Associations between COVID-19 Mortality and Public Health Factors in Texas

Michael Tsang, Kevin Jin, Mingyu Sun

02/25/2022

Dataset Information:

Our data set comes from the COVID-19 Data Repository, maintained by the Johns Hopkins Center for Systems Science and Engineering. It is a subset of the whole data set; the portion we will cover contains information about the number of cases and deaths due to COVID-19 in the different counties of Texas. Variables in the database include:

- County
- Confirmed cases
- Deaths
- Active cases (= total cases - total recovered - total deaths)
- Recovered cases
  - *Note*: JHU stopped recording state recovered cases on 3/7/2021 and global recovered cases on 8/4/2021. They determined both metrics to be estimates and generally unreliable.
- Case fatality ratio (= Number recorded deaths / Number cases)
- Incident rate (cases per 100,000 persons)
- Total number who have been tested
- Testing rate (= Total test results (Positive + Negative) per 100,000 persons)
- Hospitalization rate (= Total number hospitalized / Number cases)

Background and Motivation:

The COVID-19 pandemic began in China in December 2019. It has resulted in over 420 million cases and 5.91 million deaths around the world, and has devastated the global economy. People cannot travel as usual because of the limiting of flights. Many people lost their families and jobs. People's lives have dramatically changed. Our motivation is to understand the factors that influence COVID cases and deaths that could help us prevent a global crisis like COVID from happening again.

Locally, the impact of COVID can be felt by many throughout all counties in Texas; however, not all areas are affected equally. In addition to investigating which counties had the most cases, our group will also be looking for correlations between COVID deaths and pre-existing health conditions, as well as public health measures such as mask usage and vaccinations. If time allows, we also wish to build a model to predict future cases and deaths in the state of Texas based on existing data.

Potentially Interconnected Questions (3-4 per person):

1. Which counties in Texas were most affected by Covid-19 during a certain time period?

2. Is there a correlation between covid deaths and pre-existing health conditions (heart conditions, obesity, smoking?
    - What about sex? Age? Socioeconomic status?
    - Which county or counties in Texas have the highest number of pre-existing health conditions? How have they fared under the pandemic compared to counties with relatively healthier people?
    - What kind of people are more likely to be contagious COVID? Do they have the previous disease? Does the previous disease make them more serious?
3. Is the rollout of vaccinations correlated with a decrease in cases/hospitalizations/deaths when considering pre-existing health conditions? (Vaccine data)
    - What about mask usage?
4. "Excess deaths" refers to a phenomenon when more deaths, regardless of cause, are observed than would normally be expected during a period of time. It is commonly observed during major events such as war, pandemics, and natural disasters. Does Texas have more or fewer excess deaths compared to the rest of the country? Or the world?
5. Is there a way to possibly predict the future course of the pandemic (e.g. future case numbers, future deaths) based on modeling the cases and deaths we have seen so far?
    - What models could be used to perform such inferences?
    - By analyzing the data, can we predict the trend in Texas? Is that going to increase or decrease in the future?
6. The pandemic in the US appears to occur most aggressively in "waves" of variants (Alpha, Delta, Omicron, etc.). Judging from hospitalization and death rates compared to total case numbers, has the virus appeared to grow more or less lethal over time?
    - Does Texas ever appear to be spawning its own variant? That is, does the virus seem more lethal here than elsewhere in the US or even the world?
    - If yes or no, does one separate that possibility from the increased or decreased lethality being caused by pre-existing health conditions instead?
7. What should people do to prevent a global crisis like COVID from happening again by analyzing the data?