



# The impact of socio-economic indicators on COVID-19: an empirical multivariate analysis of sub-Saharan African countries

Phemelo Tamasiga<sup>1</sup> · Ashenafi Teshome Guta<sup>2</sup> · Helen Onyeaka<sup>3</sup> · Maureen Sindisiwe Kalane<sup>4</sup>

Accepted: 10 April 2022 / Published online: 19 May 2022

© The Author(s), under exclusive licence to Institute for Social and Economic Change 2022

## Abstract

The COVID-19 pandemic has triggered an unprecedented social and economic crisis. This study aims at investigating the impact of socio-economic indicators on the levels of COVID-19 (confirmed and death cases) in sub-Saharan Africa. The investigation makes use of the readily accessible public data: we obtain COVID-19 data from Johns Hopkins and socio-economic indicators from the World Bank. The socio-economic indicators (independent variables) used in the multilinear regression were GDP per capita, gross national income per capita, life expectancy, population density (people per sq. km of land area), the population aged 65 and above, current health expenditure per capita and total population. The dependent variables used were the COVID-19 confirmed and death cases. Amongst the seven socio-economic indicators, only 4 showed a statistically significant impact on COVID-19 cases: population density, gross national income per capita, population aged 65 and above and total population. The obtained  $R^2$  of 69% and 63% indicated that the socio-economic indicators captured and explained the variation of COVID-19 confirmed cases and COVID-19 death cases, respectively. The startling results obtained in this study were the negative but statistically significant relationship between COVID-19 deaths and population density and the positive and statistically significant relationship between gross national income per capita and COVID-19 cases (both confirmed and deaths). Both these results are at odds with literature investigating these indicators in Europe, China, India and the UK.

**Keywords** Socio-economic indicators · COVID-19 · Epidemiology · Multivariate linear regression · Sub-Saharan Africa

---

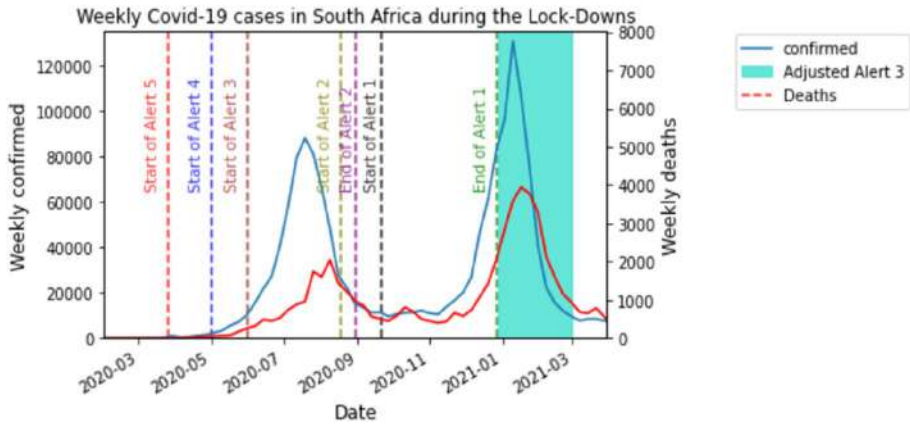
✉ Phemelo Tamasiga  
phemelo.tamasiga@gmail.com

<sup>1</sup> Duisburg, Germany

<sup>2</sup> Department of Economics, Addis Ababa University, Addis Ababa, Ethiopia

<sup>3</sup> School of Chemical Engineering, University of Birmingham, Birmingham, UK

<sup>4</sup> Communication and Study Skills Unit, University of Botswana, Private Bag UB 0022, Gaborone, Botswana



**Fig. 1** Restrictive measures and lockdowns in South Africa during COVID-19. *Source:* Authors' computation and illustration based on COVID-19 data obtained from Johns Hopkins

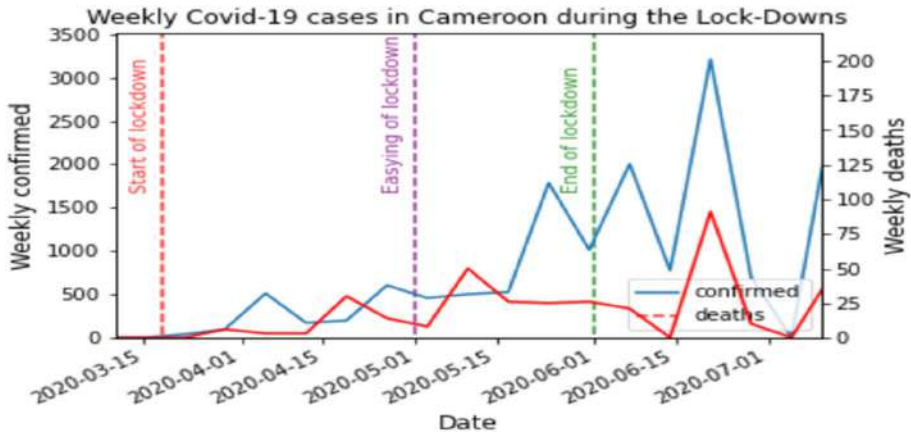
## Introduction

In December 2019, the novel coronavirus (COVID-19) was discovered in Wuhan City<sup>1</sup> and quickly spread throughout the world, bringing many countries' socio-economic systems to a standstill. In order to curb the spread of the virus, many countries implemented a variety of public health policies, such as travel bans, social distancing, quarantine in hospitals and government-designated areas, strict limits on the size of public gatherings and curfews. In the same frame of mind, Hartley and Perencevich (2020) asserted that these measures aim to reduce the spread of further infection and prevent health system strain. The implementation of these restrictive measures has produced varying results across sub-Saharan African countries. In some cases like in South Africa, the lockdowns seemed to have reduced COVID-19 cases (see Fig. 1). In other sub-Saharan African countries, for example in Cameroon, there were no clear results indicating a significant reduction in COVID-19 cases (see Fig. 2).

Given this caveat, socio-economic indicators could be used to study the differences in COVID-19 outcomes between sub-Saharan African countries. Applying socio-economic indicators can potentially aid in understanding the trends of the spread of the virus and could also assist in the formulation of strategic policies to mitigate the socio-economic consequences of the COVID-19 pandemic. We hypothesise that there is an interplay between socio-economic indicators and COVID-19 cases (confirmed and death cases). It has been suggested that the pandemic might be particularly detrimental in less-developed countries.

There are two strands of literature studying the correlation of COVID-19 cases and socio-economic indicators. The first strand is the literature focusing on the spread of COVID-19, like those of Mihoub et al. (2020). The second strand of literature describes correlations between socio-economic status, demographic factors and COVID-19. For example, the relationship between demographic factors such as old age and COVID-19 was investigated in the study of Applegate and Ouslander (2020). The socio-economic implications of COVID-19 were investigated by Alsafi et al. (2020). Other studies have

<sup>1</sup> 20,200,121-sitrep-1-2019-ncov.pdf (who.int).



**Fig. 2** Restrictive measures and lockdowns in Cameroon during COVID-19. *Source:* Authors' computation and illustration based on COVID-19 data obtained from Johns Hopkins

investigated non-socio-economic indicators and explored the interplay of COVID-19 with ethnicity, for example, in the study of Pareek et al. (2020). However, it has to be understood that these individual factors are likely to be influenced by various country-specific parameters including lockdown policies, public expenditure in health and other country-specific determinants such as level of sanitation, healthcare support, number of acute care beds, hospital beds and number of physicians.

Our paper adds to the growing body of literature on COVID-19 trends and socio-economic indicators. In comparison with existing literature, this study conducts a multilinear regression and correlation between COVID-19 cases and seven socio-economic indicators in a set of 40 sub-Saharan African countries. The 40 countries used in this study were selected based on data availability for the socio-economic indicators. We accessed publicly available COVID-19 data from Johns Hopkins. The country-specific socio-economic variables were obtained from the World Bank.

## Data and methodology

We conduct a multilinear regression analysis of all the sub-Saharan African countries. Data for the study have been taken from various secondary sources like Johns Hopkins and the World Bank. Socio-economic indicators were extracted from the World Bank website. Due to many missing values for 2019 and 2020, the 2018 data were extracted. A similar approach was adopted in a study by Mihoub et al. (2020) that applied machine learning techniques to predict COVID-19 cases in a multi-country model. Due to many missing indicators, Mihoub et al. (2020) implemented a machine learning model on COVID 2020 data using socio-economic indicators from 2017. The stages of data preparation to running the model and obtaining the results are summarised below:

*Data Collection and Assembly* We obtained data from the World Bank API from which we extracted the socio-economic indicators. We proceeded by obtaining data from Johns Hopkins to extract COVID-19 data.

**Table 1** Variables and data sources

Variables	Data sources	
<i>COVID-19 Data</i>		
COVID-19 (confirmed, deaths, active and recovered cases)	Data was extracted from Johns Hopkins and it was accessed at the following link: <a href="https://raw.githubusercontent.com/Lucas-Czarnecki/COVID-19-CLEANED-JHUCSSE/master/COVID-19_CLEAN/esse_covid_19_time_series_cleaned/time_series_covid19_cases_tidy.csv">https://raw.githubusercontent.com/Lucas-Czarnecki/COVID-19-CLEANED-JHUCSSE/master/COVID-19_CLEAN/esse_covid_19_time_series_cleaned/time_series_covid19_cases_tidy.csv</a>	
Socio-economic indicator name	World Bank indicator code	Source
<i>Socio-economic indicators data</i>		
Current health expenditure per capita (current US\$)	SH.XPD.CHEX.PC.CD	For this analysis, the following API has been selected: <a href="https://data.worldbank.org/">https://data.worldbank.org/</a>
Population density (people per sq. km of land area)	EN.POP.DNST	
Life expectancy at birth, total (years)	SP.DYN.LE00.IN	
Population aged 65 and above, total	SP.POP.65UP.TO	
GDP per capita (current US\$)	NY.GDP.PCAP.CD	
Population, total	SP.POP.TOTL	
Gross national income per capita	C1.8	

*Data pre-Processing* In both datasets of COVID-19 cases and socio-economic indicators, we treated missing values and changed the format of the feature variables to numerical. As part of pre-processing, we obtained only a subset of sub-Saharan African countries from the data sets and we merged the two datasets using the country as a unique identifier.

*Feature selection before running the model* We obtained a covariance matrix of features and produced a heatmap to get a high-level overview of the relationship between the COVID-19 cases and the socio-economic indicators.

*Model simulation in python 3.6* Once the above three stages were completed (data collection and assembly, data pre-processing and feature selection), we ran the multilinear regression model on the merged data set and obtained results.

## Data sources and variables used

The variables and data sources used are shown in Table 1.

We provide the reasons for selecting the specified variables in Table 1:

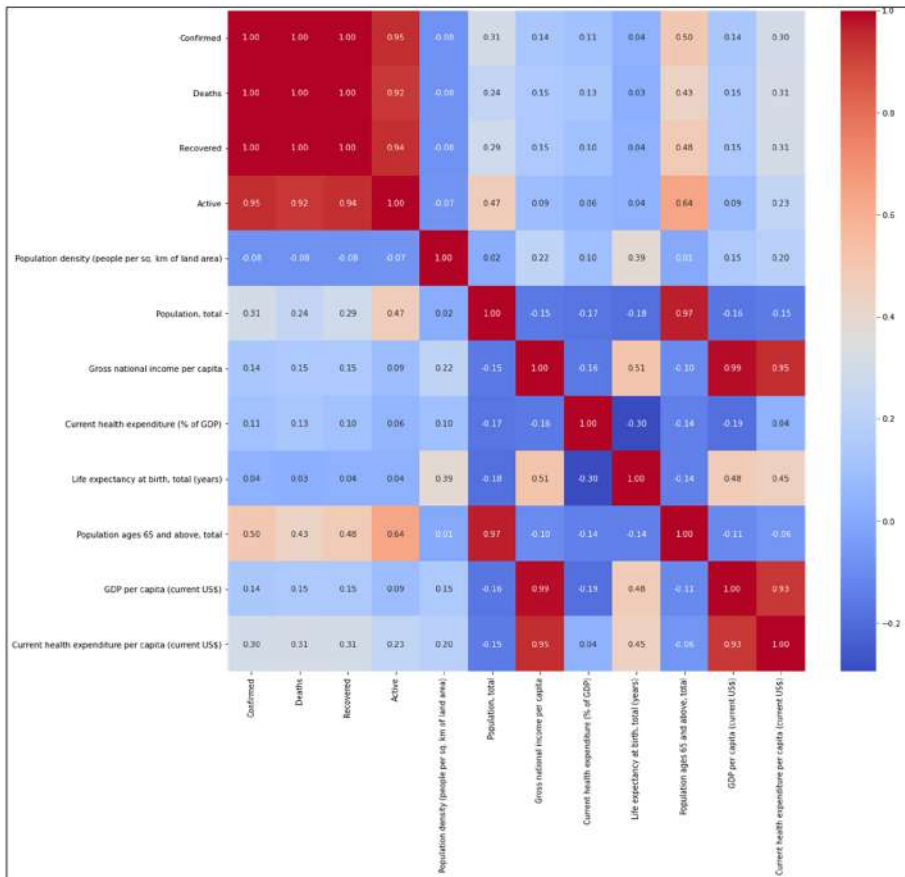
*Current health expenditure* This variable measures the preparedness of a country with respect to the COVID-19 outbreak. A high level of expenditure on health signifies that a country is health infrastructure ready to tackle the COVID-19 pandemic.

*Gross national income per capita* This variable was included to analyse the impact of income on the incidence of the COVID-19.

*Population* This variable was included because it can burden the national health facilities, especially in a health crisis.

*Life expectancy at birth* This variable indicates the number of years a newborn baby can live; it would enable an examination of the effect of higher life expectancy on health resources.

*Population aged 65 and above* There have been studies that showed that COVID-19 casualty is higher in countries with a large aging population.



**Fig. 3** Heat map showing the correlation of socio-economic indicators and COVID-19 cases for sub-Saharan African countries. *Source:* Authors’ computation and illustration based on COVID-19 data obtained from Johns Hopkins and socio-economic data obtained from the World Bank

*GDP per capita* This variable was included because it reflects access to affordable healthcare. It would be expected that countries with higher GDP per capita will have lower COVID-19 fatalities.

We proceed by finding the correlation of the defined variables above to determine how COVID-19 cases are related to the socio-economic indicators. Positive correlations would indicate that an increase in one value of one variable increases the value of the target variable and a negative correlation would indicate that an increase in one value of a variable decreases the value of the target variable. The heat map shown in Fig. 3 makes it easy to identify the socio-economic indicators which are most related to the COVID-19 cases.

The red colour indicates strong perfect correlations whereas the blue colour indicates weak correlations between the data points. For example, the correlation between COVID-19 cases and the population was 0.81 and 0.77 per the confirmed and death cases, respectively, whereas the correlations between COVID-19 cases and the population aged 65 and above were 0.84 and 0.79 for the confirmed and death cases, respectively. The population

**Table 2** Descriptive statistics

	Confirmed	Deaths	Recovered	Active	Population density (people per sq. km of land area)	Population, total	Gross national income per capita	Life expectancy at birth, total (years)	Population aged 65 and above, total	GDP per capita (current US\$)	Current health expenditure per capita (current US\$)
Count	40	40	40	40	40	40	40	40	40	40	40
Mean	14,609,023	346,725	12,375,323	1,650,602	118	20,876,495	2572	62	622,146	2765	145
Std	52,863,638	1,557,098	47,290,295	4,304,911	145	35,106,472	3307	5	1,076,711	3527	189
Min	227,314	862	0	0	2	96,762	280	52	6174	271	22
25%	1,241,547	25,495	735,244	175,497	34	2,116,489	670	59	93,663	759	40
50%	3,685,584	42,477	2,943,539	478,685	66	11,893,493	1355	63	362,350	1440	62
75%	7,603,707	119,257	5,472,617	859,952	137	23,136,270	3195	66	607,511	3178	129
Max	336,055,190	9,913,283	300,751,993	25,389,914	623	195,874,740	15,670	74	5,381,418	16,390	833

density, current health expenditure per capita, GDP per capita and gross national income per capita all have low but negative correlations with COVID-19 cases.

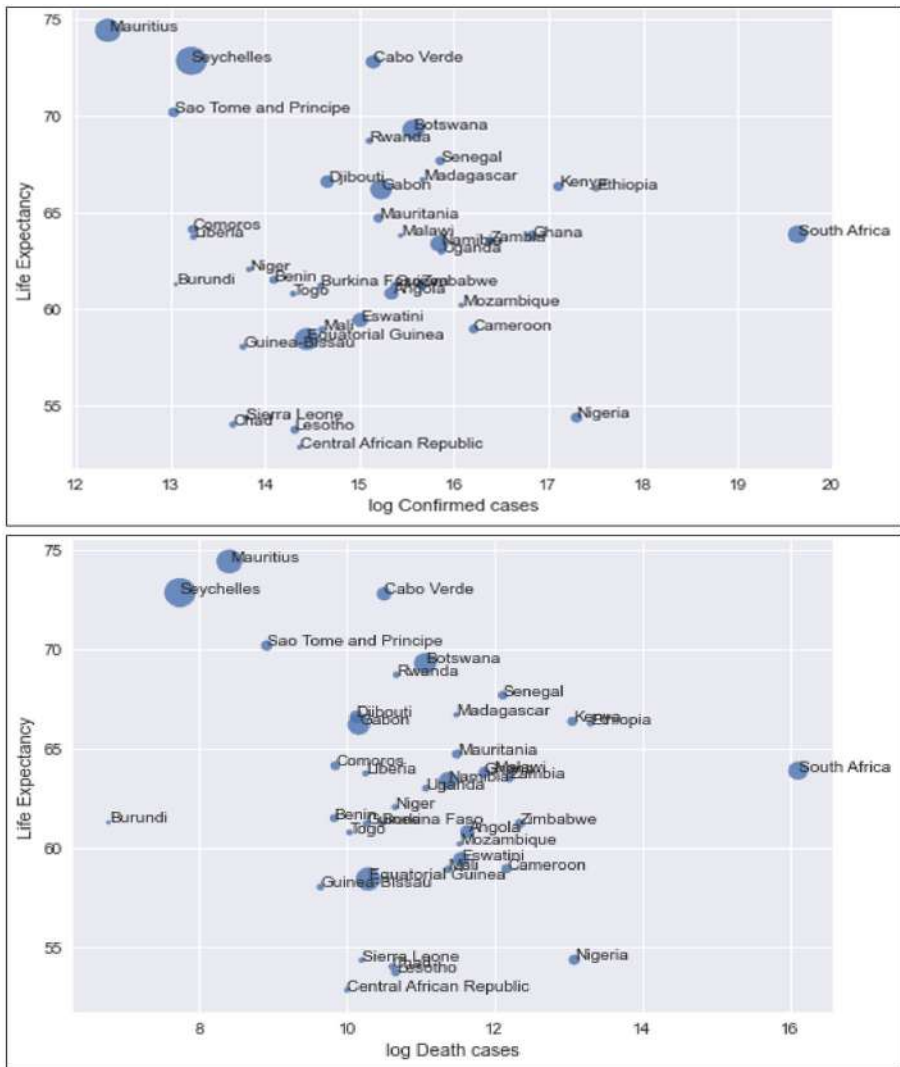
## Descriptive statistics

The software package used to produce the descriptive statistics in Table 2 was python 3.6.

The descriptive statistics in Table 2 shows that 40 sub-Saharan African countries were used in this study (Angola, Benin, Botswana, Burkina Faso, Burundi, Carpo Verde, Cameroon, Central Africa Republic, Chad, Comoros, Congo Republic, Djibouti, Equatorial Guinea, Eritrea, Eswatini, Ethiopia, Gabon, Ghana, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, South Sudan, Togo. The mean, standard deviations, max and percentiles of all the variables are shown in Table 2. Taking the number of confirmed COVID-19 cases as an example, the reported mean is 14,609,203 with a min of 227,314, a max of 336,055,190 and a standard deviation of 52,863,638.

It is critical to understand the correlation between the age demographics and the number of COVID-19 cases. The younger the population of a country, the less likely is its population to die of COVID-19 hence low reported death cases in a country because the virus takes down a larger proportion of people above 60 or 65. Therefore, we anticipate a positive correlation between COVID-19 confirmed cases and the proportion of people above 60 or 65. The scatter plot in Fig. 4 indicating the spread of sub-Saharan African countries versus COVID-19 death cases shows that South Africa, with a life expectancy of about 65 recorded the largest COVID-19 related deaths, and countries such as Seychelles, Mauritius Sao Tome, with the highest life expectancy of 70+ recorded the lowest number of deaths in sub-Saharan Africa. Countries such as Lesotho, Sierra Leone, the Central Africa Republic with a life expectancy lower than 55 recorded COVID-19 deaths which are almost equal. Moreover, Fig. 4 shows that countries with similar life expectancy values recorded very close numbers of COVID-19 deaths, and this condition underlines the notion that there is an effect of life expectancy on COVID-19 cases (deaths), and we shall explore and test the statistical significance of this relationship in the OLS estimation. This scatter plot is in the same frame of mind as Verity et al. (2020), who found out that higher COVID-19 mortality rates are reported in countries with high life expectancy (> 70 years) than those countries with lower (< 70 years) life expectancy. In the same vein, Aburto et al. (2021) asserted that the study of life expectancy in the context of the COVID-19 pandemic matters because it enables us to compare the cumulative impacts of the pandemic against past mortality shocks and recent trends across different countries.

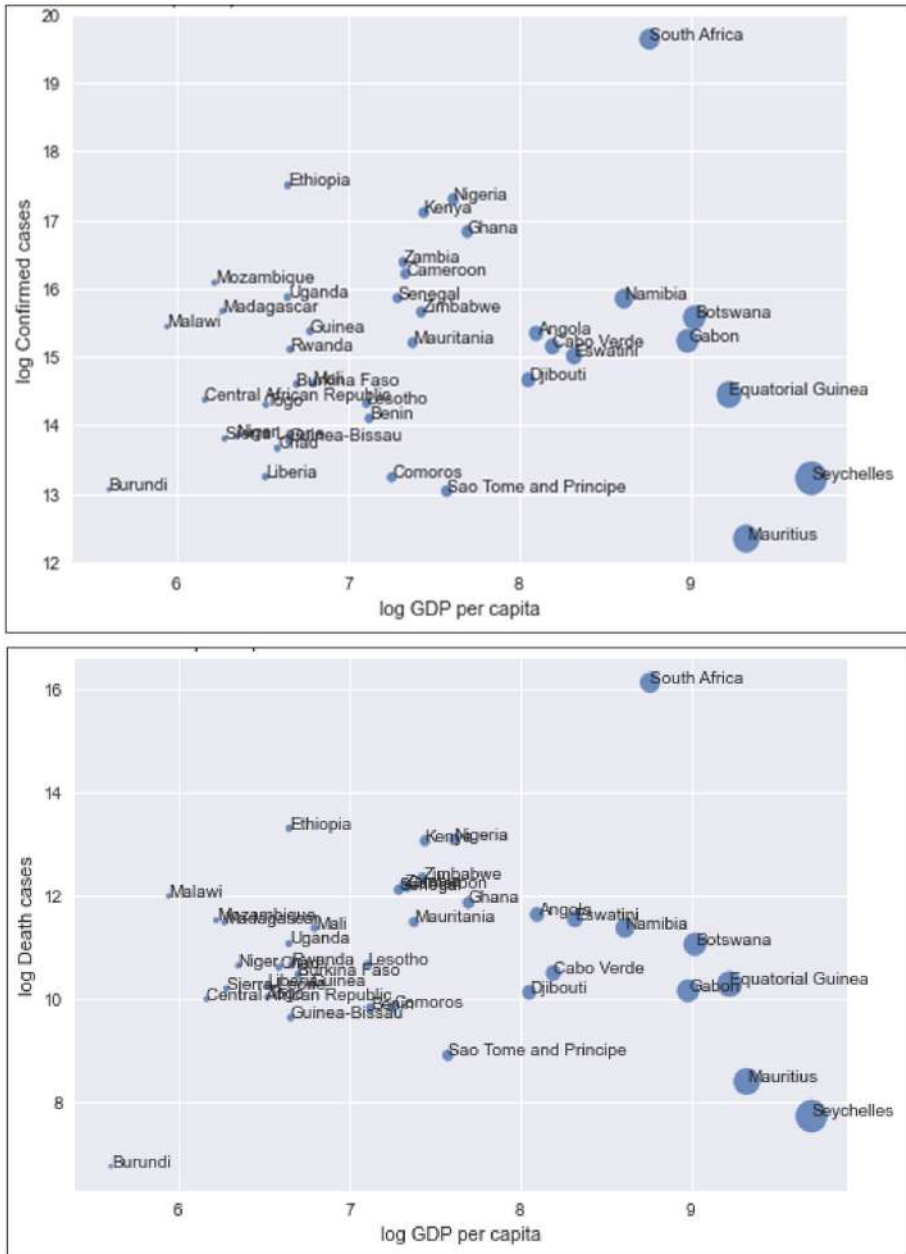
Berkman et al. (2014) and Swift (2011) documented that a two-way relationship between a country's performance and good health might exist due to the fact that health may actually drive economic performance. The scatter plot in Fig. 5 shows that SSA countries with higher GDP per capita have higher confirmed and death cases. In contrast to European countries, Pardhan and Drydakis (2021) reported a statistically significant negative association between GDP per capita and new cases of COVID-19 during the first wave (1 April–31 May 2020) in 38 European countries. Moreover, Fig. 6 shows that a few SSA countries with higher health expenditure per capita (South Africa) reported higher numbers of COVID-19 confirmed and death cases. Mauritius and Seychelles have the highest health expenditure per capita and reported amongst the



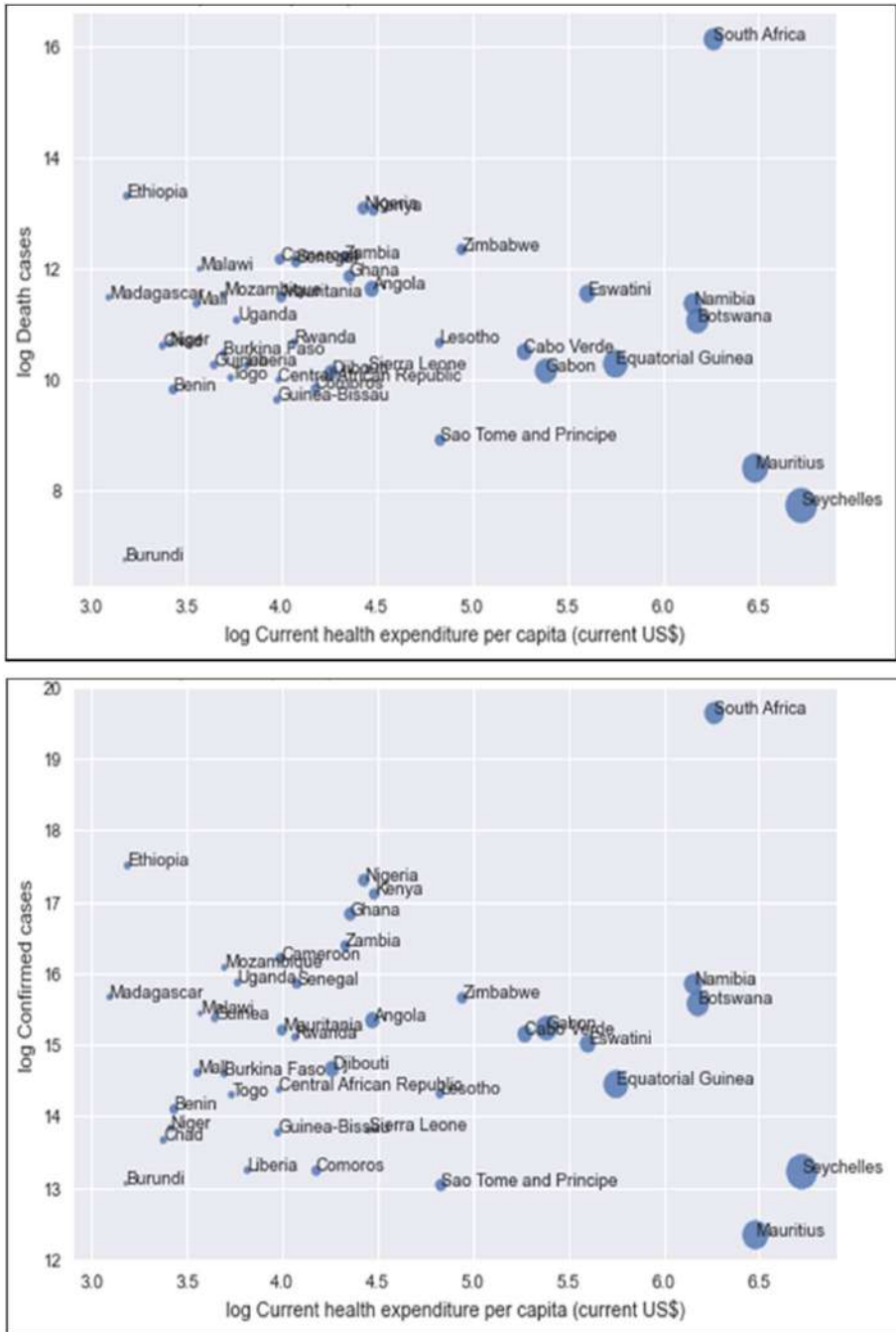
**Fig. 4** Scatter plot of life expectancy versus COVID-19 cases (confirmed and death cases) in sub-Saharan African countries. *Source:* Authors' computation and illustration based on COVID-19 data obtained from Johns Hopkins and socio-economic data obtained from the World Bank

lowest confirmed and death cases. Eurostat (2021) explained that healthcare expenditure quantifies the economic resources dedicated to health functions, excluding capital investment. Against this backfall, we argue that it would be expected that sub-Saharan African countries with higher health expenditure per capita are infrastructure ready to respond and stem the spread of COVID-19.





**Fig. 5** Scatter plot of GDP per capita versus COVID-19 (confirmed and death cases) in sub-Saharan African countries. *Source:* Authors’ computation and illustration based on COVID-19 data obtained from Johns Hopkins and socio-economic data obtained from the World Bank



**Fig. 6** Current health expenditure versus COVID-19 (confirmed and death cases) in sub-Saharan African countries. *Source:* Authors’ computation and illustration based on COVID-19 data obtained from Johns Hopkins and socio-economic data obtained from the World Bank

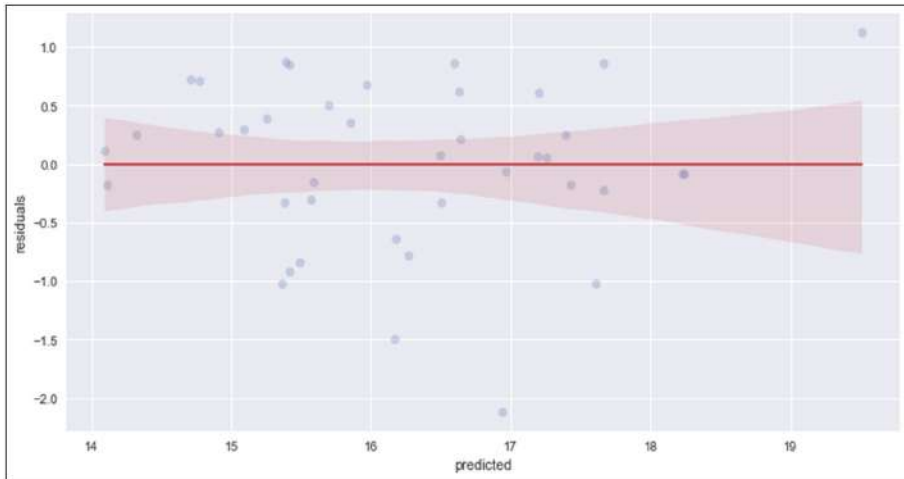


Fig. 7 Scatter plot of residuals and predicted values

## Econometric specification

Seber and Lee (2012) argued that linear regression analysis can be critical in understanding the specific dynamics of a certain phenomenon. In this study, we employ multiple linear regression on COVID-19 death and confirmed cases as the dependent variables and socio-economic indicators as independent variables. Our baseline econometric specification/model is specified in Eq. 1 for COVID-19 death cases and Eq. 2 for COVID-19 confirmed cases.

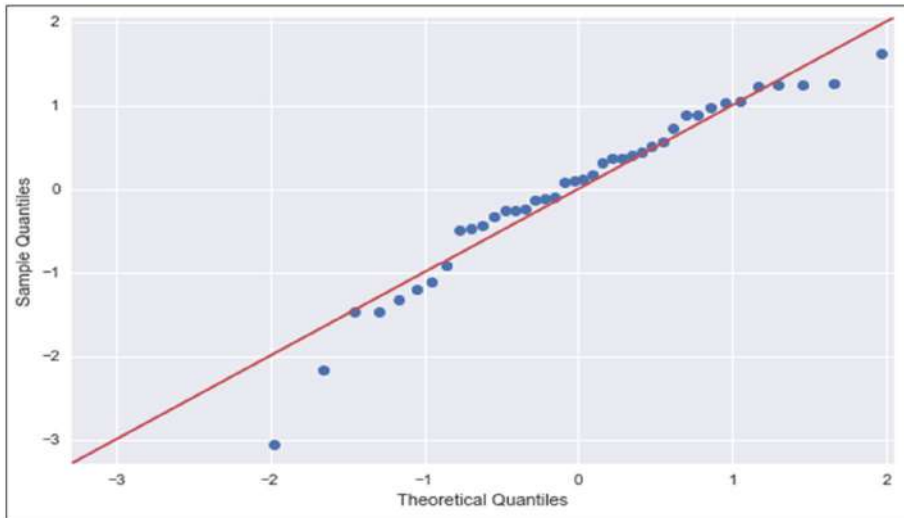
$$\text{COVID - 19}_{(\text{death cases})} = \beta_0 + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (1)$$

$$\text{COVID - 19}_{(\text{confirmed cases})} = \beta_0 + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (2)$$

where  $\beta_0$  is the constant and  $\beta_i$  refers to the regression coefficients.  $X_i$  refers to the multivariate socio-economic indicators and  $\varepsilon$  is these error term, and we make the standard assumption that the variance of the error term is similar across the values of the independent variable (Homoscedasticity).

*The null hypothesis* states that there is no impact of socio-economic variables on the number of COVID-19 cases (confirmed and deaths).

The alternative hypothesis There is an impact of socio-economic variables on the number of COVID-19 cases (confirmed and deaths).



**Fig. 8** Q–Q plot to test for the distribution of the data

## Heteroskedasticity test

There are several ways to test for heteroskedasticity. Amongst others, these include visualising a plot of residuals and predicted values and using a Breusch–Pagan test. Figure 7 shows a scatter plot of residuals and predicted values. We use the figure to detect non-linearity, error variances and outliers in the model. The illustration in Fig. 7 graph indicates that the data is homoscedastic, meaning the error terms have constant variance.

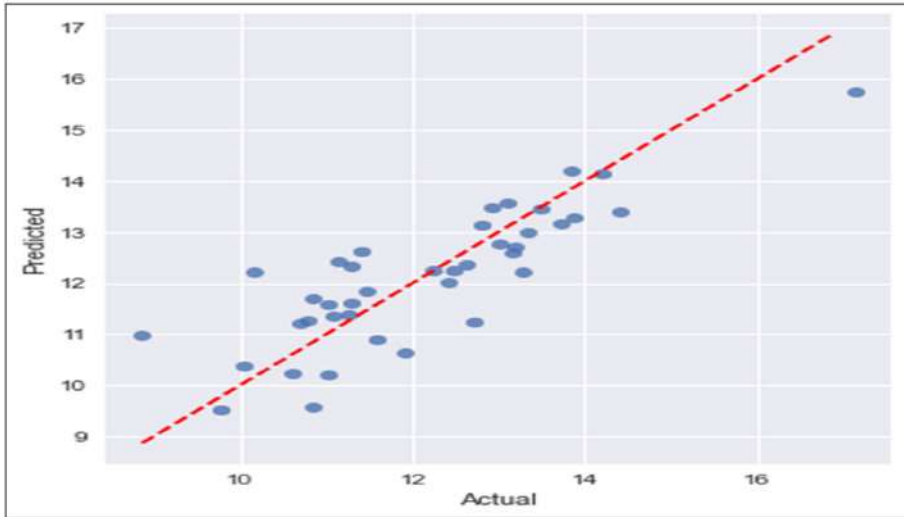
In case the data was heteroskedastic, the illustration would have shown a funnel shape which would have meant that the error terms have non-constant variance.

## Normal distribution

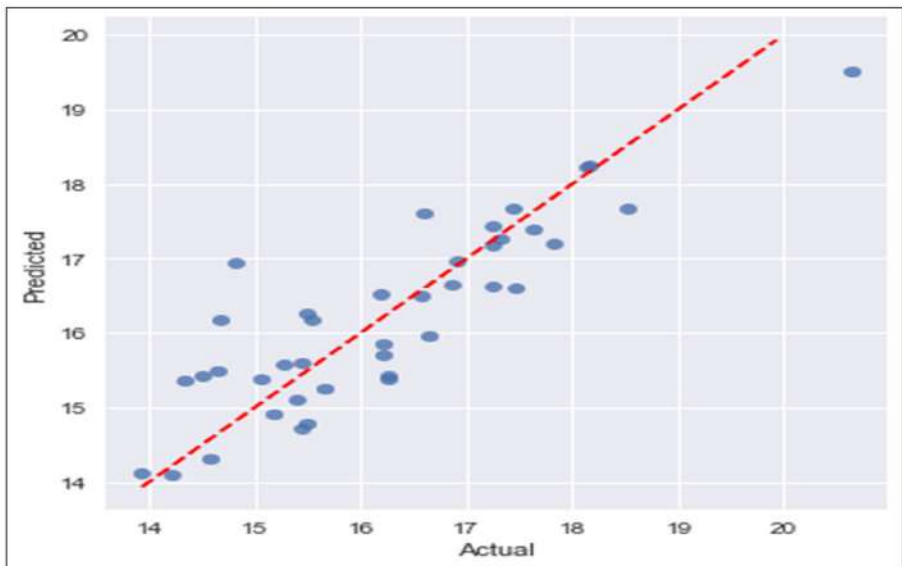
To test for the distribution of the data points we use a quantile–quantile (Q–Q) plot. The assessment of the Q–Q plot in Fig. 8 shows that the data conforms to a normal distribution. The caveat we make here is that the Q–Q plots can also be ambiguous.

## Predicted versus actual values

To test how well the model approximates the actual data, we compare the predicted values of the model against the actual response variable (in a scatter plot). Figure 9 illustrates the predicted versus the actual values of COVID-19 death cases. It shows most of the data points clustered around the red dotted 45-degree line.



**Fig. 9** COVID-19 death cases (predicted versus actuals)



**Fig. 10** COVID-19 confirmed cases (predicted versus actual)

In the same manner, Fig. 10 gives insights into the model performance by showing the scatter plot of the predicted versus the actual values of the confirmed cases. Most of the data points are clustered around the red dotted 45-degree line.

**Table 3** OLS regressions of COVID-19 confirmed cases and socio-economic indicators in sub-Saharan African countries

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Const	15.92*** (0.74)	8.10*** (1.50)	1.83 (2.15)	-4.95 (8.46)	-6.00 (8.57)	-6.03 (8.69)	-18.61* (9.70)
Log population density	-0.21 (0.17)	-0.21 (0.13)	-0.16 (0.11)	-0.20 (0.12)	-0.20 (0.12)	-0.22 (0.13)	-0.20 (0.12)
Log population aged 65 and above, total		0.63*** (0.11)	0.78*** (0.11)	0.79*** (0.11)	0.82*** (0.11)	0.81*** (0.11)	-0.77 (0.67)
Log gross national income per capita			0.58*** (0.16)	0.52*** (0.18)	0.23 (0.37)	0.69 (1.71)	2.16 (1.71)
Log life expectancy at birth, total (years)				1.75 (2.11)	2.09 (2.15)	2.14 (2.19)	3.56 (2.14)
Log health expenditure per capita					0.33 (0.38)	0.33 (0.38)	0.63 (0.38)
Log GDP per capita (current US\$)						-0.47 (1.69)	-1.91 (1.69)
Log population, total							1.57** (0.66)
R-squared	0.04	0.48	0.62	0.63	0.64	0.64	0.69
R-squared Adj	0.01	0.45	0.59	0.59	0.58	0.57	0.63
R-squared	0.04	0.48	0.62	0.63	0.64	0.64	0.69
No. observations	40	40	40	40	40	40	40

Standard errors in parentheses

\* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$

## Results and discussion

This section provides a combined results and discussion overview for the econometric specification used in this study. In this study, caution must also be taken to interpret the OLS estimation results as associations between COVID-19 and socio-economic indicators and not as causal effects.

Running the model produced results of p-values indicating statistical significance.<sup>2</sup> Based on the regression of COVID-19 confirmed cases and socio-economic indicators; population aged 65 and above reported  $p < 0.01$ , gross national income per capita reported  $p < 0.01$  and total population  $p < 0.05$  all reported statistically significant values indicating strong evidence against the null hypothesis. Therefore, we reject the null hypothesis and accept the alternative hypothesis, i.e. population aged 65 and above, gross national

<sup>2</sup> When performing a statistical test, a p-value helps determine the significance of the results in relation to the null hypothesis. In principle, a p-value, or probability value, is a number describing how likely it is that the given data would have occurred by random chance (i.e., the null hypothesis is true). A p-value less than 0.05 (typically  $\leq 0.05$ ) is statistically significant. It indicates strong evidence against the null hypothesis, as there is less than a 5% probability that the null is correct (and the results are random). Therefore, we reject the null hypothesis and accept the alternative hypothesis).

**Table 4** OLS regression of COVID-19 death cases and socio-economic indicators in sub-Saharan African countries

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Const	12.19*** (0.81)	3.54** (1.60)	-1.76 (2.45)	-2.53 (9.72)	-4.12 (9.77)	-4.06 (9.87)	-11.53 (11.71)
Log population density	-0.31 (0.19)	-0.31** (0.14)	-0.27** (0.13)	-0.27* (0.14)	-0.28** (0.14)	-0.25 (0.15)	-0.24 (0.15)
Log population aged 65 and above, total		0.70*** (0.12)	0.83*** (0.12)	0.83*** (0.12)	0.86*** (0.13)	0.87*** (0.13)	-0.07 (0.81)
Log gross national income per capita			0.49** (0.18)	0.48** (0.20)	0.05 (0.43)	-0.95 (1.94)	-0.08 (2.07)
Log life expectancy at birth, total (years)				0.20 (2.42)	0.72 (2.45)	0.62 (2.49)	1.47 (2.58)
Log health expenditure per capita					0.50 (0.43)	0.49 (0.43)	0.67 (0.46)
Log GDP per capita (current US\$)						1.01 (1.92)	0.15 (2.04)
Log population, total							0.93 (0.80)
R-squared	0.07	0.51	0.60	0.60	0.61	0.62	0.63
R-squared Adj	0.04	0.49	0.56	0.55	0.56	0.55	0.55
R-squared	0.07	0.51	0.60	0.60	0.61	0.62	0.63
No. observations	40	40	40	40	40	40	40

Standard errors in parentheses

\* $p < .1$ ; \*\* $p < .05$ ; \*\*\* $p < .01$ 

income per capita and total population have a statistically significant impact on the confirmed cases, as shown in Table 3.

Taking into account the regression of COVID-19 death cases and socio-economic indicators; population density reported  $p < 0.05$ , population aged 65 and above reported  $p < 0.01$ , gross national income per capita reported  $p < 0.05$ . Therefore, we reject the null hypothesis and accept the alternative hypothesis, i.e. population aged 65 and above, gross national income per capita and population density have a statistically significant impact on the confirmed cases, as shown in Table 4.

Model 7 (contains all socio-economic variables) shown in Table 3, explains 69% of the variability of COVID-19 confirmed cases whereas the same model explains 63% of the variability of the COVID-19 death cases (see Table 4). The  $R$  squared term increases with the addition of features (socio-economic indicators) from Model 1, which only had population density, to Model 7, which had all the seven socio-economic indicators. The results show an inverse relationship between COVID-19 cases and log GDP per capita (current US\$) and log population density although it is statistically not significant.

Table 4 shows that the variable population aged 65 and above has a statistically significant relationship with COVID 19 death cases reporting  $p < 0.01$ . Gross national income per capita reported  $p < 0.05$  and population density reported ( $p < 0.1$ ). While the population density has indicated a negative impact on the COVID-19 death cases, the population aged 65 and above and the gross national income per capita both have indicated a positive

relationship. The gross national income per capita result seems to be counter-intuitive. A positive relationship is characteristic of a developed country which implies a better testing for COVID-19, better reporting of data and a lesser number of missed deaths attributable to other causes when in fact it could have been caused by COVID-19.

A very surprising result is the negative relationship between population density and COVID-19 deaths, as shown in Table 4. This result is at odds with Feng et al. (2020), who asserted that population density increases the risk of transmission of COVID-19. We argue that sub-Saharan African countries were early in implementing a mixture of COVID-19 measures. The early lockdowns, reduction in the number of social gatherings, school closures and social distancing measures may have diluted the impact of the population density. Another startling result was the significant positive association between gross national income per capita and COVID-19 death cases ( $p < 0.05$ ) in models 3 and 4. Moreover, the non-significant positive relationship between the COVID-19 deaths and health expenditure per capita was surprising. In support of our result, a study by Khan et al. (2020) argued that healthcare capacity or functionality from an infectious disease perspective might be sensitive to decisions about allocation well in advance of the pandemic. The health expenditure per capita in developing sub-Saharan Africa cannot be compared to developed countries since they have greater expertise and relatively greater resources at a grassroots level to respond to a pandemic.

## Conclusion

The socio-economic policy of a specific country constitutes an interesting source of information and gives relevant insights into COVID-19 confirmed and death cases. The disparities in the reported COVID-19 cases during the lockdown indicate that other factors determine the trajectory of the spread of COVID-19. Some epidemiology studies from the US, UK, India and Taiwan, report that there is an impact of socio-economic indicators on COVID-19. Therefore the premise of this study was to ascertain the nature of the relationship between COVID-19 cases and socio-economic indicators (GDP per capita, gross national income per capita, life expectancy, population density, the population aged 65 and above, current health expenditure per capita and population).

Our findings indicate that a statistically significant relationship was noted between COVID-19 cases and only 4 of the socio-economic indicators (population density, gross national income per capita, the population aged 65 and above, as well as total population). We expected a positive correlation between population density and COVID-19 deaths. However, our results contradicted our expectations as we found a statistically significant but negative relationship between population density and COVID-19 deaths. Another result that was at odds with expectations was the positive correlation between gross national income per capita and COVID-19 death cases. On the one hand, it might be expected that higher gross national income per capita indicates the country's readiness for containing pandemics. On the other hand, it can be argued that countries with higher gross national income have the most economic activity and connections to international airports and ports of entry. The high movement in and out of these countries might have contributed to the spread of COVID-19 hence a higher number of infections and consequently higher mortalities. Another unexpected result was the positive relationship between the COVID-19 deaths and health expenditure per capita. Intuitively, the relationship between health expenditure per capita and deaths would be negative. High health expenditures per



capita are a proxy for a more ‘pandemic ready health system’ as investments are channelled towards obtaining sufficient hospital beds, medical professionals, emergency care units and medical laboratories.

Our findings back up the argument that socio-economic indicators can be used to elucidate scholars and scientists on non-medical factors contributing to the increasing infections in sub-Saharan Africa. We caution that COVID-19 case data may change as sub-Saharan African countries improve their testing capacities and improve on the underreported cases. The current rise that is prevalent in sub-Saharan African countries may also be attributed to countries’ ability to increase their tests. This means more COVID-19 deaths in SSA countries will likely be reported, which may change current multilinear regression results. For sub-Saharan African countries, there is hardly any analysis of gender-differentiated socio-economic indicators on COVID-19. Future studies should include gendered sociodemographic variables at the country level in sub-Saharan Africa. This will enable the exploration and recommendations of public health policies that can narrow the gender inequalities and enhance equitable access to targeted measures of infection protection.

**Data availability** The datasets generated during and/or analysed during the current study are available at the following weblinks: [https://raw.githubusercontent.com/Lucas-Czarnecki/COVID-19-CLEANED-JHUCS-SE/master/COVID-19\\_CLEAN/esse\\_covid\\_19\\_time\\_series\\_cleaned/time\\_series\\_covid19\\_cases\\_tidy.csv](https://raw.githubusercontent.com/Lucas-Czarnecki/COVID-19-CLEANED-JHUCS-SE/master/COVID-19_CLEAN/esse_covid_19_time_series_cleaned/time_series_covid19_cases_tidy.csv). This data source provides daily reported COVID-19 cases for all sub-Saharan Africa. <https://data.worldbank.org/>. This data source provides a range of aggregate measures of the socio-economic indicators used in this study.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

## References

- Aburto JM, Kashyap R, Schöley J, Angus C, Ermisch J, Mills MC, Dowd JB (2021) Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. *J Epidemiol Community Health* 75(8):735–740. <https://doi.org/10.1136/jech-2020-215505>
- Alsafi Z, Nicola M, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, Agha M, Agha R (2020) The socio-economic implications of the coronavirus pandemic (COVID-19): a review. *Int J Surg (lond, Engl)* 78:185–193. <https://doi.org/10.1016/j.ijssu.2020.04.018>
- Applegate WB, Ouslander JG (2020) COVID-19 presents high risk to older persons. *J Am Geriatr Soc* 68(4):681. <https://doi.org/10.1111/jgs.16426>
- Berkman FL, Kawachi I, Glymour MM (2014) *Social epidemiology*. Oxford University Press, Oxford
- Bhadra A, Mukherjee A, Sarkar K (2021) Impact of population density on Covid-19 infected and mortality rate in India. *Model Earth Syst Environ* 7:623–629. <https://doi.org/10.1007/s40808-020-00984-7>
- Eurostat. (2021). Healthcare expenditure across the EU: 10% of GDP. Healthcare expenditure across the EU: 10% of GDP - Products Eurostat News - Eurostat (Europa.eu). Accessed 15 Dec 2021.
- Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ (2020) Rational use of face masks in the COVID-19 pandemic. *Lancet Respir Med* 8:434–436. [https://doi.org/10.1016/S2213-2600\(20\)30134-X](https://doi.org/10.1016/S2213-2600(20)30134-X)
- Hartley DM, Perencevich EN (2020) Public health interventions for COVID-19: emerging evidence and implications for an evolving public health crisis. *JAMA* 323(19):1908–1909. <https://doi.org/10.1001/jama.2020.5910>

- Khan JR, Awan N, Islam MM, Muurlink O (2020) Healthcare capacity, health expenditure, and civil society as predictors of COVID-19 case fatalities: a global analysis. *Front Public Health* 8:347. <https://doi.org/10.3389/fpubh.2020.00347>
- Mihoub A, Snoun H, Krichen M, Salah RB, Kahia M (2020) Predicting COVID-19 spread level using socio-economic indicators and machine learning techniques. In: 2020 first international conference of smart systems and emerging technologies (SMARTTECH), pp 128–133
- Pardhan S, Drydakis N (2021) Associating the change in new COVID-19 cases to GDP per capita in 38 European countries in the first wave of the pandemic. *Front Public Health* 8:582140. <https://doi.org/10.3389/fpubh.2020.582140>
- Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS et al (2020) Ethnicity and COVID-19: an urgent public health research priority. *Lancet* 395:1421–1422. [https://doi.org/10.1016/S0140-6736\(20\)30922-3](https://doi.org/10.1016/S0140-6736(20)30922-3)
- Seber GA, Lee AJ (2012) *Linear regression analysis*, vol 329. Wiley, Hoboken
- Swift R (2011) The relationship between health and GDP in OECD countries in the very long run. *Health Econ* 20:306–322. <https://doi.org/10.1002/hec.1590>
- Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker P, Fu H, Dighe A, Griffin JT, Baguelin M, Bhatia S, Boonyasiri A, Cori A, Cucunubá Z, Fitz-John R, Gaythorpe K, Green W, Ferguson NM (2020) Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* 20(6):669–677. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7)
- Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71(3):599–607

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.