

# Global COVID-19 spread: socioeconomic determinants and lessons for future pandemics

Nicholas Ngepah (✉ [nnbal@yahoo.fr](mailto:nnbal@yahoo.fr))

School of Economics, University of Johannesburg <https://orcid.org/0000-0002-1947-0008>

---

## Research Article

**Keywords:** Social determinants of health; COVID-19; PPML; quantile regression; health systems; health policy

**Posted Date:** May 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-31764/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## **Global COVID-19 spread: socioeconomic determinants and lessons for future pandemics**

Nicholas Ngepah (PhD) School of Economics, University of Johannesburg College of Business and Economics (School of Economics, Johannesburg, South Africa; [ngepah@uj.ac.za](mailto:ngepah@uj.ac.za); [nnbal@yahoo.fr](mailto:nnbal@yahoo.fr)).

[ORCID: 0000-0002-1947-0008](https://orcid.org/0000-0002-1947-0008)

### **Abstract**

This paper examines some factors that can complement public health systems in managing the spread of the COVID-19 with implications for preparedness towards possible future pandemics. It adapts and applies two suitable econometric models in the theoretical framework of social determinants of health. The one models are the Poisson's Pseudo Maximum Likelihood (PPML) and the quantile estimator, applied to panel data for 195 countries over days of infection from first recorded case. The COVID-19-related data is from our world in data, and the socioeconomic variables are from the World Bank's World Development indicators. The results suggest that enhancing capacity for early testing complemented by reduction of international exposure; improved management of population dynamics; ensuring better sanitation and hygiene practices and reducing alcohol use are key to addressing the rapid spread of the COVID-19, and readiness for future pandemics of similar a kind.

Keywords: Social determinants of health; COVID-19; PPML; quantile regression; health systems; health policy

### **Declarations**

**Funding:** Not applicable

**Conflicts of interest/Competing interests:** None

**Availability of data and material:** 'Not applicable

**Code availability:** 'Not applicable

## Introduction

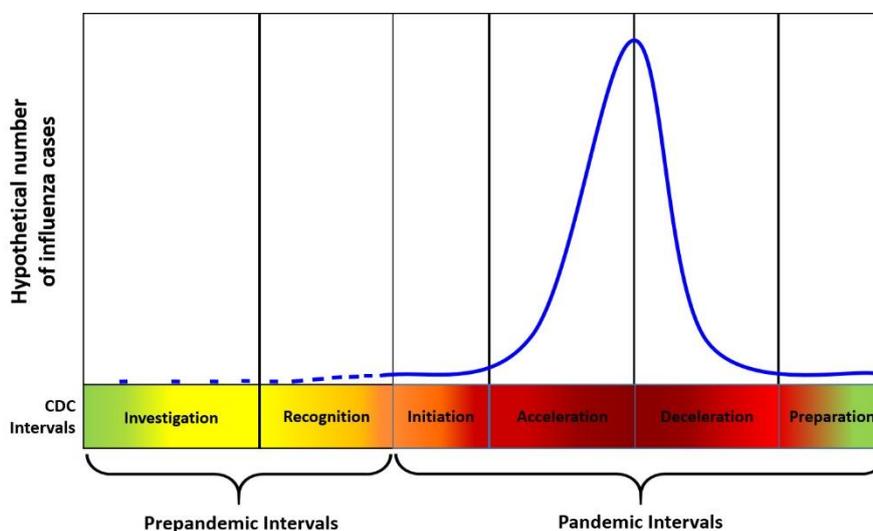
*While the nation-wide lockdown is having a devastating effect on our economy, it is nothing compared to the catastrophic human, social and economic cost if the coronavirus could spread among our people unchecked...[1].*

The above excerpt of the speech of President Ramaphosa of South Africa underscores the challenges that governments face in addressing both the public health and economic consequences of the COVID-19 outbreak. Since the announcement of the emergence of a new strand of the corona virus - severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes what has come to be known as COVID-19, the world has suddenly become a completely different place. What was initially known as a “Chinese problem”, later an “Italian problem” has quickly risen to a global issue in a very short time. According to the European Centre for Disease Prevention and Control [2], as of 21 April, 2020, the number of cases globally has risen above 2.5 million mark with over 170000 deaths, with a spread to over 210 countries globally. However, not all is gloom as the recovery cases have far outweighed the deaths, standing at over 690 000.

Such rapid rises in infections and fatalities imply that biomedical systems alone will not cope with the challenge. The broader determinants of health, which include social, economic, physical and environmental factors [3,4] can prove to be more crucial in times like these. These factors can affect not only the rate of spread but also the post-pandemic rebound of individual and population health.

As with all influenza pandemics, it is expected that the COVID-19 will go through the six intervals described by the US Centers for Disease Control (CDC). On the 30 December 2019, the Chinese Wuhan local health authority issued an epidemiological alert following the first documented case on the 8 December, which led to the closure of the Huanan Seafood market two days later [5]. However, it was only on the 11 March, 2020, that the World Health Organisation (WHO) declared the disease as a global pandemic [6]. From the hypothetical CDC curve in Figure 1, most of the cases of the pandemics globally are at the accelerating phase. It is important to mention that the rate of growth may differ from country to country such that the curve might be steeper for some than others, however, overall, all countries are expected to follow the hypothetical path.

Figure 1.: CDC’s hypothetical evolution path for the novel influenza-A virus pandemics



Source: CDC.gov, <https://www.cdc.gov/flu/pandemic-resources/national-strategy/intervals-framework.html>

Apart from a few, most governments all over the world initially downplayed the pandemics until it turned to sustained community-level transmission. Once governments noticed cases getting out of hand, the first respond became a closure of the international borders, with more stringent social distancing measures and subsequent lockdown of the socioeconomic machinery to restrict local population dynamism. The case of China has shown that social distancing, isolation of infected cases and quarantine can be effective in mitigating the pandemic [7].

While these measures aimed at flattening the evolution curve, they are inevitably leading to immediate acute and economic hardship, which may be persistent long after the pandemic. This problem is much more precarious for most developing countries, with limited resources, high levels of socioeconomic vulnerability and poverty coupled with weak governance mechanisms. Some authors such as [8] have highlighted the impossibility of such governments to be able to minimise COVID-19-infection rates and the related economic impacts. Whatever happens to the spread of the pandemic, and any associated measures to curb them would depend on the broader determinants of health which would be social, economic, physical and environmental context where the pandemic is evolving. The identification of the key socioeconomic factors that affect the spread of the disease becomes important in managing cases of the current disease and possibly future similar pandemics. It is in this manner that this paper contributes to the debate relating to the spread COVID-19. Such analyses can also derive lessons to help countries in building resilience through selections of which socioeconomic and environmental factors to strengthen in order to limit the spread of the current and future pandemics.

The rest of the paper is structured as follows: Section two gives an explanation of the empirical methodology. Section three reports and discusses the main econometric findings. Section four presents policy implications while conclusion is drawn in section five.

### **Methodological framework and econometric strategy**

The determinants of the spread of the COVID-19 pandemic can be assessed within the framework of social determinants of health. The WHO [9] classifies the social determinants of health into the conditions in which people are born, grow, work, live and age; and the wider forces and systems that define the conditions of their daily life. In today's literature, social aspects of health explain substantially more of health outcome changes than simply medical care [10, 11].

Following literature [3,4,12,13], we specify a formal framework that links disease outcome to various social, economic, physical and environmental context of a country in which individuals find themselves. The main problem we face in specifying the model is that while the socioeconomic variables are time-invariant, varying only across countries, and measured annually, the dependent variable – COVID-19 infections is measured daily, and vary both across countries and time. The only way we can explore this data is to consider the different socioeconomic variables as fixed country-specific factors. In that sense, we explore only the cross-sectional dimensions of the data. We can also assume that as the disease progresses, the different socioeconomic health determinants will affect transmissibility differently. For instance, at normal values of disease burden, one would expect some variables to behave normally, but at higher values of the disease, the effects may change. For this reason, we specify the model and estimate the cross-sectional effects. However, to take care of the evident non-linearity described above, we undertake a quantile regression to evaluate the coefficients of the determinants at different quantiles of the cumulative numbers of cases and mortality. The first model we specify is the one relating the daily cases to a

set of time-varying variables ( $X$ ) associated with the dynamics of the disease, and the annual-level time-invariant determinants ( $Z$ ) as follows:

$$H_{it} = \alpha_0 + X_{it}\beta + Z_i\gamma + \delta DAY_{it} + \varepsilon_{it} \quad (1)$$

Where  $H_{it}$  is an indicator of health in our case, cumulative infections or cumulative deaths for country  $i$  at day  $t$ ;  $X$  is a vector of time-varying variables and  $Z$  a vector of time-invariant country-specific variables;  $\alpha_0$  is the intercept and  $\beta, \delta$  and  $\gamma$  are coefficients to be estimated,  $DAY$  is the number of days from the first case detected in a country and  $\varepsilon$  is a classical mean zero disturbance. The nature of our data is such that the time-varying variables are the cumulative tests and time trend.

## Variables and data

The dependent variable for the cumulative infection model is the total number of infections at a given time. The data is a day by country panel of 196 countries with reported cases from first day of detected case in China (31 December, 2019) to current date at which data was harnessed (21 April, 2020). We expressed the cumulative infections per thousand population and use the log (LTI).

In selecting the covariates for the models, we have to be systematic. First of all, it is widely accepted that the more testing is done, the more hidden cases are discovered, hence testing rate will have positive effect on recorded cases of infections. The testing process is crucial for early detection, and isolation for treatment. Therefore, testing is an important factor in the real-time monitoring of transmission. This is vital for subsequent containment of the disease [14], and possible reduction in mortality, since early detection provides higher chances of survival. To control for testing, we use total number tested expressed per thousand of the population. The log of the variable (LTT) is used.

All COVID-19-related variables, i.e. infections and tests are taken from Our World in Data<sup>1</sup> [15]. The population data by which we transform the total infections and total tests into per thousand population, are from the World Bank's [16] World Development Indicators (WDI).

The rest of the right-hand side variables are the time-invariant category. These variables take the value of the first non-missing observation for the last five years from a given country. While these variables do not vary with time, they vary across countries. We believe that these will more or less depict the social, economic, physical and environmental conditions that prevailed in a country immediately following the disease outbreak. The selection of these variables for the model is based on literature and how we know the virus has been spreading. From the first reported case in China, all other countries have had their first case based on international exposure.

**International exposure** is captured using three variables: Registered air transport carrier departures (LAIRTC); air transport passengers carried (LAIRTP); and international tourism arrivals (LITA). We used the principal component analyses (PCA) of these three variables to generate an index of international exposure (LIEX).

**population dynamics:** After importing a few cases, there is the second stage when local inter-personal transmission kicks in. This stage depends on a number of factors. One is population dynamics, mainly density and movements including internal displacements. We use national

---

<sup>1</sup> Our World in Data is a collaborative effort between the University of Oxford, as the scientific editor and Global Change Data Lab, who publishes and maintains the website and the data tools.

population density (LNPD), urban population density (LUPD), the population share of over-55 (LP55) and internally displaced people (LIDD). Due to strong collinearity, not all these variables could enter the models at once. However, the index based on the PCA of the variables (LPD) were used in a separate model.

**Sanitation:** together with population dynamics, one of the possible determinants of propagation of communicable viruses like the SARS-CoV-2 that causes the COVID-19 is the hygiene and sanitation-related behavioural factors. In the available data, we have two kinds of sanitation-related variables. Variables for good sanitary conditions, which is share of population using safely managed sanitation services (LSAN). We also exploited information in this variable and other related variables reported in Table 2, using PCA to generate a broader index for good sanitation (LSANI). We also use water and sanitation related mortality (LWSM) to capture poor sanitation.

**Substance consumption** can expose individuals to disease and fatality in two ways. First most substances like alcohol happen in certain social setting that facilitates contact with potential sources of diseases. The second channel is the effect of these substances on the health of the individual. The Global Burden of Disease (GBD) 2016 collaborators [17] generated improved estimates of alcohol use and alcohol-attributable deaths and disability-adjusted life-years (DALYs) for 195 locations from 1990 to 2016. They concluded that Alcohol use is a topmost risk factor for global disease burden, explaining about 10% of mortality among the 15–49 years’ age group globally. However, most importantly some research [18] show that alcohol drinkers tend to have more friends and feel more engaged with, and trusting of, their local community. While this may be good for social cohesion, it can turn out to be a significant channel of transmission of a pandemic like the COVID-19. Other studies have confirmed the social connectedness effects of alcohol among young university students [19,20], implying that the effect applies for all ages. Besides the social channel of alcohol’s effect on the transmission of the pandemic, another way through which alcohol consumption heightens transmission of infections could be the disinhibition channel [21,22]. This channel postulates that that a sober individual’s behaviour is inhibited but when influenced by alcohol the inhibitions weaken, leading to less control over behaviour.

**Air pollution** is another important factor that can affect the rate of mortality from diseases like the COVID-19, especially as pollutions can be associated with upper respiratory disease and other morbidities. We use mean annual exposure to air pollution (LAPEX) and the PCA of the series of WHO related measures of exposure to air pollution and related mortality (LMAP) to generate the index of air pollution (LAPI) as shown in Tables 1 and 2.

The details of variables used in generating the different indices are in Table 2. The log of gross domestic product per capita (LGDPPC) is used to control for income levels across countries as a key economic variable. I also control for time effect, using DAY, capturing the time elapsed from the detection of first case. The one model I fitted contains the individual variables (equations 2), while the other is a model of indices generated using PCA as described above (equations 3).

### ***Econometric specifications for cumulative infections per thousand population***

$$LTI_{it} = \alpha_0 + \beta_1 LTT_{it} + \gamma_1 L ATP_i + \gamma_2 L ATC_i + \gamma_3 L ITA_i + \gamma_4 L PDEN_i + \gamma_5 L SAN_i + \gamma_6 L AC_i + \gamma_7 L CP_i + \gamma_8 L GDPPC_i + \delta DAY_{it} + \varepsilon_{it} \quad (2)$$

$$LTI_{it} = \alpha_0 + \beta_1 LTT_{it} + \gamma_1 LIEX_i + \gamma_2 LPD_i + \gamma_3 LSANI_i + \gamma_4 LSC_{it} + \gamma_5 LGDPPC_{it} + \delta DAY_{it} + \varepsilon_{it} \quad (3)$$

The different variables used in the models are captured in Tables 1 and 2 below together with the explanations of how they were each captured.

Table 1: individual variables and description

<b>Variables</b>	<b>Description (logs of)</b>	<b>Variables</b>	<b>Description (logs of)</b>
LTI	Total cases per 1000 of population*	LAPEX	mean annual exposure to air pollution, ( $\mu\text{g} / \text{m}^3$ )
LTD	Total deaths per thousand infections*	LSAN	% of population using safely managed sanitation services
LTT	Total tests per 1000 of population*	LNURS	Nurses per 1000 population
LNPD	National Population density	LMAL	Incidence of malaria per 1,000
LUPD	Urban population density	LTB	Incidence of tuberculosis per 100,000
LP55	Over-55 population share	LMNCD	Cause of death, by non-communicable diseases
LATP	Air transport passengers carried	LMAP	Mortality due to household and ambient air pollution,
LATC	Registered air transport carrier departures	LWSM	Mortality attributed to unsafe water, sanitation and lack of hygiene (per 100,000 population)
LITA	International tourism arrivals	LGDPPC	Log of GDP per capita
LAC	Alcohol consumption per capita	DAY	Number of days since the first detected case in a country
LSC	Smoking prevalence for age 15+		

Note: \*these variables are sourced from Our World in Data [15]. All other variables are from World Bank [16] WDI.

Table 2: PCA index variables and description

<b>Acronym</b>	<b>Meaning (log)</b>	<b>Description (PCA of the following variables)</b>
LIEX	International exposure index	air transport passengers carried; air transport registered carrier departures; international tourism arrivals
LPD	Population dynamics index	national population density; urban population density; population in the largest city; number of internally displaced persons
LSANI	Sanitation index	Rural and urban populations with basic drinking water services; rural and urban populations with basic sanitation services; population with safely managed drinking water services
LDP	Deprivation index	Poverty headcount ratio at \$1.90 a day (2011 PPP) (% of population); Poverty headcount ratio at \$3.20 a day (2011 PPP) (% of population); Poverty headcount ratio at \$5.50 a day (2011 PPP) (% of population); Poverty headcount ratio at national poverty lines (% of population)
LAPI	Air pollution index	Mortality rate attributed to household and ambient air pollution, age-standardized (per 100,000 population); PM2.5 air pollution, mean annual exposure (micrograms per cubic meter); PM2.5 air pollution, population exposed to levels exceeding WHO guideline value (% of total); PM2.5 pollution, population exposed to levels exceeding WHO Interim Target-1 value (% of total); PM2.5 pollution, population exposed to levels exceeding WHO Interim Target-2 value (% of total); PM2.5 pollution, population exposed to levels exceeding WHO Interim Target-3 value (% of total)

Note: all the variables used for the PCA indices are from World Bank [16] WDI.

## Estimation techniques

There are two main issues with the kind of data used in this work, which require special econometric techniques to deal with them. One is the high number of small values and/or zeros.

The second is the strongly nonlinear distribution in the data, especially the observations of cumulative infections. Based on this, it is possible that at lower values of infections, certain variables will behave differently relative to higher values of the dependent variables. On this basis, I choose to estimate the model, first, with Poisson Pseudo Maximum Likelihood (PPML), and then by means of quantile regression.

### ***Poisson Pseudo Maximum Likelihood (PPML)***

The choice of the PPML estimator can be justified on two main grounds as mentioned above. Possible heteroscedasticity and accounting for valid zero values in the observations, especially in the earlier periods before the infection. The PPML is borrowed from international trade research, which has shown its usefulness in estimating data with significant number of valid zeros and rapidly (exponentially) rising observations such as with COVID-19 infection. The estimator has been shown, in the context of international trade data analyses, to perform well in the presence of the above-named issues [23, 24]. The associated model for PPML estimation is specified as follows:

$$H_{it} = EXP[\alpha_0 + X_{it}\beta + Z_i\gamma + \delta day_{it} + \varepsilon_{it}] \quad (4)$$

These estimators are applied to panel data for 195 countries over days of infection from first recorded case. The COVID-19-related data is from our world in data, and the socioeconomic variables are from the World Bank's World Development indicators.

### ***Quantile regressions***

The other model is the quantile estimator. While the PPML estimates are averages over the distribution, it is possible to exploit the extra information provided by quantile regressions to draw inferences on how the explanatory variables behave at the top of the distribution of the dependent variable relative to the bottom or middle of the distribution. In addition to exploiting the non-linear characteristics of the data in order to derive non-biased point estimates at each quantile, the quantile estimator is also able to compare estimates for different quantiles. This is particularly useful in recommending what societal variables become most significant for curbing the spread at catastrophic levels of a pandemic like COVID-19 when existing health systems become overwhelmed. This quality is important in the context of rapidly rising rate of COVID-19 infections within a short span of time. In the current case for cumulative infections, the model for linear quantile regression for the  $q^{\text{th}}$  quantile can be expressed as:

$$H_{it} = \alpha_{q0} + X_{it}\beta_q + Z_i\gamma_q + \delta_q day_{it} + \varepsilon_{it} \quad (5)$$

The assumption is that the  $q^{\text{th}}$  quantile of  $\varepsilon_{it}$  is zero. The  $\alpha_{q0}$ ,  $\beta_q$ ,  $\gamma_q$ , and  $\delta_q$  would be derived from the solution of:

$$\min_{\varphi_q} [\sum_{\{i:H_{it} \geq (F(X,Z))\}} q |H_{it} - (\varphi(X,Z))| + \sum_{\{i:H_{it} < (F(X,Z))\}} (1-q) |H_{it} - (F(X,Z))|] \quad (6)$$

We implement (8) for 25<sup>th</sup>, 50<sup>th</sup> and 75<sup>th</sup> quantiles.

## **Key results and discussion**

In this section, I interpret, explain and discuss the results of the estimates. I begin with the summary statistics in Table 3 and proceed to correlation matrix in Table 4. I proceed with the report of the results of the PPML point estimates, and finally the quantile regression results.

### ***Summary statistics and correlations***

The summary statistics in Table 3 shows that on average, 0.24 persons per thousand of the populations of the sample countries are infected by the COVID-19-causing virus, with significant dispersions across countries. There are on average, 5.3 persons per thousands of population tested, which some countries achieving as low as zero persons tested and other as high as 128 persons tested. The fact that countries who test more tend to have higher reported cases means that countries with low levels of testing have systematic under-reporting of infected cases. This explains why we control for testing in the infections model.

Table 3: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
LTI	8,037	0.24	0.80	0.00	12.61
LTT	2,902	5.33	11.95	0.00	128.27
LATC	7,246	4.09E+05	1.26E+06	0.00	9.88E+06
LATP	7,246	4.65E+07	1.27E+08	0.00	8.89E+08
LITA	8,190	1.16E+07	1.83E+07	1.40E+04	8.93E+07
LNPD	8,537	353.45	1.50E+03	0.14	1.92E+04
LUPD	7,320	1.82E+03	2.52E+03	60.81	2.01E+04
LP55	8,584	8.82	4.33	0.00	15.74
LSAN	8,523	82.05	25.34	7.32	100.00
LAC	7,931	6.85	4.14	0.00	15.20
LSC	6,825	21.95	9.02	2.00	45.90
LAPEX	8,111	84.68	30.70	0.00	100.00
LGDPPC	8,369	2.09E+04	2.63E+04	210.80	1.96E+05
LIEX	7,044	0.82	1.69	0.00	11.80
LPD	6,365	1.42	1.56	0.00	10.25
LSANI	3,373	17.18	2.44	0.00	18.67
LAPI	7,859	3.03	1.98	0.00	7.36
DAY	8,584	26.79	19.94	1	109

There are on average 26 deaths per thousand recorded infections in our sample. Not much useful information would be expected from the correlation coefficients in Table 4 due to the strong non-linear nature of the key variables. We can only say that tests and infections are strongly and highly positively correlative as postulated earlier. Reliable results can only be inferred from appropriate econometric models that effectively control for statistical issues present in the data.

Table 4: Correlation coefficients

	LTI	LTT		LTI	LTT
LTT	0.9162*	1	LSC	0.1480*	0.1978*
LATC	-0.0500*	-0.019	LAPEX	-0.1453*	-0.2654*
LATP	-0.0366*	0.005	LGDPPC	0.4552*	0.5652*
LITA	-0.0649*	0.1463*	LIEX	-0.0156	0.0854*
LNPD	0.1179*	-0.1482*	LPD	-0.0954*	-0.2283*

LUPD	-0.3043*	-0.4829*	LSANI	0.2309*	0.4024*
LP55	0.3348*	0.4620*	LDP	-0.3559*	-0.5424*
LSAN	0.3523*	0.3857*	LAPI	-0.2199*	-0.3775*
LAC	0.1605*	0.3591*	DAY	0.4388*	0.5266*

## Econometric results

The econometric results are reported in Tables 5 and 6. The first results (in Tables 5) are point estimates for infections using the PPML estimator. For robustness check, I estimate simple pooled heteroscedacity-corrected ordinary least square models (POLS) for the point estimates. The second results are quantile estimates reported in Tables 6.

### *PPML point estimate results for cumulative infections*

The PPML estimates in Table 5 suggest good fit of the data. The high R-squares of the estimations are expected since the time series (days) dimension exceeds the number of cross-sections (countries). The sub-models estimated are individual variables in the cumulative infections model in columns 1 and 2, and the indices generated using the PCA method in columns 3 and 4. The variables are all in log form and therefore the estimated coefficients can be interpreted as elasticities.

Table 5: Average estimates for cumulative infections

VARIABLES	(1) POLS	(2) PPML	(3) POLS	(4) PPML
LTT	0.8543*** (0.0177)	0.8960*** (0.0210)	1.1748*** (0.0204)	0.9414*** (0.0358)
LATC	0.0954 (0.0762)	0.0161 (0.0477)		
LATP	0.0867 (0.0641)	0.0046 (0.0388)		
LITA	0.5429*** (0.0358)	0.1563*** (0.0305)		
LNPD	0.1494*** (0.0215)	0.0974*** (0.0130)		
LSAN	-0.4509*** (0.1234)	-1.6590*** (0.1818)		
LAC	-0.5034*** (0.0389)	0.1752*** (0.0644)		
LGDPPC	-0.3646*** (0.0428)	0.4733*** (0.0511)	-0.2179*** (0.0768)	0.2897*** (0.0894)
day	0.0272*** (0.0019)	-0.0007 (0.0017)	0.0157*** (0.0022)	0.0044* (0.0025)
LIEX			0.1585*** (0.0283)	0.1214*** (0.0327)
LPD			0.2909*** (0.0543)	0.4663*** (0.0533)
LSANI			-5.8893*** (0.3275)	-2.8294*** (0.4207)
LSUB			0.2038*** (0.0512)	0.1518* (0.0810)
LAPI			0.0644** (0.0252)	0.0299 (0.0254)
Constant	0.4150	-3.2422***	15.4883***	2.4842*

	(0.4625)	(0.6805)	(0.8168)	(1.3827)
Observations	2,234	2,234	1,138	1,138
R-squared	0.7606	0.8493	0.9179	0.8431

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

The coefficients of the variables are interpreted following the different determinants and the respective indices postulated above, which are testing, international exposure, population dynamics, sanitation, alcohol and cigarette use, air pollution and levels of per capita income.

*Testing rate:* As expected, the coefficients of the log of total testing is strongly positive and significant. The point estimate in the individual variables model of 0.9 means that 1% increase in total testing per thousands of a country's population on average leads to 0.9% increase in recorded cumulative infections per thousands of population. The coefficient in the index-based model is similar in magnitude and significance. This result suggest that testing translates into almost one to one percentage change. The finding here is consistent with the call for enhanced testing rates, since it will lead to discovery of more cases and contact-tracing [25].

*International exposure:* As earlier hypothesised, early infections and in some cases re-infections (in the case of China) come mostly from international travellers. The positive signs of coefficients of the international travel-related variables support this hypothesis. However, only the coefficient of the log of international tourism arrivals (LITA) is significant. Due to significant collinear effects among these variables, the PCA index (LIEX) of these would be more appropriate. The coefficient of LIEX suggests that a 1% increase in international exposure index of a country result on average in a 0.12% increase in cumulative infections per thousands of population. The implication is that the countries who responded by closing border would have reduced infections by 0.12% on average

*Population dynamics:* of all the variables that entered into the population dynamics, only the log of national population density (LNPD) could be allowed into the individual variables model due to strong collinearity. In accordance with expectation, a 1% increase in population density results in 0.1% increase in cumulative infections. The index of population dynamics (LPD) includes other variables like urban population density, internally displaced people and over-55 population share. A 1% increase in the index results in 0.5% increase in cumulative infections.

*Sanitation:* sanitation is the one variable that shows the strongest effect in reduction cumulative infections. A 1% increase in the population using managed basic sanitation (LSAN) results in 1.7% reduction in cumulative infections. The log of sanitation index (LSANI), which includes other sanitation related variables (Table 2) has even a bigger effect in reducing infections. A 1% increase in the index leads to 2.8% reduction in the cumulative infections. This model shows that sanitation is the single most important factor to act on in preparing a population to curb, or be less exposed to highly infectious communicable diseases like the COVID-19.

*Alcohol and substance use:* 1% increase in per capita alcohol consumption (LAC) is associated with 0.18% rise in cumulative infections per thousand of population. Cigarette prevalence was strongly collinear with most of the variables and we excluded it from the individual variables model. The index of substance use (LSUB) which includes alcohol and cigarette variables also show a strong positive effect on infections, with an elasticity of 0.15%. the fact that the combine coefficient of the index is similar to the index of alcohol consumption implies that alcohol plays more role in transmission of the disease than smoking. This is largely expected as alcohol consumption likely affects behavioural patterns of individuals and may cause them to be more exposed to vectors.

Secondly, alcohol consumption is mostly done in social settings and poses a risk in propagation of the disease. This finding lends credence to the ban of alcohol as a way of curbing the spread of the disease.

The air pollution index was not significant in propagating the disease. Higher income levels are associated with higher infections. This is only capturing the fact that richer countries with higher per capita income were the first to get the disease and were also first with to recorded high and rising cases before poorer countries.

### *Quantile estimates for cumulative infections*

The purpose of the quantile regression modelling is to determine the evolution of the effects of the social determinants of health as the infections evolve from early and low cases to catastrophic levels. We divide the distribution of the dependent variables into three quantiles (q25, q50, q75). Table 6 shows the results for the cumulative infections per thousand of population. In the models, both the dependent variables and regressors are logged hence the coefficients are elasticities. the model with individual variables are in columns 1 for q25, 2 for q50 and 3 for q75. The model that takes indices of the different determinants are in columns 4 for q25, 5 for q50 and 6 for q75.

Table 6: Quantile estimates for cumulative infections

VARIABLES	(1) q25	(2) q50	(3) q75	(4) q25	(5) q50	(6) q75
LTT	1.2386*** (0.0299)	1.1188*** (0.0202)	1.0686*** (0.0220)	1.2127*** (0.0576)	1.0700*** (0.0459)	1.0947*** (0.0572)
LATC	-0.2978*** (0.0779)	-0.0527 (0.0699)	0.1320** (0.0628)			
LATP	0.0963 (0.0607)	-0.0458 (0.0617)	-0.1240* (0.0668)			
LITA	0.4779*** (0.0598)	0.2808*** (0.0343)	0.1733*** (0.0405)			
LNPD	0.0962*** (0.0252)	0.1423*** (0.0411)	0.1854*** (0.0380)			
LSAN	0.2618** (0.1174)	0.3279*** (0.0755)	0.7354*** (0.1293)			
LAC	-0.4477*** (0.0490)	-0.5019*** (0.0367)	-0.4129*** (0.0575)			
LSC	-0.8564*** (0.0802)	-0.6173*** (0.0565)	-0.6456*** (0.0934)			
LGDPPC	0.0469 (0.0827)	0.1907*** (0.0467)	0.0957 (0.0765)	-0.5417** (0.2130)	0.1363 (0.2301)	0.0654 (0.2000)
day	-0.0127*** (0.0026)	-0.0027 (0.0018)	-0.0021 (0.0026)	0.0075 (0.0072)	0.0232*** (0.0051)	0.0144*** (0.0051)
rid1	0.4455*** (0.0909)	0.0996 (0.0686)	-0.1347* (0.0753)	0.2421 (0.2056)	0.6768*** (0.1991)	-0.1841 (0.2488)
LIEX				0.1286* (0.0719)	0.0576 (0.0609)	-0.0120 (0.0351)
LPD				0.2924*** (0.1111)	0.1759 (0.1730)	0.3506** (0.1566)
LSANI				-0.5622 (1.2307)	-5.1647*** (1.1118)	-4.8606*** (0.5881)
LSUB				-0.1085 (0.1383)	0.2397*** (0.0901)	-0.0302 (0.1609)
LAPI				0.0231 (0.0520)	0.1038 (0.0648)	0.0854 (0.0545)

Constant	-7.5555*** (0.5444)	-7.1143*** (0.3078)	-6.8476*** (0.4080)	3.4894 (2.2241)	9.4185*** (2.8166)	10.4917*** (2.3022)
Pseudo R2	0.6433	0.6508	0.6353	0.6363	0.6556	0.6575
Observations	2,206	2,206	2,206	1,138	1,138	1,138

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

*Testing:* the results show that the coefficients of the log of total testing decreases slightly as one moves from quantiles 25 to 75. While a percentage increase in testing rate results in 1.24% increase in cumulative infections in q25, the effects in q50 is 1.12% and 1.07% for q75. This depicts the fact that although more testing reveals more cases, at very high recorded cases, more testing results in marginally less new cases. This is capturing the possibility that in countries we extremely high cases, the growth rate of discovery of new cases may be decreasing and the curve is being flattened.

*International exposure:* the coefficient of the international tourism arrivals suggests that 1% increase results in 0.48% increase in infection within the first 25 percentiles of total infections. This elasticity decreases to 0.28% in q50, and 0.17% in q75. The elasticity of the index decreases from 0.13% to -0.01%, but is significant only in the 25<sup>th</sup> quantile. This result captures the fact that international exposure is significant only for the early cases of infections, after which population dynamics takes over. Of course, this is in the context of the lock-downs that countries have implemented following the first early cases. Therefore, these results only show that the closure of international borders is largely effective and the results would have been different with opened borders.

*Population dynamics:* moving from q25 to q75, the elasticities of national population density increases from 0.10% in q25, 0.14% in q50 to 0.19% in q75. The elasticities of the index behave similarly, raising from 0.30% in q25 to 0.35% in q75, and not significant in q50. We know that in most countries, authorities first began to respond by closing international borders, and subsequently restricting local travel which would have significantly affected internal population dynamics. This may explain why population dynamic coefficient is lower in q50, and also insignificant. However, in q75, the coefficient rises to 0.35 and is the highest. This is picking the fact that when a country persists long enough in restrictions of local movements, adverse socioeconomic conditions kick-in and many citizens start violating the lock-down, and hence the population dynamics variable becomes high and strongly significant again in q75. This argument is supported by the fact that the introduction of the socioeconomic variable in the model significantly reduces the elasticity of population dynamics. Overall, the elasticities of this variable confirm the hypotheses that while early cases are largely explained by international travel, later and explosive evolution of the infections is due to internal population dynamics. There is a suggestion here that measures to contain internal population dynamics has to be accompanied by some socioeconomic support especially if these measures are to last long.

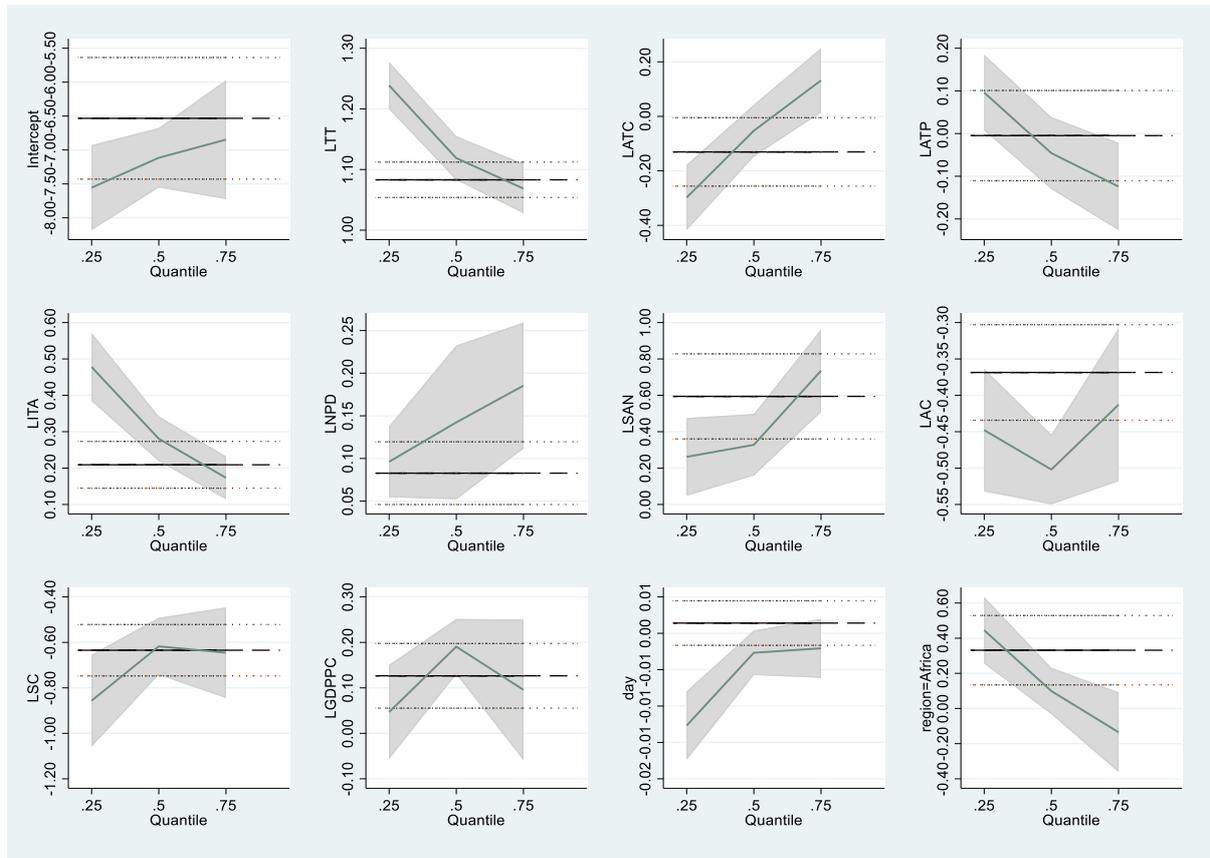
*Sanitation:* sanitation is effective in decreasing infections, and the effects increases significantly as cases rise. A 1% increase in sanitation measures and practices results in 0.26% reduction in infections in the 25<sup>th</sup> quantile, 0.33% in the 50<sup>th</sup> quantile and 0.74% reduction in the 75<sup>th</sup> quantile. The elasticities of the index have stronger effects. A 1% increase in the index brings about 0.56%, 5.17% and 4.86% reductions respectively. Sanitation is therefore the single most effective measure in reducing infections and its effects are higher and more significant at high levels of cumulative infections. This implies that at very high infection levels, countries with high levels on sanitary practices are more likely to attenuate the increases in infections.

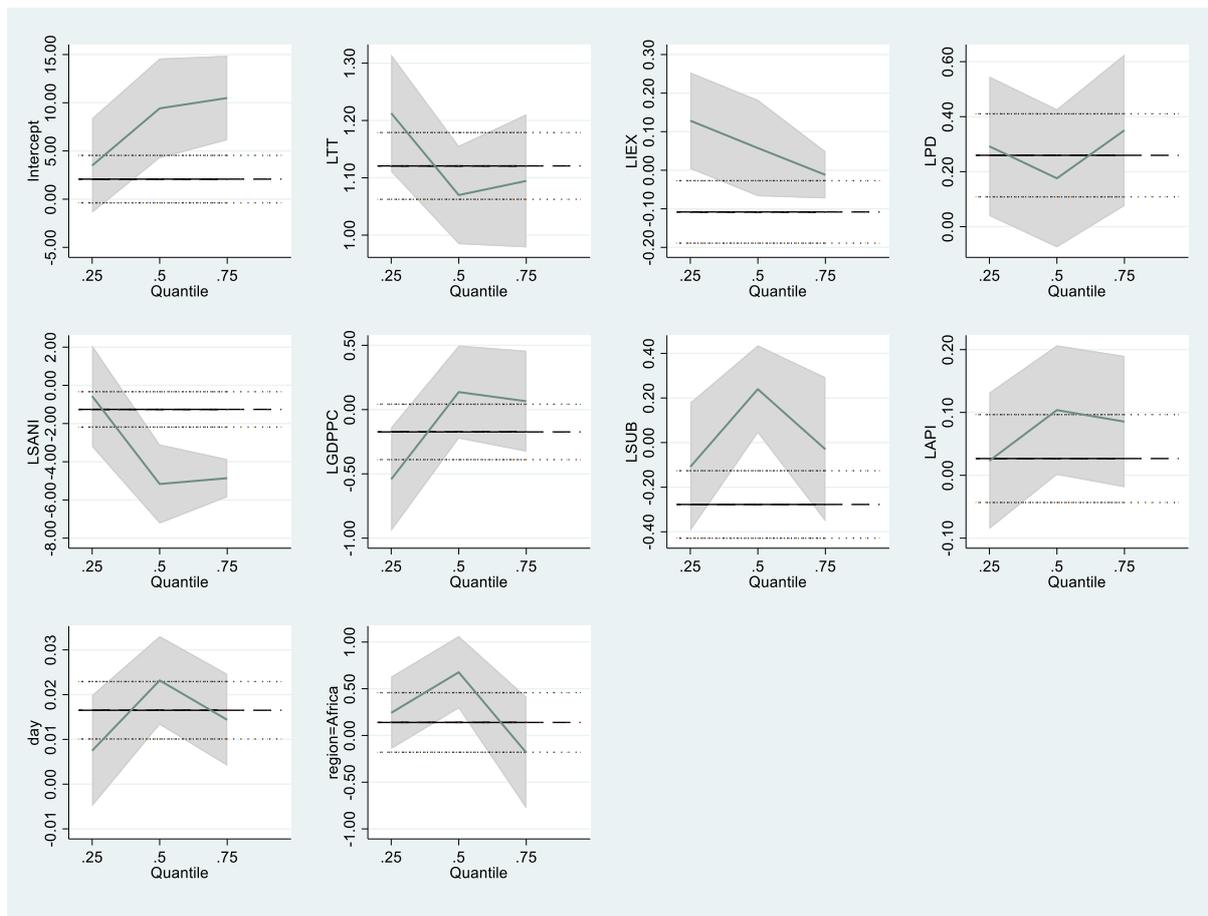
*Alcohol and substance use:* the combine effects of alcohol and cigarettes as captured by the index of their PCA shows that these substances increase rate of infections at medium infection levels (q50). A 1% increase in substance use results in 0.24% higher infections in the 50<sup>th</sup> quantile. The

coefficients are not significant for q25 and q75. We know that at the early stages of infection, local conditions are less important, so too local levels of substance use. However, as the cumulative infections rise, local dynamics kick in, and the social effects of substance use become an important aid in local transmission mechanisms. Air pollution was not significant in transmissions.

Figure 2 shows the plots of the estimated coefficients of the quantile regressions relative to the OLS average coefficients and their respective confidence interval bands. From the figures, one can clearly see the non-linear effects of the coefficients of the quintile estimators. The fact that none of those coefficients consistently lie within the confidence interval band of the average estimator validates the need and reliability of the above quantile estimators and interpretations.

Figure 2: comparative coefficients of OLS versus quantile regressions





## Policies for COVID-19 resilience and preparedness for future pandemics

The outcome of the analyses brings to light a number of findings that bear of public health policy design. Of the different variables I tested in the different models, testing coverage; international exposure; population dynamics; sanitation and substance use stand out for possible socioeconomic policy designs to complement public health policies. These can contribute in managing the current spread of the COVID-19 disease and build preparedness for possible future pandemics.

*Enhancing capacity for early testing and case management:* this policy proposition relating to testing is underscoring the fact that other socioeconomic policies are there to complement effective public health policies and systems. The degree of progress in testing is an indication of public health capacity for disease management in the sense that this it entails availability of equipment, and human and physical capacity. From the findings in this work, testing translates into almost one to one percentage change, consistent with the call for enhanced testing rates due to its role in the discovery of more cases and contact-tracing [25]. My finding that new cases decrease with testing at high cumulative infection rate show that early testing leads to better management of infections subsequently. Early testing helps to discover more cases, but also helps in isolation of those cases and tracing contacts in time for effective control of the spread, so that later on, less infections may be recorded. Testing therefore plays some role in flattening the curve.

*Preparedness for managing of international movements though reduction of international exposure:* managing international borders with effective preparedness for early and prompt action can prove very effective in managing the rate of spread and giving room for health systems to cope. Countries that acted early enough in restricting international movement would have gain significantly in curbing not only the initial rate of international transmission of the disease, but also the subsequent

rate of citizen-to-citizen transmission. The findings on this confirm that international exposure significantly raises cases only in the early periods of infections, due to the later international movement restrictions. Effective and early actions in this regard can go a long way. However, the fact that international tourism arrivals is the main channel of impact show that there is a great deal of socioeconomic trade-off to be made in respect of international travel restrictions. The tourism sector is an important income-generating sector for most of the citizens in poor countries. This underscores the importance of accompanying economic support measures together with measures in curbing the spread of the current and possible future pandemics.

*Managing population dynamics:* upon successful border management to curb infections due to international movements, the second and most important element is internal population dynamics. The lockdown measures that almost all countries have implemented is aimed at managing the spread of the pandemic by limiting movements of the local population. These population-restricting measures have been effective at first, but later turns ineffective. Restricting population movements without considering other unintended socioeconomic effects will weaken the impact of lowering population dynamics. If a country persists long enough in restrictions of local movements, adverse socioeconomic conditions kick-in and many citizens start violating the lockdown measures unless effective socioeconomic support is given. Therefore, measures to contain internal population dynamics has to be accompanied by some socioeconomic support especially if these measures are to last long.

*Ensuring better sanitation and hygiene practices:* This work reveals that adoption of good sanitation and hygiene practices is the single most important factor in curbing the spread of a pandemic like COVID-19. The more sanitation measure is taken at higher levels of infections, the more reduction in infections are obtained. Since the future of global health may be marked quite periodically by various infectious disease, this research recommends that in normal times, good hygiene practices should be inculcated into the populations. Policy measures should ensure that such good sanitation programmes should become a lifestyle of a society. This is the kind of measure that can prove very effective in flattening the infectious curve of the current pandemic, and reduce the exposure of a population to future similar pandemics. This finding has significant implication in raising the awareness of governments in respect of provision of clean water to every community, together with programmes for lifestyle change towards good hygiene and sanitation practices.

*Reducing alcohol and substance use:* the use of alcohol can be strongly associated with social gatherings and other social practices that expose a population to the disease. While cigarette smoking prevalence has not shown a strong and significant effects in the transmission of the pandemic, alcohol seems to strongly do. In terms of spread of an infectious pandemic like COVID-19, countries with less alcohol consumption per capita will tend to have a relatively slower spread of the disease. Alcohol consumption can alter an individual's sense of caution, and lead such to a riskier behaviour that would lead to higher exposure to the disease. Alcohol can also lead to higher levels of contacts through associated social relationships, hence exposing a population to higher levels of infection. Despite this finding, on the one hand, banning alcohol altogether may lead to a more agitation in the population due to the addictive nature of alcohol, and hence lead to higher effect of population dynamics on the spread of disease. On the other hand, allowing the uncontrolled use of alcohol would clearly lead to higher infection rates. The solution for countries could therefore be an optimally controlled allowance of alcohol accompanied by restrictions in public places of drinking.

## Conclusion

This paper's aim has been to investigate which socioeconomic factors could be valuable in complementing public health systems in the face of rising COVID-19 infections and to ensure preparedness for future similar pandemics. Within the framework of social determinants of health, a model of COVID-19 infections was developed. The COVID-19-related data capturing infections and tests are from Our World in Data. The population dynamics data and other variables depicting the socioeconomic and sanitary context that prevailed in a country immediately following the disease outbreak are from the World Bank's World Development Indicators. The COVID-19 dataset is a daily panel of 195 countries from the first recorded COVID-19 case to current (21 April, 2020). This data is matched with country-level regressors relating to social determinants of health. These are the first non-missing annual values for each country beginning 2015.

There are two main issues in the data, requiring special econometric techniques to deal with them. One is the high number of small values and/or zeros. The second is the strongly nonlinear distribution of the observations relating to infections observations. For this reason, it is possible that at lower values of cumulative infections, certain variables will behave differently compared to higher values of the dependent variables. There are two suitable econometric techniques applicable in this context. The one is the Poisson Pseudo Maximum Likelihood (PPML), and the other is quantile regression.

The results of the analyses highlight a number of findings that bear on public health policy design. *The findings show that: enhancing for early testing* helps to discover more cases, but also helps in isolation of those cases and tracing contacts in time for effective control of the spread, so that later on, less infections may be recorded. Testing therefore plays some role in flattening the curve. *Preparedness for managing of international movements through reduction of international exposure* can prove very effective in managing the rate of spread and giving room for health systems to cope. However, the fact that international tourism arrivals is the main channel of impact show that there is a great deal of socioeconomic trade-off to be made in respect of international travel restrictions. *Managing population dynamics* is good for curbing infections, but restricting population movements without considering other unintended socioeconomic effects will weaken the impact of lowering population dynamics. *Ensuring better sanitation and hygiene practices* is the single most important factor in curbing the spread of a pandemic like COVID-19. Policy measures that towards good sanitation programmes, if they become a lifestyle of a society, will prove very effective in flattening the infectious curve of the current pandemic, and reduce the exposure of a population to future similar pandemics. *Reducing alcohol and substance use* also helps in curbing infections. Alcohol in particular, can also lead to higher levels of contacts through associated social relationships, hence exposing a population to higher levels of infection. The finding calls for a very fine balancing act in the control of alcohol use.

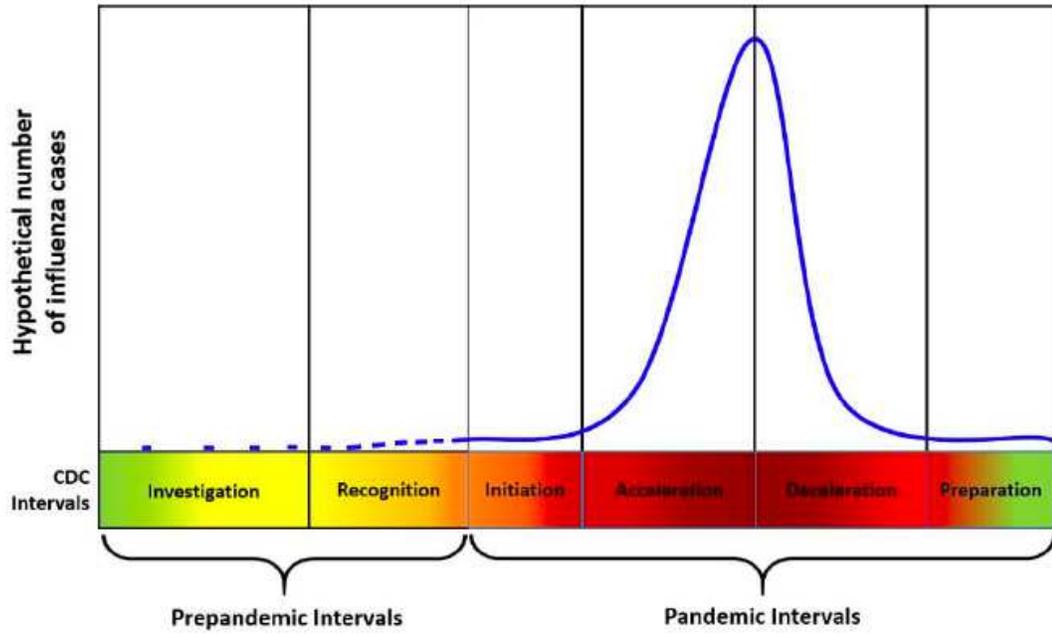
## References

- 1 Ramaphosa C. Statement by President Cyril Ramaphosa on further economic and social measures in response to the COVID-19 epidemic. The Presidency of South Africa. 2020. <http://www.thepresidency.gov.za/speeches/statement-president-cyril-ramaphosa-further-economic-and-social-measures-response-covid-19>. Accessed 21 April 2020.

- 2 ECDC. Situation update worldwide, as of 21 April 2020. European Centre for Disease Control and Prevention. 2020. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>. Accessed 21 April 2020.
- 3 de Andrade L, Filho A, Solar O, Rigoli F, de Salazar L, Serrate P, Ribeiro K, Koller T, Cruz F, Atun R. Universal health coverage in Latin America 3: Social determinants of health, universal health coverage, and sustainable development: case studies from Latin American countries, *Lancet* 2015; 385: 1343–5.
- 4 Horton R. (2013). Offline: Four principles of social medicine. *Lancet* 382: 192.
- 5 Huang C, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395(10223), 497-506.
- 6 Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta Biomed* 2020; 91(1): 157-160.
- 7 WHO. Coronavirus disease 2019 (COVID-19) situation report—44. World Health Organisation. 2020. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200304-sit-rep-44-covid-19.pdf?sfvrsn=783b4c9d\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200304-sit-rep-44-covid-19.pdf?sfvrsn=783b4c9d_2). Accessed April 27 2020.
- 8 Anderson R, Heesterbeek H, Klinkenberg D, Hollingsworth D. How will country-based mitigation measures influence the course of the COVID-19 epidemic?, *Lancet* 2020; 395(10228), 931-934, February
- 9 WHO. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. World Health Organization. 2008. [http://whqlibdoc.who.int/publications/2008/9789241563703\\_eng.pdf](http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf). Accessed April 21 2020.
- 10 Horwitz L, Chang C, Arcilla H, Knickman J. Quantifying health systems' investment in social determinants of health, by sector, 2017–19, *Health Affairs* 2020; 39(2): 192–198.
- 11 Marmot M, Wilkinson R. Social determinants of health. 2<sup>nd</sup> ed. Oxford: Oxford University Press; 2006.
- 12 Marmot, M., Allen, J., Bell, R., Bloomer, E. and Goldblatt, P. WHO European review of social determinants of health and the health divide, *Lancet* 2013; 380: 1011–29.
- 13 Viner R, Ozer E, Denny S, Marmot M, Resnick M, Fatusi A. and Currie, C. Adolescent Health 2: Adolescence and the social determinants of health *Lancet* 2012; 379: 1641–52
- 14 Yuan J, Li M, Lv G, Lu K. Monitoring Transmissibility and Mortality of COVID-19 in Europe. *International journal of infectious diseases* 2020; <https://doi.org/10.1016/j.ijid.2020.03.050>
- 15 Ritchie H. Coronavirus Source Data, Our World in Data. 2020. <https://ourworldindata.org/coronavirus-source-data>. Accessed 27 April 2020.
- 16 World Bank. World Development Indicators. Washington, D.C.: The World Bank. 2020. <https://databank.worldbank.org/source/world-development-indicators>. Accessed 27 April 2020.
- 17 GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2018; 392(10152): 1015-1035.
- 18 Dunbar M, Launay J, Wlodarski R. et al. Functional Benefits of (Modest) Alcohol Consumption. *Adaptive Human Behavior and Physiology* 2017; **3**: 118–133.
- 19 Sarah M. [Risk rituals and the female life-course: negotiating uncertainty in the transitions to womanhood and motherhood](#). *Health* 2020; *Risk & Society* 22(1): 15-30.

- 20 Lizabeth A, Crawford, Katherine B, Novak R., Jayasekare R. Volunteerism, Alcohol Beliefs, and First-Year College Students' Drinking Behaviors: Implications for Prevention. *The Journal of Primary Prevention* 2019; 40(4): 429-448.
- 21 Källmén H1, Gustafson R. Alcohol and disinhibition. *Eur Addict Res.* 1998; 4(4):150-62.
- 22 Steele M, and Southwick L. Alcohol and social behavior I: The psychology of drunken excess. *J Pers Soc Psychol.* 1985; 48(1):18-34.
- 23 Santos Silva JMC, Tenreyro S. The Log of Gravity. *The Review of Economics and Statistics* 2006; 88(4): 641-658.
- 24 Santos Silva JMC, Tenreyro S. Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator. *Economics Letters* 2011; 112(2): 220-222.
- 25 Hellewell J Abbott S, Gimma A, Bosse N, Jarvis C, Russell T, Munday J, Kucharski A, Edmunds J, Funk S, Eggo R. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, *Lancet Glob Health* 2020; 8: e488–96

# Figures



Source: CDC.gov, <https://www.cdc.gov/flu/pandemic-resources/national-strategy/intervals-framework.html>

Figure 1

CDC's hypothetical evolution path for the novel influenza-A virus pandemics

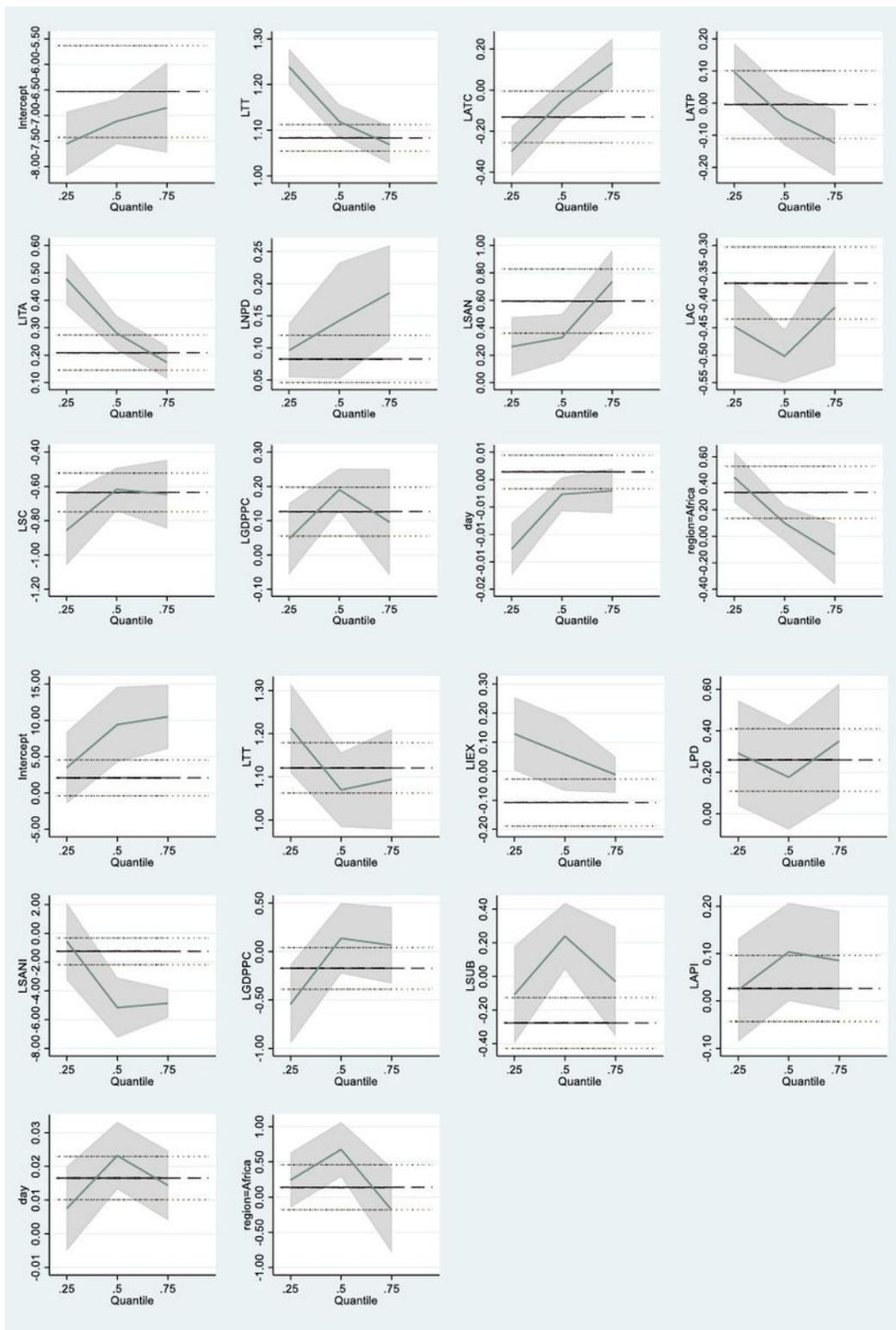


Figure 2

comparative coefficients of OLS versus quantile regressions