Impact of R&D Spending on COVID-19 Mortality by Country

Ramzi Chariag

Central European University, Department of Economics

Abstract

This paper aims to study the impact of average R&D spending on COVID-19 mortality rate by country. I would like to know whether an increase in R&D as a percentage of GDP in a given country has a significant impact on the mortality rate of the country in question, which has not been examined so far. The premise is that countries that invest in R&D give more weight to the input of scientists when making policy. Additionally, such countries should be more capable of adapting quickly to the new conditions imposed by the pandemic, and thus should have a lower mortality rate. Regressing mortality on our variable of interest using OLS yields that R&D spending is insignificant. However, by using the number of published research papers per year as an instrument for R&D spending as a percentage of GDP, it is possible to eliminate endogeneity. I surprisingly found that an opposite relationship exists.

1. Introduction and literature review

The pandemic that hit the world by the end of 2019 raised many questions about inequality within and across countries. One of the most important among these questions was "who was better prepared?" and "how can we react more efficiently to such sudden and unexpected shocks in the future?" Only during these times have expressions such as "phase 3 trials" and "statistically significant evidence" become so ubiquitous among the general public. Many different approaches and strategies have emerged in managing the pandemic. Governments and experts got creative while trying to figure out the best solutions that work for them, given their income, healthcare infrastructure. They also have to consider other demographic factors like the age of the population and the prevalence of certain health conditions suspected to have a detrimental effect on the development of the disease amongst individuals. "An understanding of factors at the national level associated with a higher population risk for more widespread infection, severity of illness, and mortality is critical. The impact of existing national policies, and the association of specific country-level factors with outcomes, is urgently required as many jurisdictions have begun the process of relaxing public health interventions with an accompanying risk of subsequent waves of infection" (Chaudhry et al., 2020). Elements such as infrastructure and deeply rooted cultural and institutional norms are not easily malleable. However, understanding their impact on the development of the pandemic could have a pragmatic use in the future. To this end, Chaudhry et al. (2020) conduct a "A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcome." They use various sources of publicly available data and indices put into the following categories: Socioeconomic characteristics, health related characteristics, health care capacity, Global Health Security Capabilities (GHS)[1], border closure, and lockdown. Their aim was to evaluate the effect of both timing and type of government policy on COVID-19 mortality rates. They find that full lockdowns and higher GHS scores positively predicted recovery cases per million. Also, they report that early border closures reduced the peak transmission. However, more testing had no effect on cases per million i.e. more information did not help governments react better. On the negative side, they found that GDP per capita and smoking prevalence were associated with higher mortality rates. In similar efforts, de Siqueira, et al. (2020) looked at the effect of obesity on hospitalization and mortality by conducting a review study. They found a positive correlation between higher BMIs and worse outcomes amongst COVID-19 patients. In addition, obesity (BMI>30 kg/m²) increased the need for medical care.

As this paper is being written, vaccines are already being rolled out across most of the world. Though COVID-19 seems like it will eventually fade out thanks to vaccines, it still remains a rather big issue in most nations. While it is probably not something that a government could dramatically impact through policy at this advanced stage of the crisis, there is a potentially useful inference to make about the impact of trusting scientists and funding them so that scientific advancement can prevent calamities, or at least mitigate them in the future.

2. Data and Descriptive Statistics

The data used comes from various sources such as governmental statistics in the case of mortality, UNESCO, WorldBank, and Ourworldindata. All of the datasets used are publicly available. The variables that were employed in this study are:

COVID-19 mortality rate (lmortality): Log of mortality rates of COVID-19 by country, scraped from Worldometer on the 10th of May, 2021.

R&D spending as a percentage of GDP(lrd): Log of average R&D spending as a percentage of GDP between 2010 and 2020. Data about every single country on every single year was not available, but that should not be a problem, because the variable is actually quite consistent throughout the years. This was expected, since governments usually do not often alter such choices drastically. This data was obtained from a UNESCO dataset which has less observations (countries) compared to the other datasets. Note that there are only 93 observations from the source that match with the rest of the variables. If I exclude lrd, the rest of the variables match on more than 160 observations.

Number of research papers published per year: Log of the weighted average of the number of research papers published between 2000 and 2018. Newer data points were given larger weights. The discount factor used to derive the weights was 0.95 with the year 2018 having a weight of 100 (which brings the weight of the year 2000 to 39.72).

[1] Measured on a scale from 0 to 100 and presents a country's overall preparedness in the event of a global pandemic. Higher scores indicate a greater level of national preparedness.

GDP per capita (lGDPPC): Log of GDP per capita, adjusted for purchasing power parity (PPP) using constant 2017 US Dollars.

Median age: median age per country.

Healthcare access and quality index (HAQ) index: Constructed by GBD collaborators (2016) in 195 countries and territories in order to quantify the accessibility and reliability of healthcare services, and ranges from 0 to 100; 0 as the first percentile (worst) observed between 1990 and 2016, and 100 as the 99th percentile (best). Healthcare quality and access is crucial when it comes to dealing with a global pandemic.

BMI: Mean BMI for all of both genders, assuming men and women each represent 50% of society, since the raw data from ourworldindata.org has the values segregated by gender.

Diabetes: prevalence of diabetes as a % of population between the ages of 20 and 79.

Logs were taken, because COVID mortality rates are relatively small numbers that have an even smaller variation, while variables such as the number of papers published and GDP per capita are in the order of thousands. Another reason to take the logarithm of R&D spending is because variation amongst countries can be quite large. This transformation also allows for the interpretation of the coefficient as a relative change in R&D spending.

The following table gives us an idea about why taking the logs was useful in terms of both scale and variation of the variables.

Table 1: Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Mortality	93	.02	.013	.001	.093
Lrd	93	-.576	1.266	-4.353	1.497
Lpapers	93	7.675	2.508	2.334	12.9
lGDPPC	93	9.853	.995	7.105	11.613
m_age	93	35.582	6.078	20.894	44.081
HAQ	93	65.749	14.435	37.217	88.117
BMI	93	24.724	1.783	20.221	28.381
Diabetes	93	7.688	3.638	1.9	22

There are two variables that could be interpreted as "how much a certain country invests in research," which are lrd and lpapers. I begin by running a regression of our dependent variable on both of them independently to see which one is a better fit.

I first run the following regression:

$$mortality = \beta 0 + lrd \times \beta 1 + X\beta 2$$

X: vector of covariates that is incrementally expanded.

Table 2: R&D spending as a variable of interest

VARIABLES	(1) lmortality	(2) lmortality	(3) lmortality	(4) lmortality	(5) lmortality
lrd	0.0529	0.00502	0.0112	0.0987	0.0976
	(0.0717)	(0.0723)	(0.0860)	(0.0767)	(0.0771)
lGDPPC	-0.184	-0.356**	-0.334	-0.572**	-0.577**
	(0.117)	(0.142)	(0.224)	(0.250)	(0.272)
m_age		0.0436**	0.0454**	0.0526***	0.0529**
		(0.0197)	(0.0193)	(0.0199)	(0.0205)
HAQ			-0.00273	-0.0133	-0.0129
			(0.0170)	(0.0144)	(0.0156)
BMI				0.219***	0.217***
				(0.0695)	(0.0665)
diabetes					0.00169
					(0.0208)
Constant	-2.288*	-2.176*	-2.267*	-4.835***	-4.812***
	(1.178)	(1.132)	(1.353)	(1.292)	(1.304)
Observations	93	93	93	93	93
R-squared	0.038	0.080	0.080	0.223	0.223

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From table 2, note that the variable of interest is never significant, even at 10% with seemingly arbitrary coefficients.

Then, I run a similar second model where I change the log of R&D spending by the log of the actual number of papers published.

$$mortality = \beta 0 + lpapers \times \beta 1 + X\beta 2$$

From table 3 below, note that the variable of interest actually has a positive coefficient that is significant at 1% i.e. the number of research papers published per year positively correlates with mortality rates. The magnitude of the impact decreases as more controls are added, which suggests that the initial estimate is biased, but it suddenly jumps up in magnitude when BMI is included.

Now, for the sake of argument, let's say that the number of papers published conveys information that is structurally different from that conveyed by R&D spending as a percentage of GDP. Let's pretend that the two variables are not mechanically connected to each other and proceed to regress mortality on both of them, using the same vector of covariates. i.e. this model assumes that the number of papers published per year does not depend on R&D spending, and that even if a correlation exists, it is spurious. This is admittedly a strong assumption, but I do it to observe whether the estimate for the coefficient for lrd and lpapers present omitted variables bias as I add covariates.

Table 3: Number of published papers as a variable of interest

	(1)	(2)	(3)	(4)	(5)
VARIABLES	lmortality	lmortality	lmortality	lmortality	lmortality
lpapers	0.0853***	0.0718**	0.0675**	0.107***	0.106***
	(0.0257)	(0.0275)	(0.0276)	(0.0260)	(0.0261)
lGDPPC	-0.233***	-0.343***	-0.389***	-0.627***	-0.591***
	(0.0711)	(0.103)	(0.144)	(0.175)	(0.176)
m_age		0.0261*	0.0188	0.0266	0.0274
		(0.0142)	(0.0195)	(0.0189)	(0.0187)
HAQ			0.00707	-0.00959	-0.0134
			(0.0133)	(0.0119)	(0.0123)
BMI				0.255***	0.277***
				(0.0540)	(0.0541)
diabetes					-0.0223
					(0.0162)
Constant	-2.543***	-2.265***	-1.987**	-5.414***	-5.907***
	(0.617)	(0.672)	(0.902)	(0.975)	(1.032)
Observations	166	166	166	166	166
R-squared	0.080	0.092	0.095	0.255	0.262

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Third hypothetical model:

$$mortality = \beta 0 + lrd \times \beta 1 + lpapers \times \beta 2 + X\beta 3$$

Note from table 5 below that both variables have a significant coefficient. However, the coefficient on lrd decreases in magnitude and in significance as more controls are added, while that of lpapers remains relatively unchanged.

In this model, I assumed that there is no underlying mechanism affecting both lrd and lpapers, which is incorrect. In fact, these two variables do not belong together in the same model, because they both represent a measure of how much an economy spends on science. Conceptually, they should be measuring the same quantity. This indicates that they should have a strong correlation. Thus, I proceed to examine the relationship between these two variables.

Table 4: Correlation between lrd and lpapers

Variables	(1)	(2)
(1) lrd	1.000	
(2) lpapers	0.792	1.000

Table 5 Including both variables in the model

VARIABLES	(1) lmortality	(2) lmortality	(3) lmortality	(4) lmortality	(5) lmortality	(6) lmortality
lrd	-0.300***	-0.199**	-0.229**	-0.255**	-0.174*	-0.172*
	(0.0857)	(0.0903)	(0.0904)	(0.120)	(0.0973)	(0.0979)
lpapers	0.159***	0.158***	0.150***	0.156***	0.161***	0.163***
	(0.0439)	(0.0435)	(0.0435)	(0.0449)	(0.0415)	(0.0407)
lGDPPC		-0.176	-0.326**	-0.385*	-0.630**	-0.611**
		(0.111)	(0.139)	(0.217)	(0.250)	(0.270)
m_age			0.0379*	0.0325*	0.0394*	0.0381*
			(0.0194)	(0.0193)	(0.0200)	(0.0206)
HAQ				0.00774	-0.00274	-0.00419
				(0.0172)	(0.0136)	(0.0150)
BMI					0.223***	0.227***
					(0.0679)	(0.0656)
diabetes						-0.00668
						(0.0211)
Constant	-5.527***	-3.717***	-3.554***	-3.349**	-6.006***	-6.108***
	(0.393)	(1.139)	(1.138)	(1.287)	(1.329)	(1.305)
Observations	93	93	93	93	93	93
R-squared	0.117	0.145	0.177	0.179	0.329	0.329

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3. IV model and results

R&D spending as a percentage of GDP suffers from endogeneity issues. It is mechanically related to GDP, which in itself is mechanically related to one of the explanatory variables. It is also related to all sorts of macroeconomic variables that are not included in the model such as unemployment and inflation. These macroeconomic variables can have a significant impact on GDP per capita and HAQ. On the other hand, the number of research papers published per year was not impacted much by adding GDP and HAQ. This suggests that it is in fact exogenous, which confirms that are no mechanisms by which lpapers is correlates with the residual from either models in which it was included. And since the variables have an almost 80% correlation, it makes sense to use lpapers as an instrument for lrd.

First stage regression:

$$lrd = \beta 0 + lpapers \times \beta 1 + X\beta 2$$

I obtain F_{stat}=53.66 (R^2=0.7892) which suggests that lpapers is a strong instrument for lrd. Consequently, our instrument is both valid and relevant.

Second stage regression:

$$mortality = \beta 0 + lrd \times \beta 1 + X\beta 2$$

The results are presented in the tables below:

Table 6: First stage regression

VARIABLES	(1) lrd
lpapers	0.257*** (0.0321)
lGDPPC	0.00983 (0.169)
m_age	-0.0106 (0.0206)
HAQ	0.0521*** (0.0123)
BMI	-0.0808 (0.0499)
diabetes	0.00853 (0.0199)
Constant	-3.764*** (1.167)
Observations	93
R-squared	0.789

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Second stage regression

VARIABLES	(1) lmortality
lrd	0.460*** (0.141)
lGDPPC	-0.617** (0.288)
m_age	0.0448** (0.0222)
HAQ	-0.0372** (0.0173)
BMI	0.279*** (0.0721)
diabetes	-0.0121 (0.0206)
Constant	-3.727** (1.592)
Observations	93
R-squared	0.080

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

From the second regression, summarized in table 7, note that a 1% increase in R&D spending as a percentage of GDP is associated with 0.4% increase in mortality rate i.e. if a nation doubles its R&D expenditure as a percentage of GDP, COVID-19 mortality rate increases by 40% (of the already existing rate, not in percentage points). This result is significant at 1% level, and all standard errors are relatively small. A possible explanation of this is that most countries with high R&D spending are not spending it on advancing healthcare and medicine but rather on potentially more profitable research such as tech. This remains to be examined.

Another noteworthy observation to make is that in comparison to the first two models I had (summarized in tables 2 and 3), the coefficients of the covariates are all significant with signs that actually make sense.

By using an IV, I get rid of the exogeneity problem that R&D spending as a percentage of GDP suffers from. This issue is a result of the fact that the variability in R&D spending as a percentage of GDP, per capita GDP and HAQ is generated by the same macroeconomic mechanism. However, the number of published research papers per year should not have this issue. Since most of these papers are written by tenured professors or those seeking tenure, it makes sense that the actual number of papers produced by these individuals would not be connected to things like unemployment, GDP, exchange rates, etc. I proceed to use it as an instrument for R&D spending. It is possible to safely assume that the yearly number of papers published in a country is exogenous. The results of the first stage regression suggest that the instrumental variable is also relevant. The second stage regression shows a counterintuitive finding, that is: R&D spending has a positive effect on COVID-19 mortality. This can be explained by the fact that R&D spending is in fact done mostly for profit. Another puzzling result is the coefficient on GDP per capita being negative, which is counter to what Chaudhry et al. (2020) had found. This actually makes sense, because one would expect richer economies to be able to afford better living conditions and treatments. A possible conclusion from both of these observations that richer countries can afford treatment when the sickness hits, but do not necessarily invest in it when there is no pandemic taking place. Instead, research money is being spend on projects with more profitable prospects.

Conclusion

Countless trends have been spotted with respect to COVID-19 mortality, most of which focus on health issues that might exacerbate the disease. This paper instead analysed the economic aspect of the pandemic, as it studied the impact of R&D spending on COVID-19 mortality rate. To identify the level of this impact, an instrumental variable was employed to eliminate the endogeneity caused by the fact that R&D as a percentage of GDP is structurally connected to GDP and to the HAQ index. It was found that surprisingly, R&D spending has a positive impact on COVID-19 mortality. The proposed explanation is that even though richer countries spend more on research, most of that research goes into subjects that bring the most profit, and not necessarily the least COVID related deaths.

Sources

Chaudhry R., Dranitsaris G., Mubashir T., Bratoszko J., Riazi S. (2020) *A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes*, EClinicalMedicine, Volume 25, 100464, August 01, 2020

De Siqueira J., Almeida L., Zica B., Brum., Barcelo A., Galil A., (2020) *Impact of obesity on hospitalizations and mortality, due to COVID-19: A systematic review*, Obesity Research & Clinical Practice, Volume 14, Issue 5, September–October 2020, Pages 398-403

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*