

# The effect of high risk groups in Covid-19 death cases.

Student Name: Christina Pavlou

Student ID: 190041644

**Abstract** - The paper will explore Covid-19 cases between the months of January until July 2020 from a dataset found in Kaggle. The purpose of the paper is to relate some health concerns which put people in higher risk when contracting the virus (smoking, asthma, cancer and old age). Subsets will be created to represent every country and broken down further for every month, from the new tables, percentages will be calculated on the number of deaths per cases during the pandemic. The paper aims to answer whether these health concerns had an effect on the death toll. The findings indicate that there is some relation between high rated countries in some, but not all of the health concerns selected for the paper. For more conclusive results, further research needs to be done, and a number of other variables should be taken into consideration.

## I. INTRODUCTION

The paper focuses on Covid-19 cases, more precisely the first wave of the pandemic, January to July 2020. Covid-19 is a virus that attacks mainly the respiratory system, in some cases people that have been infected are asymptomatic, or can develop mild to severe symptoms. Even though most Covid-19 patients will recover without any intense treatment from health care professionals, there are some underlying health concerns which put people at higher risk of developing severe symptoms. Some major concerns are people that have respiratory diseases, cancer, diabetes, obesity, and smokers. [1]. Governments, World Health Organisation (WHO) and medical professionals encourage people to minimise their exposure in any public spaces, minimise contact with other people and isolate as much as possible. In the case of people that are under high risk categories in case of contracting Covid-19, more measures have been set in place, for example in many countries people are not allowed to enter the facilities of nursing homes in order to minimise exposure to people as much as possible. Also, people that have underlying respiratory diseases, heart diseases, cancer etc. have the choice to work from home at all times. The purpose of the paper is to analyse how countries that are rated among the highest in some of the high-risk categories (smoking, old population, cancer and asthma) are responding to the pandemic in terms of the rate of deaths/cases.

## II. ANALYTICAL QUESTIONS AND DATA

Since the beginning of the pandemic health professional and WHO have indicated that some people are at higher risk. The purpose of the paper is to take into consideration 4 of these subgroups of people (oldest population, highest rates in smokers, asthma patients diagnosed by a doctor and cancer patients), patients that fall under these categories are thought to be more vulnerable in the case of contracting the virus which can lead them to develop more severe symptoms. Data from a variety of datasets will be collected to select which countries will be part of the analysis.

The main dataset that will be used in the paper was found on Kaggle. The datasets mentioned above explores Covid-19

cases from the beginning of the year until mid-summer 2020. The paper will focus on the analysis of 24 countries each selected due to their rating on some health concern that is considered important in the case of Covid-19 patients [2].

Covid- 19 Dataset variables		
Column Name	Type	Description
Date	Datetime	Marks the day of input
Country/Region	String	Name of the countries included in the dataset
Confirmed	Integer	Number of confirmed cases
Deaths	Integer	Number of deaths
Recovered	Integer	Number of patients that have recovered
Active	Integer	Number of active cases of Covid-19
New cases	Integer	Additional cases of Covid-19 for the day
New deaths	Integer	Additional deaths of Covid-19 for the day
New recovered	Integer	Additional recovery cases of Covid-19 for the day

Fig. 1. Variables used in the dataset

Recordings of Covid-19 at the mentioned dataset begin on the 22-01-2020 and end on the 27-07-2020. In total the rows of the dataset go up to 35156 and is a sum of 187 different countries. Each row in the dataset represents another day for each country in the span of the 7 months where records were documented.

## III. ANALYSIS

Analysis Plan:

1. Clean data: check data for any missing values, entry mistakes.
2. Reshape the data to accommodate best the purpose of the paper, create smaller dataframes to isolate the data for each country selected.
3. Find correlation between variables.

- Calculate percentage of deaths per cases for each month.
- Create boxplots and line plots to visualise the calculations.
- Repeat for all the categories.
- Bar plots to illustrate the last month available in the data (July) were all entries of deaths and cases are added together.
- Draw conclusions from the findings.

As part of the data cleaning process the dataset was searched for any missing values or mistakes. Under some entries it is clear there are some mistakes, for examples in some entries under new deaths input is a negative value, also there are days were new cases or new deaths have a spike of more than a 1000 cases compared to all the other days of the month. In the case of the first example the rows have been taken out of the analysis. In terms of the second example the days were included in the analysis and are considered to be outliers.

Among the countries with the highest old population are Japan, Italy and Finland. The dataset counts citizens that are older than 65. Japan is the leading country with 28.2%, Italy comes second with 22.8% and Finland with 21.9% of the population being older than 65. The information was retrieved from Population Reference Bureau website. [3, 3] Next for the highest recorded smokers the following countries will be included in the paper. Greece has a 43% smokers, next country chosen was Montenegro with 45.9% smokers and last Russia with 39%. The dataset recording which countries have the highest percentage in smokers was Our World in Data [4]. For the next health concern data was collected from World Cancer Research Fund International. The data is presented counting the number of people that have been diagnosed with cancer in 100.000. Australia is the leading country in cancer patients with 468.0 in 100.000, New Zealand with 438.1 in 100.000 and Ireland with 373.7 in 100.000 [5]. The last health risk that will be analysed is the number of people that have been diagnosed by a health care professional that have asthma. Countries that have been selected for the purpose of the analysis have a higher than 15% citizens with asthma. The three countries that have been selected are, Netherlands with 15.17%, United Kingdom which records at 17.5% and Sweden with 20% [6]. The countries that each category will be compared to were chosen based on 3 factors, first the country in comparison had to be low rated on the same category, not have a significant high rate in all other health concerns mentioned in the paper and have a GDP per capita relating to the greatest degree possible given the constraints set by the other two factors.

Taken from a dataset available in Our world in data the following table indicates GDP per capita of the countries represented in the paper.

GDP per capita			
Highest ranked		Lowest ranked	
Countries	Dollars	Countries	Dollars
Japan	39.002	Iceland	46.483
Italy	35.220	South Korea	35.938
Finland	40.586	Georgia	9.745
Greece	24.524	Canada	44.018
Russia	24.766	Mexico	17.336
Montenegro	16.409	Iran	19.083
Australia	44.649	Israel	33.132
New Zealand	36.086	Lebanon	13.368
Hungary	26.778	Austria	45.437
Netherlands	48.473	Estonia	29.481
Sweden	46.949	Georgia	9.745
United Kingdom	39.753	China	15.309

Fig. 2. GDP per capita

In the dataset countries are not ranked on their health care system, although it is an important factor in the case of analysing a virus where patients will rely heavily to the support provided. The income per capital is an indication of the money the government is contributing for health care. [7] Countries that have a high GDP per capita suggests that more money was allocated to health care expenses of the country leading to the assumption more people can receive care from professionals and also receive better treatment. [8]. Using income per capita was a way to keep countries on each category within a range.

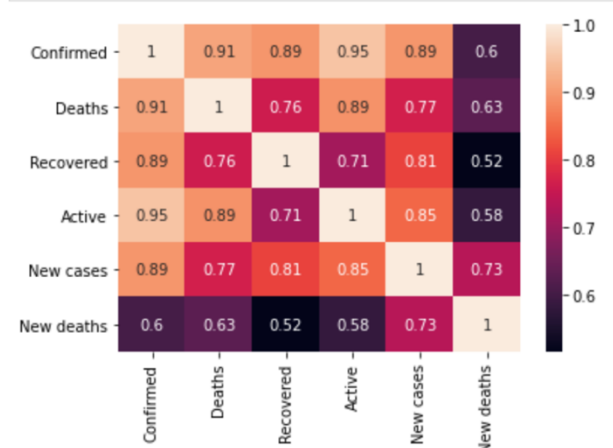


Fig3. Correlation Matrix on a heatmap between Confirmed cases, Deaths, Recovered, Active cases, New cases and deaths.

As it is expected the rise of new cases daily has a notable effect on most other variables as well. For the purpose of the project the relation between deaths and number of cases will be further analysed for each selected country.

Each country intended to be used in the project was isolated and separate DataFrames were created which were further divided by month. Using the new dataframes, percentages were calculated for deaths per confirmed cases. To visualise the progression of deaths per cases for every country boxplots and line plots were created. Box plots help us observe the ranges in percentages each country falls under and also we have an indication as to which percentage was the most persistent for the country. Each boxplot figure displays 6 countries, the first 3 are the highest ranked in the category and the last 3 are the lowest ranked. In support of the boxplots, line plots show monthly percentages of the countries with a clearer indication of the progression of deaths per case during January to July.

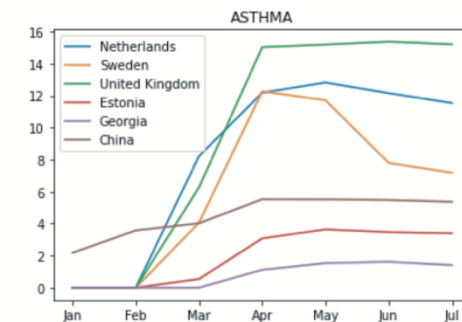
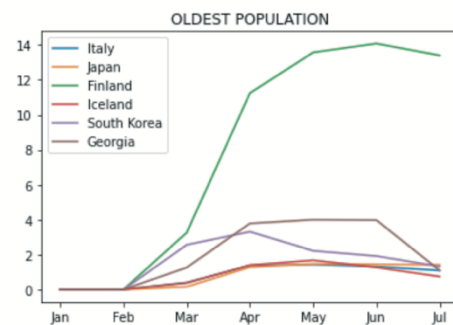
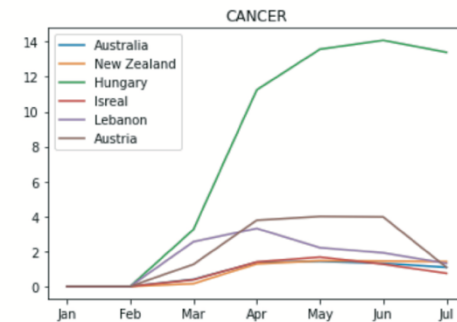
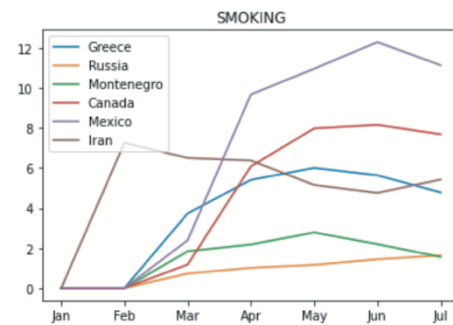
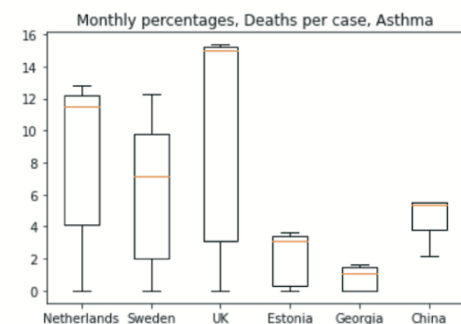
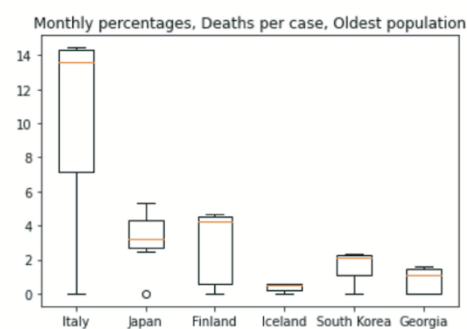
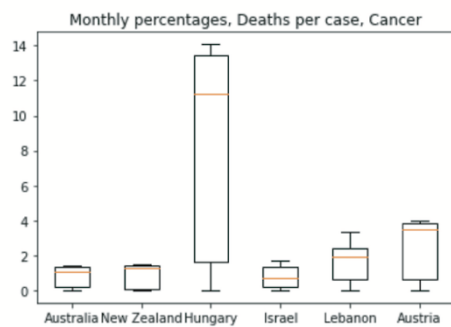
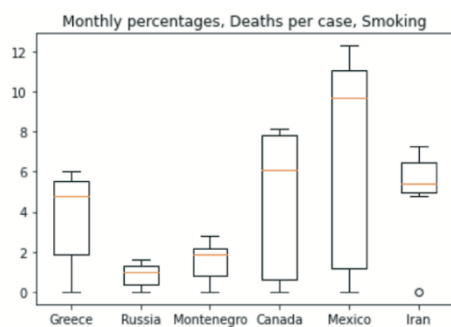


Fig 4. Monthly percentages of each health concern

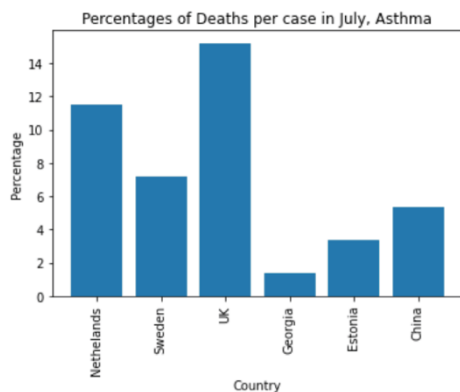
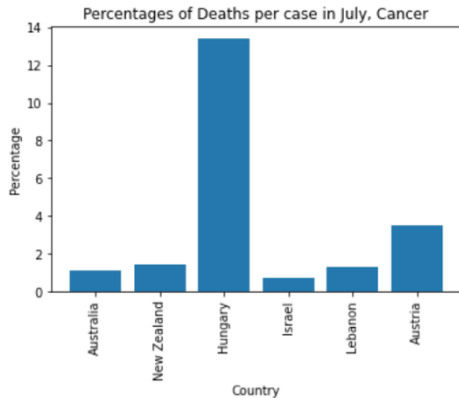
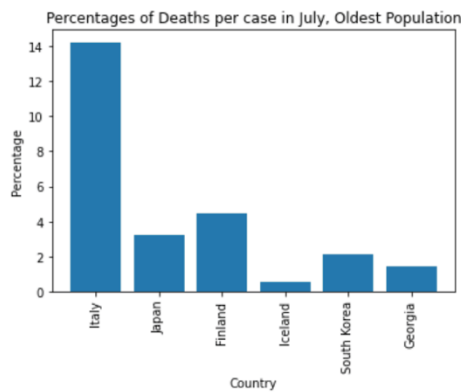
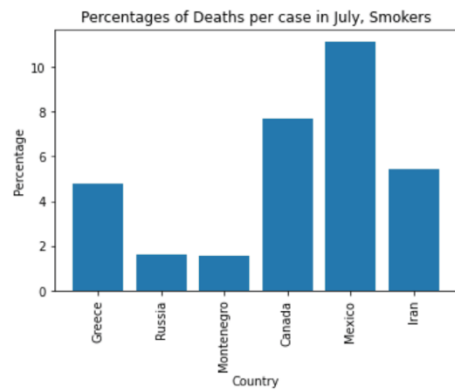


Fig5. Deaths/ per cases in July

Smoking			Cancer		
Country		July	Country		July
0	Greece	4.778803	0	Australia	1.091289
1	Russia	1.632708	1	New Zealand	1.412974
2	Montenegro	1.555479	2	Hungary	13.399281
3	Canada	7.680022	3	Israel	0.740799
4	Mexico	11.131030	4	Lebanon	1.313756
5	Iran	5.419508	5	Austria	3.468236

Old Age			Asthma		
Country		July	Country		July
0	Italy	14.256596	0	Netherlands	11.532773
1	Japan	3.204675	1	Sweden	7.179293
2	Finland	4.447148	2	United Kingdom	15.194824
3	Iceland	0.539374	3	Georgia	1.407212
4	South Korea	2.112230	4	Estonia	3.392330
5	Geogria	1.407212	5	China	5.365106

Fig6. Tables, percentage of Deaths per case in July

In the last month (July) deaths cases and confirmed cases are added together between January and July to display cumulative of death toll.

The bar plots and the tables with the exact perctantage will help us to draw conclusions and answer the question that was set at the beginning of the paper. Similarly to the box plots the first 3 countries represent the highest ranked in the category and the last 3 represent the lowest rated. For every health concern category there is a separate plot and table above.

#### IV. FINDINGS, REFLECTIONS AND FURTHER WORK

From the comparison between countries in each category for 2 of them (Asthma and Old Age) seems that the countries ranked highest performed worse compared to the ones on the other side of the spectrum. Cancer dominant countries and their opposite countries are all on a similar range but one, Hungary which is significantly higher than the rest. Final category, smoking it is evident that countries rated low on the mentioned health category have performed noticeably worse. To apply the findings to the analytical question set at the beginning, whether health concerns, in the case of this paper, old age, smoking, asthma and cancer played a role on the number of deaths each country had per confirmed case more extensive research needs to be done. There are a number of factors which play a role in the number of deaths per cases, for example a detailed worldwide rate of health care provided in each country rather than just using alone the income per capita of each country. Another factor that plays a role is the timeline each country was on lockdown and population compliance on the regulations. Treatments can also make a difference in recovery of patients, in the dataset there are no indications as to what treatment each country offered. Daily testing done would also affect the number of confirmed cases per day.

Furthermore, it is important to notice the time each country had an outbreak since the virus has been subjected to mutations, some versions of the virus were more contagious, some had a higher mortality rate etc. [9]. Another important element would be to include patient records indicating percentage of confirmed cases how many are smokers, have diabetes or are obese, heart diseases etc. Counting of cases should be the same between countries in order to have the best comparison between them. For the first 4 months of the pandemic countries had differences in the way deaths were counted. For example, France and Germany were adding deaths in care homes compared to the UK where deaths in care homes started to add up to daily number death tolls after the 29th of April. [10]

Even though more analysis and different data are needed to confidently answer the question, the findings of the analysis can be an indication that there is a relation between countries with population vulnerable due to underlying health reasons but at the same time not every category is on the same level of severity when contracting Covid -19. Based on the findings of the paper it can be said that old age and asthma in people contracting Covid-19 had a more major effect on their health in comparison to people with cancer or smokers.

To conclude based on the findings some influence seems to exist between health risks and death toll per case, but not every risk has the same level of impact. Two of the categories positively support the assumption that there is an effect on death tolls in the case of vulnerable population, one category where countries are mostly on the same scale and one category which the claim is not positively supported.

Word Count	
Abstract	142 /150
Introduction	231/300
Analytical questions and data	304/300
Analysis	996/1000
Findings, reflections and further work	521/600

## V. BIBLIOGRAPHY

- [1] World Health Organisation, "Coronavirus," 13 Dec 2020. [Online]. Available: [https://www.who.int/a-topics/coronavirus#tab=tab\\_1](https://www.who.int/a-topics/coronavirus#tab=tab_1). [Accessed 7 Dec 2020].
- [2] World Health Organisation Western Pacific, "COVID-19: vulnerable and high risk groups," 2020. [Online]. Available: <https://www.who.int/westernpacific/emergencies/covid-19/information/high-risk-groups>. [Accessed 2 12 2020].
- [3] Population Reference Bureau, "Countries With the Oldest Populations in the World," 23 03 2020. [Online]. Available: <https://www.prb.org/countries-with-the-oldest-populations/>. [Accessed 2 12 2020].
- [4] H. Ritchie, "In which countries do people smoke the most?," 2 9 2019. [Online]. Available: <https://ourworldindata.org/which-countries-smoke-most>. [Accessed 2 12 2020].
- [5] WCRF International, "Global cancer data by country Exploring which countries have the highest cancer rates," 2018. [Online]. Available: <https://www.wcrf.org/dietandcancer/cancer-trends/data-cancer-frequency-country>. [Accessed 2 12 2020].
- [6] Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada, "Global asthma prevalence in adults: findings from the cross-sectional world health survey," 19 Mar 2012.
- [7] Our World in Data, "Healthcare Expenditure vs. GDP, 2014," 2014. [Online]. Available: <https://ourworldindata.org/grapher/healthcare-expenditure-vs-gdp>. [Accessed 7 12 2020].
- [8] N. B. G. M. N. M. N. Y. J. Elizabeth Asiedu, *On the Impact of Income per Capita on Health Outcomes: Is Africa Different?*, 2015.
- [9] University of Texas at Austin, "Coronavirus mutation may have made it more contagious," 30 Oct 2020. [Online]. Available: <https://www.sciencedaily.com/releases/2020/10/201030142123.htm>. [Accessed 14 Dec 2020].
- [10] C. M. & A. Reuben, "Coronavirus: Why are international comparisons difficult?," 17 Jun 2020. [Online]. Available: <https://www.bbc.com/news/52311014>. [Accessed 17 Dec 2020].