差分隐私和人口普查区数据来源对监测健康不平等的影响

# Impact of Differential Privacy and Census Tract Data Source (Decennial Census Versus American Community Survey) for Monitoring Health Inequities

*Nancy Krieger, PhD, Rachel C. Nethery, PhD, Jarvis T. Chen, ScD, Pamela D. Waterman, MPH, Emily Wright, BA, Tamara Rushovich, MPH, and Brent A. Coull, PhD*

**Objectives.** To investigate how census tract (CT) estimates of mortality rates and inequities are affected by (1) differential privacy (DP), whereby the public decennial census (DC) data are injected with statistical "noise" to protect individual privacy, and (2) uncertainty arising from the small number of different persons surveyed each year in a given CT for the American Community Survey (ACS).

**Methods.** We compared estimates of the 2008–2012 average annual premature mortality rate (death before age 65 years) in Massachusetts using CT data from the 2010 DC, 2010 DC with DP, and 2008–2012 ACS 5-year estimate data.

**Results.** For these 3 denominator sources, the age-standardized premature mortality rates (per 100 000) for the total population respectively equaled 166.4 (95% confidence interval [CI] = 162.2, 170.6), 166.4 (95% CI = 162.2, 170.6), and 166.3 (95% CI = 162.1, 170.5), and inequities in the range from best to worst quintile for CT racialized economic segregation were from 103.4 to 260.1, 102.9 to 258.7, and 102.8 to 262.4. Similarity of results across CT denominator sources held for analyses stratified by gender and race/ethnicity.

**Conclusions.** Estimates of health inequities at the CT level may not be affected by use of 2020 DP data and uncertainty in the ACS data. (*Am J Public Health*. 2021;111:265–268. https://doi.org/10.2105/AJPH.2020.305989)

Despite the importance of accurate census data for public health—for denominators, for characterizing areas, and for allocating political representation and resources[1]—little is known about how census tract (CT) estimates of health rates and inequities—critical for local health monitoring and analysis[2,3]—will be affected by the new use of differential privacy (DP) with the 2020 decennial census (DC).[4] In brief, DP refers to a procedure whereby statistical "noise" is injected into the publicly released DC data to protect individual privacy.[4] New research has raised concerns that DP combined with census postprocessing of these data may bias substate population counts (e.g., counties, CTs), deflating population counts in urban and American Indian areas and inflating them in other areas, and thus affecting computation of rates.[5]

Also still poorly understood are impacts of the 2008 federal shift from collecting detailed social and economic data in the DC long form to the annually conducted American Community Survey (ACS).[6] Of particular concern is the uncertainty arising from the small number of different persons surveyed each year in a given CT, producing wide margins of error for population counts.[7]

To our knowledge, no research has assessed the potential impact of DP on population health estimates computed from CT data or compared this impact with that of sampling-related error in the ACS. In November 2019, the US Census Bureau released its first-ever DP

demonstration product, comprising the 2010 DC data with DP applied, enabling research to address this issue.[4] We empirically evaluated the impact of using CT population counts from the 2010 DC, 2010 DC with DP, and 2008–2012 ACS on estimating inequities in premature mortality in Massachusetts.

## METHODS

Our 3 CT population sources were (1) the most recent DC file with DP, produced by the US Census Bureau in November 2019 for the 2010 DC[4,8–10]; (2) the original 2010 DC; and (3) the 2008–2012 5-year estimates from the ACS.[6]

### Mortality Data

We obtained individual-level mortality data for 2008 to 2012 for all premature deaths (younger than 65 years; n = 55 836 deaths) from the Massachusetts Department of Public Health[11] (Table A, available as a supplement to the online version of this article at http://www.ajph.org). We geocoded the residential address at death to the corresponding CT[2]; only 0.4% of deaths could not be geocoded with this level of precision, yielding an analytic data set with 55 560 deaths. We focused on premature mortality because this outcome is a widely used population health indicator that manifests strong social gradients and is not affected by misclassification of cause of death.[2,3]

### Metric for Health Inequities

We used the index of concentration at the extremes (ICE) for racialized economic segregation, which we developed in 2014, building on Massey's initial use of the ICE for solely economic measures,[12] with our measure shown in

numerous studies to be more sensitive to health inequities than metrics employing solely economic or racial data (Table B, available as a supplement to the online version of this article at http://www.ajph.org). The ICE delineates people's concentration, in an area, in the extremes of the selected measure and ranges from −1 (all in the deprived group) to 1 (all in the most privileged group).[12] Its formula is

$$(1) \qquad ICE_i = (A_i - P_i)/T_i$$

where $A_i$, $P_i$, and $T_i$ correspond, respectively, to the number of persons in the $i$th geographic area categorized as belonging to the most privileged extreme, the most deprived extreme, and the total population whose privilege level was measured.[12] For our analyses, we set these extremes as (1) high-income White (alone) population versus (2) low-income Black (alone) population[12] (see Table B for the census variables used). Missing data precluded computing the ICE for 19 (1.3%) of the Massachusetts CTs.

### Statistical Methods

We computed, for the total population and also stratified by race/ethnicity and gender, the 2008–2012 average annual age-standardized premature mortality rate (death before age 65 years per 100 000 persons, standardized to the year 2000 standard million[2]) and associated 95% confidence interval (CI) in Massachusetts using CT population counts from the 2010 DC, 2010 DC with DP, and 2008–2012 ACS 5-year estimate data. We then categorized the CT in quintiles of the ICE for racialized economic segregation, aggregated the mortality and population count data across tracts within each quintile (without taking into account spatial correlations), and computed premature mortality rates

by ICE quintile, overall and by race/ethnicity and gender. We then plotted and compared the point estimates and their 95% CIs for each source of population count data. We also conducted sensitivity analyses using the percentage of persons below poverty (Table B).

## RESULTS

In 2010, the population of Massachusetts included 5 644 905 persons younger than 65 years (based on the 2010 DC) and 1478 CTs. The age-standardized premature mortality rates (per 100 000) for the total population were highly similar across the 3 denominator sources (DC, DP, and ACS) and respectively equaled 166.4 (95% CI = 162.2, 170.6), 166.3 (95% CI = 162.2, 170.5), and 166.4 (95% CI = 162.1, 170.6; Figure 1). Also similar across denominator sources was the range from best to worst quintile for CT racialized economic segregation (103.4–260.1, 102.9–258.7, and 102.8–262.4; Table C, available as a supplement to the online version of this article at http://www.ajph.org).

Robustness across CT denominator sources held for analyses stratified by race/ethnicity and by gender (Table C), with results for the non-Hispanic White population closely paralleling those for the total population (reflecting that they constituted 74.0% of the 2010 Massachusetts population younger than 65 years). Among the Black population (7.1% of the total population aged younger than 65 years), these rates respectively equaled 230.5 (95% CI = 210.5, 250.6), 229.8 (95% CI = 209.8, 249.8), and 226.4 (95% CI = 206.4, 245.8)—and the range across the ICE quintiles was 173.0 to 258.6, 161.3 to 260.4, and 177.5 to 249.6 (Table C). These rates for women were identical across the 3 CT denominator sources (118.6; 95% CI = 107.6, 129.6)

Outcome
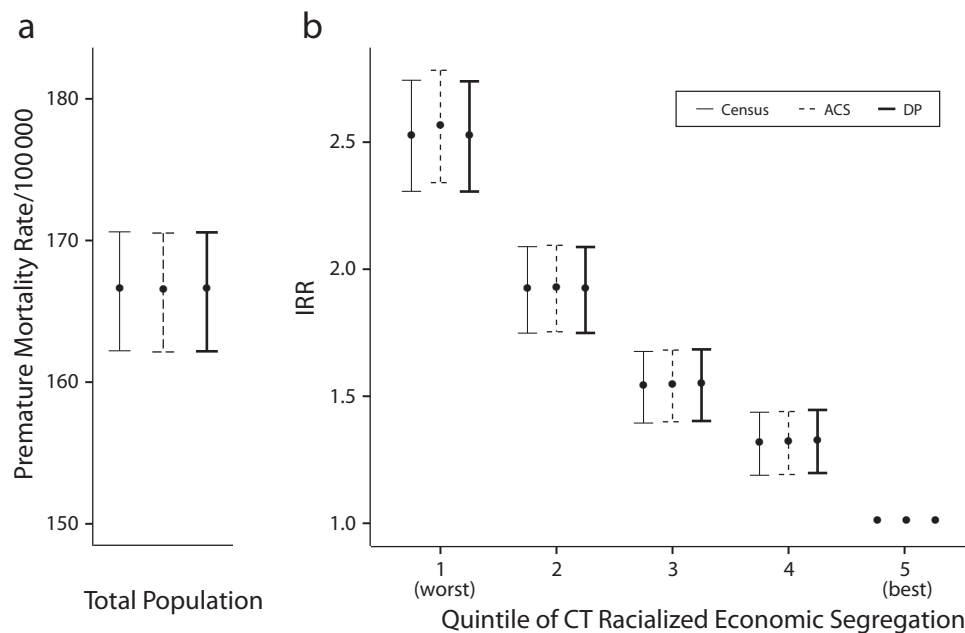
**FIGURE 1—** Population Health Estimates of (a) Premature Mortality Rate (Death Before Age 65 Years) and (b) Incidence Rate Ratio (IRR) for Premature Mortality by Quintile for Racialized Economic Segregation: Massachusetts, 2008–2012, Using 3 Different Sources of Census Tract (CT) Data: 2010 Decennial Census (DC), 2010 DC With Differential Privacy (DP), and 2008–2012 5-Year Estimate From the American Community Survey (ACS)

and virtually identical for men (214.2 [95% CI = 199.0, 229.4]; 214.2 [95% CI = 198.9, 229.4]; and 214.0 [95% CI = 198.8, 229.3]); for both groups, the range in rates across the ICE quintiles was likewise similar across the 3 denominator sources (Table C). Sensitivity analyses of inequities by the CT poverty level yielded similar results across the 3 denominator sources (Table C).

## DISCUSSION

Our study, the first, to our knowledge, to compare estimates of premature mortality rates and inequities in this outcome using CT denominators obtained from the 2010 DC, the 2010 DC with DP, and the 2008–2012 5-year estimate ACS data, provides novel evidence that these estimates—at least in the state of Massachusetts—are robust to the source of denominator data employed. This finding held when we aggregated

across the total population, and also when we stratified by race/ethnicity, and by gender.

One key limitation of our study concerns generalizability. Additional research should investigate whether similar results are obtained for other states, for other small geographic units (especially those not nested within counties; e.g., American Indian areas), and different health outcomes (e.g., morbidity, health practices, and cause-specific mortality) as expressed across the life course (e.g., from infancy to among the elderly). An additional limitation is that our study did not statistically account for spatial correlation among CTs or the available margins of error for ACS estimates[7]; this is a focus of our ongoing work.

In summary, our results provide initial evidence that monitoring of population health and health inequities using aggregated CT-level population denominators may not be adversely affected by

the impending shift to use of differentially private census data, starting with the 2020 decennial census. AJPH

## ABOUT THE AUTHORS

Nancy Krieger, Jarvis T. Chen, Pamela D. Waterman, Emily Wright, and Tamara Rushovich are with the Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA. Rachel C. Nethery is with the Department of Biostatistics, Harvard T. H. Chan School of Public Health. Brent A. Coull is with the Department of Biostatistics and Department of Environmental Health, Harvard T. H. Chan School of Public Health.

## CORRESPONDENCE

Correspondence should be sent to Nancy Krieger, PhD, Professor of Social Epidemiology, American Cancer Society Clinical Research Professor, Department of Social and Behavioral Sciences (Kresge 717), Harvard T. H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115 (e-mail: nkrieger@hsph.harvard.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

community survey) for monitoring health inequities. *Am J Public Health*. 2021;111(2):265–268.

## CONTRIBUTORS

N. Krieger conceptualized the study and led the drafting of the article. N. Krieger, R. C. Nethery, J. T. Chen, and B. A. Coull designed the statistical analyses. P. D. Waterman assembled the data. R. C. Nethery and J. T. Chen led conduct of the analyses. E. Wright and T. Rushovich assisted with analyses. All authors contributed to the article's content and approved the final version for submission.

## ACKNOWLEDGMENTS

## CONFLICTS OF INTEREST

None of the authors have any conflicts of interest to declare.

## HUMAN PARTICIPANT PROTECTION

Massachusetts Department of Public Health institutional review board approval: 946302; Harvard T. H. Chan School of Public Health institutional review board approval: IRB16-1325 (expedited).

## REFERENCES

1. Krieger N. The US Census and the people's health: public health engagement from enslavement and "Indians not taxed" to census tracts and health equity (1790–2018). *Am J Public Health*. 2019;109(8):1092–1100. https://doi.org/10.2105/AJPH.2019.305017

2. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*. 2005;95(2):312–323. https://doi.org/10.2105/AJPH.2003.032482

3. Gourevitch MN, Athens JK, Levine SE, Kleiman N, Thorpe LE. City-level measures of health, health determinants, and equity to foster population health improvement: the City Health Dashboard. *Am J Public Health*. 2019;109(4):585–592. https://doi.org/10.2105/AJPH.2018.304903

4. US Census Bureau. Disclosure avoidance and the 2020 Census. Available at: https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html. Accessed September 3, 2020.

5. Committee on National Statistics, The National Academies of Sciences, Engineering, and Medicine. Workshop on 2020 Census Data Products: Data Needs and Privacy Considerations. Washington, DC: December 11–12, 2019. Available at: https://sites.nationalacademies.org/DBASSE/CNSTAT/DBASSE_196518. Accessed September 3, 2020.

6. US Census Bureau. American Community Survey. Available at: https://www.census.gov/programs-surveys/acs. Accessed September 3, 2020.

7. Folch DC, Arribas-Bel D, Koschinsky J, Spielman SE. Spatial variation in the quality of American Community Survey estimates. *Demography*. 2016;53(5):1535–1554. https://doi.org/10.1007/s13524-016-0499-1

8. US Census. 2010 Census data products: United States–at a glance. Available at: https://www.census.gov/programs-surveys/decennial-census/guidance/2010/2010-data-products-at-a-glance.html. Accessed September 3, 2020.

9. Arguillas F. CISER Census 2010 Demographic and Housing Demonstration: Version 1.0. Ithaca, NY: CISER; 2019. https://doi.org/10.6077/fe6a-f789

10. Walker K. Tidycensus: load US Census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames. R package version 0.9.9.2. 2020. Available at: https://CRAN.R-project.org/package=tidycensus. Accessed September 3, 2020.

11. Massachusetts Department of Public Health. Registry of Vital Records and Statistics. Available at: https://www.mass.gov/orgs/registry-of-vital-records-and-statistics. Accessed September 3, 2020.

12. Krieger N, Waterman PD, Spasojevic J, Li W, Maduro G, Van Wye G. Public health monitoring of privilege and deprivation using the index of concentration at the extremes. *Am J Public Health*. 2016;106(2):256–263. https://doi.org/10.2105/AJPH.2015.302955