#### **[GLBL550] Final Report: Healthcare System Resilience and COVID-19**

Kendra Nealon, Ryohei Suzuki

**1. Introduction**

Today, the world is facing a difficult time with the COVID-19 outbreak, but the severity of the crisis varies from country to country and region to region. While some countries are beginning to see cause for hope, others are still suffering from a fast-increasing death rate. Even within a set of countries or regions with similar cultural backgrounds and social norms, the numbers of cases and death toll still differ greatly. There is no established consensus on what causes this variation, but one of the main determinants is the preparedness and resilience of each country.

In this project, we discuss the relationship between healthcare system resilience and COVID-19 outbreak in each country. While doing our research, we answered the following three questions.

1) What is the current scope of the pandemic?

* This section compares the number of known cases as of April 17 to population size, GDP per capita, as well as percentage of elderly people, smokers, and overweight adults.

2) How prepared are the countries for when the pandemic gets worse?

* We looked at the number of available health care workers and hospital beds available in comparison to those required to effectively treat the virus.

3) How does the capacity of the medical system relate to the number of deaths due to coronavirus?

* This section consists of a regression analysis of correlation between medical capacity (number of ICU beds and healthcare workers available) and the number of deaths to date.

We based our project on a dataset from the World Health Organization and use Python in visualizing and analysing the data. To narrow our research, we focused on countries in the EU.

**2. Data**

2-1. Dataset

Our primary dataset is the WHO Essential Equipment Forecasting Tool, which forecasts essential equipment and workers needs by country, as well as the number of cases that each country could expect at its peak given a set of user inputs. In order to retrieve data from these countries, dataset users type in a country and number of cumulative known cases. The dataset then forecasts the number of cases over 4-6 weeks and number of beds, healthcare workers, and tests needed, as well as a long list of equipment requirements. The dataset provides information on each country’s population and current capacity of its healthcare system.

We also used Novel Coronavirus time series data on GitHub, which is based on the Johns Hopkins COVID trackers to input the WHO forecasting tool. As a supplement, some others on basic EU demographics can be referred to.

Primary Dataset:

* The World Health Organization’s COVID-19 Essential Resource Planning
* Novel Coronavirus 2019 time series data on cases on GitHub

Supplemental Datasets:

* World Bank GDP per capita levels
* Eurostat: European Statistics
* Johns Hopkins: Coronavirus Research Center

2-2. Sanity Checks and Limitations of the Dataset

The WHO dataset was developed to gauge the need for supplies/equipment and health workforce requirements during the COVID-19 pandemic, not to implement precise estimations. In particular, this dataset outputs estimated numbers of cases based on an epidemic curve with input from a user. So this dataset has some limitations particularly when used as a tool for estimating numbers of cases.

To build the curve, statisticians make many assumptions. Users must follow the default settings, even if there are some potential errors. For purposes of this project, we left the peak at 20% of the population size assuming that government policies may be issued before the virus gets much beyond that point. However, the true peak of the virus will likely differ greatly. The dataset also estimates a peak number of viruses in the short-term, within 4-6 weeks. This means that the dataset is not useful for long-term analysis.

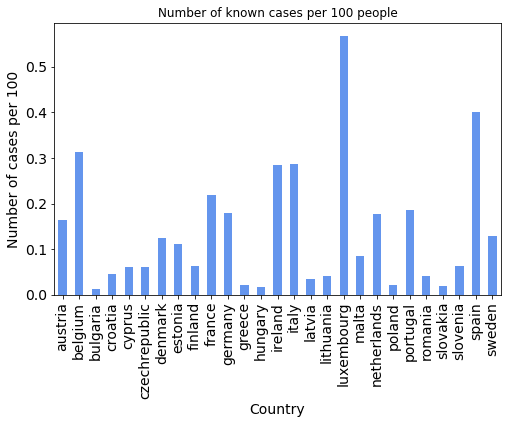
Lastly, the scope of the dataset is limited. It does not take into account demographics other than population size, so when doing more holistic analysis, we should not rely solely on this dataset. In addition, it does not reflect policy updates such as social isolation, closure of businesses, additional construction of hospitals and beds, and so on.

**3. Background Analysis**

3-1. Current Scope of the Pandemic

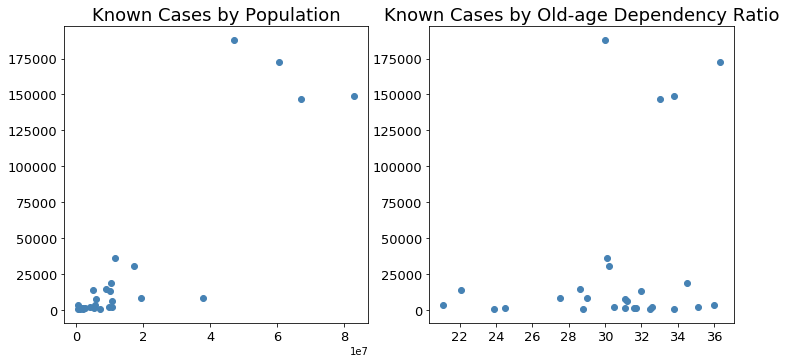
As of April 17 when we conducted this analysis, there were 834,894 known COVID-19 cases in the EU. 78.7% of those were concentrated in Spain, Italy, Germany, and France (in descending order). Though these countries had the highest number of cases, cases per capita are a more robust way to look at how the virus is affecting each country (Table 1). When considering this measure, Luxembourg, Belgium, and Ireland in addition to the aforementioned four, are having the most severe outbreaks. To note, we used the python library pandas for analysis and data visualization in this section and the following.

(Table 1)



We looked at the number of known cases as compared with each country’s population size, GDP per capita, and old-age dependency ratio to determine if these top four countries had any similar characteristics (table 2). Old-age dependency ratio is the ratio between the number of persons aged 65 and over per 100 persons of working age (15-64). We wanted to look into this because the number of known cases is equivalent to the number of tests performed. As of mid-April, tests were only being performed for people with severe cases, and the elderly are more likely to have serious cases of COVID-19.

(Table 2)

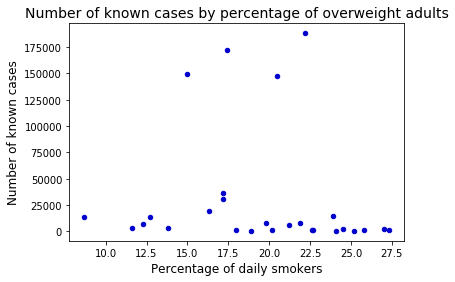


As expected, our analysis found that countries with the highest number of cases are also the most populous. These four countries tend to have a higher percentage of elderly people, but as do many other countries in the EU. We also created a similar scatter plot for GDP per capita, but we found no evidence that richer or poorer countries are correlated with having more cases.

There are hypotheses in the news and media that COVID-19 is more prevalent among those who are overweight and smoke often. Therefore, we performed two additional analyses:

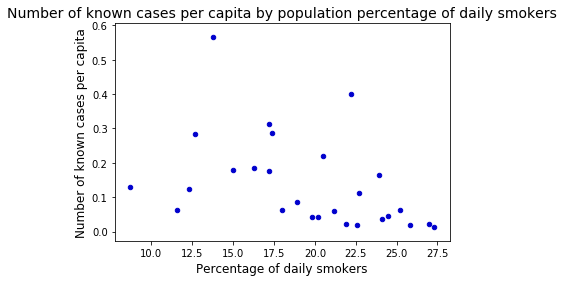
1. Number of cases and cases per capita to percentage of overweight adults in each country (Table 3): This data was inconclusive -- in fact two of the countries with the most number of cases (Italy, France) have the lowest percentage of overweight adults amongst all EU countries.[[1]](#footnote-1) It is important to note that many adults across all countries in the EU are overweight, so it may not be a differentiating factor. To add color, at the low end, 44.9% of Italy’s adults are overweight. At the high end, this is 61 percent for Malta.

(Table 3)



1. Number of cases and cases per capita to percentage of adults who admit to being daily smokers (Table 4): The percentage of adults who have at least one cigarette per day range from 8.7% in Sweden to 27.3% in Bulgaria. Similarly, this analysis did not show any correlation to the number of known cases.

(Table 4)

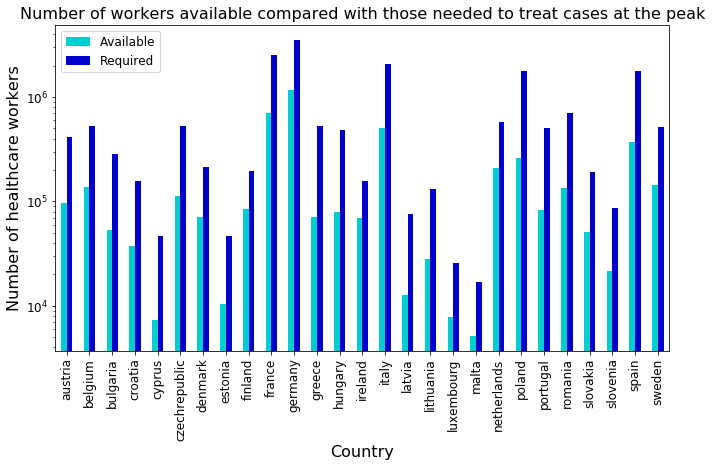


3-2. Preparedness

The WHO Essential Resource Planning dataset includes information on the current and required capacity of each country’s healthcare system in terms of number of workers available to treat COVID-19 cases and number of critical care beds. To note, they estimate the number of workers required by forecasting the number of severe and critical cases and using a WHO recommended worker to patient ratio. For our report, we examined:

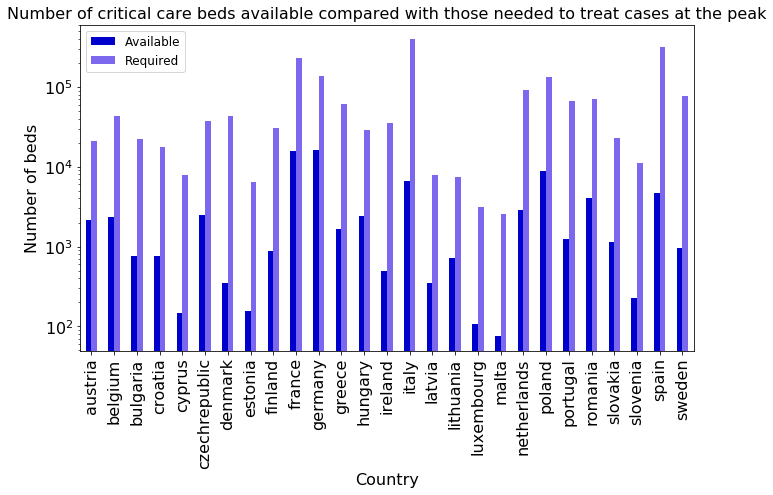
1. The number of healthcare workers required to treat the WHO forecasted peak number of cases to the number that are currently employed: All EU countries had far fewer healthcare workers than would be required at the peak. Table 5 is in log form to better display discrepancies, but the country that is closest to having the necessary number of workers is Ireland, and it still needs 2.2 times as many workers as it has. On the other end, Greece needs 7.3 times as many workers as it currently has.

(Table 5)



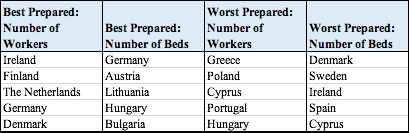
1. The number of critical care beds required to treat patients at the forecasted peak: Once again, all EU countries have fewer beds than would be required to handle the WHO forecasted peak. Table 6 demonstrates how many beds each country is lacking compared with how many it has, in log form. At the low end, Malta would require 2,570 more beds. On the high end, Italy needs close to 40,000 more. Compared with existing (as of April 17) capacity of beds, Denmark would need 3.9 times as many critical care beds as it has. On the low end, Germany needs 1.2 times as many.

(Table 6)



The five countries that are best and worst prepared in terms of required resource to existing capacity are:

(Table 7)



However, all countries in the EU lack both beds and healthcare workers. This paints a grim picture, as no country is truly prepared to handle the peak number of cases.

**4. Advanced Analysis**

4-1. Measures

As shown in the previous chapter, there are gaps between healthcare capacity available and that required. Among various factors which can be attributed to large numbers of deaths, healthcare capacity seems one of the most critical. In this chapter, we explore regression analyses to see the correlation or causal inference between healthcare capacity and numbers of deaths.

As for a statistical model, we used Ordinary Least Squares (OLS) regression, which is the most commonly used type of linear regression.

Using OLS, we looked for a coefficient (β) for each explanatory variable (x) so that the combinations of β’s and x’s best explain the dependent variable (y). To implement OLS regression, we use Statsmodels, a Python module for statistical analyses.

We implement two kinds of analyses. First, we focus on healthcare capacity itself without considering actual needs. Second, we analyze the impact of gaps between availability and needs on actual numbers of deaths.

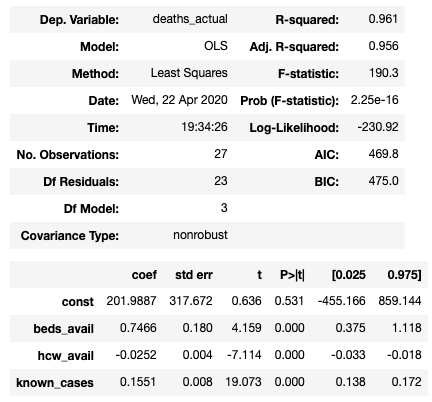
4-2. Analysis 1: Healthcare Capacity

In our first regression analysis, we regress the actual number of deaths on the indicators of healthcare capacity without considering what would be required for the peak of the crisis. We pick up the numbers of available beds and available healthcare workers as explanatory variables from the WHO dataset. We also choose known cases as of April 17 as a control variable.

* Dependent variable: actual numbers of deaths as of April 17 (death\_actual)
* Explanatory variable 1: Number of available beds (beds\_avail)
* Explanatory variable 2: Number of available healthcare workers (hcw\_avail)
* Control variable: Number of known cases as of April 17 (known\_cases)

The result is shown in table 8. The availability of healthcare workers decreases numbers of deaths, while availability of beds increases it. Given that t-statistic for beds\_avail is above 1.96 and that for hcw\_avail is below -1.96, the result is significant. Also, R-squared is above 90 percent, so this model explains the observatory data well. The result for healthcare worker availability is expected, but bed availability is against our intuition. This unnatural result might be attributed to our choice of explanatory variables, so we explored another regression analysis.

(Table 8)



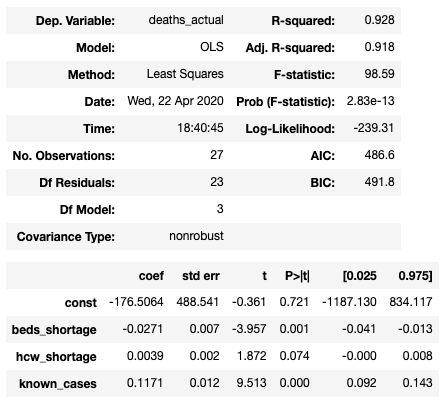
4-3. Analysis 2: Gap between Availability and Needs

Next, we regressed the actual number of deaths on the indicators of shortage of healthcare capacity. While we kept the same dependent variable and control variable as the previous analysis, we used the shortage of beds and healthcare workers as explanatory variables. This describes the differences between the available numbers of beds and workers and the required numbers to handle the peak.

* Dependent variable: actual numbers of deaths as of April 17 (death\_actual)
* Explanatory variable 1: Number of shortage of beds (beds\_shortage)
* Explanatory variable 2: Number of shortage of healthcare workers (hcw\_shortage)
* Control variable: Number of known cases as of April 17 (known\_cases)

The result is shown in table 9, but the conclusion is similar to the previous one: a shortage of healthcare workers increases numbers of deaths, while shortage of beds decreases it. With the high R-squared, the model is well explanatory, but t-statistic for hcw\_shortage is below 1.96, meaning the result about healthcare workers is significant.

(Table 9)



4-4. Limitations on our Analyses

These unexpected results can be explained by some factors associated with limitations of our analyses. Particularly, data reliability and omitted variable bias are the most decisive.

*Data Reliability*

There are some potential problems which result from the nature of our dataset. First, some data items are estimations based on user inputs and assumptions. This is significantly problematic in our second regression. The shortage amounts of beds and healthcare workers are calculated by taking the differences between the available numbers and estimated numbers of needs. The latter is susceptible to data fluctuations caused by the input/assumption problem. Second, as mentioned above, it does not take emergent enhancement measures into consideration, so the data on healthcare capacity might not represent the actual situation with government intervention. Lastly, the current outbreak is ongoing, and countries can be in different stages (growth, plateau, decline). Aware that this dataset is not appropriate in historical comparison, we decide not to do historical comparisons. This might be one of the main causes of the unnaturality.

*Omitted variables*

As there are a wide range of factors that affect the numbers of deaths, omitted variable bias is inevitable, especially when relying on a single dataset as we do in this project. For example, given the high mortality rate of the coronavirus in the elderly generation, elderly population seems decisive, but it is not within the scope of the dataset and our analyses. Even the ratio of people who wear a mask can be an important variable. Those variables, however, are omitted in our regression analyses.

Similarly, a problem of multicollinearity can exist. The availability of beds and healthcare workers is likely to be highly linearly related. As this project does not focus on a statistical aspect, we make our model very simple, but a more elaborated model would be helpful in making our regression analyses more convincing.

**5. Conclusion**

As COVID-19 leaves a destructive path through Europe, it is clear why these countries are struggling to such an extreme extent: Their healthcare systems do not have adequate capacity. Within the scope of our current analysis of demographics, it is not possible to explain why certain countries have been hardest hit -- among the variables that we studied (population, GDP per capita, old-age dependency ratio, percentage of daily smokers and overweight adults), only population size correlates with the number of known cases.

However, the preparedness of each country in terms of beds and healthcare workers able to treat the peak number of COVID-19 patients is grim. No country in the EU has enough of these resources. Since performing this analysis, we hope that policy changes have since been issued, in terms of social isolation measures and capacity building of healthcare systems.

From our regression analyses, we found that while the availability of healthcare workers plays an important role in reducing the number of death tolls, the availability of beds has an adverse impact. This unintuitive result suggests that our dataset and/or regression method are likely to have some limitations. Python is a powerful tool in visualizing and analyzing data, but it is important for convincing and reasonable analyses that analysts themselves are familiar with statistical techniques as well as the characteristics of the dataset they are using.

*Sources*

Datasets:

* The World Health Organization’s Coronavirus disease (COVID-19) technical guidance: Essential resource planning:  
  <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/covid-19-critical-items>
* Novel Coronavirus 2019 time series data on cases on GitHub:  
  <https://github.com/datasets/covid-19>
* World Bank GDP per capita levels: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
* Eurostat: European Statistics: <https://ec.europa.eu/eurostat/data/database>
  + Data on overweight adults: <https://ec.europa.eu/eurostat/web/products-datasets/product?code=sdg_02_10>
* Johns Hopkins Coronavirus Research Center: <https://coronavirus.jhu.edu/map.html>
* WHO Tobacco Free Initiative (TFI)

<https://www.who.int/tobacco/publications/surveillance/reportontrendstobaccosmoking/en/>

1. Overweight described as having a BMI above 25. Data as of 2014. [↑](#footnote-ref-1)