Check for updates

SOCIAL SCIENCES

Evaluating bias and noise induced by the U.S. Census Bureau's privacy protection methods

Christopher T. Kenny[1]†, Cory McCartan[2]†, Shiro Kuriwaki[3], Tyler Simko[1], Kosuke Imai[1,4]*

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

The U.S. Census Bureau faces a difficult trade-off between the accuracy of Census statistics and the protection of individual information. We conduct an independent evaluation of bias and noise induced by the Bureau's two main disclosure avoidance systems: the TopDown algorithm used for the 2020 Census and the swapping algorithm implemented for the three previous Censuses. Our evaluation leverages the Noisy Measurement File (NMF) as well as two independent runs of the TopDown algorithm applied to the 2010 decennial Census. We find that the NMF contains too much noise to be directly useful without measurement error modeling, especially for Hispanic and multiracial populations. TopDown's postprocessing reduces the NMF noise and produces data whose accuracy is similar to that of swapping. While the estimated errors for both TopDown and swapping algorithms are generally no greater than other sources of Census error, they can be relatively substantial for geographies with small total populations.

INTRODUCTION

Population statistics produced by the U.S. Census serve as the basis for many consequential public policy decisions, such as legislative redistricting and the disbursement of over $675 billion in federal funds (*1*, *2*). At the same time, the U.S. Census Bureau is legally obligated to protect the confidentiality of individual responses to the Census questionnaires (US Code Title 13). This means that the Bureau faces a difficult trade-off between data accuracy and privacy protection.

The Bureau introduced a new disclosure avoidance system (DAS) for the 2020 decennial Census named TopDown, based on a mathematical definition of privacy called differential privacy (*3*). The TopDown algorithm first injects mean-zero, independent random noise to tabulations of the confidential dataset (the Census Edited File or CEF), which contains individual information about every enumerated resident in the United States. The resulting Noisy Measurement File (NMF) is then postprocessed to ensure that the final census data satisfy a set of requirements for straightforward data use; for example, all the resulting tabulations are non-negative integers and are consistent across different census geographic levels.

The TopDown algorithm in the 2020 decennial Census replaced the old DAS, which was based on a swapping algorithm and used in the 1990, 2000, and 2010 Censuses (*4*). The swapping algorithm randomly selects a household record in a small census block and interchanges it with a similar record in another block. Households with higher disclosure risk, because of their unique attributes or the small population of their block, are swapped with higher probability.

This change to the DAS led to debates among scholars and stakeholders regarding the appropriate balance between data accuracy and privacy protection (*2*, *5*–*11*). While the Bureau has released a series of demonstration datasets and error metrics of the TopDown algorithm, the NMF was not released until April 2023 (*12*, *13*). The Bureau also has yet to release metrics that directly compare the TopDown error to the NMF and swapping algorithm errors.

We provide an independent assessment of the bias and noise induced by the TopDown algorithm and the swapping algorithm, relative to the CEF. Although we cannot directly observe the CEF, we use the known statistical properties of the noise added to the CEF to evaluate bias and uncertainty of the entire TopDown algorithm. Averaging the discrepancies between the TopDown and NMF statistics over a set of geographies forms an unbiased estimator of the TopDown bias, because the NMF is an unbiased estimate of the CEF itself. We compare the bias and noise caused by the swapping algorithm, the TopDown algorithm, and the noise induced in the NMF before postprocessing. Errors from TopDown can be put in context with the estimated sizes of coverage and nonsampling errors by Census researchers (*14*).

Because this approach is fundamentally statistical, we are limited to estimating the average bias, i.e., the errors in the counts averaged over thousands or millions of census geographies. Such aggregation may mask biases that potentially exist in some census geographies. However, our estimates of the TopDown and swapping noise, as measured by their root mean square error (RMSE), capture important distributional information about these errors that can be useful to practitioners in evaluating the reliability of census statistics. We also study how the noise and bias depend on the overall population of census geographies, providing additional insight into how the DAS-induced errors may vary across these geographies.

Our evaluation is possible because of the fortuitous availability of three distinct data sources based on the 2010 decennial Census. The first is the NMF and corresponding TopDown output (including postprocessing), contained in the 2010 Redistricting and DHC Production Settings Demonstration Data published in April 2023 (*15*). Second, we use the official release of the 2010 decennial census, which represents the output of the swapping algorithm. We transform the information in the NMF to match the traditional Census geographies in the official census. Third, we also use another 2010 Demonstration Data (June 2021 vintage) released in August 2021 (*16*). This demonstration data is based on a different run of the TopDown algorithm, providing us with an additional independent output of the new DAS [The Bureau had initially refused to release the NMF but recently made it publicly available. Unfortunately, the Bureau deleted the NMF used to generate the August 2021

[1]Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138, USA. [2]Center for Data Science, New York University, New York, NY 10011, USA. [3]Department of Political Science and Institution for Social and Policy Studies, Yale University, New Haven, CT 06511, USA. [4]Department of Statistics, Harvard University, Cambridge, MA 02138, USA.
*Corresponding author. Email: imai@harvard.edu
†These authors contributed equally to this work.

Downloaded from https://www.science.org on May 01, 2024

Demonstration Data (*17*). This means that the April 2023 Demonstration Data are based on a new, independent run of the TopDown algorithm.] Because these two releases are statistically independent, we can use them alongside known statistical properties of the noise-injection stage of the TopDown algorithm to quantify the uncertainty of our estimates. The details of our methodology are described in Materials and Methods.

Our key contribution is to evaluate the bias and noise of the TopDown algorithm relative to the CEF: the nonprivatized microdata. Prior work attempting to evaluate the bias induced by the TopDown algorithm has been limited to applying an experimental version of differential privacy (*18*), often to simulated data the researcher creates (*19*). Other work compares the TopDown algorithm against the swapping algorithm only (*2*, *20*). The methodology we develop may therefore also be useful for practitioners who wish to characterize the bias and noise of particular published statistics beyond those considered here.

Our findings, detailed in Results and based on the methods described in Materials and Methods, can be summarized as follows. Overall, both the TopDown and swapping algorithms produce nearly unbiased population and racial group counts, on average. Variance is generally small in absolute terms for most geographies, but errors are substantially larger for certain racial/ethnic groups, particularly for Hispanic and multiracial populations. TopDown's postprocessing reduces the noise in the NMF, and ultimately the variance from swapping appears to be roughly similar to the variance from TopDown. These patterns hold across census geographies with varying population sizes and racial diversity. In general, the TopDown and swapping errors are no greater than other sources of errors, such as undercounting and missing responses. For small-population geographies, errors from both TopDown and swapping can be large relative to the total population as well as other sources of error, warranting caution in conducting statistical analyses which rely on these small counts.

MATERIALS AND METHODS

In this section, we describe our methodology used to evaluate the bias and uncertainty of the TopDown and swapping algorithms. This methodology may be useful for other researchers who wish to quantify the average bias and variance of a set of statistics published after TopDown or swapping have been applied.

Census and NMF geographies

The Census Bureau partitions the entire geographic area of the United States into census blocks, of which there were 11,078,297 in 2010. As shown in Fig. 1, sets of blocks can be hierarchically grouped together to form a geographic spine. The most familiar geographic spine is the standard Census spine, where blocks belong to block groups, which in turn belong to tracts, counties, and states. The NMF is built around a slightly different NMF spine.

Figure 1 displays the census spines for the standard census geographies and for the NMF geographies. Geographies larger than blocks are split in the NMF according to whether they include American Indian/Alaska Native (AI/AN) areas, in an effort to improve accuracy in these areas. For example, a tract in a typical census release is on the left spine but is broken into AI/AN and non-AI/AN areas on the right spine. Blocks retain a one-to-one correspondence, but each is designated as part of the AI/AN or non-AI/AN spine. Block groups are on-spine within each respective hierarchy, but they represent different geographies. In other words, optimized block groups on the NMF spine do not necessarily correspond to block groups on the traditional census spine. All analyses herein use the released block groups, not optimized block groups. Note that block groups, the geography in Census releases, are distinct from optimized block groups used in TopDown (see www.census.gov/about/training-workshops/2021/2021-07-01-das-presentation.pdf for further details). As a result, we refer to block groups as "off-spine" for TopDown, even though block groups are on the traditional Census spine.

We transformed the NMF into traditional Census geographies, such as block groups and voting districts (VTDs), by building a mapping from the unique identifiers of NMF on-spine geographies, geocodes, to traditional Census GEOID identifiers. We describe this cross-walking process in detail in the Supplementary Materials, section S1.

Setup and notation

We use Γ to denote the set of all geographies on the NMF spine—blocks, AI/AN tracts, non-AI/AN counties, etc. The CEF can be equivalently represented as a set of histograms, $\{\mathbf{x}_\gamma\}_{\gamma\in\Gamma}$, one for every geography $\gamma \in \Gamma$, which count the number of people in that geography with every possible combination of measured attributes. For the redistricting NMF, the total number of bins for each histogram is $2016 = (2 \text{ voting-age categories}) \times (2 \text{ Hispanic categories}) \times (63 \text{ race categories}) \times (8 \text{ housing types})$.

In our analysis, we regard the CEF as the ground truth and estimate average bias and RMSE relative to it. The detailed histograms $\{\mathbf{x}_\gamma\}_{\gamma\in\Gamma}$ can be aggregated to a set S of statistics of interest $\{\mathbf{y}^*_\gamma\}_{\gamma\in\Gamma}$, like the total voting-age population, or Black population, using a $|S| \times 2016$ aggregation matrix A

$$\mathbf{y}^*_\gamma := A\mathbf{x}_\gamma$$

However, simply aggregating the confidential CEF $\{\mathbf{x}_\gamma\}_{\gamma\in\Gamma}$ does not necessarily provide any privacy guarantee to respondents. To protect respondent confidentiality, the Bureau has developed disclosure avoidance methods, which range from simple (suppressing parts of $\{\mathbf{y}^*_\gamma\}_{\gamma\in\Gamma}$) to complex. We examine the TopDown algorithm, developed for the 2020 Census, and compare its properties with the swapping method, used for the 1990, 2000, and 2010 Censuses.

TopDown algorithm

To produce a differentially private version of $\{\mathbf{y}^*_\gamma\}_{\gamma\in\Gamma}$, the Bureau applies the TopDown algorithm, which proceeds in two main steps.

First, in each geography $\gamma \in \Gamma$, noisy measurements are made of $\mathbf{x}_\gamma$ through a series of queries encoded in a known query matrix Q

$$\tilde{\mathbf{M}}_\gamma := Q\mathbf{x}_\gamma + \eta_\gamma, \quad \eta_\gamma \sim N_Z(0, \Sigma_\gamma)$$

where N_Z is the discrete Gaussian distribution, and $\Sigma_\gamma = \text{diag}(\varsigma^2_\gamma)$ is a diagonal covariance matrix, with each query's variance $\varsigma^2_{\gamma j}$ determined by a privacy-loss budget schedule set in advance by the Bureau. It is this noisy measurement step which provides zero-Concentrated Differential Privacy guarantees (*21*).

Second, the noisy measurements are hierarchically postprocessed using a multipass optimization and rounding routine. The postprocessing ensures that published statistics are consistent across and within geographies and enforces policy constraints. For example, these constraints ensure that all statistics will be non-negative and that the

Downloaded from https://www.science.org on May 01, 2024

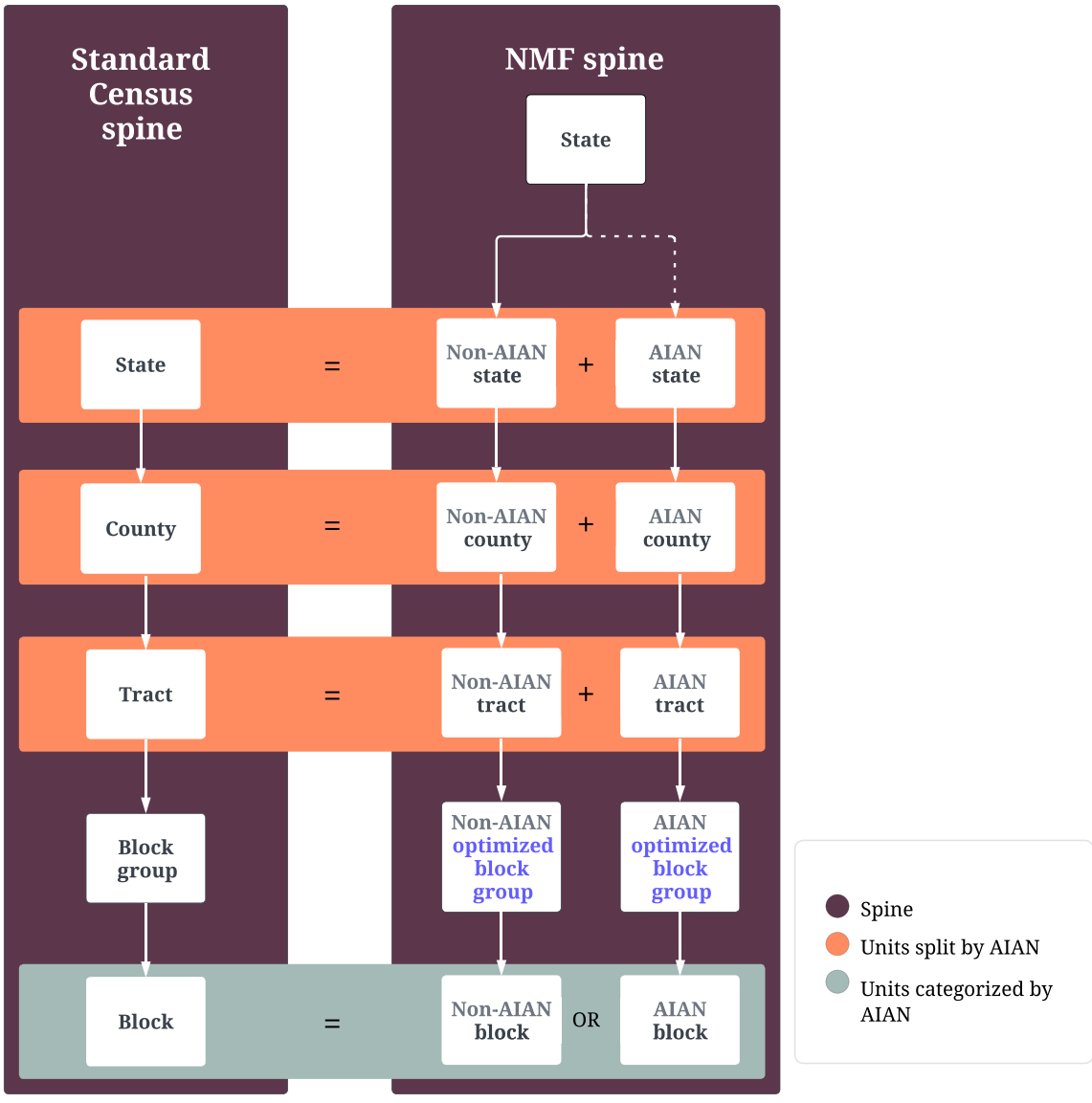

Fig. 1. The Census geographic hierarchy (spines). Spines for standard census geographies and for the hierarchy in the NMF. Higher units indicate enclosing units. In the right spine, substate geographic units are split into American Indian/Alaska Native (AI/AN) and non-AI/AN portions. The dotted arrow indicates that not all states have an AI/AN segment. The teal area indicates that the units can be matched one-to-one across spines. For example, a single block is never split into a non-AI/AN versus AI/AN fragment, so a block from the NMF spine matches 1:1 to a block in the standard spine.

published state population totals exactly match the population totals from the CEF. The study in (*3*) presents more information on the postprocessing including the justification for these constraints.

For our purposes, we can write the output of the postprocessing step as a new set of noise-infused microdata

$$\{\mathbf{X}_\gamma^{\text{td}}\}_{\gamma\in\Gamma} = g(\{(\tilde{\mathbf{M}}_\gamma, \mathbf{x}_\gamma)\}_{\gamma\in\Gamma})$$

where the dependence on the CEF counts $\mathbf{x}_\gamma$ is necessary because certain constraints are functions of the original microdata. The new microdata can be aggregated, yielding differentially private statistics

$$\mathbf{Y}_\gamma^{\text{td}} \coloneqq A\mathbf{X}_\gamma^{\text{td}}$$

for every geography $\gamma \in \Gamma$. We write the error produced by the Top-Down algorithm as

$$\varepsilon_\gamma^{\text{td}} \coloneqq \mathbf{Y}_\gamma^{\text{td}} - \mathbf{y}_\gamma^* = A(\mathbf{X}_\gamma^{\text{td}} - \mathbf{x}_\gamma)$$

The postprocessed statistics $\mathbf{Y}^{\text{td}}$ cannot be directly compared to the noisy measurements $\tilde{\mathbf{M}}$, because the latter are at the query level rather than the statistic level. This means that a single statistic, like the total population, may be computed separately from multiple different queries, each producing a different answer. Fortunately, using the workflow described in (*22*), it is possible to combine the noisy measurements from all the queries to form unbiased, minimum-variance estimates of $\mathbf{y}_\gamma^*$. This is accomplished by aggregating the

Downloaded from https://www.science.org on May 01, 2024

noisy measurements to form multiple noisy estimates of each statistic and then combining these multiple estimates with an inverse-variance–weighted mean.

Formally, we define an aggregation-weighting matrix $B(\Sigma)$ which performs both these steps at once, yielding

$$\mathbf{Y}^{\text{nm}}_{\gamma} \coloneqq B(\Sigma_{\gamma})\tilde{\mathbf{M}}_{\gamma} = B(\Sigma_{\gamma})Q\mathbf{x}_{\gamma} + B(\Sigma_{\gamma})\eta_{\gamma}$$

for every geography $\gamma \in \Gamma$. The noisy measurement error

$$\varepsilon^{\text{nm}}_{\gamma} \coloneqq \mathbf{Y}^{\text{nm}}_{\gamma} - \mathbf{y}^{*}_{\gamma} = B(\Sigma_{\gamma})\eta_{\gamma}$$

therefore has the following approximately known distribution

$$\varepsilon^{\text{nm}}_{\gamma} \sim N\left[0, B(\Sigma_{\gamma})^{\top}\Sigma_{\gamma}B(\Sigma_{\gamma})\right]$$

independently across $\gamma \in \Gamma$. We denote the variance of each statistic in $\mathbf{Y}^{\text{nm}}_{\gamma}$ by $\sigma^2_{\gamma} = \text{diag}\,[B(\Sigma_{\gamma})^{\top}\Sigma_{\gamma}B(\Sigma_{\gamma})]$. The distribution would be exact but for discreteness in η_{γ}. However, for most queries and geographies, the query variance $\varsigma^2_{\gamma j}$ is relatively large, and so after combining multiple queries together the discreteness artifacts are minimal. The values of $\varsigma^2_{\gamma j}$ are known from the specifications of the TopDown algorithm, and we use the values provided in the NMF.

Swapping algorithm

Swapping randomly moves records in the CEF from one geography to another at a certain, but unknown, rate, which depends on a set of confidential rules applied to the records. While opaque, swapping was applied to the 2010 census in a way that provides several benefits for end users. First, swapping preserves the total and voting-age population of each census block (*23*). Second, most swaps occur within small areas, so at coarser geographic scales, most data are close to exact (*4*).

Here, we represent swapping as an unknown function applied to the CEF and a generic source of randomness $\mathbf{U}$, which can then be aggregated in the same was as the CEF to produce publishable statistics

$$\{\mathbf{X}^{\text{sw}}_{\gamma}\}_{\gamma\in\Gamma} = f(\{\mathbf{x}_{\gamma}\}, \mathbf{U})$$

$$\mathbf{Y}^{\text{sw}}_{\gamma} \coloneqq A\mathbf{X}^{\text{sw}}_{\gamma} \text{ for } \gamma \in \Gamma$$

We write the error produced by swapping as

$$\varepsilon^{\text{sw}}_{\gamma}\mathbf{Y}^{\text{sw}}_{\gamma} - \mathbf{y}^{*}_{\gamma} = A(\mathbf{X}^{\text{sw}}_{\gamma} - \mathbf{x}_{\gamma})$$

Independence relations among the Census-induced errors

Our goal is to construct estimators of the average bias and variance of the postprocessed data $\mathbf{Y}^{\text{td}}_{\gamma s}$ for some set of geographies $I \subseteq \Gamma$ and some statistic $s \in \mathcal{S}$. We then would like to compare these properties with those of the swapping data $Y^{\text{sw}}_{\gamma s}$. To do so, it will be useful to state several independence relations among the various Census-induced errors defined above. Throughout the paper, all randomness comes from the Bureau-injected noise $\{\eta_{\gamma}\}_{\gamma\in\Gamma}$ and $\mathbf{U}$. The underlying CEF counts $\{\mathbf{x}_{\gamma}\}_{\gamma\in\Gamma}$ are treated as fixed. The proofs of all results appear in the Supplementary Materials, section S2.

First, some but not all of the errors are independent of each other.

Proposition 2.1. *For a single run of the TopDown Algorithm, the following independence relations hold:*

a. $\{\varepsilon^{\text{nm}}_{\gamma}\}_{\gamma\in\Gamma}$ *are mutually independent, and*

b. $\{\varepsilon^{\text{nm}}_{\gamma}\}_{\gamma\in\Gamma}$ *and* $\{\varepsilon^{\text{sw}}_{\gamma}\}_{\gamma\in\Gamma}$ *are independent.*

Note that $\{\varepsilon^{\text{nm}}_{\gamma}\}_{\gamma\in\Gamma}$ and $\{\varepsilon^{\text{td}}_{\gamma}\}_{\gamma\in\Gamma}$ are not independent because the latter is a function of the former. We expect the two sets of statistics to be positively correlated.

Furthermore, we take advantage of the fact that the Bureau released another set of the postprocessed demonstration data $\tilde{Y}^{\text{td}}_{\gamma s}$ based on a different, independent, run of the TopDown Algorithm. Although the Bureau did not release the NMF used to produce $\tilde{Y}^{\text{td}}_{\gamma s}$, we know that the released NMF, and hence $\varepsilon^{\text{nm}}_{\gamma}$, is independent of $\tilde{Y}^{\text{td}}_{\gamma s}$.

Proposition 2.2. *Let* $\tilde{Y}^{\text{td}}_{\gamma s}$ *be the second set of postprocessed demonstration data, with corresponding noisy measurements* $\{\tilde{\eta}_{\gamma}\}_{\gamma\in\Gamma}$ *and errors* $\{\tilde{\varepsilon}^{\text{td}}_{\gamma}\}_{\gamma\in\Gamma}$. *Then,*

a. $\{\varepsilon^{\text{nm}}_{\gamma}\}_{\gamma\in\Gamma}$ *and* $\{\tilde{\varepsilon}^{\text{td}}_{\gamma}\}_{\gamma\in\Gamma}$ *are independent, and*

b. $\{\varepsilon^{\text{td}}_{\gamma}\}_{\gamma\in\Gamma}$ *and* $\{\tilde{\varepsilon}^{\text{td}}_{\gamma}\}_{\gamma\in\Gamma}$ *are independent.*

Proposed estimators for TopDown bias and variance

We wish to estimate the average bias in the TopDown statistics $Y^{\text{td}}_{\gamma s}$ for some set of geographies $I \subseteq \Gamma$ (e.g., census tracts within a county) and some statistic s (e.g., the total number of voting-age Hispanics). When s is the total population, the true CEF counts $Y_{\gamma s}$ are known exactly because of swapping's population invariance. Therefore, Fig. 2 shows the exact distribution of $|\varepsilon^{\text{td}}_{\gamma s}|$.

Average bias

When s is the population of racial groups as in Figs. 3 to 5, the true CEF counts are unknown so we must estimate the bias from comparing the TopDown and swapping counts to the NMF. The average bias of TopDown across these geographies may be written as

$$\mu^{\text{td}}_{Is} = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[\varepsilon^{\text{td}}_{\gamma s}\right]$$

where the average bias in the postprocessing is the same as the average bias of the entire TopDown algorithm, because the NMF adds mean-zero noise and has no bias.

We propose the following simple unweighted estimator of the average bias

$$\hat{\mu}^{\text{td}}_{Is} \coloneqq \frac{1}{|I|}\sum_{\gamma\in I}\left(Y^{\text{td}}_{\gamma s} - Y^{\text{nm}}_{\gamma s}\right)$$

Subtracting the noisy statistic from the postprocessed statistic cancels the unknown true value, so that the estimator is unbiased

$$\mathbb{E}[\hat{\mu}^{\text{td}}_{Is}] = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}[Y^{\text{td}}_{\gamma s} - Y^{\text{nm}}_{\gamma s}]$$

$$= \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}[\varepsilon^{\text{td}}_{\gamma s} - \varepsilon^{\text{nm}}_{\gamma s}] = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}[\varepsilon^{\text{td}}_{\gamma s}] = \mu^{\text{td}}_{Is}$$

Unfortunately, the variance of this estimator is not identifiable because of the unknown correlation between $\varepsilon^{\text{td}}_{\gamma s}$ and $\varepsilon^{\text{td}}_{\gamma' s}$. Therefore, we propose using a similar bias estimator based on an independent set of postprocessed counts $\tilde{Y}^{\text{td}}_{\gamma s}$ instead of $Y^{\text{td}}_{\gamma s}$

$$\hat{\tilde{\mu}}^{\text{td}}_{Is} \coloneqq \frac{1}{|I|}\sum_{\gamma\in I}\left(\tilde{Y}^{\text{td}}_{\gamma s} - Y^{\text{nm}}_{\gamma s}\right)$$

Downloaded from https://www.science.org on May 01, 2024

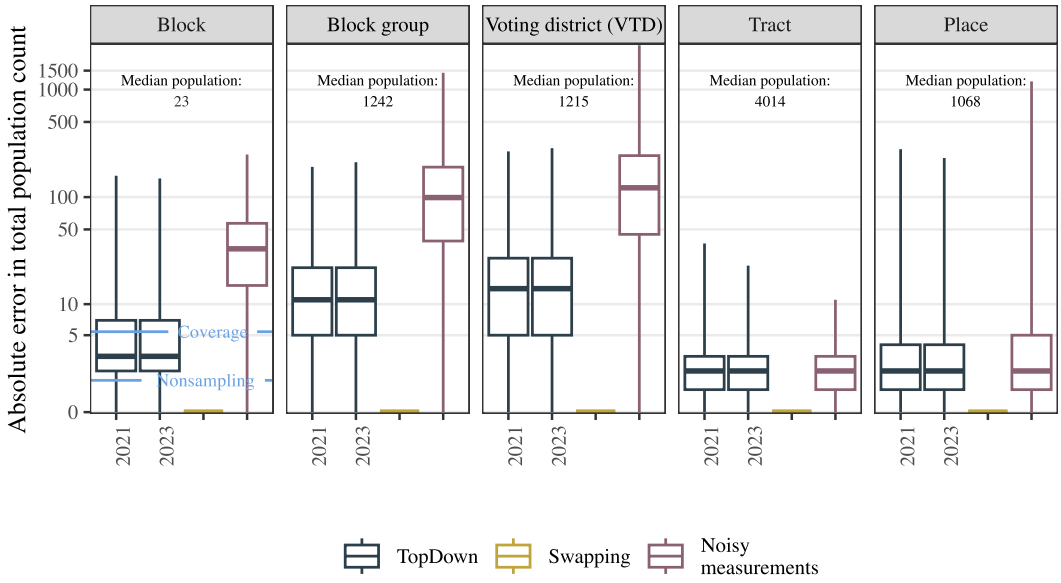

Fig. 2. Distribution of absolute count error in total population. The figure displays the absolute error in total populations as enumerated in the CEF. The *y* axis is shown on a pseudo-$\log_{10}$ scale. Each panel depicts a geographical level where a boxplot shows the nationwide distribution of population error at that geographical level, with horizontal lines for the first quartile, median, and third quartile. Swapping errors are always zero for census geographies, as total population remains invariant. The Bureau-estimated mean absolute error due to coverage and nonsampling errors at the block level are indicated on the leftmost pattern by horizontal lines. Coverage and nonsampling error affect the CEF and thus are present in addition to DAS errors from TopDown, swapping, or noisy measurements.

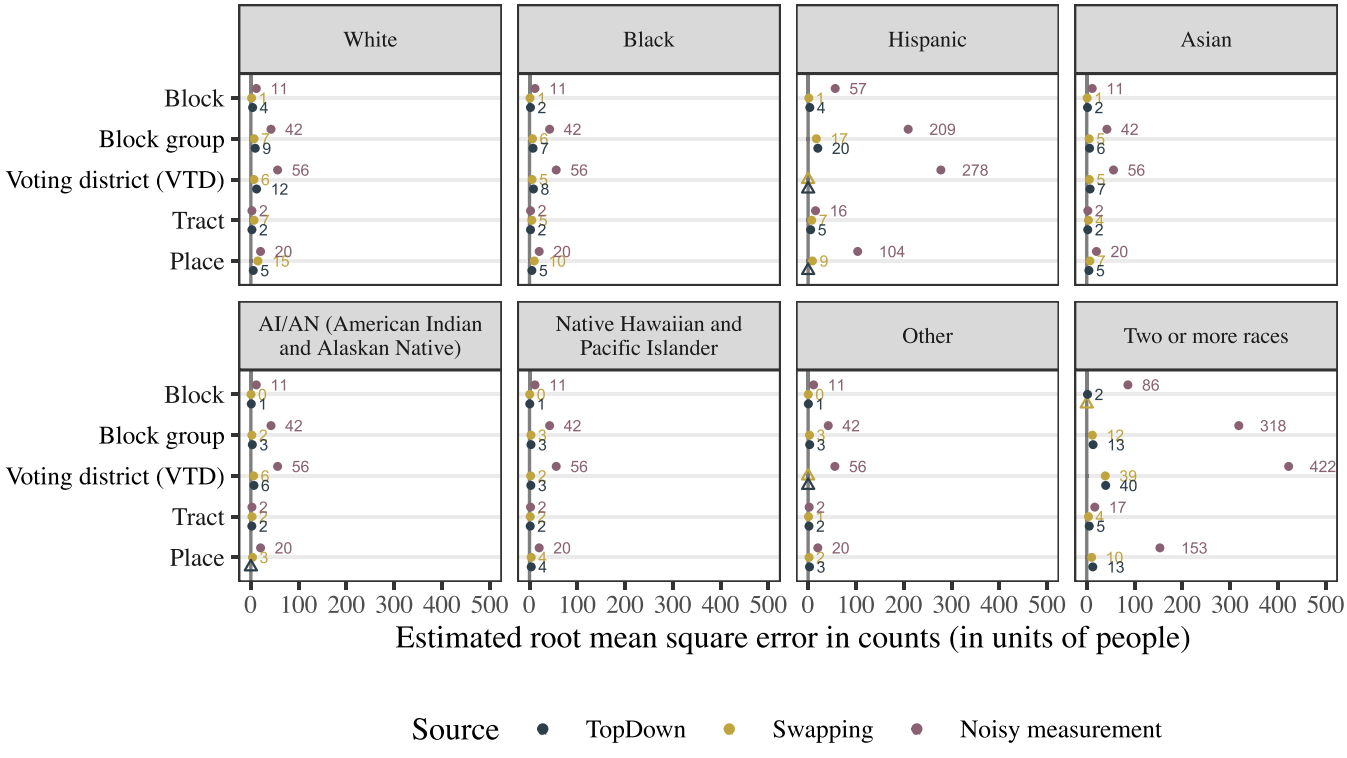

Fig. 3. Estimated RMSE for population counts of a race/ethnicity group, at each geographic level. The RMSE quantifies the average magnitude of error for a given geography for a particular geographic unit (see Materials and Methods for estimators). Triangles for RMSE indicate that the estimated mean square error was negative and hence was set to zero.

Downloaded from https://www.science.org on May 01, 2024

This estimator is similarly unbiased, i.e., $\mathbb{E}[\hat{\tilde{\mu}}^{td}_{Is}] = \mu^{td}_{Is}$, but it has an important advantage over the other estimator $\hat{\mu}^{td}_{Is}$—its variance is identifiable because of the independence with the noisy measurements [Proposition 2.2(a)]. The following proposition introduces an unbiased estimator of this variance:

Proposition 2.3. *Define the following variance estimator for the TopDown average bias estimator of $s \in \mathcal{S}$ across geographies $I \subseteq \Gamma$:*

$$\hat{V}\left(\hat{\tilde{\mu}}^{td}_{Is}\right) := \frac{1}{2|I|^2}\left(\sum_{\gamma\in I} Y^{td}_{\gamma s} - \tilde{Y}^{td}_{\gamma s}\right)^2 + \frac{1}{|I|^2}\sum_{\gamma\in I}\sigma^2_{\gamma s}$$

Then $\hat{V}(\hat{\mu}^{td}_{Is})$ is unbiased

$$\mathbb{E}\left[\hat{V}\left(\hat{\tilde{\mu}}^{td}_{Is}\right)\right] = \mathbb{V}\left[\hat{\tilde{\mu}}^{td}_{Is}\right]$$

Average variance

In addition to the average bias in the TopDown counts $Y^{td}_{\gamma s}$, it is also possible to obtain an unbiased estimate of their average variance for some set of geographies $I \subseteq \Gamma$ and some statistic s. Unlike with the bias estimation, however, we can no longer form an unbiased estimate with the TopDown counts $Y^{td}_{\gamma s}$ alone. Instead, we must use $\tilde{Y}^{td}_{\gamma s}$, which are independent of $Y^{nm}_{\gamma s}$. Our proposed estimator is

$$\hat{V}^{td}_{Is} := \frac{1}{2|I|}\sum_{\gamma\in I}\left(Y^{td}_{\gamma s} - \tilde{Y}^{td}_{\gamma s}\right)^2$$

This estimator is unbiased for $V^{td}_{Is} := |I|-1\sum_{\gamma\in I}\mathrm{V}[\varepsilon^{td}_{\gamma s}]$ according to the proof of Proposition 2.3. We use this estimator to compute approximate confidence intervals in Fig. 4. Note that the confidence intervals for the total population column in Fig. 4 is estimated differently, because the unit-level error of those from a draw of the TopDown algorithm are known exactly. We have two TopDown draws, and each provides one estimate. For each population bin, we use the average of these two draws as our point estimate and compute its SE as $\hat{V}(\hat{\mu}^{td}) = \frac{1}{2}|Y^{td} - Y^{nm}|$.

Mean square error

Last, we can obtain an unbiased estimate of the mean square error (MSE) for TopDown counts, which is defined as

$$\mathrm{MSE}^{td}_{Is} = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[\varepsilon^{td^2}_{\gamma s}\right]$$

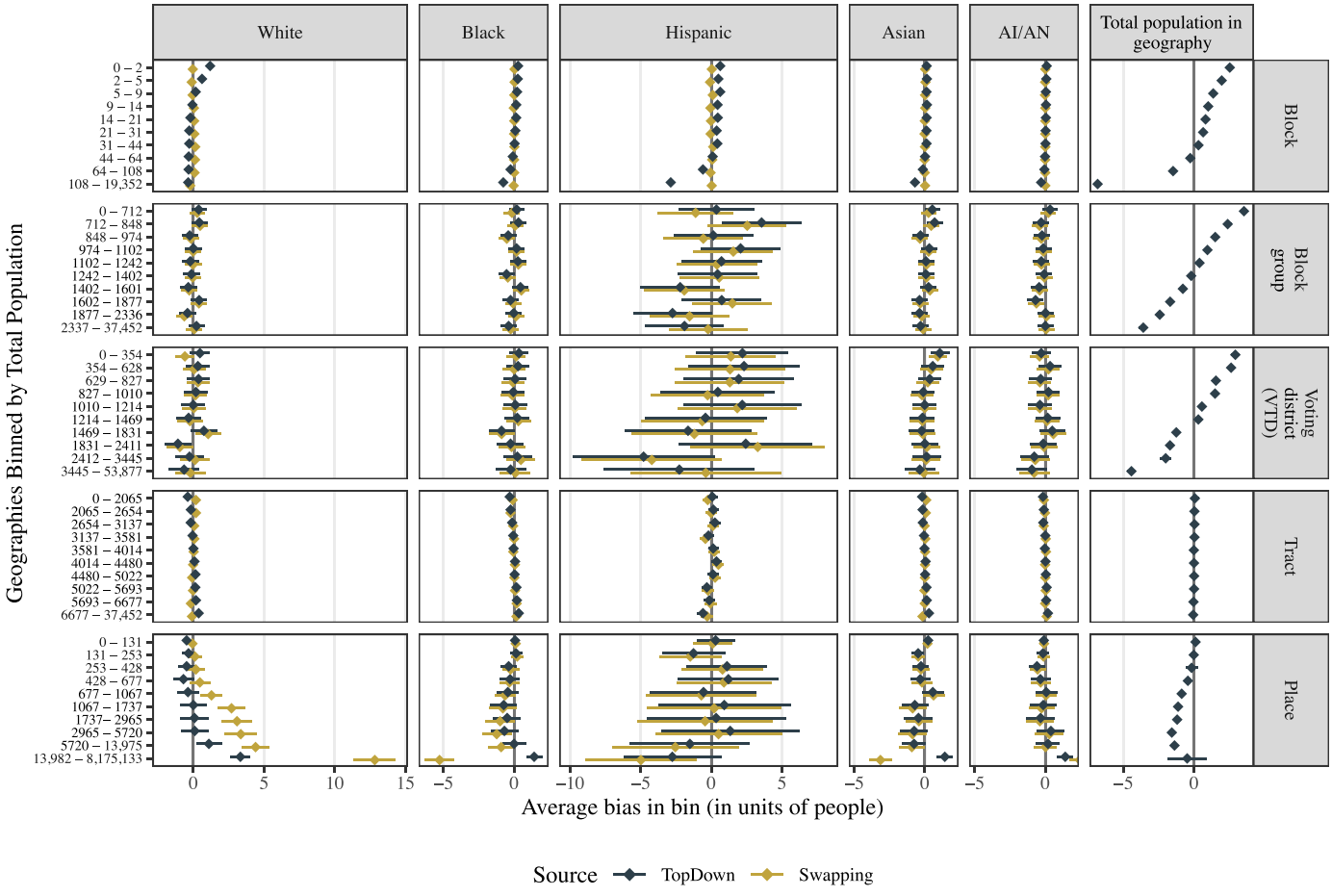

Fig. 4. Average bias for race/ethnicity population counts at each geographic level by its total population. The figure estimates the average overcounting or undercounting in a group of geographies, separately for five geographic levels and five race/ethnicity groups. See Materials and Methods for estimators. Bins on the *y* axis are deciles of total population of the geographic level points show the estimated bias, and lines show estimated 95% confidence intervals.

Downloaded from https://www.science.org on May 01, 2024

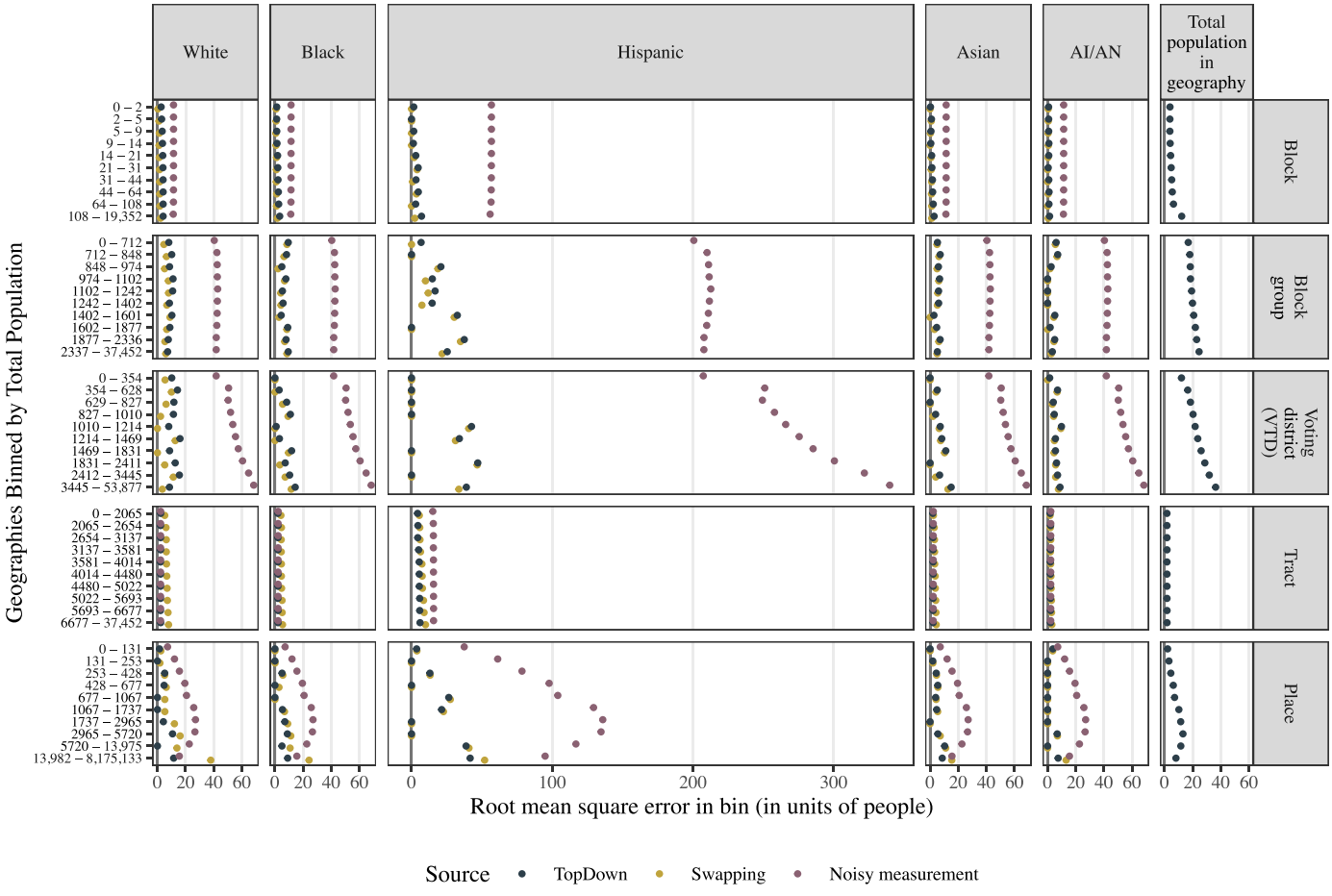

Fig. 5. Estimated RMSE of race/ethnicity counts at each geographic level, by its total population. Unlike Fig. 3, it estimates RMSE for a subset of geographies. See Materials and Methods for estimators. Bins on the *y* axis are deciles by total population of the geographic level.

and represents the combination of bias squared and variance. We propose the following estimator

$$\widehat{\mathrm{MSE}}^{\mathrm{td}}_{Is} := \frac{1}{|I|}\sum_{\gamma\in I}\left[\left(\tilde{Y}^{\mathrm{td}}_{\gamma s} - Y^{\mathrm{nm}}_{\gamma s}\right)^2 - \sigma^2_{\gamma s}\right]$$

which is unbiased for the MSE

$$\mathbb{E}\left[\widehat{\mathrm{MSE}}^{\mathrm{td}}_{Is}\right] = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[\left(\tilde{Y}^{\mathrm{td}}_{\gamma s} - Y^{\mathrm{nm}}_{\gamma s}\right)^2\right] - \sigma^2_{\gamma s}$$
$$= \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[\left(\tilde{\varepsilon}^{\mathrm{td}}_{\gamma s}\right)^2\right] + \sigma^2_{\gamma s} - \sigma^2_{\gamma s} = \mathrm{MSE}^{\mathrm{td}}_{Is}$$

In practice, this estimator may take a negative value due to a high degree of estimation noise. In such cases, we replace a negative estimate with zero in our analysis. This occurs in 4 of the 40 TopDown estimates in Fig. 3 but occurs in none of the TopDown estimates in Fig. 5. We take the square root of this estimator to obtain the estimated RMSE in Figs. 3 and 5.

Proposed estimators for swapping bias and variance

Estimating the average bias of the swapped counts $Y^{\mathrm{sw}}_{\gamma s}$ is also straightforward, and we can simply apply the same strategy as the one used for the TopDown algorithm.

$$\mathbb{E}\left[\hat{\mu}^{\mathrm{sw}}_{Is}\right] = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[Y^{\mathrm{sw}}_{\gamma s} - Y^{\mathrm{nm}}_{\gamma s}\right]$$
$$= \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[\varepsilon^{\mathrm{sw}}_{\gamma s} - \varepsilon^{\mathrm{nm}}_{\gamma s}\right] = \frac{1}{|I|}\sum_{\gamma\in I}\mathbb{E}\left[\varepsilon^{\mathrm{sw}}_{\gamma s}\right] = \mu^{\mathrm{sw}}_{Is}$$

Unfortunately, the valid SE is difficult to obtain because, unlike the case of the TopDown algorithm, we do not have access to an independent second draw. To address this unidentifiability, we propose the following estimator, which is conservative under an additional assumption.

Proposition 2.4. *Define the following variance estimator for the swapping average bias estimator of $s \in S$ across geographies $I \subseteq \Gamma$*

$$\hat{V}(\hat{\mu}^{\mathrm{sw}}_{Is}) := \frac{1}{|I|^2}\sum_{\gamma\in I}(Y^{\mathrm{sw}}_{\gamma s} - Y^{\mathrm{nm}}_{\gamma s})^2$$

If $\sum_{\gamma\neq\gamma'\in I}\mathrm{Cov}(\varepsilon^{\mathrm{sw}}_{\gamma s}, \varepsilon^{\mathrm{sw}}_{\gamma' s}) \leq 0$, *then*

$$\mathbb{E}\left[\hat{V}(\hat{\mu}^{\mathrm{sw}}_{Is})\right] \geq \mathbb{V}[\hat{\mu}^{\mathrm{sw}}_{Is}]$$

The assumed weakly negative correlation across geographies is plausible, because errors in one geography must be matched by opposite errors in other geographies for the error at the higher

Downloaded from https://www.science.org on May 01, 2024

geographic level to be controlled. The lack of detailed information about the swapping algorithm, however, makes it difficult to know whether this additional assumption holds.

We can estimate the MSE of the swapped counts $Y^{sw}_{\gamma s}$ similarly as with the TopDown counts, using the following estimator

$$\widehat{\mathrm{MSE}}^{sw}_{Is} := \frac{1}{|I|}\sum_{\gamma\in I}\left(Y^{sw}_{\gamma s} - Y^{nm}_{\gamma s}\right)^2 - \sigma^2_{\gamma s}$$

which is unbiased for MSE^{sw}_{Is} by an identical argument to that for the unbiasedness of $\widehat{\mathrm{MSE}}^{td}_{Is}$. Unfortunately, like the case of the TopDown algorithm, this estimate may take a negative value in practice. This occurs in 3 of the 40 Swapping estimates in Fig. 3 but occurs in none of the TopDown estimates in Fig. 5. Again, a negative value will be replaced with zero in our analysis.

RESULTS

Absolute errors in total population counts are small on average

We first examine errors in the total population counts in different census geographies. To do so, we compare counts from the postprocessed and NMFs against the counts from the swapping algorithm in the original 2010 decennial Census release. The Bureau's particular swapping algorithm preserved a geographic unit's total population from the CEF, because it swapped households with the "same number of adults [and] same number of children" [(*24*), p. 518]. Therefore, counts for total population released after swapping serves as the ground truth against which we measure TopDown algorithm's error exactly (*25*). No statistical assumptions are necessary to compute this error.

We reconstructed the NMF to create a national dataset for five census geographies: blocks, (tabulation) block groups, tracts, VTDs, and census places (see Materials and Methods). Total population quantiles for these geographies among populated geographic units are reported in Table 1. Blocks, block groups, and tracts are statistical units used by the Census Bureau, while VTDs generally represent voting precincts and census places typically represent towns or cities.

A challenge in comparing these counts to the NMF is that geographic hierarchies, or spines in Census terminology, are defined differently in the NMF than in the standard Census data (see Fig. 1). We therefore parsed the NMF geography identifiers and transformed them into their regular Census spine equivalents. This results in several modifications, detailed in Materials and Methods. For example, the NMF treats portions of a tract that are designated as AI/AN separately from the rest of the tract by assigning it different degrees of privacy-protecting noise. In these cases, we combine the estimates of both the AI/AN and non-AI/AN fragments to a single tract-level NMF estimate. Thus, the estimates we use match the geographies for which traditional census data is released.

When multiple estimates are produced for each geography in the NMF, we combine them by following the method of (*22*). We expect that on-spine geographies, i.e., geographies part of the NMF spine (blocks, tracts, and places), will suffer from less bias and noise than off-spine geographies (nonoptimized block groups and VTDs) that can receive more noise injection by virtue of containing fragments of on-spine geographies.

Figure 2 shows the absolute error in counts, that is, the absolute difference between a geographic unit's total population from each algorithm and the corresponding count in the CEF. We find consistently that the NMF has substantially larger errors than TopDown for both small and off-spine geographies. These NMF errors shrink to become comparable to the TopDown errors for larger geographies. The estimated absolute TopDown error for blocks is typically three to seven people after postprocessing, though there is a substantial tail of larger errors. Noisy measurements often have much larger errors, frequently exceeding the size of the median populated block, which is only 23 people. Thus in relative terms, the TopDown errors are on the order of 10% of the typical block's population. Swapping, as noted above, preserves the total population of every census geography as an invariant and so its error is exactly zero for all census geographies. Note that any geography built from whole blocks, which includes all census geographies, will then have zero total population error from swapping (for the hierarchy of census geographies which can be built from whole blocks, see www2.census.gov/geo/pdfs/reference/geodiagram.pdf).

Block groups and VTDs exhibit similar patterns, where the TopDown errors are smaller than the noisy measurement errors. However, tracts show nearly equivalent errors across both TopDown and noisy measurements. Notably, tracts are on-spine geographies, so each tract receives dedicated queries in the noising process. Geographies that are small and off-spine, like block groups and VTDs, have larger errors in the NMF. Absolute errors are typically very small for census places, as displayed in the rightmost panel of Fig. 2. Census places vary greatly—from small incorporated places to New York City—and we explore possible heterogeneity across population size in the "Counts of racial and ethnic groups are accurate under TopDown and swapping" section.

Table 1. Population summary statistics for Census geographies. Summaries across populated geographic units studied in this paper. Blocks are nested in block groups, which are nested in tracts. Place stands for Census Designated Place. For example, the median Census Block is populated by 23 people.

	Block	Block group	Voting district (VTD)	Tract	Place
Minimum	1	1	1	1	1
25th percentile	8	912	734	2,901	331
Median	23	1,242	1,215	4,014	1,068
75th percentile	54	1,726	2,084	5,333	4,038
Maximum	19,352	37,452	53,877	37,452	8,175,133

Downloaded from https://www.science.org on May 01, 2024

In sum, the typical absolute error in total population is small in postprocessed data, while the noisy measurement errors are much greater. These errors are largest in relative terms for small places, especially those which are off-spine geographies. Errors are likely larger for off-spine geographies because they do not benefit from the same efforts to minimize errors that TopDown incorporates for on-spine geographies. Notably, many other geographies used in practice will also be off-spine, including legislative districts, school districts, neighborhoods, and individual Native American reservations.

These estimates of DAS bias are roughly similar in magnitude to the two main sources of non-DAS error—coverage error (omissions and incorrect enumerations) and nonsampling error (such as data entry errors or incorrect locations) at the block level. Nonsampling errors can come from sources like transcription errors and incorrectly locating housing units. Coverage error results from omissions or additional incorrect enumerations; i.e., differences between the target population and population actually counted by the Census. The Bureau has estimated approximate sizes of these errors in originally internal memos at the block and county levels (*14*). At the block level, the mean absolute coverage error is approximately 5.5 people; the mean absolute nonsampling error is 1.5 people [(*14*), table 2, p. 23]. In comparison, the block-level mean absolute error for TopDown is 4.9 and for the noisy measurements is 39.1. This means that the block-level variance in population counts is increased by about 75% because of the new TopDown algorithm (because the TopDown and coverage errors are statistically independent, the new variance is $1.5^2 + 5.5^2 + 4.9^2 \approx 56.5$, versus the non-DAS error variance of $1.5^2 + 5.5^2 \approx 32.5$—a roughly 75% increase).

The share of the total census error which is attributable to privacy protections is highest at the block level, and will decrease for larger geographies. Note that the Bureau has not reported the results of their error simulations for geographies other than counties and census blocks. However, the magnitude of coverage and nonsampling error would increase with population. At the same time, larger Census geographies have smaller DAS error because of the design of the TopDown algorithm.

While the noise added to tabulations to produce the NMF is completely random, we note that the TopDown errors shown in Fig. 2 are nonrandom because of postprocessing. The error produced by one run of the TopDown algorithm for a given geography will typically be similar to the error produced for that same geography on a completely independent run of the TopDown algorithm. This is due to the postprocessing step, which incorporates the complex interplay of population statistics, the structure of the geographic spine, and the constraints and invariants imposed by the Bureau. Table 2 reports the sample correlation in the population error between the two independent runs of the TopDown algorithm (the 2021 and 2023 demonstration data), across geographies, for a given level. Even for typically populous geographic units like Census places, the sample correlation in errors across geographies can be as high as 0.63. Roughly, this means that around $0.63^2 \approx 40\%$ of the variance introduced by the TopDown algorithm is systematic rather than random. This does not appear to be an artifact of the necessary upward bias for geographies with zero population: The pattern remains consistent even among geographies with at least 100 or 1000 people.

Counts of racial and ethnic groups are accurate under TopDown and swapping

Unlike the total population counts, we do not have access to the ground truth counts for each racial and ethnic group because they are not held invariant during swapping. We thus must construct estimators for the average bias and RMSE of both TopDown and swapping, using the unbiased but noisy NMF. The validity of our estimators only relies on the fact that swapping, one run of the NMF and TopDown algorithm, and another run of the NMF and TopDown algorithm are all based on entirely independent sources of randomness. This proposition, formalized in the "Proposed estimators for TopDown bias and variance" section in Materials and Methods, is justified by official documentation by the Bureau.

The RMSE of TopDown counts represents the square root of the combination of the bias (squared) and variance of TopDown. The bias originates from the postprocessing, whereas the variance originates from both the NMF and the post processing. The "Proposed estimators for TopDown bias and variance" section in Materials and Methods shows how we leverage the NMF to construct an unbiased estimator for the MSE.

Figure 3 demonstrates that the estimated RMSEs for TopDown and swapping are small in absolute terms for most race and ethnicity groups across census geographic levels. We find little evidence that TopDown and swapping are substantially different in terms of RMSE. We do, however, caution that the RMSE estimates for TopDown and swapping are particularly noisy because of the high variability of the noisy measurements and lack uncertainty quantification that is available for our average bias estimates.

On the other hand, because the variance of the NMF counts is reported by the Bureau and known exactly, the RMSE of the NMF can be calculated exactly. The results show that the NMF counts are much noisier than the postprocessed TopDown and swapping counts. The errors are particularly substantial for off-spine geographies such as block groups, VTDs, and places as well as for Hispanics and multiracial groups. This is because these groups are a combination of multiple queries (*22*). In the Census classification, Hispanic voters consist of Hispanic whites, Hispanic Blacks, and other racial

Table 2. Correlation of TopDown errors between runs. This table reports the sample correlation between the TopDown errors in the 2021 and 2023 demonstration data across each kind of Census geography. Correlations are calculated for errors in total and voting-age population because the exact counts for those statistics are invariant under swapping.

	Block	Block group	Voting district (VTD)	Tract	Place
Voting-age population	0.280	0.409	0.393	0.104	0.634
Total population	0.449	0.619	0.600	−0.007	0.572

Downloaded from https://www.science.org on May 01, 2024

groups, whereas non-Hispanic whites consist only of one racial group. Computing the Hispanic population in an off-spine geography would therefore involve identifying a set of on-spine geographies that consist of the target off-spine geography and summing the population counts of subgroups that form Hispanics. This summation process across racial categories and geographies leads to a greater variance of the resulting population count (within each on-spine geography, each race query is given the same variance). Postprocessing of the NMF reduces this variance substantially.

Next, we turn to separating out the bias component of the RMSE. The average bias represents the average of biases across geographic units. Because each bias could be either positive or negative, the average bias can mask important variation of biases that may exist in some geographies. For example, although the average of biases across all geographic units in the nation is necessarily zero, that does not mean the bias is negligible for any given unit.

The average bias is likely much smaller than the absolute average bias shown in Fig. 2. We can compute the average bias by leveraging the NMF as a comparison. In addition, we can compute the estimation uncertainty around our average bias estimate by leveraging the fortuitous existence of two draws of the TopDown algorithm (Materials and Methods, "Proposed estimators for TopDown bias and variance" section).

To test whether the bias tends to be positive in some geographies but not others, we classify geographic units into 10 bins of total population and estimate the average bias of a racial group count within that subset of geographic units. Accurate binning of this sort is possible because the total population counts are reported without privacy-induced error under swapping.

The result, Fig. 4, shows that estimated average biases are consistently small across all racial groups, geographic levels, and data sources. Figure S1 in the Supplementary Materials demonstrates these patterns for the race groups that are not shown in Fig. 4. For on-spine geographies (block and tract), average biases are estimated to be consistently near zero in almost all cases. Off-spine geographies (block group, VTD, and place) have estimated average biases with slightly higher magnitudes, but they are rarely statistically distinguishable from zero. These small average biases appear to address patterns found during previous releases of DAS demonstration data [further details on the updates made to the DAS after the April 2021 Demonstration Data are available at (*26*)]. The improvements could be due to increases in the privacy budget or changes to the TopDown algorithm.

Although the magnitude of bias is not large, small racial counts are biased systematically. In the largest Census Places, swapping tends to overcount white people, undercount Hispanic, Black, and Asian people. TopDown has smaller average bias in these geographies.

TopDown does induce a detectable, though again small, systematic pattern of counting bias for total population. It tends to overcount total populations in small blocks, block groups, and VTDs by about two to three people, and undercount them in large blocks, block groups, and VTDs. We computed average biases for total population using the average of two TopDown draws compared with the Swapping algorithm and without relying on the NMF.

We also calculate RMSE by population size bin (Fig. 5 and fig. S3 in the Supplementary Materials). As expected, we find largely constant RMSE for NMF across on-spine geographies (block and tracts). RMSEs for off-spine geographies (block group, VTD, and place) generally stay constant or increase with the size of the population decile. As in Fig. 3, we find consistently larger RMSEs for the Hispanic population.

NMF consistently has the largest errors, but these errors are substantially reduced by postprocessing. While our estimates are necessarily noisy given that the Bureau has only released two independent runs of the TopDown algorithm under the final set of parameters, we find no evidence of large systematic errors. Instead, our findings suggest that postprocessing improves the data accuracy. The resulting TopDown estimates are roughly as variable as those under swapping in nearly all instances. Of course, the postprocessing introduces an intractable dependency structure to the errors, while the NMF errors are independent and have a known distribution. Thus, despite the higher variance, the NMF counts can be used to produce statistically valid downstream estimates.

Further, these small bias and uncertainty estimates are largely constant across population size. We stress that this is unlike other sources of Census error, which Census researchers have found tend to scale with population size (*14*). For example, Census researchers estimate that coverage error in blocks with populations over 1000 have mean absolute deviations dozens of times larger than blocks with one to nine people (14.5 versus 0.4) [(*14*), table 5, p. 15]. In contrast, we estimate that errors from TopDown, NMF, and swapping are approximately constant across all population deciles, as shown in Fig. 5. While errors from NMF tend to be larger than TopDown, swapping, and other sources of error for small geographies, error sizes are roughly constant across population size for on-spine geographies. We find NMF errors in off-spine geographies can broadly increase across population sizes, but even these are generally reduced to near zero for TopDown and swapping estimates.

These findings imply that for large-population geographies at all levels, variation from the TopDown and swapping algorithms is likely much smaller than variation from coverage and nonsampling errors. However, for small-population geographies, where coverage and nonsampling errors are much smaller, variation from the TopDown and swapping algorithms can be much greater relative to both the census errors and the size of the population. For example, the typical RMSE for race counts at the block level is about three people. Because the median populated block has only 23 people, this means that the variation in race counts is on the same order of magnitude as the actual population of each race. In addition, for smaller racial groups, the RMSE can be larger than the population of the racial group itself.

Census data users should thus take note: For small-population geographies, the size of the DAS-introduced error can be large relative to the true CEF counts of interest as well as other census errors. This warning applies not only for the new TopDown algorithm but also for the swapping algorithm used in the three previous decennial censuses.

Accuracy does not substantially vary with racial heterogeneity

Last, we assess whether bias in total population varies by the demographics of the geography, instead of the size of the geography. We focus on the percentage of the geography that identifies as nonwhite. We use this proportion, rather than a direct measure of diversity like the Herfindahl index, to avoid transformations that may introduce bias. This relationship is difficult to assess because racial group counts (unlike total population) themselves are measured with swapping or noisy-measurement error. The above analysis has

Downloaded from https://www.science.org on May 01, 2024

shown that the average magnitude of the noisy-measurement error is an order of magnitude larger than the average noise introduced by the TopDown algorithm.

To overcome this challenge, we fit a flexible nonparametric regression model that accounts for the known noisy measurement error structure in racial group counts using the method in (*27*) as implemented in (*28*). This allows us to properly estimate the average bias in each geography (measured without error) as a function of the percentage of the population that is non-white (which is measured with error). The details about implementation are described in the Supplementary Materials, section S4. We note that the original method in (*27*) does not allow for the between-geography correlation that is present here. However, we expect this correlation to affect the variance, rather than the bias estimates we present. While this method does not provide measures of uncertainty, simple binned averages also yields the same pattern, where undercounting biases in block groups and voting tabulation districts are statistically distinguishable (see fig. S3 in the Supplementary Materials). This approach ignores the noise in the percentage non-white residents, but the noise is on the order of a few percentage points, and if anything would be expected to attenuate the observed pattern, not contribute to it.

Figure 6 shows the fitted local polynomial regression lines for four geographies. For tracts, which is an on-spine geography, the pattern of average bias is essentially flat with no bias. Off-spine geographies exhibit a systematic relationship where geographies that are 30 to 70% non-white are systematically undercounted [consistent with (*2*)]. However, the magnitude of average undercounting is only around one to three people, consistent with earlier results. Recall that the swapping algorithm preserves total population and hence has no error for any census geographical unit.

The modest biases may be due to various changes the Bureau made to the DAS postprocessing procedure and privacy budget throughout feedback received during several "demonstration data" releases. For example, the work in (*26*) documents that the DAS team made changes to the postprocessing system to reduce biases undercounting population in racially and ethnically diverse areas, as documented in (*2*). Earlier work in (*2*) found that population biases likely decreased between iterative demonstration releases of the DAS after the Bureau incorporated several changes from stakeholder feedback.

DISCUSSION

We have shown how to use the recently released NMF to evaluate the bias and noise caused by the Census Bureau's DASs. The primary advantage of our methodology is the ability to characterize the DAS-induced errors relative to the CEF, which is the nonprivatized confidential census microdata. Although the CEF remains private, we are able to evaluate bias and noise by exploiting known statistical properties of the noise induced to create the NMF.

Previous studies, without access to the NMF, were only able to compare the datasets processed by two different DAS: TopDown and swapping algorithms. These past comparisons remain useful for policy evaluation and allow researchers to compare how policy outcomes may change when basing decisions on a new dataset. However, they could not assess the relative magnitude of bias and noise induced by the TopDown and swapping algorithms because their analyses depend on comparisons between versions of the DAS (*25*, *29*).

Our evaluation has shown that the average biases introduced by both the TopDown and swapping algorithms are overall small. The variances of the TopDown and swapping error are also similar in magnitude. These DAS-induced errors are usually smaller than other sources of census error and also tend to be small relative to the population of each geography. These findings can reassure users of Census data that population counts from large-population geographies, like counties, generally include only minor errors induced by both swapping and TopDown.

For small-population geographies like, however, both TopDown and swapping algorithms introduce larger relative errors. We emphasize two points about counts from small-population areas for Census data users to keep in mind. First, estimates from small-population geographies like census blocks have the largest variance among census geographies. We find these deviations can be as large or larger than other sources of census error like coverage and nonsampling error. Second, this variance at small levels can be large relative to the geographies' populations. Consequently, analysts of Census data since 1990 should proceed carefully in analyzing small areas and counts, because they may contain more relative noise from the DAS than previously understood.

The errors in the NMF are substantially larger than that from TopDown and swapping, but they come with statistical guarantees that enable researchers to conduct valid statistical inference relative to the confidential CEF. The estimates presented here were only

Downloaded from https://www.science.org on May 01, 2024

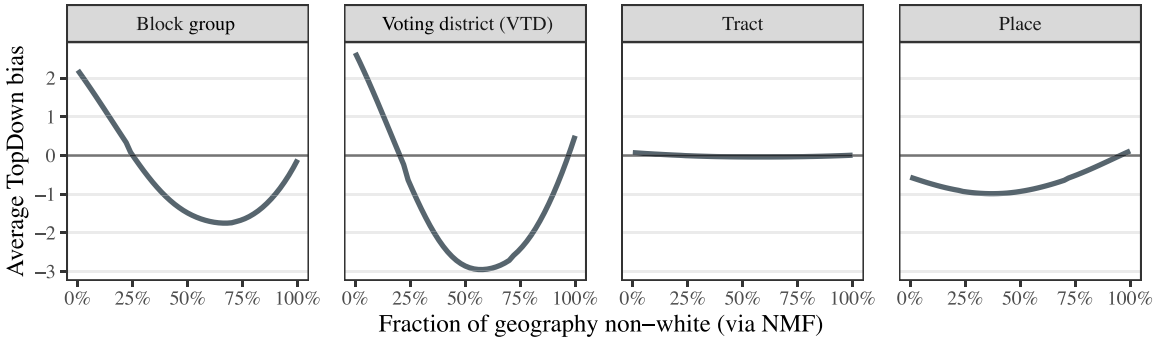

Fig. 6. Average bias in total population count by percent non-white. The figure estimates the degree of average TopDown overcounting or undercounting by the racial demographics of the geographic unit. Lines are estimated from a local polynomial regression model that adjusts for noise in the non-white percentage estimate. Swapping would have no error for total population.

made possible by these statistical properties of the NMF. Thus, while analyses that treat the NMF counts as ground truth (i.e., as a drop-in replacement for the published Census statistics) will likely result in much higher bias and variance, the NMF should play an important role for researchers working with Census data.

These results have several policy implications for future releases of decennial Census data. First, postprocessing plays an important role in preserving the accuracy of Census data for direct use. The NMF, while unbiased, may not be sufficiently accurate for public policy decisions without measurement error modeling and using it naively could have differential impacts on racial and ethnic groups. The final small biases and noise come after rounds of iteration between the Bureau and data stakeholders, which our results here and elsewhere suggest likely improved the accuracy of the final released data (*2*).

Second, while the Bureau replaced swapping with TopDown, citing the need for formal privacy guarantees to address increasing vulnerability to privacy leaks, we find that the bias and noise induced by two algorithms are similar. Our finding is consistent with our earlier result that an individual's race can be predicted accurately under both swapping and TopDown (*2*). It is also consistent with the recent analysis that shows swapping can be differentially private, albeit with a different differential privacy definition (*30*).

Together, these results provide additional empirical context for ongoing conversations about the Bureau's trade-off between data accuracy and privacy protection. Biases from data produced by the TopDown and swapping algorithms are smaller than found in the released NMF, but the extent to which privacy is meaningfully protected is disputed and dependent on the definition of privacy used (*9*, *25*). Further research and discussion about the privacy-accuracy trade-off and the real-world privacy protections offered by a formally private system such as TopDown are warranted.

Last, while our analysis shows how to rigorously evaluate the DAS without access to the CEF, the Bureau has all of the necessary information—including the CEF—to much more precisely assess the bias and noise of the DAS than we are able to do here. Although the Bureau has released some summary statistics about the error induced by TopDown (*12*), we believe that our independent assessment, combined with a methodology for computing the bias and variance, is useful for policy makers and researchers.

Ideally, the Bureau could produce margins of error for decennial census statistics that hold on average across all geographies nationwide and/or runs of the TopDown algorithm. At a minimum, the Bureau should release a wide variety of error metrics including mean error and MSE of the TopDown counts for all on- and off-spine geographies in a machine-readable format. These error metrics should also be broken out by total population of each geography, because as Figs. 4 and 5 show, bias and variance can vary with total population. Given our findings of error variability in swapped counts, the Bureau should also release a similar set of metrics for runs of their 2010 swapping algorithm under production settings. No error metrics of any sort have been publicly reported for the 1990, 2000, and 2010 swapping algorithms, to our knowledge. For these metrics to be useful to researchers, especially in areas with small populations or small racial groups, they would need to be quite accurate. This means that releasing such error metrics would require Census Bureau decision makers to again weigh the benefits of such a release against the potential privacy risks.

Supplementary Materials

This PDF file includes:

Sections S1 to S4
Figs. S1 to S3
Table S1

REFERENCES AND NOTES

1. M. Hotchkiss, J. Phelan, Uses of census bureau data in federal funds distribution: A new design for the 21st century (U.S. Census Bureau, 2017).
2. C. T. Kenny, S. Kuriwaki, C. McCartan, E. T. R. Rosenman, T. Simko, K. Imai, The impact of the US Census disclosure avoidance system on redistricting and voting rights analysis. *Sci. Adv.* **7**, 1–17 (2021).
3. J. Abowd, R. Ashmead, R. Cumings-Menon, S. Garfinkel, M. Heineck, C. Heiss, R. Johns, D. Kifer, P. Leclerc, A. Machanavajjhala, B. Moran, W. Sexton, M. Spence, P. Zhuravlev, The 2020 Census Disclosure Avoidance System TopDown Algorithm, *Harvard Data Science Review* (2022); https://hdsr.mitpress.mit.edu/pub/7evz361i.
4. L. McKenna, "Disclosure avoidance techniques used for the 1970 through 2010 decennial censuses of population and housing," (Tech. Rep. CES-18-47, Research and Methodology Directorate, U.S. Census Bureau, 2018).
5. S. Ruggles, C. Fitch, D. Magnuson, J. Schroeder, Differential privacy and census data: Implications for social and economic research. *AEA Papers Proc.* **109**, 403–408 (2019).
6. A. R. Santos-Lozada, J. T. Howard, A. M. Verdery, How differential privacy will affect our understanding of health disparities in the United States. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13405–13412 (2020).
7. National Congress of American Indians, Letter to Steven D. Dillingham from Kevin Allis, Chief Executive Officer (2020).
8. J. Scariano, I. Youngs, Balancing utility versus privacy in the 2020 census: Sentiments from data users, Available at SSRN 4089888 (2022).
9. V. J. Hotz, C. R. Bollinger, T. Komarova, C. F. Manski, R. A. Moffitt, D. Nekipelov, A. Sojourner, B. D. Spencer, Balancing data privacy and usability in the federal statistical system. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2104906119 (2022).
10. V. J. Hotz, J. Salvo, A Chronicle of the Application of Differential Privacy to the 2020 Census, *Harvard Data Science Review* (2022); https://hdsr.mitpress.mit.edu/pub/ql9z7ehf.
11. S. Ruggles, D. V. Riper, The role of chance in the census bureau database reconstruction experiment. *Popul. Res. Policy Rev.* **41**, 781–788 (2022).
12. U.S. Census Bureau, "2020 Census Disclosure Avoidance System Detailed Summary Metrics," (Tech. Rep., 2023). May 25, 2023; https://perma.cc/QM5R-BJ5J.
13. *Phillips* v. *Census Bureau* (2022). 1:2022cv09304, US District Court for the Southern District of New York.
14. W. R. Bell, J. L. Schafer, April 19, 2021 memorandum for john abowd: Simulation studies to investigate variation in census counts and in census coverage error using 2010 sf-1 data and 2010 ccm results (2021).
15. J. M. Abowd, R. Ashmead, R. Cumings-Menon, S. Garfinkel, M. Heineck, C. Heiss, R. Johns, D. Kifer, P. Leclerc, A. Machanavajjhala, B. Moran, W. Sexton, M. Spence, P. Zhuravlev, 2010 census production settings redistricting data (P.L. 94-171) demonstration noisy measurement file (ICPSR 38777) (2023); www.icpsr.umich.edu/web/ICPSR/studies/38777, www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/, and https://doi.org/10.7910/DVN/1OR2A6.
16. U.S. Census Bureau, "2010 Demonstration Data Products Suite: Redistricting and DHC (Tech. Rep., 2021). June 2021 vintage; www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/01-Redistricting_File–PL_94-171/2021-06-08_ppmf_Production_Settings/.
17. J. M. McKenzie, Freedom of information act denial letter (2022). In Phillips v. Census Bureau.
18. B. Asquith, B. Hershbein, T. Kugler, S. Reed, S. Ruggles, J. Schroeder, S. Yesiltepe, D. Van Riper, Assessing the Impact of Differential Privacy on Measures of Population and Racial Residential Segregation, *Harvard Data Science Review* (2022); https://hdsr.mitpress.mit.edu/pub/1rsg867y.
19. M. Christ, S. Radway, S. M. Bellovin, *2022 IEEE Symposium on Security and Privacy (SP)* (IEEE, 2022), pp. 457–472.
20. A. Cohen, M. Duchin, J. Matthews, B. Suwal, *2nd Symposium on Foundations of Responsible Computing (FORC 2021)*, K. Ligett, S. Gupta, eds. (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2021), vol. 192 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 5:1–5:22.
21. C. L. Canonne, G. Kamath, T. Steinke, The discrete gaussian for differential privacy. *Adv. Neural Inf. Process. Syst.* **33**, 15676–15688 (2020).
22. C. McCartan, T. Simko, K. Imai, Making Differential Privacy Work for Census Data Users, *Harvard Data Science Review* **5** (2023); https://hdsr.mitpress.mit.edu/pub/y6oilhhe.
23. L. Zayatz, J. Lucero, P. Massell, A. Ramanayake, Disclosure avoidance for census 2010 and american community survey five-year tabular data products, Statistics 2009–10 (Statistical Research Division, U.S. Census Bureau, 2009).

Downloaded from https://www.science.org on May 01, 2024

24. U.S. Census Bureau, 2010 Census Summary File 1 Technical Documentatione (2011); www2.census.gov/programs-surveys/decennial/2010/technical-documentation/complete-tech-docs/summary-file/sf1.pdf.
25. C. T. Kenny, S. Kuriwaki, C. M. Cartan, E. T. R. Rosenman, T. Simko, K. Imai, Comment: The essential role of policy evaluation for the 2020 census disclosure avoidance system. *Harvard Data Science Review* **Special Issue 2**, 1 (2023).
26. U.S. Census Bureau, Census bureau sets key parameters to protect privacy in 2020 census results (2021); https://perma.cc/E2ZR-ZAXK.
27. A. Delaigle, J. Fan, R. J. Carroll, A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Am. Stat. Assoc.* **104**, 348–359 (2009).
28. X. Huang, H. Zhou, Conditional density estimation with covariate measurement error. *Electron. J. Stat.* **14**, 970–1023 (2020).
29. d. boyd, J. Sarathy, Differential perspectives: Epistemic disconnects surrounding the US Census Bureau's use of differential privacy, *Harvard Data Science Review* (2022).
30. J. Bailie, R. Gong, X.-L. Meng, "Can swapping be differentially private? A refreshment stirred, not shaken," (National Bureau of Economic Research, 2023).

Acknowledgments: We thank reviewers of the Alexander and Diviya Magaro Peer Pre-Review Program for useful comments and M. B. Hawes of the Census Bureau's Research and Methodology Directorate for sharing information about various details of the Noisy Measurement File. **Funding:** The authors acknowledge that they received no funding in support of this research. **Author contributions:** Conceptualization: C.T.K., S.K., C.M., T.S., and K.I. Methodology: S.K., C.M., and K.I. Investigation: C.T.K., S.K., C.M., T.S., and K.I. Visualization: C.T.K., S.K., C.M., and T.S. Supervision: K.I. Writing—original draft: C.T.K., S.K., C.M., T.S., and K.I. Writing—review and editing: C.T.K., S.K., C.M., T.S., and K.I. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data, as well as code to reproduce the analysis, are available as part of the supplemental repository at https://doi.org/10.7910/DVN/TMIN3H.

Submitted 7 October 2023
Accepted 28 March 2024
Published 1 May 2024
10.1126/sciadv.adl2524

Downloaded from https://www.science.org on May 01, 2024