THE UNIVERSITY OF SYDNEY

COVID-19 Analysis Report

Content pages: 5

Introduction

As the Novel Coronavirus COVID-19 (2019-nCoV) occurred across the world, a phenomenon has been found that death rates are vastly distinct in different countries. Therefore, the objective of this report is to demonstrate "The country-specific determinants of death rates by COVID-19".

Assuming the data is reliable and accurate. But in the real world, the reality of data is skeptical for several reasons. Firstly, the virus is primarily found and spreads unpredictable fast throughout countries, so data can be recorded inefficiency based on this situation. Secondly, due to different unfavorable conditions of each country, some deaths were not considered as dying from COVID and were not recorded in time. While some deaths were categorized into COVID, but it may because of other illness. Thirdly, testing methods as one of the determinants are distinctive across most countries. The number of testing positive cases will directly affect the death rate, and details will be discussed later.

In terms of death rates, Case Fatality Rate (CFR) is going to be discussed here. According to the research, CFR is the ratio between the number of confirmed deaths from the disease and the number of confirmed cases. In other words, CFR is related to the number of reported deaths per number of reported cases. Carefully, it has not the same meaning as the risk of death for an infected person. (ourworldindata.org)

$$\text{Case Fatality Rate (CFR, in \%)} = \frac{\text{Number of deaths from disease}}{\text{Number of diagnosed cases of disease}} * 100$$

As the CFR calculation formula demonstrates, it is clear that CFR is determined by the number of reported deaths from disease and the number of reported diagnosed cases of disease. Thus, based on these two aspects, there are three major determinants has been identified: Population structure, ICU capability and differences in number of people tested (Johns Hopkins Coronavirus Resource, 2020).

Firstly, regarding the determinant of population structure, according to the WHO statement: older people are at highest risk from COVID-19. More than 50% of all deaths were people aged 80 years or older. In addition, it is very crucial to know that the top 30 countries in European region have the largest percentage of older people, in other words, countries with the population ageing where they are suffering the higher risks of death from COVID-19. Because, most older people have at least one underlying comorbidity such as cardiovascular diseases/hypertension and diabetes etc.(euro.who.int)

Secondly, ICU capability is also a significant degerminator towards death rates of COVID-19. According to the research, the CFR are generally higher in strained health-care systems. As we know, ICU admissions are dependent on the severity of illness and the ICU capacity of the health-

care system (Jean-Louis Vincent, Fabio S Taccone). Therefore, it is highly possible that people died due to the lack of ICU beds. Importantly, increasing ICU capacity requires more equipment (e.g.Ventilators), consumables, and pharmaceuticals, which might be in short supply.

Last but not least, according to Dietrich Rothenbacher, who is the director of the Institute of Epidemiology and Medical Biometry at the University of Ulm in Germany. The major source of difference in death rates across the world is the lack of widespread, systematic testing in most countries (BBC.com). For instance, the death rate would be high when some countries only test patients with obvious symptoms and do not test the less-ill or even without symptoms Covid-19 patients who do not go to hospital. Conversely, the death rate can be lower in countries where testing is widespread (such as Germany or South Korea).

Data Management

The data of COVID-19 are sourced from The Coronavirus resource centre of Johns Hopkins University that are originally collected through the government official COVID-19 website of each country. There is a total of 24,024 observations reported daily at country level, some at state level, and the unit of data is positive and consistent here, measured by people. All points shown on the map are based on geographic centroids and are not representative of a specific address. Furthermore, according to the CDC guidelines in JHU data explanatory notes, total confirmed cases include presumptive positive and probable cases, while death cases in the US include confirmed and probable.

Firstly, pivoting the date features into values, because the original dataset shows every date in a feature structure. The accumulated COVID-19 data on April 21st at country level are using to do the analysis. The confirmed and death cases data are joined into one worksheet depending on the unique key, Country/Region. Then, provinces that do not belong to those countries has been deleted. For instance, the two cruises, Diamond Princess and Grand Princess, do not have geographical coordinates and are listed as provinces of Canada. Further, regarding the countries that have negative confirmed and death cases values have been ignored since those have no meaning according to our objective. Also, countries with 0 confirmed cases are dropped, because they are inconsistent with the death rate measure method. Finally, several countries have common characters in their name, and this will affect the accuracy of the searching result in the dashboard, thus we put country code after one of them to achieve uniqueness. For example, Nigeria includes the characters Niger, which is also a country, so we change Niger into Niger(NER), the result will be unique when searching for any one of them.

The population distribution by age groups data is sourced from 2020 Population estimates and projections. The metadata is estimated by The World Bank staff based on age or sex distributions of United Nations Population Division's World Population Prospects: 2019 Revision. In order to compare the population distribution of older age groups in different countries, we extract the

population of 174 countries that are common in COVID-19 dataset with series by 5-year age groups of 2019, in unit of percentage, a result of 5,742 observations selected. Within the data source, there is no data for the Female age group 00-04. Since the purpose is to compare older age groups of different countries, the missing values will not affect the accuracy of our result. Therefore, we keep it in the graph. In addition, the original data has long string values for each age group. Therefore, extracting the useful integers represents each age group using regular expressions. Also, the name for the age group "80 and older" is inconsistent in gender, so we unify them to "80 and above".

It seems fair to compare the ICU capacity in countries that have enough testing, sufficient statistics, and public critical care beds data. This filter includes 19 countries, as described in the excel data source file. To reflect the real situation of COVID19 treatment in various countries, the latest critical care beds data has been used.

The critical care beds data comes from journals, news, government statistical data. The Asian countries' data is from the US National Library of Medicine National Institutes of Health. Europe countries data is from the journal The variability of critical care bed numbers in Europe. Australia data is from the Medical Journal of Australia(MJA). Canada data is from the Canadian Medical Association Journal. United Kingdom data is from the National Health Service England(NHS). United States data is from United States Resource Availability for COVID-19 on The Society of Critical Care Medicine(SCCM). Sweden data is from journal Swedish emergency hospital surgical surge capacity to mass casualty incidents.

By observing the data, there is a finding that there could be a correlation between the confirmed cases divided by the critical care beds per capita and the death rate. Thus, we generate a feature named confirmed cases per ICU equaling to confirmed cases divided by the critical care beds.

In order to construct the interactive dashboard, we also join the population distribution data and critical care beds data to the existing worksheet. Therefore, there are a total of 10 features in our dataset, including Country/Region, Confirmed Cases, Death cases, Death rate, Lat and Long, Critical care beds, Confirmed cases per critical care beds, female and male population percentage and the age group. Note that the floats or percentages have been rounded to 2 decimals, and people in integers.

Interactive Dashboard (see Appendix)

An interactive dashboard is applied in the Tableau dashboard. Avoiding cluttering the dashboard, there is one map, two graphs and three BANs ("big-ass-numbers") created to encode six different insights. The map briefly shows a geographic information and the level of death rate of different countries; Two graphs, Population pyramid and the scatter plot, these illustrate the two major determinants of distinct death rates that discussed in this report: capacity of ICU beds and the

distribution of various age groups in a population. Finally, the three BANs display the Number of Confirmed cases, Number of deaths as well as Average death rate of COVID-19 according to each country.

All related information will automatically display by selecting any country on the map or scatter plot, and readers can compare different countries either in this way or using a filter button on the top right search bar, which allows looking for a specific country as well. Tooktip also gives relevant messages for quickly encoding. And the main color yellow is used to draw attention to a specific piece of information, which provides a more powerful visual effect to the audience.

Visual Implantations and retinal variables

Spatial Geography and 2-Dimension graphs are major encoding methods in clarifying our objective in this report. Polygons of map shows the area of countries, longitude and latitude gives a geographic location, and 1 colour hue x 5 values is used in map to exhibit the comparable death rate, and here, red indicates serve, yellow means mild.

The scatter plot uses consistent marker and colour but uses the value of Confirmed Cases per ICU bed to determine the size. As shown, the group of European countries on top right has both a high death rate and great lack of ICU beds.

Moreover, red colour is used for BANs to help quick read the main message of the whole dashboard. However, since there are many colors already used, we choose default grey and dark navy on the population pyramid to avoid messy but still tell the two categories of gender. Each bar represents a categorical variable, which is age group, and the length of bar indicates the level of percentage of population, longer is larger.

Graph Identification and graph enhancement tools

We only use the dashboard title to decode the dashboard objective and another subtitle to decode the scatter graph objective, axes labels are hidden simultaneously due to duplicate information. Furthermore, as the population pyramid is a traditional way to illustrate the population distribution by age groups, thus, the title would be redundant here. Besides, only one legend applied to our dashboard gives brief information about the death rate of each colour on the map. Once a country is selected, the legend would show a more accurate Death rate. In addition, since detailed information is only available when selecting the dots on the scatter graph, we label them by the purpose of encoding fast. The font size and color are all legible, and we use default color and size

to keep visual prominence low. Also, we use serif for small and crowded texts by reason of reducing visual fatigue, and others are san-serif.

There is a note showing at the bottom of the graph, which gives further explanations about the Data source, Missing values, Point Level, and Cases Facility Rate calculation formula. The note provides the audience a more convenient way to understand and encoding the graph. Thus, appropriate encoding in the dashboard can make the audience decode it fast and effortless.

Conclusion

In consequence, for those countries with a high death rate that are shown in the scatter plot, the older people take a large portion of their population. The representative country is Italy, which has a 13.4% death rate with over 9% of females falling in the "80 and above" age group. Others can be the United Kingdom, France, and so on. Moreover, the Critical Care beds are scary for those with high death rate countries, a phenomena of inefficiency treatment. For example, in Italy, each ICU bed has to service 24 confirmed cases in order to reduce the pressure of the medical system due to COVID-19, otherwise, confirmed cases with critical condition would have high probability to die due to the inadequate treatment. However, each ICU bed needs to serve 12 people in Canada, which has a death rate of 4.84%.

Some research also provides the same pattern as we found. Old people are more likely to be infected by the COVID19 virus than younger people (Abbatecola & Antonelli-Incalzi, 2020) and old people who get COVID19 are also more at risk to result in severe and death (Zhou et al., 2020). On the other hand, COVID19 has a high critical care rate, which is predicted to be 30% (Willan, King, Jeffery & Bienz, 2020). Furthermore, one study shows that some patients admitted to intensive care less than one day after hospital admission (Wu et al., 2020).

In conclusion, the countries with an outbreak of COVID19 lack the intensive care beds to treat severe patients can result in a high death rate. Although there is much evidence to prove that population structure could cause CFR distinct, we still need sufficient and accurate data of confirmed and death cases in different age groups to provide a more reliable conclusion. And the test method is definitely an important determinant causing a distinct death rate.

Reference

Abbatecola, A., & Antonelli-Incalzi, R. (2020). COVID-19 Spiraling of Frailty in Older Italian Patients. The Journal Of Nutrition, Health & Aging. doi: 10.1007/s12603-020-1357-9

Henriques, M. (2020). Coronavirus: Why death and mortality rates differ. Retrieved 28 April 2020, from https://www.bbc.com/future/article/20200401-coronavirus-why-death-and-mortality-rates-differ

Jean-Louis Vincent, Fabio S Taccone Understanding pathways to death in patients with COVID-19 The Lancet Respiratory Medicine, Available online 6 April 2020, Pages

Johns Hopkins University(2020). Johns Hopkins Coronavirus Resource Center. Mortality Analyses. Retrieved from http://coronavirus.jhu.edu/data/mortality

Statement – Older people are at highest risk from COVID-19, but all must act to prevent community spread. (2020). Retrieved 28 April 2020, from http://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/statements/statement-older-people-are-at-highest-risk-from-covid-19,-but-all-must-act-to-prevent-community-spread

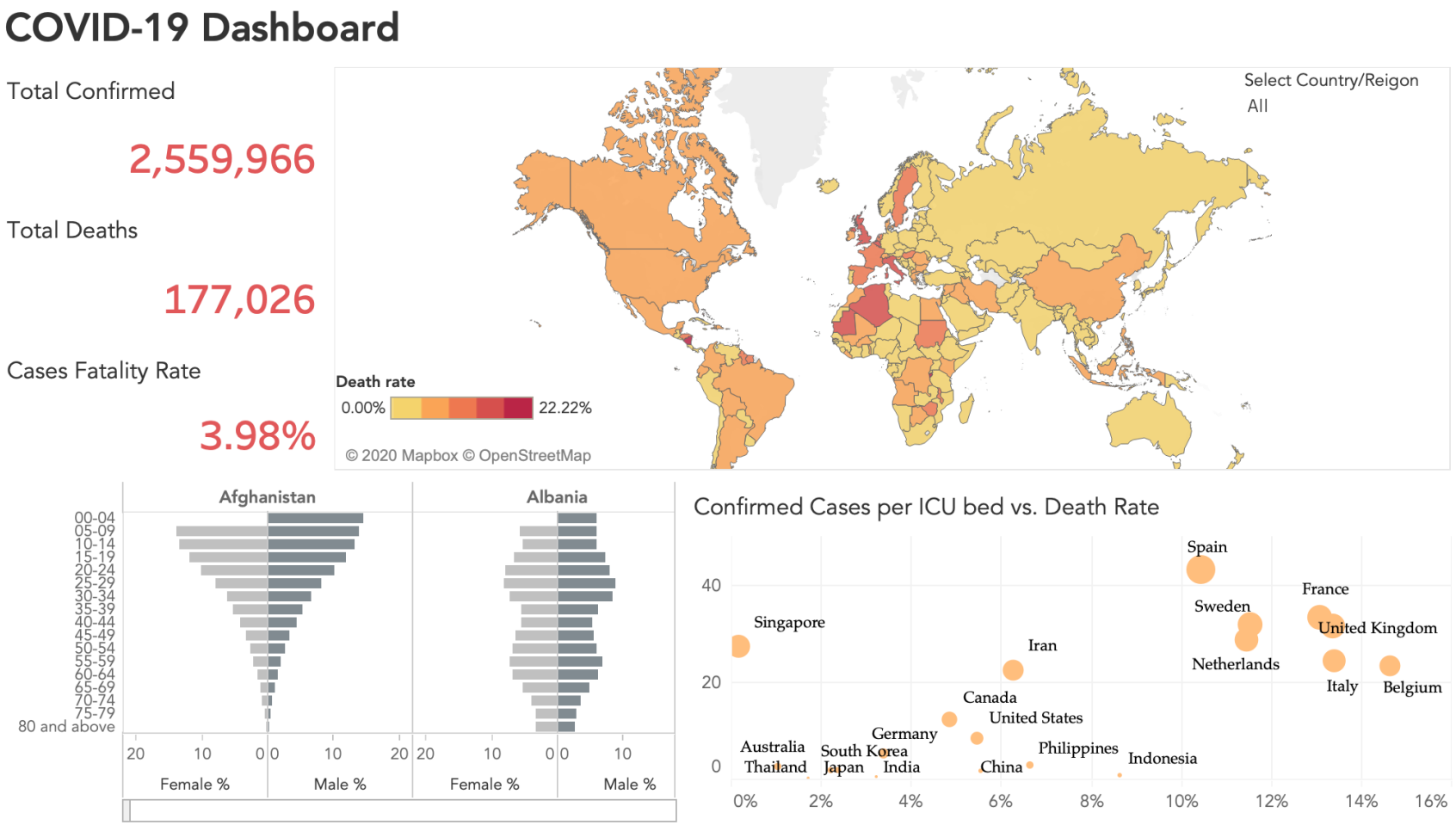
What do we know about the risk of dying from COVID-19?. (2020). Retrieved 28 April 2020, from https://ourworldindata.org/covid-mortality-risk

Willan, J., King, A., Jeffery, K., & Bienz, N. (2020). Challenges for NHS hospitals during covid-19 epidemic. BMJ, m1117. doi: 10.1136/bmj.m1117

Wu, C., Chen, X., Cai, Y., Xia, J., Zhou, X., & Xu, S. et al. (2020). Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Internal Medicine. doi: 10.1001/jamainternmed.2020.0994

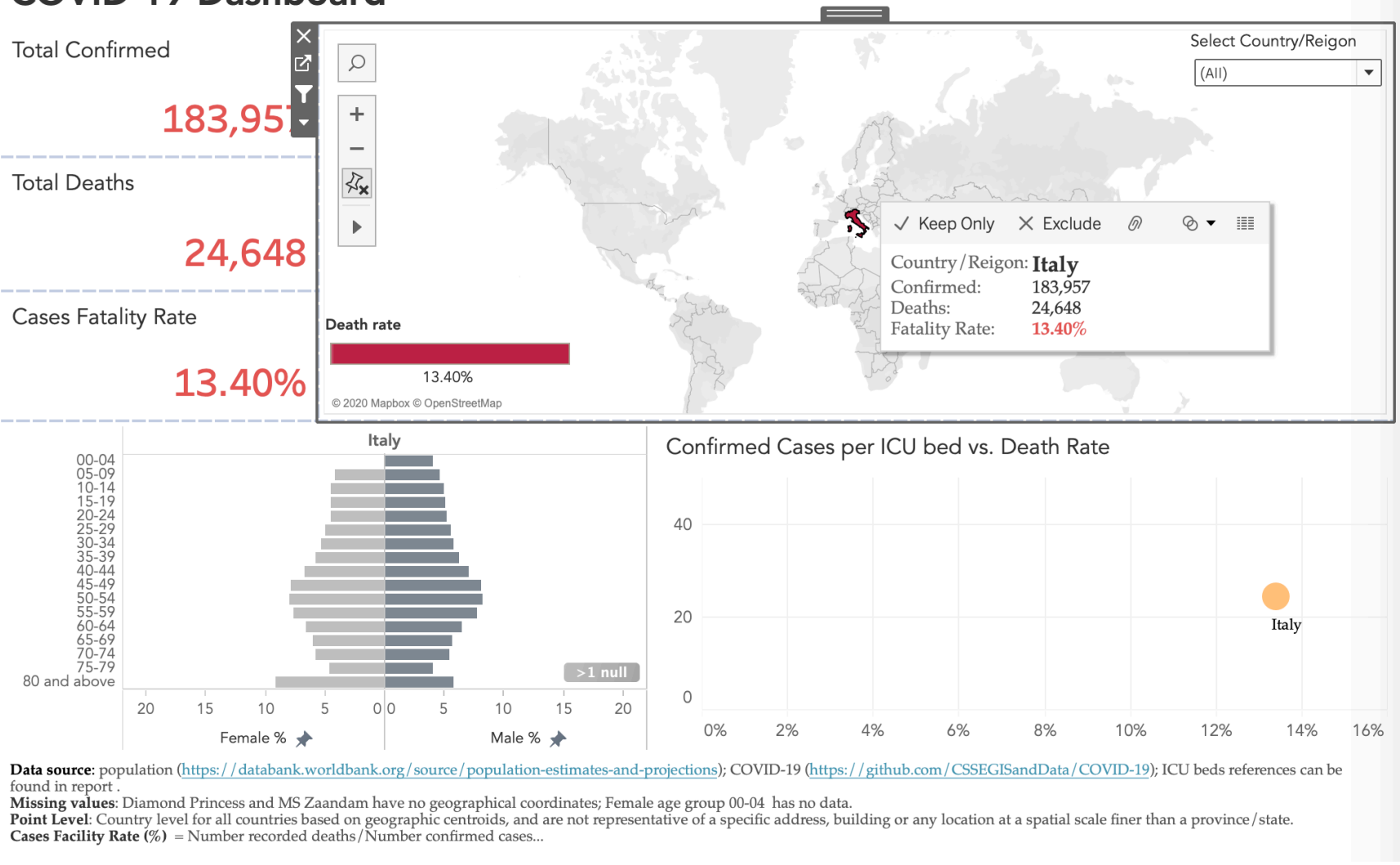
Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., & Liu, Z. et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet, 395(10229), 1054-1062. doi: 10.1016/s0140-6736(20)30566-3

Appendix



Data source: population (https://databank.worldbank.org/source/population-estimates-and-projections); COVID-19 (https://github.com/CSSEGISandData/COVID-19); ICU beds references can be found in report .
Missing values: Diamond Princess and MS Zaandam have no geographical coordinates; Female age group 00-04 has no data.
Point Level: Country level for all countries based on geographic centroids, and are not representative of a specific address, building or any location at a spatial scale finer than a province/state.
Cases Facility Rate (%) = Number recorded deaths/Number confirmed cases...

COVID-19 Dashboard



Data source: population (https://databank.worldbank.org/source/population-estimates-and-projections); COVID-19 (https://github.com/CSSEGISandData/COVID-19); ICU beds references can be found in report .
Missing values: Diamond Princess and MS Zaandam have no geographical coordinates; Female age group 00-04 has no data.
Point Level: Country level for all countries based on geographic centroids, and are not representative of a specific address, building or any location at a spatial scale finer than a province/state.
Cases Facility Rate (%) = Number recorded deaths/Number confirmed cases...