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

Mathew E. Hauer1,2\* & Alexis R. Santos3,4

**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.

3Penn State Population Center, Pennsylvania State University.

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

**Abstract**

Scientists, policy makers, and journalists rely on accurate population and mortality data to address and inform mitigation efforts regarding the coronavirus disease 2019 (COVID-19) pandemic, with age-specific mortality rates of high importance due to the concentration of COVID-19 deaths at older ages. Population counts – the principal denominators for calculating age-specific mortality rates –will be subject to noise injection in the United States with the 2020 Census via differential privacy. We highlight how the planned, noise-infused U.S. Census data will significantly alter our understanding of the COVID-19 pandemic 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 COVID-19 mortality amid this global pandemic.

**Main**

As the 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 [CITES] and varying age structures play a role in how the pandemic will unfold. 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 simply divides the count of deaths by the count of population. Mortality rates rely on the accuracy of both the death count and the population count. The US Census Bureau will implement a new privacy protection algorithm, called Differential Privacy (DP) to infuse noise into decennial census products – the very products used to calculate COVID-19 incidence and mortality rates – in order to further protect the data of Census respondents. The extent to which DP will distort COVID-19 mortality rates is currently untested.

To estimate the extent to which DP will distort COVID-19 mortality rates, we combine the Census Bureau’s DP demonstration product [CITES] with empirical COVID-19 age and sex mortality curves from Italy [CITES]. This allows us to simulate how different mortality rate calculations using noise infused population counts could be from using true population counts.

**Figure 1** shows the distortion of COVID-19 age-sex specific mortality rates for US counties.

A picture containing text, map, table, bird

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Pop. Size** | **MAPE** | **n** | **% of n** |
| < 1K | 40.9% | 17,950 | 42.9% |
| < 2K | 29.8% | 26,372 | 63.1% |
| < 3K | 26.1% | 30,635 | 73.3% |
| < 4K | 24.5% | 32,936 | 78.8% |
| < 5K | 23.6% | 34,379 | 82.2% |
| < 10K | 21.6% | 37,711 | 90.2% |
| < 25K | 20.3% | 40,171 | 96.1% |
| < 50K | 19.9% | 41,139 | 98.4% |
| < 100K | 19.6% | 41,619 | 99.5% |
| < 500K | 19.5% | 41,805 | 100.0% |
| < 771K | 19.5% | 41,812 | 100.0% |