Science Advances
AAAS

Supplementary Materials for

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

Christopher T. Kenny *et al.*

Corresponding author: Kosuke Imai, imai@harvard.edu

Sci. Adv. **10**, eadl2524 (2024)
DOI: 10.1126/sciadv.adl2524

This PDF file includes:

Sections S1 to S4
Figs. S1 to S3
Table S1

S1 Mapping the NMF geographies to the standard Census geographies

To construct noisy measurements for standard census geographies, we first map the geocode of the NMF to traditional GEOIDs. Queries for the NMF are produced along a non-traditional spine that treat American Indian / Alaska Native (AI/AN) and non-AI/AN areas as separate components. This means that the NMF does not contain noisy measurements for some of standard census "spine" geographies such as census block groups and census tracts. Although this problem does not apply to a minority of states that contain no AI/AN area (AR, DC, DE, IL, IN, KY, MD, MO, NH, NJ, OH, PA, TN, VT, WV), most states and hence the census geographies within them are split into AI/AN and non-AI/AN areas. For the sake of completeness, this section describes in detail how we mapped the NMF geographies to the standard Census geographies.

Detailed Geocode–GEOID correspondence. An NMF geocode is a unique identifier for each geography. The NMF contains basic documentation on the correspondence between geocodes and GEOIDs (*1*). Like in the usual decennial census releases, the lowest level of geography is the census block, for which the NMF geocodes are 31 digits long. All census geographies can be built by aggregating census blocks. Most blocks are included in the TopDown algorithm's process with the exception of the blocks that have zero population. Every included block has a one-to-one mapping between geocodes and traditional GEOIDs. Other geographies may or may not have a unique mapping between them. This is in large part a function of if a state or county has any AI/AN areas or not, where units are more similar if there are no AI/AN areas.

An example of a valid geocode for a census block is `0531000100011065300195010011010`. The first digit indicates whether this block is on the AI/AN spine (`1`) or not (`0`) (*1, item 14*). The final 16 digits is related to the GEOID of the block (*1, item 15*). Specifically, the GEOID for a block is the 16th to the 26th digit of the geocode concatenated with the 28th to 31th digits. The 27th digit, which refers to the traditional block group, is replicated in geocodes, though it is (correctly) never replicated in traditional GEOIDs. As such, it can be dropped to create traditional GEOIDs. Besides the first and final 16 digits, the other digits have substantive meanings. These are explained in the table below.

Table S1: Explanation of digits of the NMF geocodes.

Component	Start	End	Description
1	1	1	AI/AN digit; 0 = non-AI/AN , 1 = AI/AN
2	2	3	state FIPS code
3	4	5	spine optimization code; 10 = on spine, county or below
4	6	8	county FIPS
5	9	12	tract FIPS-equivalent
6	13	15	optimize block group FIPS-equivalent
7	16	17	state FIPS (traditional GEOID component)
8	18	20	county FIPS (traditional GEOID component)
9	21	26	tract FIPS (traditional GEOID component)
10	27	27	block group FIPS; repeat of first digit of component 10 (block FIPS)
11	28	31	block FIPS (traditional GEOID component)

Creating Crosswalks to Standard Census Geographies. As discussed above, each standard census geography corresponds to a collection of (non-overlapping) AI/AN and non-AI/AN pieces. A census block is never split between AI/AN and non-AI/AN pieces. Instead, it belongs to either the AI/AN spine or non-AI/AN spine. Spines "do not subdivide tabulation blocks" (*1, item 12*).

A census block-group is a collection of census blocks and hence may be split between the AI/AN and non-AI/AN spines. The same applies to a census tract. Optimized block groups are special groups of blocks which are only used in the TopDown algorithm. These block groups are designed to better nest into off-spine geographies of interest, such as census places. Traditional block groups do not necessarily optimize this, but are used frequently in research, as the American Community Survey often releases data at the block group level. These relationships are shown in Figure 1 in the main text.

We use a simple algorithm to produce noisy measurements for standard census geographies that are not on the NMF spines. Our algorithm is designed to produce minimum-variance estimates for off-spine geographies subject to the constraint that estimates be produced by summing component geographies. This constraint is sufficient (though perhaps not necessary) to ensure that resulting estimates from disjoint geographies are statistically independent, a critical requirement for further analysis.

There often exist more than one set of non-overlapping NMF geographies that exactly correspond to any given standard census geography. To minimize the variance of the resulting noisy measurements,

we compose the desired area from a small number of large non-overlapping NMF geographies. This is because large NMF geographies have smaller variances and adding more NMF geographies leads to a greater variance.

We start with the largest NMF geography that fits within a given target geography, working through the NMF spine hierarchically from counties down to census blocks. For example, for a given voting district (VTD), we find all of the optimized block groups that are fully contained within it (after checking both counties and tracts, which are generally never fully contained or coterminous with VTDs). The noisy measurements for these optimized block groups are used in the VTD estimate. We then include the noisy measurements for the census blocks covering the remaining areas of the VTD. The noisy measurements and their variances for this VTD can be obtained by simply summing the corresponding noisy measurements of the on-spine NMF geographies that are used to fill the VTD.

Procedures may exist which produce smaller-variance estimates for a set of geographies or a specific geography. One way to do so involves subtracting small NMF geographies from larger ones to fill a standard census geography. However, this approach (and other alternatives) will tend to create dependencies between noisy measurements because the same noisy measurement will be used to produce estimates for different off-spine geographies. This significantly complicates subsequent statistical analysis. As a result, we do not use this approach.

S2 Proofs of Propositions

In this section, we provide proofs for all the propositions of this section.

Proposition 2.1.

Proof. Since $\{\boldsymbol{\eta}_\gamma\}_{\gamma\in\Gamma}$ are mutually independent across $\gamma \in \Gamma$, and $\boldsymbol{\varepsilon}_\gamma^{\mathrm{nm}} \coloneqq B(\Sigma_\gamma)\boldsymbol{\eta}_\gamma$ is a function of $\boldsymbol{\eta}_\gamma$ alone ($B(\Sigma_\gamma)$ is fixed *a priori*), $\{\boldsymbol{\varepsilon}_\gamma^{\mathrm{nm}}\}_{\gamma\in\Gamma}$ are therefore also mutually independent, proving (a).

Similarly, we must also have that $\{\boldsymbol{\varepsilon}_\gamma^{\mathrm{nm}}\}_{\gamma\in\Gamma}$ and $\{\boldsymbol{\varepsilon}_\gamma^{\mathrm{sw}}\}_{\gamma\in\Gamma}$, since each $\boldsymbol{\eta}_\gamma^{\mathrm{nm}}$ is drawn independently, proving (b). □

Proposition 2.2.

Proof. As in the statement of the proposition, let $\widetilde{Y}^{\mathrm{td}}_{\gamma s}$ be the second set of post-processed demonstration data, with corresponding noisy measurements $\{\tilde{\boldsymbol{\eta}}_\gamma\}_{\gamma\in\Gamma}$ and errors $\{\tilde{\boldsymbol{\varepsilon}}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$.

Then the $\{\tilde{\boldsymbol{\varepsilon}}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$ are a function of $\{\tilde{\boldsymbol{\eta}}_\gamma\}_{\gamma\in\Gamma}$ alone, and so are independent of $\{\boldsymbol{\varepsilon}^{\mathrm{nm}}_\gamma\}_{\gamma\in\Gamma}$, which are a function of $\{\boldsymbol{\eta}_\gamma\}_{\gamma\in\Gamma}$ alone, proving (a).

Similarly, the $\{\tilde{\boldsymbol{\varepsilon}}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$ are independent of $\{\boldsymbol{\varepsilon}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$, which are a function of $\{\boldsymbol{\eta}_\gamma\}_{\gamma\in\Gamma}$ alone, proving (b). □

Proposition 2.3.

Proof. The variance of $\hat{\tilde{\mu}}^{\mathrm{td}}_{\mathcal{I}s}$ is

$$\begin{aligned}\mathbb{V}[\hat{\tilde{\mu}}^{\mathrm{td}}_{\mathcal{I}s}] &= \frac{1}{|\mathcal{I}|^2}\left(\sum_{\gamma\in\mathcal{I}}\mathbb{V}[\widetilde{Y}^{\mathrm{td}}_{\gamma s}-Y^{\mathrm{nm}}_{\gamma s}]+\sum_{\gamma,\gamma'\in\mathcal{I},\ \gamma\neq\gamma'}\mathrm{Cov}(\widetilde{Y}^{\mathrm{td}}_{\gamma s}-Y^{\mathrm{nm}}_{\gamma s},\widetilde{Y}^{\mathrm{td}}_{\gamma' s}-Y^{\mathrm{nm}}_{\gamma' s})\right)\\ &= \frac{1}{|\mathcal{I}|^2}\left(\sum_{\gamma\in\mathcal{I}}\left(\mathbb{V}[\tilde{\varepsilon}^{\mathrm{td}}_{\gamma s}]+\sigma^2_{\gamma s}\right)+\sum_{\gamma,\gamma'\in\mathcal{I},\ \gamma\neq\gamma'}\mathrm{Cov}(\tilde{\varepsilon}^{\mathrm{td}}_{\gamma s},\tilde{\varepsilon}^{\mathrm{td}}_{\gamma' s})\right),\end{aligned}$$

since by Proposition 2.2(a) $\{\boldsymbol{\varepsilon}^{\mathrm{nm}}_\gamma\}_{\gamma\in\Gamma}$ and $\{\tilde{\boldsymbol{\varepsilon}}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$ are independent, and by Proposition 2.1(a) the $\{\tilde{\boldsymbol{\varepsilon}}^{\mathrm{nm}}_\gamma\}_{\gamma\in\Gamma}$ are mutually independent.

For any $\gamma,\gamma'\in\mathcal{I}$, we have

$$\begin{aligned}\mathbb{E}[(Y^{\mathrm{td}}_{\gamma s}-\widetilde{Y}^{\mathrm{td}}_{\gamma s})(Y^{\mathrm{td}}_{\gamma' s}-\widetilde{Y}^{\mathrm{td}}_{\gamma' s})] &= \mathbb{E}[(Y+\varepsilon^{\mathrm{td}}_{\gamma s}-Y-\tilde{\varepsilon}^{\mathrm{td}}_{\gamma s})(Y+\varepsilon^{\mathrm{td}}_{\gamma' s}-Y-\tilde{\varepsilon}^{\mathrm{td}}_{\gamma' s})]\\ &= \mathbb{E}[(\varepsilon^{\mathrm{td}}_{\gamma s}-\tilde{\varepsilon}^{\mathrm{td}}_{\gamma s})(\varepsilon^{\mathrm{td}}_{\gamma' s}-\tilde{\varepsilon}^{\mathrm{td}}_{\gamma' s})]\\ &= 2\,\mathrm{Cov}(\varepsilon^{\mathrm{td}}_{\gamma s},\varepsilon^{\mathrm{td}}_{\gamma' s}),\end{aligned}$$

where the final line follows because by Proposition 2.2(b), $\{\boldsymbol{\varepsilon}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$ and $\{\tilde{\boldsymbol{\varepsilon}}^{\mathrm{td}}_\gamma\}_{\gamma\in\Gamma}$ are independent.

Then

$$\begin{aligned}\mathbb{E}\left[\frac{1}{2}\left(\sum_{\gamma\in\mathcal{I}}Y^{\mathrm{td}}_{\gamma s}-\widetilde{Y}^{\mathrm{td}}_{\gamma s}\right)^2\right] &= \frac{1}{2}\sum_{\gamma,\gamma'\in\mathcal{I}}\mathbb{E}\left[(Y^{\mathrm{td}}_{\gamma s}-\widetilde{Y}^{\mathrm{td}}_{\gamma' s})(Y^{\mathrm{td}}_{\gamma s}-\widetilde{Y}^{\mathrm{td}}_{\gamma' s})\right]\\ &= \frac{1}{2}\sum_{\gamma,\gamma'\in\mathcal{I}}2\,\mathrm{Cov}(\varepsilon^{\mathrm{td}}_{\gamma s},\varepsilon^{\mathrm{td}}_{\gamma' s})\\ &= \sum_{\gamma\in\mathcal{I}}\mathbb{V}[\varepsilon^{\mathrm{td}}_{\gamma s}]+\sum_{\gamma,\gamma'\in\mathcal{I},\ \gamma\neq\gamma'}\mathrm{Cov}(\varepsilon^{\mathrm{td}}_{\gamma s},\varepsilon^{\mathrm{td}}_{\gamma' s}).\end{aligned}$$

Thus adding in $\sum_{\gamma\in\mathcal{I}}\sigma^2_{\gamma s}$ and dividing by $|\mathcal{I}|^2$, we find that

$$\mathbb{E}[\widehat{V}(\hat{\tilde{\mu}}^{\mathrm{td}}_{\mathcal{I}s})] = \mathbb{V}[\hat{\tilde{\mu}}^{\mathrm{td}}_{\mathcal{I}s}].$$

□

Proposition 2.4.

Proof. The variance of $\hat{\mu}^{\text{sw}}_{\mathcal{I}s}$ is

$$\begin{aligned}
\mathbb{V}[\hat{\mu}^{\text{sw}}_{\mathcal{I}s}] &= \frac{1}{|\mathcal{I}|^2}\left(\sum_{\gamma\in\mathcal{I}}\mathbb{V}[Y^{\text{sw}}_{\gamma s} - Y^{\text{nm}}_{\gamma s}] + \sum_{\gamma,\gamma'\in\mathcal{I},\ \gamma\neq\gamma'}\text{Cov}(Y^{\text{sw}}_{\gamma s} - Y^{\text{nm}}_{\gamma s}, Y^{\text{sw}}_{\gamma' s} - Y^{\text{nm}}_{\gamma' s})\right) \\
&= \frac{1}{|\mathcal{I}|^2}\left(\sum_{\gamma\in\mathcal{I}}\left(\mathbb{V}[\varepsilon^{\text{sw}}_{\gamma s}] + \sigma^2_{\gamma s}\right) + \sum_{\gamma,\gamma'\in\mathcal{I},\ \gamma\neq\gamma'}\text{Cov}(\varepsilon^{\text{sw}}_{\gamma s}, \varepsilon^{\text{sw}}_{\gamma' s})\right), \\
&\leq \frac{1}{|\mathcal{I}|^2}\sum_{\gamma\in\mathcal{I}}\left(\mathbb{V}[\varepsilon^{\text{sw}}_{\gamma s}] + \sigma^2_{\gamma s}\right),
\end{aligned}$$

with the final line following by assumption.

For any $\gamma \in \mathcal{I}$,

$$\begin{aligned}
\mathbb{E}\left[(Y^{\text{sw}}_{\gamma s} - Y^{\text{nm}}_{\gamma s})^2\right] &= \mathbb{E}\left[(\varepsilon^{\text{sw}}_{\gamma s} - \varepsilon^{\text{nm}}_{\gamma s})^2\right] \\
&= \mathbb{E}[\varepsilon^{\text{sw}2}_{\gamma s}] + \sigma^2_{\gamma s} \\
&\geq \mathbb{V}[\varepsilon^{\text{sw}}_{\gamma s}] + \sigma^2_{\gamma s},
\end{aligned}$$

where the third line follows because $\varepsilon^{\text{nm}}_{\gamma s}$ is zero-mean and independent of $\varepsilon^{\text{sw}}_{\gamma s}$ by Proposition 2.1(b).

Combining these two results, by the linearity of expectation

$$\mathbb{E}[\widehat{V}(\hat{\mu}^{\text{sw}}_{\mathcal{I}s})] \geq \mathbb{V}[\hat{\mu}^{\text{sw}}_{\mathcal{I}s}].$$

□

S3 Average Bias and Root Mean Square Error for NH/PI, Other, and Two or More Races

We replicate Figures 4 and 5 in the main text for the remaining three top-level race groups. These are Native Hawaiian and Pacific Islander (NH/PI), some other race, and those who identify as two or more races.

As seen in Figure S1, the average biases for NH/PI and Other race groups are estimated to be small. This is similar to the patterns for White, Black, Asian, and AI/AN racial groups. For members of two or

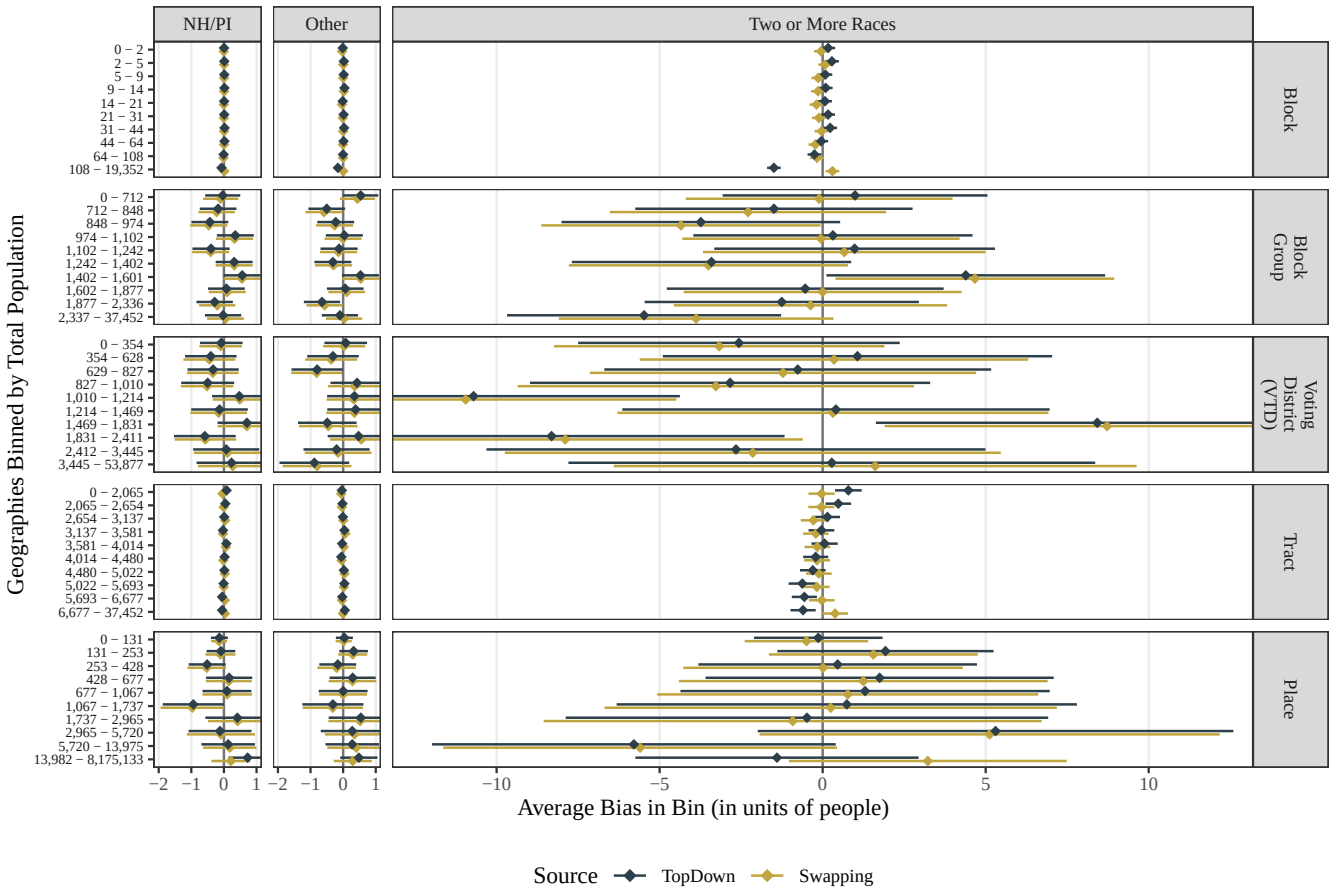


Figure S1: **Average bias for race/ethnicity population counts at each geographic level by its total population.** The figure estimates the average overcounting or undercounting in a group of geographies, separately for five geographic levels and the remaining three race/ethnicity groups. See "Proposed Estimators" in the main text 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.

more race groups, the estimated average biases vary more. For two or more race groups, higher population blocks and tracts tend to display downward average bias, while the smallest blocks and tracts are, on average, biased towards having more population. Estimates for off-spine geographies are noisier and do not display a clear pattern of average bias across populations.

Figure S2 similarly replicates the trends in Figure 5. The RMSE for NH/PI and Other are smallest and relatively constant across geography sizes for on-spine geographies. The RMSEs are largest for the two or more race groups, likely following a similar logic to Hispanics, where separate queries are used for this grouping. As in Figure 5, the NMF is clearly the noisiest, with the post-processing step of the TopDown algorithm reducing the variation to about the same scale as the noise from swapping.

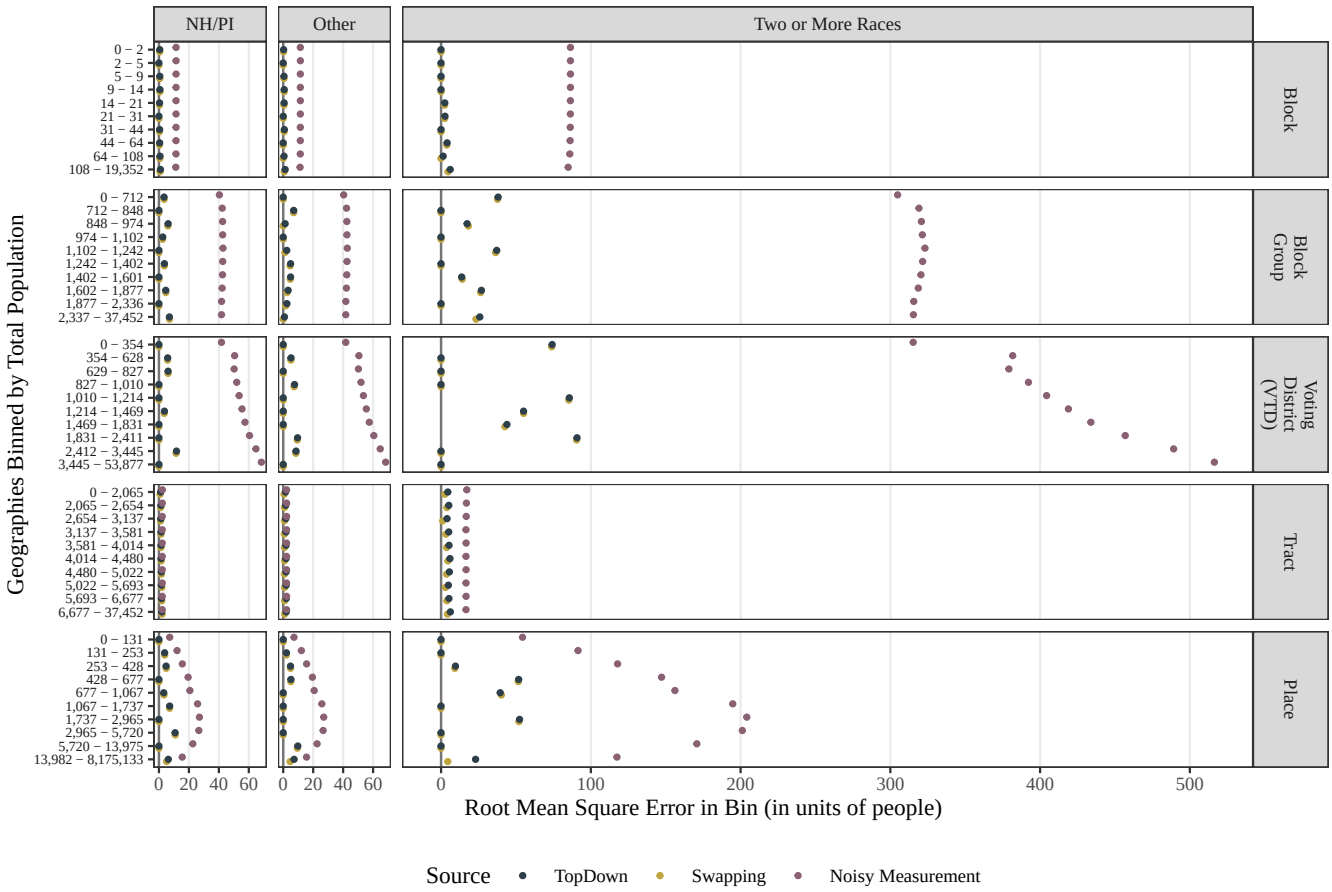


Figure S2: **Estimated root mean square error of race/ethnicity counts at each geographic level, by its total population.** This estimates RMSE for a subset of geographies. See "Proposed Estimators" in the main text for estimators. The RMSE estimates for Post-Processing and Swapping error are occasionally negative, and these are imputed at 0. Bins on the *y*-axis are deciles by total population of the geographic level.

S4 Local Linear Regression with Measurement Error

Figure 6 in the main text shows results of a local linear regression where we regress the error in the counting of total population, which is known exactly, on the percentage of people who are non-White. We use the method proposed by (*2*) to fit this regression based on the implementation available in the R package `lpme`. We use 50 equally separated evaluation points $\{0, 0.02, 0.04, \cdots, 1\}$ along the values of percentage non-white, and a bandwidth of +/- 5 percentage points. The sequence of 50 fitted values from the models form a local linear regression that approximates the true nonparametric model. We fit this model for block groups, tracts, VTDs, and places. We do not fit it to blocks because the values of the fraction non-White are too concentrated (e.g., in values such as 1/3 or 1/2). This concentration is due to

the large left skew of the distribution of block populations, where many blocks have near-zero populations.

This nonparametric regression approach is designed to account for measurement error by incorporating the standard deviation of the injected noise. As implemented in the `lpme` package, the method only allows for the homoskedastic variance. Although not ideal, we use the median variance of noisy measurements across geographies (rescaled by the square of the total population for the fraction) as the input to the model.

Additionally, the existing methodology assumes that each observation is independent of the others. In fact, the TopDown errors are correlated across geographies. Just as with OLS, we expect this correlation to affect the variance of the nonparametric estimate, not the bias although we have not rigorously proved this conjecture. In addition, statistical results for the method we use are all asymptotic; here, we have a large but fixed number of Census geographies and so the results should be interpreted with some caution.

We also implemented an alternative approach to model this relationship that enables some quantification of uncertainty. Figure S3 uses mutually exclusive bins of percentage non-White measure, instead of a moving average. We binned data between 0 and 100% non-White by bins 4 percentage points wide. Negative values and values above 100% — which arise due to the noisy measurement of the non-White population — are placed in their own bin. Together, for each level of geography, we partition the units into 27 bins. We then compute the simple average of the total population counting error of the geographic units in each bin, with a standard error of the mean computed by the standard deviation of the error divided by the square root of the number of geographic units in the bin. This method cannot account for the substantial measurement error in inherent in the percentage non-White measure, but we can provide a rough measure of uncertainty. The s.d. in Figure S3 represents the median of the standard deviation inherent in the measure of percentage non-White for a unit in each level. For example, the noisy measurement induces a noise of about 3 percentage points for the median block group. The pattern of local averages here mirrors the trend in Figure 6. Moreover, it shows that the undercounting bias is in off-spine geographies is statistically significant from zero.

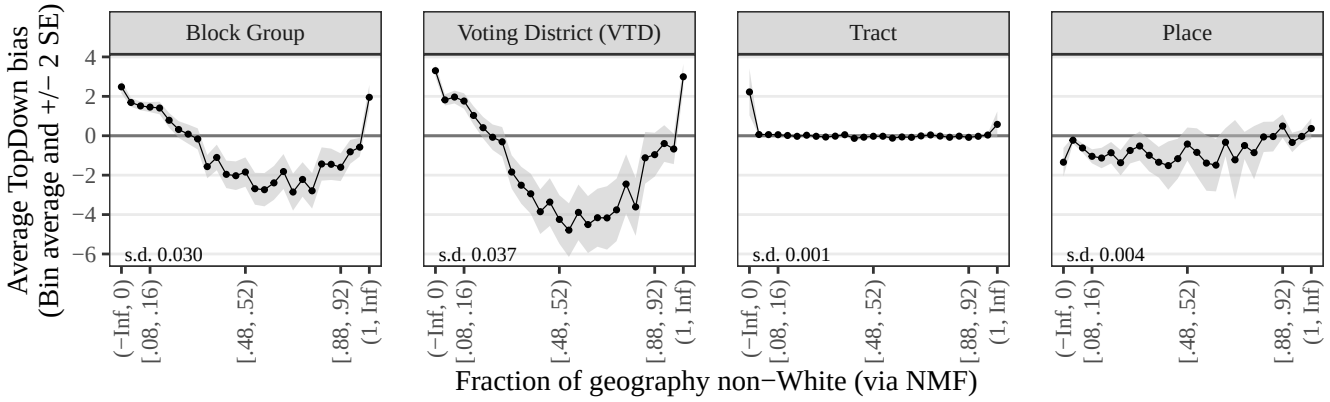


Figure S3: **Average bias in total population count by percent non-white, with mutually exclusive bins.** Similar to Figure 6, the figure estimates the degree of average TopDown overcounting or undercounting by the racial demographics of the geographic unit. We show the average error in each mutually exclusive bin of the noisily measured percentage non-White. Bands show twice the standard error around each average.

Other Supplementary Materials for this manuscript include the following:

Data and replication code, https://doi.org/10.7910/DVN/TMIN3H.