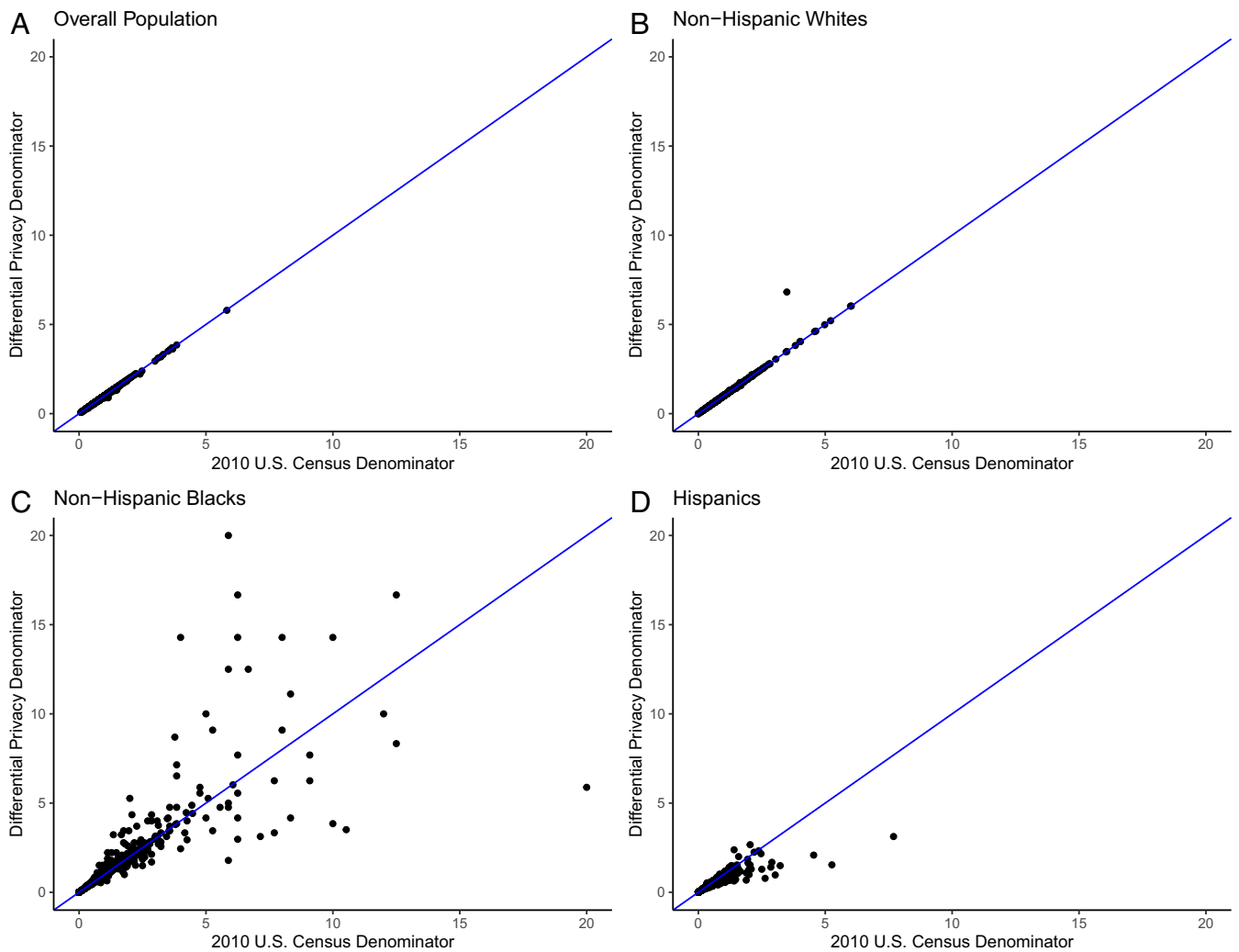


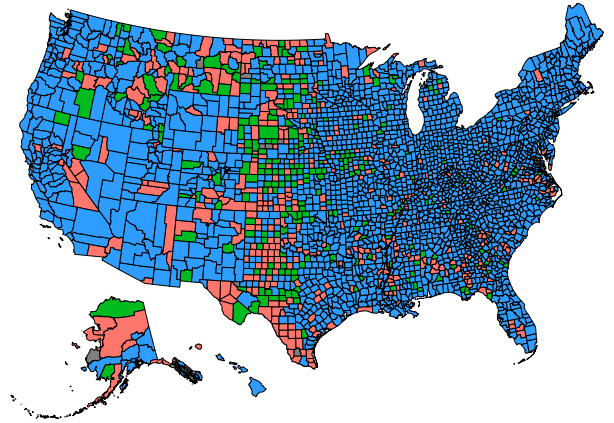
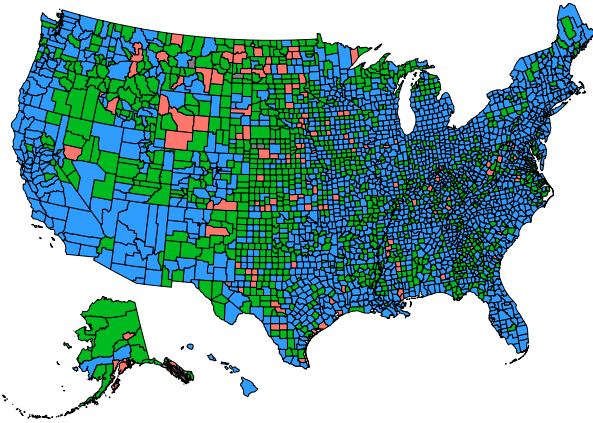
Fig. 2. Comparison between mortality rates using the official 2010 US Census counts (x axis) and counts with noise infused by DP (y axis) as denominators: overall population (A), non-Hispanic whites (B), non-Hispanic blacks (C), and Hispanics (D).

transformations of the US population, which is information used by the public and the private sector in their decision-making processes. Because population counts derived from the census inform intercensal population estimates and projections, inaccurate data releases in the 2020 Census could affect our knowledge of the population into the future and as long as the proposed disclosure avoidance system remains in place. Further, this may bias analyses that use these data sources to study population change across time and space, potentially yielding the appearance of discontinuities around 2020 even in the absence of such discontinuities. Second, emerging destinations of minority populations will not be identifiable because these populations will be moved to other areas. Infusing noise in the data, in comparison to the current disclosure avoidance system, will produce inaccurate patterns of demographic change with higher levels of error found in the calculations for non-Hispanic blacks and Hispanics. At the same time, these counts are bound to impact post-2020 districting for both federal and state elections, as well as evaluations of that redistricting. Fortunately, we know how the districts were drawn in 2010, and the resulting infusion of noise in the demographic composition of these districts should be the subject of future studies. Likewise, noise-infused data products will pose challenges for the work of state and local governments because many of the demographic and economic

analyses performed in this sector are informed by data sources produced by the US Census and focus on smaller geographic areas and sometimes less populous ones. Third, these changes in population counts will affect understandings of health disparities in the nation, leading to overestimates of population-level health metrics of minority populations in smaller areas and underestimates of mortality levels in more populated ones. Here, the effects are dramatic. For example, in McCulloch County, Texas, the mortality rate ratio for non-Hispanic blacks is 75.9, indicating the mortality rate would be 24% lower under the current methodology compared with the differential privacy methodology. Similarly, in Clarke County, Virginia, the mortality rate ratio for Hispanics is 121.4, indicating the mortality rate would be 21% higher under the current methodology compared with the differential privacy methodology. At the same time, the non-Hispanic white mortality rate ratios were essentially unchanged for these two counties, at 100.3 and 99.8, respectively, meaning substantial biases may enter into understandings of disparities. The infusion of noise into the data are more pronounced for areas with smaller populations and areas at the higher RUCC codes. This method could then have implications for the allocation of funds to tackle health disparities across the nation, leading to overspending in some areas and underspending in others based on inaccurate needs assessments.

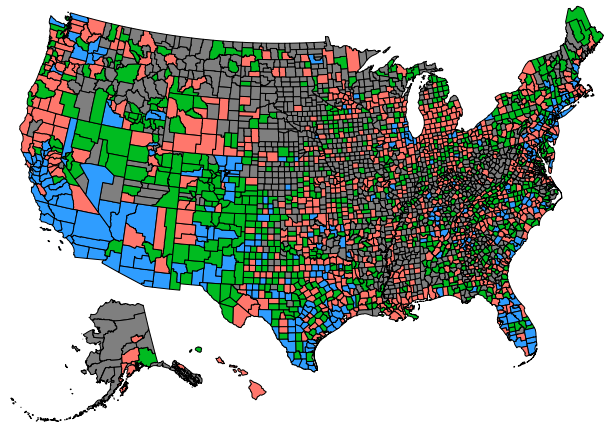
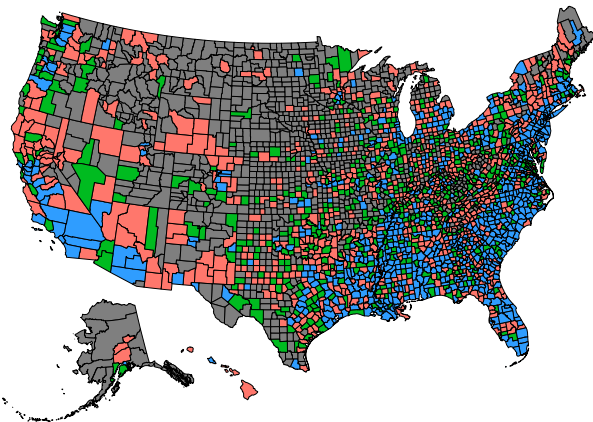
A Overall Population

B Non-Hispanic Whites



C Non-Hispanic Blacks

D Hispanics



MRR < 99.75 >100.25 99.75,100.25 NA

Fig. 3. County-level mortality rate ratios for the overall population and three major racial/ethnic groups indicating whether the MRR was below 99.75 (red), within 99.75–100.25 (blue), or over 100.25 (green): overall population (A), non-Hispanic whites (B), non-Hispanic blacks (C), and Hispanics (D). MRR comparisons are between originally published 2010 data and the 2010 data with noise infused by DP. Counties that do not have enough information for the calculation of these rates are presented in gray.

Our results focus on the changes in population counts and their effects on the accuracy of population-level health metrics. Nevertheless, the proposed change in disclosure avoidance methodology has additional effects. Given that census products are used in complex survey design, sampling and weighting that rely on these products will also be biased, potentially leading to cascading effects in other layers of knowledge production about the nation. While Complete Count Committees, a coalition of local governments, community-based organizations, faith-based groups, schools, businesses, the media, and others, strive for a complete count of the population across the nation, they will not receive a data release that accurately represents the population they serve. If these data releases result in inaccurate population-level metrics, local and state government and businesses will be burdened with finding solutions to this by investing their resources in population registries or leveraging administrative records to obtain data that are available under the current disclosure avoidance method. For example, California has already expressed concern about the accuracy of population counts produced by the US Census and are dedicating resources to produce the California Neighborhood Counts, a minicensus funded by the state government (48). However, different attitudes

regarding the accuracy of population counts are likely to vary between states as well as within them. For example, California has announced they will spend more than \$187 million in census outreach, while the Texas and 24 other state legislatures have declined to spend any money in complete count efforts (49).

Approach, Methods, Data, and Measures

We obtained 2010 county-level population counts released under the traditional disclosure avoidance techniques and the ones produced with the proposed differential privacy procedures (50, 51). We accessed counts for the total population, non-Hispanic whites, non-Hispanic blacks, and Hispanics. Any difference between these counts is due to changes in the disclosure avoidance methodologies (traditional techniques vs. currently proposed). Traditional disclosure techniques applied to the 2010 census counts include record-swapping, item imputation, whole household imputation, rounding, and top- and bottom-coding (52). The 2010 Demonstration Data Products, which implement DP, work by allocating a “privacy-loss budget” or ϵ (53). The 2010 DP counts were produced under a global $\epsilon = 6$, where personal records use $\epsilon = 4$ and housing records use $\epsilon = 2$ (54). We do not have demonstration data across different levels of privacy loss. We produced a descriptive analysis that includes the minimum, average, and maximum values for changes in population counts and how these changes compared to 2010 official population counts for the groups of interest. In Fig. 1, we present the county-level percent change due to the change in disclosure

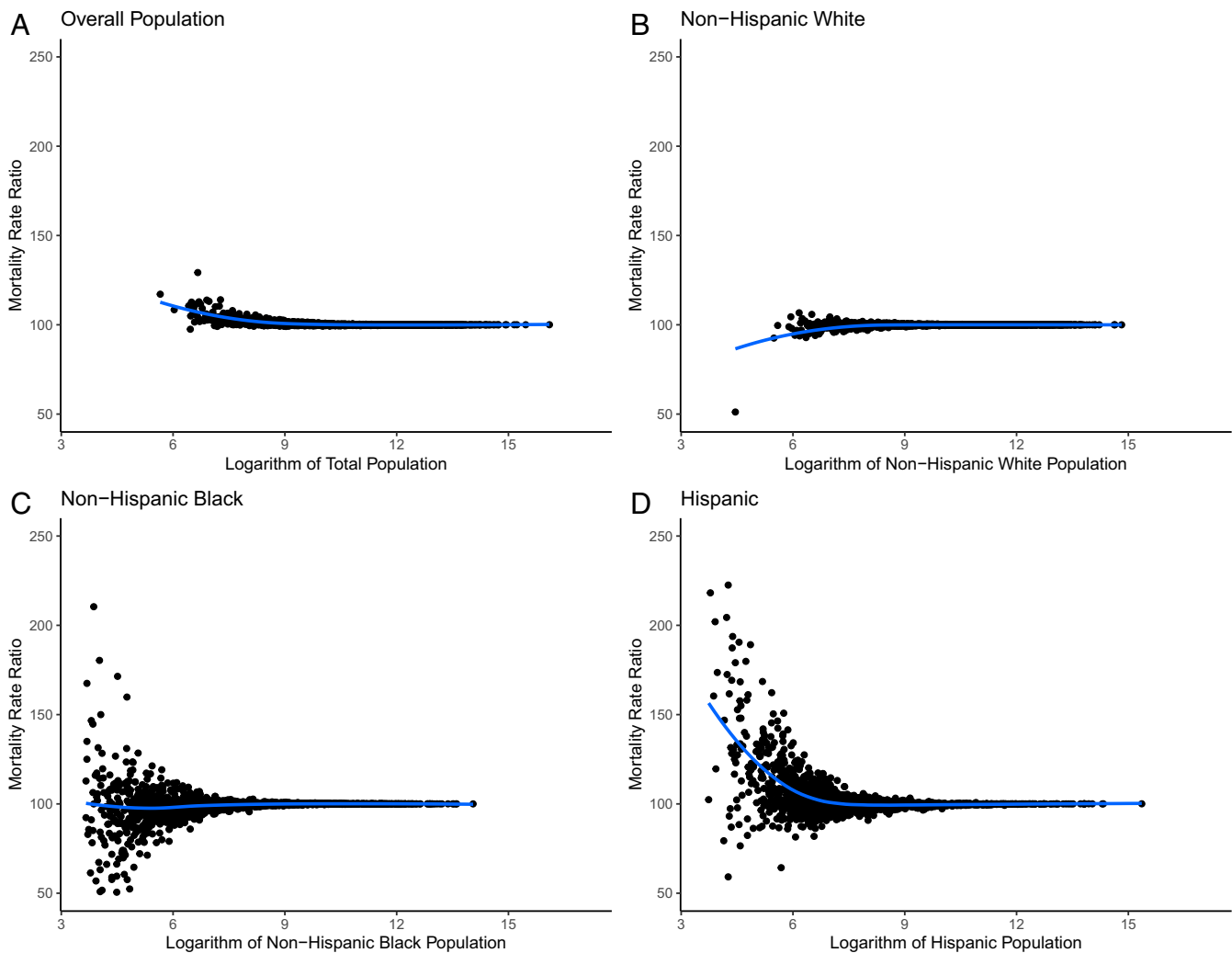


Fig. 4. County-level mortality rate ratio (MRR) and population size (the logarithm of the originally published population size for 2010): overall population (A), non-Hispanic whites (B), non-Hispanic blacks (C), and Hispanics (D). An MRR of 100 is equivalent to equal mortality rates, an MRR below 100 indicates that the mortality rate under the DP is higher than the one calculated using 2010 US Census results, and an MRR above 100 indicates the contrary.

avoidance methodologies for the overall population and the three major racial/ethnic groups. County-level population changes over or below $\pm 100\%$ were omitted from the visualization (number of omitted counties by race, non-Hispanic blacks, $n = 13$; Hispanics, $n = 5$). We present the association between percent change and population size, and present maps for percent change are provided in *SI Appendix, Fig. S1*.

Given that our primary interest was in how a population-level health metric would be affected by the implementation of differential privacy, we produced county-level death counts for the aforementioned population groups by exploring changes in crude mortality rate estimates. We obtained individual-level death records from the All County Multiple Cause of Death Mortality Microdata files provided by the National Center of Health Statistics (NCHS) through a collaborative data use agreement (55). Using these death counts, we estimated county-level 2010 crude mortality rates for the overall population, non-Hispanic whites, non-Hispanic blacks, and Hispanics using the 2010 official population counts (M_1) and those produced using the differential privacy methodology (M_2). We also produced a mortality rate ratio (MRR), a ratio of these mortality rates, to explore how the differences in denominators produced different mortality rates. The MRR, calculated as $MRR = (M_1/M_2) * 100$, indicates whether M_2 was higher than ($MRR < 100$), the same as ($MRR = 100$), or lower than ($MRR > 100$) the M_1 (43). We accessed the 2013 Rural-Urban Continuum codes (RUCCs) published by the USDA to incorporate this measure of urbanization and adjacency to a metro area to our analysis (56). We matched each county with its corresponding RUCC and performed an analysis of the MRR by RUCC (*SI Appendix, Fig. S2*, provides an explanation of the RUCC). This provides the opportunity to analyze the

changes beyond the standard metro and nonmetro classification and distinguish among diverse residential groups (56). Finally, we produced maps of the MRR to visualize the spatial distribution of MRR differences, allowing for a small range of variability, categorized as below 99.75, within 99.75–100.25, and over 100.25; an alternative range of variability was also explored (*SI Appendix, Fig. S3*). We also present the association between MRR and population size in Fig. 4. MRRs over 250 were suppressed in the visualizations presented in Fig. 4 (number of counties, OP, $n = 0$; NH whites, $n = 0$; NH blacks, $n = 6$; Hispanics, $n = 4$). We used R (57) for data manipulation and ggplot2 to produce data visualizations (58).

Data Availability. All data and code used to produce this analysis are available through an online repository, and access will be granted upon request to the corresponding author following NCHS guidelines for data disclosure.

ACKNOWLEDGMENTS. We thank the Integrated Public Use Microdata Series (IPUMS) and the National Historical Geographic Information System (NHGIS) teams for sharing the 2010 demonstration products in a format accessible for the data user community. This work was supported by the Population Research Institute (Grant R24-HD041025 and P2CHD041025), the Data Accelerator, and the Social Science Research Institute at the Pennsylvania State University. This work was also supported by the Center for Community Based and Applied Health Research at the University of Texas at San Antonio. We thank David Van Riper for his support in the early stages of this work. Finally, we thank the employees of the US Census Bureau for allowing the data-user community to provide timely feedback regarding this crucial issue.

1. J. Mervis, Can a set of equations keep U.S. census data private? *Science*, 10.1126/science.aaw5470 (2019).
2. S. L. Garfinkel, Deploying Differential Privacy for the 2020 Census of Population and Housing in Joint Statistical Meetings (US Census Bureau, Washington, DC, 2019).
3. S. L. Garfinkel, J. M. Abowd, S. Powazek, "Issues encountered deploying differential privacy" in *Proceedings of the ACM Conference on Computer and Communications Security* (ACM, New York, NY, 2018), pp. 133–137.
4. Committee on National Statistics, Workshop on 2020 Census data products: Data needs and privacy considerations. <https://www.nationalacademies.org/event/12-11-2019/workshop-on-2020-census-data-products-data-needs-and-privacy-considerations>. Accessed 13 February 2020.
5. S. Ruggles, C. Fitch, D. Magnuson, J. Schroeder, Differential privacy and census data: Implications for social and economic research. *AEA Pap. Proc.* **109**, 403–408 (2019).
6. A. Acquisti, R. Gross, Predicting social security numbers from public data. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10975–10980 (2009).
7. J. M. Abowd, Stepping-up: The U.S. Census Bureau Tries to Be a Good Data Steward in the 21st Century (US Census Bureau, Washington, DC, 2019).
8. P. Leclerc, "Results from a consolidated database reconstruction and intruder re-identification attack on the 2010 decennial census in challenges and new approaches for protecting privacy" in Federal Statistical Programs. <https://www.nationalacademies.org/event/12-11-2019/workshop-on-2020-census-data-products-data-needs-and-privacy-considerations>. Accessed 17 April 2020.
9. J. M. Abowd, Preparing for the 2020 Census: Disclosure Avoidance in Annual Meeting of the American Association of Geographers (US Census Bureau, Washington, DC, 2019).
10. M. Hawes, P. Leclerc, "Background on differential privacy at the U.S. Census Bureau and, 1940 census application" in Harvard Data Science Review Symposium. <https://hdr.mitpress.mit.edu/pub/h7kdirec/release/5>. Accessed 15 April 2020.
11. S. L. Garfinkel, J. M. Abowd, C. Martindale, Understanding database reconstruction attacks on public data. *ACM Queue* **16**, 1–26 (2016).
12. S. Ruggles, et al, Implications of differential privacy for Census Bureau data and scientific research (Minnesota Population Center, Working Paper Series No. 2018-6, 2018).
13. J. Abowd, "The U.S. Census Bureau adopts differential privacy" in *24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, 2018).
14. R. Jarmin, Census Bureau Adopts Cutting Edge Privacy Protections for 2020 Census (US Census Bureau, Washington, DC, 2019).
15. J. M. Abowd, I. M. Schmutte, An economic analysis of privacy protection and statistical accuracy as social choices. *Am. Econ. Rev.* **109**, 171–202 (2019).
16. J. M. Abowd, I. M. Schmutte, W. N. Sexton, L. Vilhuber, Why the economics profession must actively participate in the privacy protection debate. *AEA Pap. Proc.* **109**, 397–402 (2019).
17. J. M. Abowd, V. A. Velkoff, Modernizing disclosure avoidance: What we've learned, where we are now. *Census Blogs* (2020). https://www.census.gov/newsroom/blogs/research-matters/2020/03/modernizing_discluso.html. Accessed 25 March 2020.
18. S. K. Smith, J. Tayman, D. A. Swanson, *A Practitioner's Guide to State and Local Population Projections* (Springer, 2013).
19. L. Pol, R. Thomas, *The Demography of Health and Healthcare* (Springer Science & Business Media, 2000).
20. J. M. Martins, F. Yusuf, D. A. Swanson, *Consumer Demographics and Behaviour: Markets are People*. Springer Series on Demographic Methods and Population Analysis. (Springer, Dordrecht, The Netherlands, 2012), vol. 30.
21. D. A. Swanson, P. J. Walashek, *CEMAF as a Method A Proposal for a Re-Designed Census and an Independent U.S. Census Bureau* (Springer, Dordrecht, The Netherlands, 2011).
22. M. Hotchkiss, J. Phelan, Uses of Census Bureau Data in Federal Funds Distribution (US Census Bureau, Washington, DC, 2017).
23. P. A. Rogerson, D. Kim, Population distribution and redistribution of the baby-boom cohort in the United States: Recent trends and implications. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15319–15324 (2005).
24. D. Almond, L. Edlund, Son-biased sex ratios in the 2000 United States Census. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5681–5682 (2008).
25. K. G. Manton, L. Corder, E. Stallard, Chronic disability trends in elderly United States populations: 1982–1994. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2593–2598 (1997).
26. S. H. Woolf, H. Schoemaker, Life expectancy and mortality rates in the United States, 1959–2017. *JAMA* **322**, 1996–2016 (2019).
27. S. L. Cutter, C. Finch, Temporal and spatial changes in social vulnerability to natural hazards. *Plan. Clim. Chang. A Read. Green Infrastruct. Sustain. Des. Resilient Cities* **105**, 129–137 (2018).
28. P. A. Morrison, T. M. Bryan, Targeting spatial clusters of elderly consumers in the U.S.A. *Popul. Res. Policy Rev.* **29**, 33–46 (2010).
29. J. Baker, D. A. Swanson, J. Tayman, L. Tedrow, "Forecasting school enrollment size and composition" in *Cohort Change Ratios and Their Applications* (Springer, Dordrecht, The Netherlands, 2017), pp. 107–118.
30. M. K. Beckett, P. A. Morrison, Assessing the need for a new medical school: A case study in applied demography. *Popul. Res. Policy Rev.* **29**, 19–32 (2010).
31. J. Esser, K. Nagel, "Census-based travel demand generation for transportation simulations" in *Traffic and Mobility* (Springer, Dordrecht, The Netherlands, 1999), pp. 133–148.
32. P. A. Morrison, W. A. V. Clark, Local redistricting: The demographic context of boundary drawing. *Natl. Civ. Rev.* **81**, 57–63 (1992).
33. W. A. V. Clark, P. A. Morrison, Demographic foundations of political empowerment in multiminority cities. *Demography* **32**, 183–201 (1995).
34. S. J. Hill, D. J. Hopkins, G. A. Huber, Local demographic changes and US presidential voting, 2012 to 2016. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25023–25028 (2019).
35. P. A. Morrison, A. F. Abrahamse, Applying demographic analysis to store site selection. *Popul. Res. Policy Rev.* **15**, 479–489 (1996).
36. G. R. Webster, The census, reapportionment, and redistricting. *Geogr. Teach.* **16**, 89–94 (2019).
37. S. Hirsch, Unpacking page v. Bartels: A fresh redistricting paradigm emerges in new jersey. *Elect. Law J. Rules. Polit. Policy* **1**, 7–23 (2002).
38. G. Herschlag et al, Quantifying Gerrymandering in North Carolina. arXiv:1801.03783 (10 January 2018).
39. M. Spallek, M. Haynes, J. Baxter, N. Kapelle, The value of administrative data for longitudinal social research: A case study investigating income support receipt and relationship separation in Australia. *Int. J. Soc. Res. Methodol.*, 1–15 (2020).
40. B. Jarosz, J. Hofmocker, Research note: What counts as a house? Comparing 2010 census counts and administrative records. *Popul. Res. Policy Rev.* **32**, 753–765 (2013).
41. J. R. Goldstein, A. J. Morning, The multiple-race population of the United States: issues and estimates. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6230–6235 (2000).
42. B. Jarosz, Poisson distribution: A model for estimating households by household size. *Popul. Res. Policy Rev.*, 10.1007/s11113-020-09575-x (2020).
43. D. T. Rowland, *Demographic Methods and Concepts* (Oxford University Press, 2003).
44. A. Case, A. Deaton, Rising morbidity and mortality in midlife among white non-hispanic Americans in the 21st century. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15078–15083 (2015).
45. A. Case, A. Deaton, *Deaths of Despair and the Future of Capitalism* (Princeton University Press, 2020).
46. S. M. Monnat, D. L. Brown, More than a rural revolt: Landscapes of despair and the 2016 presidential election. *J. Rural Stud.* **55**, 227–236 (2017).
47. J. H. Glasser, The quality and utility of death certificate data. *Am. J. Public Health* **71**, 231–233 (1981).
48. H. Lo Wang, Outspending every other state on the census, California starts its own count too. *National Public Radio*, 13 January 2020. <https://www.npr.org/2020/01/13/795897141/outspending-every-other-state-on-the-census-california-starts-its-own-count-too>. Accessed 14 March 2020.
49. M. Wines, J. A. Del Real, In 2020 Census, big efforts in some states. In others, not so much. *NY Times*, 15 December 2019. <https://www.nytimes.com/2019/12/15/us/census-california-texas-undercount.html>. Accessed 18 December 2019.
50. US Census Bureau, *Profile of the general population and housing characteristics: 2010*, (US Census Bureau, Washington, DC, 2013).
51. National Historical GIS, Differentially Private 2010 Census Data (2019). <https://www.nhgis.org/differentially-private-2010-census-data>. Accessed 30 November 2019.
52. L. Zayatz, Disclosure avoidance practices and research at the U.S. Census Bureau: An update. *J. Off. Stat.* **23**, 253–265 (2007).
53. S. Petti, A. Flaxman, Differential privacy in the 2020 US census: What will it do? Quantifying the accuracy/privacy tradeoff. *Gates Open Res.* **3**, 1722 (2019).
54. US Census Bureau, *Frequently Asked Questions for the Demonstration Data Products* (US Census Bureau, Washington, DC, 2010).
55. National Center for Health Statistics, All County Multiple Cause of Death Mortality Microdata File, 2010, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program (National Center for Health Statistics, Hyattsville, MD, 2010).
56. US Department of Agriculture, Rural-Urban Continuum Codes (2013). <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>. Accessed 2 February 2020.
57. R Core Team, *R: A Language and Environment for Statistical Computing, Version 1.2.5033* (R Foundation for Statistical Computing, Vienna, Austria, 2018).
58. H. Wickman, *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).