**Differential privacy in the 2020 Census and the distortion of COVID-19 mortality rates**

Mathew E. Hauer1,2\* & Alexis R. Santos-Lozada3,4

**Affiliations:**

1Department of Sociology, Florida State University, Tallahassee, Florida, USA.

2Center for Demography and Population Health, Florida State University.

3Human Development and Family Studies, Pennsylvania State University, State College, PA, USA.

4Population Research Institute, Pennsylvania State University.

\*[mehauer@fsu.edu](mailto:hauer@uga.edu)

As the coronavirus disease 2019 (COVID-19) grips the globe, scientists, policy makers, and journalists calculate mortality rates to better understand, communicate, address, and inform mitigation efforts of the COVID-19 pandemic. Because of these mortality rate calculations, we know that the elderly are more susceptible to COVID-19 related mortality and varying age structures play a role in how the pandemic will unfold (Dowd et al., 2020). Age-specific mortality rates of high importance due to the concentration of COVID-19 deaths at older ages. Accurate mortality rate calculations and estimates are thus paramount to managing this pandemic and illuminating how to manage future pandemics.

The calculation of mortality rates is relatively straightforward: one divides the count of deaths by the population counts. Mortality rates rely on the accuracy of both of these counts. Population counts - the denominators for calculating mortality rates - will be subject to noise injection in the United States when differential privacy (DP) is implemented, as initially proposed, as a disclosure avoidance system starting with the 2020 Census tabulations (Mervis, 2019). These are the very products used to calculated COVID-19 incidence and mortality rates. A recent study has concluded that the implementation of DP will substantially reduce our understanding of mortality dynamics particularly for small and rural areas, and racial/ethnic minorities (Santos-Lozada, Howard, & Verdery, 2020) The extent to which the implementation of DP, as initially proposed, would distort the calculation of pandemic related mortality rates is currently untested. Given how crucial population counts are for the evaluation, tracking and \_\_\_\_.

To estimate the extent to which DP will distort pandemic related mortality rates, we combine the US Census Bureau’s demonstration products (National Historical GIS, 2019) with empirical COVID-19 age and sex mortality curves from Italy (Dowd et al., 2020). This allows us to simulate how different mortality rate calculations using counts produce with DP could be from population counts produced using current methods.

Figure 1 shows age-sex structures from 2010 population counts (gray pyramids) and counts resulting from the implementation of DP for six counties (in red). In Panels A and B, the changes in age-sex structure is negligible or marginal. The rest of the cases are instances where if not the whole pyramid, then at least one category is substantially altered. In Panel C, this is the case for persons aged 60-69 where there is an under reporting of females and a substantial over reporting of males, the contrary happens to persons aged 80 + and 30-39 years. In Panel D, we observe shifts for all except for those aged 50-59 years, with higher differences observed for older age groups - with un under representation of persons aged 60-69 years and over representation of those aged 70-79 years and 80 +. In Panels E and F, we observe shifts in the population at oldest age group. We ask how the implementation of DP would affect an assessment of a pandemic with similar patterns to those as COVID-19.

Figure 2 shows the distortion of COVID-19 age-sex specific mortality rates by population size for US counties, using the 2010 demonstration products. We calculate absolute error for each county-age-sex combination. The absolute error is a measure provides a measure of overall accuracy in the estimation process by not allowing negative and positive errors to balance each other (Murdock & Ellis, 1991). We find that counties with \_\_\_\_\_\_\_\_ population have higher absolute errors than bigger populations. Of particular concern is that these errors are not limited to small areas or a single age group. We find errors exceeding 100% for every age-sex combination.

**BALANCING DATA PRIVACY AND UTILITY**

We highlight how the planned, noise-infused U.S. Census data will significantly alter our understanding of a pandemic similar to COVID-19 via noisy mortality rates. Using age-sex specific COVID-19 mortality curves from Italy and Wuhan, we show that differential privacy will introduce significant errors in COVID-19 expected age-sex specific mortality rates – sometimes causing age-specific mortality rates to exceed 100% -- hindering our ability to understand the pandemic. These errors are particularly large for the nearly 43% of county age-sex groupings containing fewer than 1000 persons. Overall, differential privacy will introduce significant challenges in our understanding of mortality amid a global pandemic. Despite the

In order to implement a disclosure avoidance system the US Census needs to consider how this change will affect our understanding of population-level phenomenon.

**References**

Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., … Mills, M. C. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9696–9698. https://doi.org/10.1073/pnas.2004911117

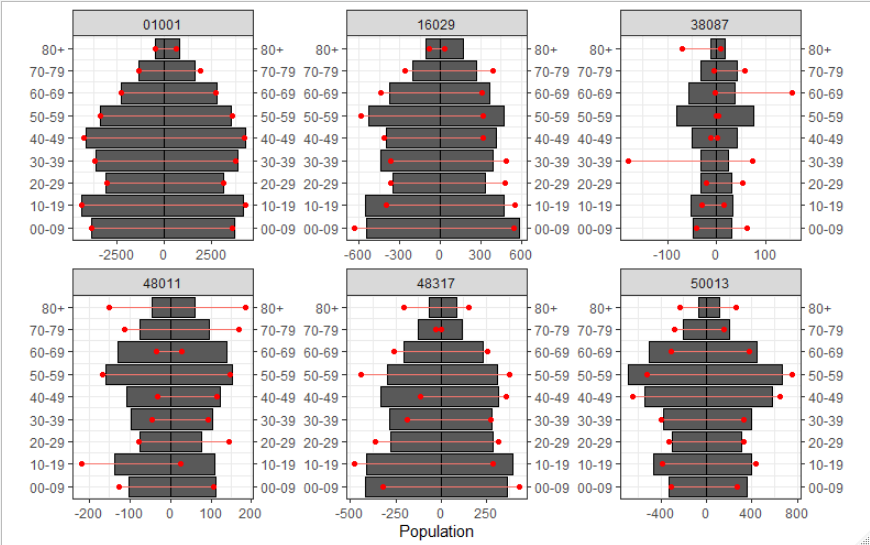
Mervis, J. (2019). Can a set of equations keep U.S. census data private? *Science*. https://doi.org/10.1126/science.aaw5470

Murdock, S. H., & Ellis, D. R. (1991). *Applied Demography: An Introduction to Basic Concepts, Methods, and Data* (1st ed.). Westview Press.

National Historical GIS. (2019). Differentially Private 2010 Census Data. Retrieved from https://www.nhgis.org/differentially-private-2010-census-data

Santos-Lozada, A. R., Howard, J. T., & Verdery, A. M. (2020). How differential privacy will affect our understanding of health disparities in the United States, 1–8. https://doi.org/10.1073/pnas.2003714117

**Figure 1** Changes in age-sex structures due to the implementation of differential privacy in 2010 U.S. Census data



**Figure 2** shows the distortion of COVID-19 age-sex specific mortality rates for US counties.

A picture containing text, map, table, bird

Description automatically generated