**Read me file for project: Impact of R&D Spending on COVID-19 Mortality by Country by Ramzi CHARIAG**

# Data and Sources for each variable

The data used for this project is entirely public, and was downloaded from the following sources.

Mortality:

The number of deaths as a percentage of total cases. This dataset is being updated frequently, and I used the latest version available at the time of writing the paper which was on 10th of May, 2021.

R&d: Research and development spending by country averaged since the year 2000

lpapers: log of Number of research papers published per country

Control variables:

HAQ: Healthcare access and quality index

Other factors such as underlying health conditions and income:

BMI: average body mass index per country

Lgdp: log of PPP income by country in 2017 US dollars

Median age: median age by country

# Transformations

The file “Transformed” includes the datasets from raw but renamed, and metadata sheets deleted. Other than that there was only one change made which was to the dataset of R&D spending by country.

The dataset for R&D spending by country goes back a number of years, I have averaged the latest number of years since 2000 and used average R&D in my analysis. Both averaged and raw files are included in the folder “transformed”

# Replication:

To replicate the tables from the paper, first run “clean\_data.do”

Then to generate each table run the corresponding do file.

The whole project can be replicated in one go by running “master\_do.do”

Table 1: table containing descriptive statistics generated by “descriptive\_statistics.do”

Table 2: regression of mortality rate on R&D generated by “regress\_mortality\_on\_r&d.do”

Table 3: regression of mortality rate on lpapers generated by “regress\_mortality\_on\_papers.do”

Table 5: regression of mortality rate on both R&D and lpapers generated by “regress\_mortality\_on\_r&d\_and\_papers.do”

Tables 6 and 7: first stage and second stage regressions generated by “IV\_model.do”

Note: “IV\_model.do” will only work after running “descriptive statistics.do” as it uses its output.

Each regression requires a slightly different merging procedure which is why merging is not done separately by its own do file.

# References

Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org/coronavirus>

The Lancet in May 2017 in "Healthcare Access and Quality Index based on mortality from causes amenable to personal healthcare in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015." <https://ourworldindata.org/grapher/healthcare-access-and-quality-index>

United Nations Educational, Scientific and Cultural Organization,*The UNESCO Institute for Statistics (UIS),* [Science,technology and innovation (unesco.org)](http://data.uis.unesco.org/Index.aspx?DataSetCode=SCN_DS&lang=en)

World Bank, *World Development Indicators,* <https://data.worldbank.org/indicator/IP.JRN.ARTC.SC>

International Comparison Program, World Bank | World Development Indicators database, World Bank | Eurostat-OECD PPP Programme. <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>

GBD 2016 Healthcare Access and Quality Collaborators, (2018) Measuring performance on the Healthcare Access and Quality Index for 195 countries and territories and selected subnational locations: a systematic analysis from the Global Burden of Disease Study 2016, TheLancet, Volume 391, Issue 10136, P2236-2271, June 02, 2018

<https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31859-6/fulltext>

United Nations, Department of Economic and Social Affairs: Population Dynamics, <https://population.un.org/wpp/DataQuery/>