An In-Depth Analysis of the Spread of COVID-19 and Risk Modeling

By: Connor Schwarz

Abstract:

SARS-COV-2 is a highly contagious, deadly virus currently on the rise in the United States and around the world. To understand this virus and its effects, a probit model was chosen to be applied to a data simulation to outline the most important risk factors. This report utilizes current data supplied from around the world to model both the spread of COVID-19 in communities across the United States, the factors that play into the spread, and the deadliest factors of patient presentation.

Introduction:

Current data of the spread of SARS-COV-2 shows exponential growth in communities, especially in the United States. This report focuses on the reasoning behind the spread and model creation for mortality rates given presenting conditions.

Data:

In this report, the data utilized come from reputable sources such as John Hopkins, OWID data collection, and data given by Chinese scientists early on in the pandemic outlining patient presentation.

Three problems were confronted: Total spread across the United States, reasons for higher transmission in certain communities, and the indicators for COVID mortality. John Hopkins and OWID data sets consisted of over 50 variables, however, the interest was in total cases, collection date, and health care ranking (outside sourcing). The data set from various independent Chinese studies was presented as a collection that outlined the percentage of patients presenting with a broad range of conditions.

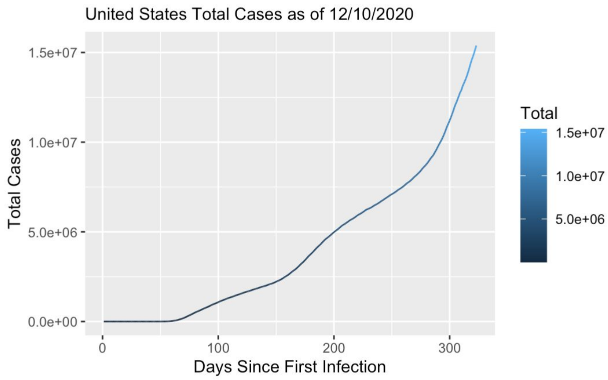
Model:

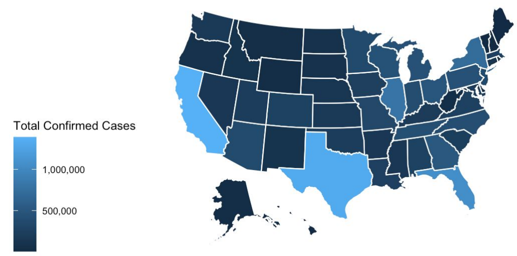
Given the binary nature of the dependent variable, a probit regression model was used to determine the statistical significance of each independent variable. A logit regression model could have also been used, however, the difference between models is negligible.

No changes were made to the model. Following its completion, the probit model allowed for an accurate distribution of mortality given age.

Results:

The first portion of the code is to give an accurate representation of total cases over time in the United States. The graph shows the difference in the rise of the first wave and the current second wave (fig. 1). The first wave showed a slower, drawn-out rise in cases. The second wave showed an exponential increase in the first days. The reasoning for the significant difference is the relaxed restrictions and increased travel due to the holiday season. Figure 2 outlines the states with the most cases (light blue).

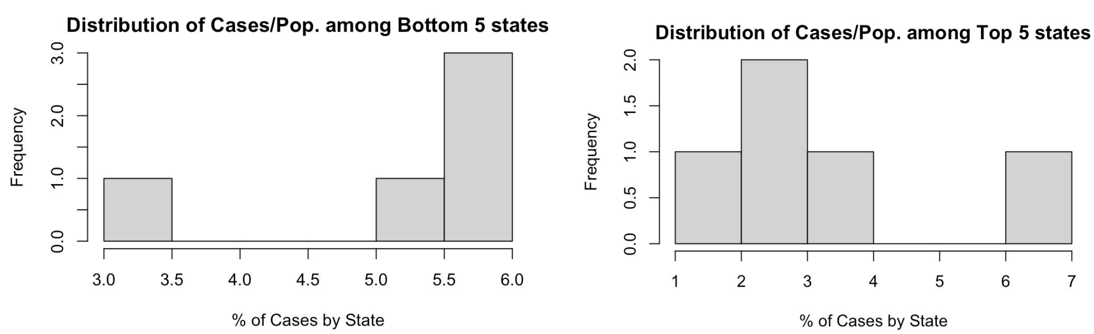




(Figure 1) (Figure 2)

The second part of the report details the correlation between the ability to provide quality health care to the general population and the spread of SARS-COV-2. Using health care ranking as a base comparison, the top 5 states have a lower case percentage (total cases/pop. size), compared to the bottom 5 states. The top 5 had a mean of 3.34% compared to the bottom 5 of 5.1%. The difference is statistically significant, demonstrating the effects of poor health care contributing to the spread of COVID-19. See Figure 3.

Figure 3



The final part analyzes clinical data sourced from Chinese studies at the beginning of the SARS-COV-2 epidemic. Each study outlined the percentage of patients presenting with certain conditions as well as mortality. These percentages were used as a method of data simulation, with each percentage allowing for data randomization and population. For variables presenting with a mean, a Poisson distribution[lambda = mean, n =1000] was applied to create an accurate distribution.

Using the data simulation, a probit regression analysis was used to determine the greatest risk factors in patient presentation. Following its completion, the greatest risk factors were:

-Hypertension

-Age

-Invasive Mechanical Ventilation

-ARDS

-White Blood Cell Count

To represent the data, Age was compared to Mortality using the PDF determined by the model (Figure 4). To describe ARDS, a table was used to show the relationship between patient presentation with ARDS and mortality. (Figure 5)

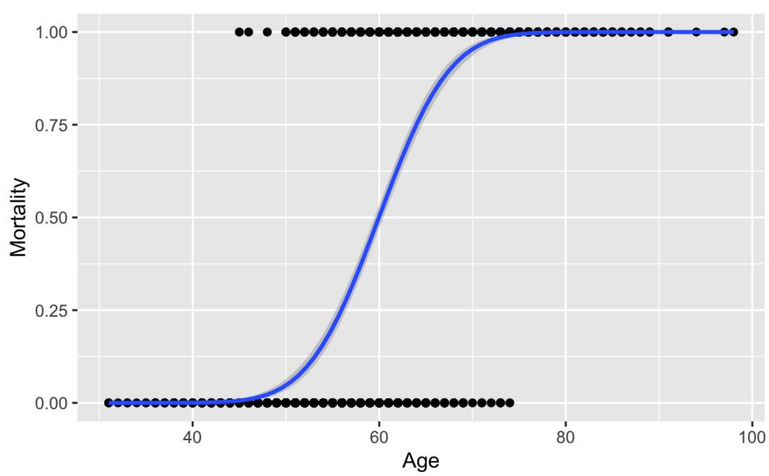


Figure 4

Mortality given ARDS Presentation (Figure 5)

	ARDS = 0	ARDS = 1
Mortality =0	933	67
Mortality =1	80	920

Average Mortality with ARDS = 93%

Ablation:

The introduction of a larger, more comprehensive data set may provide a stronger statistical significance to patient presentation pre existing conditions. For example, Diabetes is considered to be a condition influencing mortality outcomes in patients. However, the probit analysis failed to return any statistical significance for Diabetes. The reasoning may be due to the lower prevalence of diabetes in the sample group which consisted of only Asians (11.2% prevalence in the general population).

Literature Review:

All literature reviewed in this report was data supplied directly from the CDC, John Hopkins, and Chinese Peer Review. In my review, most variables tested remained consistent across all sources. Specifically, studies of contagious periods for lab-confirmed positive cases showed the highest rates of transmission on day 6 but dropped dramatically on day 7-8. This data is the reasoning behind the 10 day isolation period for non-hospitalized cases.

Chinese peer review supplied the most detailed information regarding patient presentation with its sources corroborating its claims.

Conclusion:

SARS-COV-2 can spread rapidly and discretely. With high numbers of asymptomatic cases, which do not prompt an urge to seek testing in a COVID positive person, it can spread in low restriction communities. These factors are the causes of the data presented, especially in regards to the exponential growth of COVID cases in the second wave.

The use of probit modeling has shown that COVID patient mortality can be largely associated with the following conditions: ARDS, hypertension, WBCC, Age, and IMV. Age can be shown to have an extreme influence on mortality, with significant probability changes in the age range of 60-70. Lower age groups seem to be largely unaffected and present to hospitals at a significantly lower rate than older age groups (50+).