Factors that most expose countries to COVID-19: a composite indicators-based approach

Matheus Pereira Libório · Petr Yakovlevitch Ekel · João Francisco de Abreu · Sandro Laudaes

Accepted: 17 November 2021 / Published online: 2 December 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Studies carried out in different countries correlate social, economic, environmental, and health factors with the number of cases and deaths from COVID-19. However, such studies do not reveal which factors make one country more exposed to COVID-19 than other. Based on the composite indicators approach, this research identifies the factors that most impact the number of cases and deaths of COVID-19 worldwide and measures countries' exposure to COVID-19. Three composite indicators of exposure to COVID-19 were constructed through Principal Component Analysis, Simple Additive Weighting, and k-means clustering. The number of cases and deaths from COVID-19 is strongly correlated ($R > 0.60$) with composite indicator scores and moderately concordant ($K > 0.4$) with country clusters. Factors directly or indirectly associated with the age of the population are the ones that most expose countries to COVID-19. The population of countries

most exposed to COVID-19 is 12 years older on average. The proportion of the elderly population in these countries is at least twice that of countries less exposed to COVID-19. Factors that can increase the population's life expectancy, such as Gross Domestic Product per capita and the Human Development Index, are four times and 1.3 times higher in more exposed countries to COVID-19. Providing better living conditions increases both the population's life expectancy and the country's exposure to COVID-19.

Keywords COVID-19 · Composite indicator · Generalized reduced gradient · Principal component analysis · K-means clustering

Introduction

Several studies analyze the geography of the COVID-19 pandemic caused by the SARS-CoV-2 virus at municipal, national, continental, and global scales (Maiti et al., 2021; Tang et al., 2021). On a global scale, studies have found a significant correlation between healthy life expectancy and the number of cases and deaths from COVID-19 (Azarpazhooh et al., 2020). The confirmed cases of COVID-19 have a strong positive effect on the number of deaths globally and regionally (Appiah-Otoo & Kursah, 2021).

M. P. Libório (✉) · P. Y. Ekel · J. F. de Abreu · S. Laudaes
Pontifical Catholic University of Minas Gerais,
Belo Horizonte 30535-012, Brazil
e-mail: m4th32s@gmail.com

P. Y. Ekel
e-mail: petr.ekel2709@gmail.com

J. F. de Abreu
e-mail: jofabreu@hotmail.com

S. Laudaes
e-mail: sandrolaudaes@gmail.com

Continental-scale studies showed that the spread and impact of the first wave of COVID-19 were lower in European countries that adopted anti-pandemic measures quickly and consistently than in countries that responded later or with milder restrictive measures (Dzúrová & Květoň, 2021). High population densities catalyze the spread of COVID-19 in Africa (Olusola et al., 2020). The spread of COVID-19 showed significant spatial and temporal differences in Africa countries, being influenced by the annual average of air transport passengers and the population density of each country (Onafeso et al., 2021).

A greater number of studies are related to a national scale. Forati and Ghose (2021) found a strong spatial relationship between activity on social networks and the spread of Covid-19 in the United States. Benita and Gasca-Sanchez (2021) showed positive correlations between income inequality, the prevalence of obesity, diabetes, and the concentration of fine particles with the cases and deaths by COVID-19 in Mexico. Paez et al. (2020) showed that the incidence of COVID-19 in Spain is negatively related to temperature and humidity and positively related to Gross Domestic Product per capita and the presence of mass transport systems. In Germany, Mitze and Kosfeld (2021) demonstrated how commuting patterns to work are decisive in defining the spatial dynamics of daily cases of COVID-19. Gupta et al. (2021) showed that the incidence of COVID-19 is higher in Indian districts with a higher level of urbanization. In South America, a positive correlation was found between population density and cases of COVID-19 and a negative correlation between Gross Domestic Product per capita and deaths from COVID-19 (Oyedotun & Moonsammy, 2021). The neighborhood is a spatial element that directly affects the spread of COVID-19 cases in Asian countries (Shabani & Shahnazi, 2020).

The number of studies at the municipal scale is also significant. Nasiri et al. (2021) analyzed the number of cases and deaths from COVID-19 in Tehran, Iran. The authors revealed that the number of cases and deaths from COVID-19 is higher among men, but the number of deaths per thousand is higher among women. Kalla et al. (2021) showed that the number of cases of COVID-19 occurs in areas of greater social vulnerability in Batna, Algeria. Tribby and Hartmann (2021) revealed a positive correlation between COVID-19 cases and the number of people per household, African-American and Hispanic populations, and

people over 65 in New York City, United States. Also, in New York, Yang et al. (2021) proved a positive association between violations of the use of a face mask reported by the police and mortality from COVID-19.

Many of these studies are aimed at establishing relationships between the number of cases and deaths from COVID-19 and social, demographic, and economic factors (Dowd et al., 2020; Sannigrahi et al., 2020; Sarmadi et al., 2020; Zhou et al., 2020). Generally, these studies analyze the impact of each factor on the number of cases and deaths from COVID-19 separately. This separate analysis allows recognizing statistically consistent relationships, but it does not provide a reliable basis for comparisons. From this gap, it is possible to state two important questions. The first one is: which factors most impact the number of cases and deaths from COVID-19: social, demographic, or economic? The second question is the following: which countries are most exposed to COVID-19, considering the impacts of each factor?

Driven by these questions, this research has a twofold objective. First, to identify the factors that most impact the number of cases and deaths of COVID-19 worldwide. Second, to provide a measure of countries' exposure to COVID-19. The focus of this study is the relationship between the Indicator COVID-19 Cases per million (I-Cases) and the Indicator COVID-19 Deaths per million (I-Deaths) with social, demographic, and economic factors. The analysis of these relationships from the composite indicators framework brings two contributions. First, the weights of the sub-indicators reveal the factors that most expose countries to I-Cases and I-Deaths by COVID-19. Secondly, it offers a one-dimensional measure of countries' exposure to COVID-19. It is beyond the scope of this research to analyze the temporal dynamics of the propagation of COVID-19. It is also not the scope of this research to analyze the effects of vaccination on the temporal dynamics of COVID-19.

Multidimensional analysis of COVID-19

The number of cases and deaths from COVID-19 is associated with social, demographic, and health factors (Drefahl et al., 2020; Karmakar et al., 2021;

Markovič et al., 2021). Although this multidimensionality is widely recognized, not many studies propose to build composite indicators for the analysis of COVID-19. A search performed in Google Scholar and Web of Science on August 28, 2021, with the terms “COVID-19” AND “Composite Ind*” resulted in five and thirty-seven publications, respectively. Table 1 shows that sixty percent of these publications are associated with the stock market and the socio-economic impacts of COVID-19.

Studies on the Exposure measurement topic can be subdivided into three groups. The first one includes studies that analyze exposure to COVID-19 over time in a geographic area (Pang et al., 2021). The second group is associated with studies that analyze exposure factors to COVID-19 in different geographic areas (Martines et al., 2021; Weinstein et al., 2021). Finally, the third group covers studies that map the areas of exposure to COVID-19 from demographic and socio-economic composite indicators (Acharya et al., 2020; Dlamini et al., 2020; Nicodemo et al., 2020; Joshua

et al., 2021; Karácsonyi et al., 2021; Sarkar & Chouhan, 2021; Yigitcanlar et al., 2021).

This research is similar to the studies in group two. These studies show that the cases of COVID-19 are associated with factors such as inequality, socioeconomic vulnerability (Martines et al., 2021), morbidity, health risk behaviors, access to health, population mobility, social distance, and education (Weinstein et al., 2021).

Operational framework

The development of this research is divided into data collection and processing, construction of composite indicators, internal and external validation of the composite indicators, and bias analysis.

Data collection and processing

Data on COVID-19 cases, deaths, hospitalizations, tests until May 5, 2021, of 193 countries were

Table 1 The application of composite indicators in COVID-19 analysis

Topic	Papers	Country	Reference
Exposure measurement	10	Australia, Brazil, China, England, Eswatini, India, and South Korea	Acharya et al. (2020), Dlamini et al. (2020), Nicodemo et al. (2020), Joshua et al. (2021), Karácsonyi et al. (2021), Martines et al. (2021), Pang et al. (2021), Sarkar and Chouhan (2021), Weinstein et al. (2021), and Yigitcanlar et al. (2021)
Health care structure	4	Brazil, European Union Countries, India, and United States	Ferraz et al. (2021), Luca et al. (2021), Pandey et al. (2021) and Tripathi et al. (2021)
Socioeconomic impacts	10	China, European Countries, Ghana, India, Romanian, Russia, Ukraine, and United States	Benzell et al. (2020), Funke et al. (2020), Ignat and Constantin (2020), Bukari et al. (2021), Erić et al. (2021), Hryhoruk et al. (2021), Kitrar (2021), Kujur and Goswami (2021), Liu et al. (2021), and Suh and Alhaery (2021)
Stock markets	12	China, Germany, Indonesia, Italy, Japan, Malaysia, Pakistan, and United States	Khan et al. (2020), Lee et al. (2020), Shehzad et al., (2020a, 2020b, 2021), Cristofaro et al. (2021), Goh et al. (2021), Li et al. (2021), Liu (2021), Salisu et al. (2021), Shehzad et al. (2021), Vukovic et al. (2021) and, Wen et al. (2021)
Transmission control	3	World, China, and Mexico	Li et al. (2020), Kinnunen et al. (2021) and Knaul et al. (2021)
Others	3		Wyper et al. (2020), Boyd and Wilson (2021), and Kaiser et al. (2021)

In the topic Others, Boyd and Wilson (2021) employed composite indicators to identify countries of refuge to safeguard humanity's survival from the threat of the COVID-19 pandemic. Kaiser et al. (2021) discussed the pitfalls of employing composite indicators such as the Global Health Security Index in policy formation. Wyper et al. (2020) suggested that rapid patient assessments can be enhanced from a composite indicator of vulnerability to severe health consequences from COVID-19

extracted from the Our World in Data. The corresponding data is available at the link <https://github.com/owid/covid-19-data/blob/master/public/data/owid-covid-codebook.csv>.

Data processing was carried out in five stages. In stage one, data not related to countries were excluded, such as data grouped by continent. In stage two, data from countries that met two criteria were maintained. Criterion one, the countries with less than 5% of outliers in the sub-indicators were maintained. The outliers were calculated by applying the 3σ z-score rule. Criterion two, countries with data for 50% of sub-indicators or more were maintained. Data availability was calculated by the ratio between the available sub-indicators and the total number of sub-indicators. In stage three, sub-indicators that met three criteria were maintained. Criterion one, sub-indicators on comparable scales. Criterion two, significant correlation with I-Cases. Criterion three, data availability greater than 0.80. In stage four, the forty-three missing data were imputed by the Rule of Three. The base value used in the Rule of Three was obtained from the correlation matrix between the sub-indicators. In stage five, the polarity of the sub-indicators was standardized so that the sign of the correlations between each sub-indicator and the I-Cases assumes the same sign (Mazziotta & Pareto, 2017). The countries and sub-indicators employed in the analysis are presented in the Annex.

Construction of the composite indicators

Composite indicators are one-dimensional measures of complex phenomena that typically involve many sub-indicators (Otoiu et al., 2021). This property is advantageous in geography because it allows representing many sub-indicators in a single map (Libório et al., 2020). The advantages of representing complex phenomena simply through composite indicators have conquered researchers from the most different areas of knowledge (Greco et al., 2019). The specialized literature has evolved in recent years and offers numerous methods for constructing composite indicators (El Gibari et al., 2019).

Regardless of the method, the construction of composite indicators usually involves the normalization of scale, weighting, and aggregation of sub-indicators (Kuc-Czarneck et al., 2020). The most commonly used scale normalization of sub-indicators is normalization by minimum–maximum values and

standardization by the mean and standard deviation (Cinelli et al., 2021). The weighting of the sub-indicators can be carried out by data-driven, expert opinion or equal weights (Becker et al., 2017). There are two aggregation schemes, with compensatory aggregation a much more popular approach than non-compensatory aggregation (Dialga et al., 2017).

Three strategies based on the operational framework of composite indicators were developed (El Gibari et al., 2019; Nardo et al., 2005). The first strategy employs Abadie and Carpentier's (1969) Generalized Reduced Gradient (GRG) method. The second strategy employs Pearson's (1901) Principal Component Analysis (PCA). The third strategy combines k-means clustering (Rousseeuw, 1987).

The process of constructing composite indicators applying GRG and PCA involves the normalization of scale, weighting, and aggregation of sub-indicators. The composite indicator constructed by k-means follows another procedure where geographical units are grouped into categories considering the Euclidean distance of the n-sub-indicators (Libório et al., 2021; Libório et al., 2021).

Generalized Reduced Gradient (GRG) strategy: aims at maximizing the correlation between I-Cases with the composite indicator of exposure to COVID-19, so-called CI-Exposure-GRG. GRG is an optimization algorithm that allows finding quasi-optimal solutions for calculations with powers (Lasdon et al., 1974) as in the correlation calculation. The CI-Exposure-GRG can be defined as a composite indicator built from Simple Additive Weighting. In other words, the compensatory aggregation of normalized sub-indicators with data-driven weights. In particular, the weights are obtained to maximize the correlation between the I-Cases and the CI-Exposure-GRG. The GRG strategy was developed in Excel (Powell & Batt, 2008) in three stages. First, a base-composite indicator was constructed from the aggregation of nine sub-indicators normalized with Equal Weights. Second, the correlation between the base-composite indicator and the I-Case was calculated. Third, the correlation between the I-Cases and the base-composite indicator was maximized from the GRG algorithm. In short, the algorithm calculates the correlation between the I-Cases and the CI-Exposure-GRG for different combinations of weights. The combination that maximizes the correlation is used as a solution. Two constraints were considered. Constraints one: all sub-indicators

impact COVID-19 with a weight greater than 0.10. Constraints two: the sum of the square of the weights is equal to one so that the relative importance of the sub-indicators in the composite indicator is guaranteed (Becker et al., 2017).

Principal Component Analysis (PCA) strategy: aims at creating the CI-Exposure-PCA to synthesize the sub-indicators into a new index that captures most of the variance of the original data. The coordinates of the eigenvectors in the principal component represent the weights of the sub-indicators in CI-Exposure-PCA (Jolliffe, 2005). These weights represent the impact of each factor on Covid-19's I-Cases and I-Deaths worldwide. The CI-Exposure-PCA can be defined as a composite indicator constructed from the compensatory aggregation of nine standardized sub-indicators with data-driven weights (Nardo et al., 2005). The PCA strategy was developed in three stages. In stage one, the CI-Exposure-PCA is constructed using the Software R. In stage two, the Variance Extracted (VE) in the Principal Component and the Kaiser–Meyer–Olkin (KMO, Kaiser, 1974) test are checked. The CI-Exposure-PCA is considered statistically consistent when the VE and the KMO exceed 0.50 and 0.60, respectively (Libório et al., 2020).

K-means strategy: aims at constructing a categorical composite indicator based on the performance of each country's sub-indicators so-called CI-Exposure-K-means. The Euclidean distance between the performance of the sub-indicators in each country is used in the definition of clusters (Krishna et al., 2018). The K-means strategy was developed in four stages. In stage one, the most correlated sub-indicators with the most significant weights in the CI-Exposure-PCA were selected. The exclusion of poorly correlated sub-indicators is a procedure used to improve the quality of clusters (Libório et al., 2021; Libório et al., 2021). In stage two, the Average silhouette, Elbow, and Gap Statistic methods were employed to define the number of clusters (Rousseeuw, 1987). In stage three, the measures of cohesion quality and resolution of the clusters generated in stage two were verified from the average width of the silhouette of each observation within the group (Bernardes et al., 2021). The cluster configuration with the best cohesion and resolution was selected as CI-Exposure-K-means. In stage four, the categories of exposure to COVID-19 were defined based on the Two Factor Analysis of Variance (ANOVA) with Replication (Quirk, 2012).

Composite indicators internal and external validation and bias analysis

Two approaches were applied to validate the composite indicators of exposure to COVID-19 internally. The first approach is based on the Two Factor ANOVA with Replication results. In the first approach, the averages of the sub-indicators, CI-Exposure-GRG, CI-Exposure-PCA, I-Cases, and I-Deaths, were analyzed following the groups of countries established by CI-Exposure-K-means. The second approach is based on the Kappa coefficient (Landis & Koch, 1977). The Kappa coefficient provides a measure of the compatibility of composite indicators results. The calculation was performed in three stages. In stage one, the number of countries per class in CI-Exposure-K-means was computed. In stage two, the CI-Exposure-GRG and CI-Exposure-PCA scores were positioned in descending order. In stage three, the scores are transformed into categories considering the number of countries in each class according to stage one. In stage four, the Kappa K coefficient was calculated. K is interpreted as follows: $K < 0$ No agreement; $0.00 < K < 0.20$ Slight agreement; $0.21 < K < 0.40$ Fair agreement; $0.41 < K < 0.60$ Moderate agreement; $0.61 < K < 0.80$ Substantial agreement (Landis & Koch, 1977).

The external validation of the CI-Exposure-GRG and CI-Exposure-PCA was performed using the correlation coefficients with the I-Cases and I-Deaths. The validation of the CI-Exposure-K-means was performed using the Kappa coefficient following the same process as the internal validity based on the scores of the I-Cases and I-Deaths. Finally, the bias analysis of the results was performed using test data per million people.

Factors that most impact the countries' exposure to COVID-19

Table 2 shows that the GRG algorithm improves the correlation between I-Cases and the base-composite indicator from 0.46 to 0.67 in the so-called CI-Exposure-GRG. This improvement was achieved by combining the weights of the sub-indicators. In particular, the weights changed from 0.33 in the base-composite indicator to values ranging between 0.10 and 0.89.

Table 2 Weights of sub-indicators in I-cases and I-deaths

Code	Sub-indicator	Base-composite indicator	CI-Exposure-GRG
STI	Stringency index	0.33	0.10
MDA	Median age of the population	0.33	0.89
65Y	Proportion of population aged 65 years	0.33	0.13
70Y	Proportion of population aged 70 years	0.33	0.37
GDP	Gross domestic product per capita	0.33	0.10
DCD	Deaths from cardiovascular diseases (per 1000 people)	0.33	0.10
HOB	Hospital beds (per 1000 people)	0.33	0.10
PLE	Population life expectancy	0.33	0.10
HDI	Human development index	0.33	0.10
	Sum of the squared weights	1.00	1.00
	Correlation with I-cases	0.46	0.67
	Correlation with I-deaths	0.54	0.67

The GRG strategy allows us to conclude that the sub-indicators Median age of the population and Proportion of the population aged 65 and 70 years impact I-Cases and I-Deaths more than the other sub-indicators. In particular, the weight of Median age of the population in I-Cases and I-Deaths is almost nine times greater than non-age-related sub-indicators. These results suggest that the factor that most impacts COVID-19 globally is the age of the population. The results indicate that social and economic factors do not have much weight in the countries' CI-Exposure-GRG.

Table 3 Results of the PCA model

Sub-indicator	Eigenvector	Weights
STI	0.06	0.00
MDA	0.41	0.17
65Y	0.39	0.15
70Y	0.39	0.15
GDP	0.33	0.11
DCD	– 0.19	0.04
HOB	0.29	0.08
PLE	0.38	0.14
HDI	0.40	0.16
VE	$t > 0.50$	0.62
KMO	$t > 0.60$	0.82

t = test threshold (Libório et al., 2020)

Fig. 1 Variance extracted bar plot

Table 3 and Fig. 1 provide information on the VE in the Principal Component and the KMO of the PCA model.

The variance extracted in the Principal Component was 0.62, surpassing the test threshold of 0.50. The KMO sample adequacy test was 0.82, surpassing the acceptance threshold of 0.60. Based on the results of these tests, it is possible to state that the PCA model is statistically consistent.

The analysis of CI-Exposure-PCA confirms that sub-indicators on the age of the population have significant weights to explain the I-Cases. Once again, Median age presents considered the most critical sub-indicator to explain I-Cases. CI-Exposure-GRG and CI-Exposure-PCA have similarities and differences. The Stringency index and Deaths from cardiovascular diseases (per 1000 people) sub-indicators have low weight in the CI-Exposure-GRG and CI-Exposure-PCA. The Population life expectancy and Human

Development Index sub-indicators have significant weights in the CI-Exposure-GRG and CI-Exposure-PCA. The maps presented in Fig. 2 offer a visual understanding of the similarities and differences between CI-Exposure-GRG and CI-Exposure-PCA.

The result of defining the number of categories of CI-Exposure-K-means can be seen in the graphs in Fig. 3. The graphs at the top of the figure indicate that the ideal number of groups can be $k = 2$, $k = 5$, or $k = 4$. The graphs at the bottom of the figure indicate that the average silhouette width of the three cluster arrangements exceeds the threshold of 0.50. In other words, the $k = 2$, $k = 5$, and $k = 4$ arrays lead to clusters that are internally homogeneous and heterogeneous with each other (Rousseeuw, 1987).

It is observed that the cohesion and resolution of the groups measured by the average silhouette width are greater when $k = 2$. Thus, the CI-Exposure-K-means groups countries from a group with 38 countries and another with 119 countries. Two Factor ANOVA with Replication was computed from this grouping.

The analysis included the 38 countries classified as Group 2 and 38 countries drawn from the universe of 119 countries classified as Group 1. The level of error was established at α equal to 0.05. The Stringency Index and Deaths from cardiovascular diseases (per 1000 people) sub-indicators were excluded from the

CI-Exposure-K-means. These sub-indicators weigh less than 0.50 in the PCA and are poorly correlated with the other sub-indicators. Countries in groups 1 and 2 tend to have similar scores on these sub-indicators. Table 4 shows the values of the means of the sub-indicators, CI-Exposure-GRG, CI-Exposure-PCA, I-Cases, and I-Deaths for groups 1 and 2 obtained through Two Factor ANOVA with Replication.

The data in Table 4 confirm the previous evidence of the negative relationship between Median age of the population, Proportion of population aged 65 years, and Proportion of population aged 70 years with CI-Exposure-GRG, CI-Exposure-PCA, I-Cases, and I-Deaths. In turn, it shows that good performances in social, demographic, and economic sub-indicators seen in Group 2 are associated with higher CI-Exposure-PCA, CI-Exposure-GRG, I-Cases, and I-Deaths. Countries that offer better living conditions increase life expectancy and become more exposed to COVID-19.

On the one hand, Population life expectancy in Group 1 countries is 9.86 years less than in Group 2 countries. On the other hand, I-Cases, I-Deaths, CI-Exposure-GRG, and CI-Exposure-PCA of the countries of the Group 1 are 0.61, 0.27, 0.41, and 0.41, respectively, lower than in Group 2 countries. The

Fig. 2 The maps of the composite indicators, I-cases and I-deaths

Fig. 3 The definition of the number of clusters

Table 4 Difference between the means and variances of sub-indicators and composite indicators of country groups

Sub-indicator code	Means group 1	Means group 2	G1/G2	Variances group 1	Variances group 2
I-Cases	22,533	57,902	0.39	567,703,189	1,124,795,628
I-Deaths	558	892	0.63	521,084	509,860
MDA	27.29	39.33	0.69	75.78	24.16
65Y	6.88	14.52	0.47	25.58	41.29
70Y	4.34	9.51	0.46	11.47	20.90
GDP	10,487	42,138	0.25	59,051,674	163,103,472
PLE	70.12	79.98	0.88	51.58	10.25
HDI	0.67	0.89	0.74	0.02	0.00
CI-exposure-GRG	0.62	1.05	0.59	0.11	0.08
CI-exposure-PCA	0.62	1.05	0.59	0.11	0.08

The sources of variation by sample, columns, and interactions presented p -value < 0.05 . Group 1 ($N = 119$, $n = 38$): Iran, Costa Rica, Equatorial Guinea, Belize, Uruguay, Ethiopia, Kyrgyzstan, Turkey, Senegal, Peru, Cape Verde, Mexico, Bosnia and Herzegovina, Gambia, Colombia, Congo, El Salvador, Grenada, Ukraine, Iraq, Libya, Mauritania, Somalia, Chile, Cote d'Ivoire, Sierra Leone, Liberia, Cuba, Paraguay, Burkina Faso, Thailand, Mozambique, Malawi, Romania, Uganda, Fiji, Pakistan, Bulgaria. Group 2 ($n = 38$): Australia, Austria, Bahamas, Bahrain, Belgium, Brunei, Canada, Cyprus, Denmark, Estonia, Finland, France, Germany, Iceland, Ireland, Israel, Italy, Kuwait, Lithuania, Malaysia, Malta, Netherlands, New Zealand, Norway, Oman, Poland, Portugal, Saudi Arabia, Seychelles, Slovakia, Slovenia, Spain, Sweden, Switzerland, Trinidad and Tobago, United Arab Emirates, United Kingdom, United States

maps in Fig. 4 show the Kappa coefficients of the CI-Exposure-K-means with the CI-Exposure-GRG and the CI-Exposure-PCA.

The results presented in Fig. 4 show a substantial agreement between CI-Exposure-K-means and CI-

Exposure-GRG and an almost perfect agreement between CI-Exposure-K-means and CI-Exposure-PCA. Despite a different calculation structure, the three composite indicators produce very similar

Fig. 4 Internal validity of the composed indicators according to the kappa coefficient

results, suggesting that the evidence presented in this research is internally consistent.

Regarding external validity, the correlation and agreement coefficients shown in Table 5 suggest that the composite indicators are consistent with the countries' I-Cases and I-Deaths and provide a good measure of the countries' exposure to COVID-19.

The strength of the correlations between CI-Exposure-GRG and CI-Exposure-PCA with I-Cases and I-Deaths are both moderate to strong. The CI-Exposure-K-means has a moderate agreement with I-Cases and I-Deaths.

Open issues

This study reveals that the variables that most expose a country to Covid-19 are directly and indirectly related to the age of the population. I-Cases and I-Deaths are positively correlated with sub-indicators that increase mean age and the proportion of the elderly population. The quality of life of countries with higher Gross Domestic Product per capita and Human Development Index is reflected in populations with higher Median age, Proportion of aged 65 and 70 years, and the Human Development Index. This result suggests that the positive correlation between Gross Domestic Product per capita and the number of COVID-19 cases (Oyedotun & Moonsammy, 2021; Paez et al., 2020) is indirect and associated with population size.

Table 5 Correlation and concordance coefficients between CI-exposure-GRG, CI-exposure-PCA, and CI-exposure-K-means with I-cases and I-deaths

Composite indicator	Coefficient	I-Cases	I-Deaths
CI-Exposure-GRG	Correlation	0.67	0.67
CI-Exposure-PCA	Correlation	0.72	0.64
CI-Exposure-K-means	Concordance	0.45	0.48

However, it is necessary to consider the number of tests carried out in the countries. Countries that carry out mass testing to control transmission may have higher percentages of I-Cases than countries that do not do this control. The population's access to tests can also increase the I-Cases. Table 4 shows that the number of I-Cases in Group 2 countries is 2.56 times greater than in Group 1. In turn, Fig. 5 shows that the average number of tests per million in Group 2 is 5.9 times higher than in Group 1.

Although this difference is significant, it is unknown to what extent the number of tests reflects the number of cases or the other way around. The identification of this cause-and-effect relationship is relevant because it allows answering important questions. First, populations from Group 1 countries are not tested because younger people are less likely to have symptoms? Second, populations of Group 1 countries are not tested because they are less likely to be contaminated by COVID-19?

Fig. 5 Box plot of the number of tests per million people per group

Conclusions

This research reveals that the factors that most expose countries to COVID-19 are directly or indirectly associated with the age of the population. Countries that have Gross Domestic Product per capita and Human Development Index 4 times and 1.2 times higher have I-Cases and I-Deaths 2.6 and 1.6 times higher. The average age of the population of countries most exposed to COVID-19 is, on average, 12 years older. The proportion of the population aged 60 and 70 in these countries is also 2.1 and 2.2 times higher.

Factors directly associated with the population age have a combined weight of 0.95 and 0.61 in the CI-Exposure-GRG and CI-Exposure-PCA, respectively. The correlation of these composite indicators with the I-Cases was $R = 0.67$ and $R = 0.72$ and with the I-Deaths of $R = 0.67$ and $R = 0.64$. The Kappa concordance coefficient between country scores with country groups was $K = 0.45$ and $K = 0.48$. These results indicate that composite indicators are externally and internally consistent and provide reliable information about COVID-19 in the 157 countries analyzed.

The results suggest that the global inequality scale and socioeconomic vulnerability are not determinants of exposure to COVID-19, as shown by other national-scale studies. Life expectancy can satisfactorily substitute healthy life expectancy as an explanatory factor for the number of cases of COVID-19. Population density is not a factor that exposes countries to COVID-19 on a global scale. This result suggests that the correlation between COVID-19 and population density can be restricted to analyzes of continents with similar countries as in Africa or nationally as in the United States. Positive correlations between the prevalence of diabetes and COVID-19 cases and deaths occur only nationally as in Mexico, but not on a global scale. These results suggest that correlations between COVID-19 with socioeconomic vulnerability, population density, the prevalence of diabetes, and other indicators should be performed by age group.

This research also suggests that I-Cases in countries less exposed to COVID-19 could be higher than those indicated by statistics. The average number of tests per thousand people in countries less exposed to COVID-19 is 5.9 times less than those exposed to COVID-19. However, this problem does not change to reality about I-Deaths than it is 37% higher in countries more

exposed to COVID-19. These results indicate that the composite indicator of exposure to COVID-19 should be analyzed alongside factors that can distort the statistics. Factors on vaccination, social distancing, and health infrastructure can be added to a composite indicator and enrich the analyzes.

Acknowledgment The scope of this research was proposed by Belo Investment Research.

Funding This work was carried out with the support of The Coordination for the Improvement of Higher Education Personnel–Brazil (CAPES)–Financing Code 001; and National Council for Scientific and Technological Development of Brazil (CNPq) Productivity Grant, Grant 311032/2016–8.

Human or animals rights No Human Participants and/or Animals are involved in this research.

Financial interest The authors declare they have no financial interests.

Appendix: Annex–Countries and sub-indicators employed in the analysis

Continents and countries with more than 5% outliers and more than 50% missing data have been discarded. The following data were discarded: Africa, Andorra, Anguilla, Aruba, Asia, Belarus, Bermuda, Burundi, Cayman Islands, Curacao, Czechia, Democratic Republic of Congo, Dominica, Europe, European Union, Faeroe Islands, Falkland Islands, Gibraltar, Greenland, Hong Kong, Hungary, Indonesia, Isle of Man, Japan, Kosovo, Laos, Liechtenstein, Luxembourg, Macao, Madagascar, Marshall Islands, Mauritius, Micronesia (country), Monaco, Montenegro, Montserrat, Nauru, Nicaragua, North America, Oceania, Qatar, Saint Helena, Saint Kitts and Nevis, Samoa, San Marino, Singapore, Solomon Islands, South America, South Korea, Taiwan, Tanzania, Timor, Tonga, Turks and Caicos Islands, Uzbekistan, Vatican and, Vietnam.

The countries analyzed in this study were: Afghanistan, Albania, Algeria, Angola, Antigua and Barbuda, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belgium, Belize, Benin, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Brunei, Bulgaria, Burkina Faso, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Chad, Chile, China,

Colombia, Comoros, Congo, Costa Rica, Cote d'Ivoire, Croatia, Cuba, Cyprus, Denmark, Djibouti, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Eswatini, Ethiopia, Fiji, Finland, France, Gabon, Gambia, Georgia, Germany, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, Iceland, India, Iran, Iraq, Ireland, Israel, Italy, Jamaica, Jordan, Kazakhstan, Kenya, Kuwait, Kyrgyzstan, Latvia, Lebanon, Lesotho, Liberia, Libya, Lithuania, Malawi, Malaysia, Maldives, Mali, Malta, Mauritania, Mexico, Moldova, Mongolia, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Niger, Nigeria, North Macedonia, Norway, Oman, Pakistan, Palestine, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Romania, Russia, Rwanda, Saint Lucia, Saint Vincent and the Grenadines, Sao Tome and Principe, Saudi Arabia, Senegal, Serbia, Seychelles, Sierra Leone, Slovakia, Slovenia, Somalia, South Africa, South Sudan, Spain, Sri Lanka, Sudan, Suriname, Sweden, Switzerland, Syria, Tajikistan, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, Ukraine,

United Arab Emirates, United Kingdom, United States, Uruguay, Vanuatu, Venezuela, Yemen, Zambia and, Zimbabwe. The sub-indicators considered in the analysis of these countries are listed in Table 6.

Four sub-indicators have a percentage of missing data above 20%: Population in extreme poverty, Percentage of female smokers, Percentage of male smokers, and Handwashing facilities. The p -value–1 and p -value–2 indicate that Population Density, Diabetes prevalence, Percentage of male smokers are not correlated with I-Cases or I-Deaths. As shown by Paez et al. (2020) and Oyedotun and Moonsammy (2021), the Gross Domestic Product per capita is positively correlated with I-Cases and or with I-Deaths.

Nine of the fifteen original sub-indicators and 157 of the 193 original countries were maintained, considering the percentages of outliers, missing data, and significance of correlations. The exclusion of sub-indicators and countries allowed to reduce the missing data from twenty percent to three percent.

Table 6 Data availability of sub-indicators and their correlations with I-cases (1) and I-deaths (2)

Sub-indicator	gl ($n-2$)	$R-1$	$R-2$	p - value–1	p - value–2	Data availability	Maintain	Polarity
Stringency index	178	0.19	0.19	0.01	0.01	0.92	Yes	+ 1
Population density	198	– 0.01	0.07	0.94	0.33	0.98	No	– 1
Median age of the population	185	0.60	0.64	0.00	0.00	0.99	Yes	– 1
Proportion of population aged 65 years	183	0.64	0.61	0.00	0.00	0.99	Yes	– 1
Proportion of population aged 70 years	184	0.63	0.61	0.00	0.00	0.99	Yes	– 1
Gross domestic product per capita	188	0.29	0.50	0.00	0.00	0.97	Yes	+ 1
Population in extreme poverty	123	– 0.46	– 0.51	0.00	0.00	0.68	No	– 1
Deaths from cardiovascular diseases (per 1000 people)	186	– 0.19	– 0.28	0.01	0.00	0.99	Yes	– 1
Diabetes prevalence	193	– 0.02	0.01	0.79	0.85	0.99	No	– 1
Percentage of female smokers	143	0.69	0.68	0.00	0.00	0.78	No	– 1
Percentage of male smokers	141	0.08	0.11	0.33	0.21	0.77	No	– 1
Handwashing facilities	93	0.46	0.55	0.00	0.00	0.53	No	+ 1
Hospital beds (per 1000 people)	167	0.34	0.35	0.00	0.00	0.89	Yes	+ 1
Population life expectancy	205	0.51	0.60	0.00	0.00	0.99	Yes	+ 1
Human development index	185	0.55	0.64	0.00	0.00	0.99	Yes	+ 1

$R-1$ and p -value–1 refer to I-cases and, $R-2$ and p -value–2 refer to I-deaths

References

- Abadie, J., & Carpentier, J. (1969). Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints. In R. Fletcher (Ed.), *Optimization* (pp. 37–47). Academic Press.
- Acharya, R., & Porwal, A. (2020). A vulnerability index for the management of and response to the COVID-19 epidemic in India: An ecological study. *The Lancet Global Health*, 8(9), e1142–e1151.
- Appiah-Otoo, I., & Kursah, M. B. (2021). Modelling spatial variations of novel coronavirus disease (COVID-19): Evidence from a global perspective. *Geojournal*. <https://doi.org/10.1007/s10708-021-10427-0>
- Azarpazhooh, M. R., Morovatdar, N., Avan, A., Phan, T. G., Divani, A. A., Yassi, N., & Di Napoli, M. (2020). COVID-19 pandemic and burden of non-communicable diseases: an ecological study on data of 185 countries. *Journal of Stroke and Cerebrovascular Diseases*, 29(9), 105089.
- Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, 80, 12–22.
- Benita, F., & Gasca-Sanchez, F. (2021). The main factors influencing COVID-19 spread and deaths in Mexico: A comparison between phases I and II. *Applied Geography*, 134, 102523.
- Benzell, S. G., Collis, A., & Nicolaidis, C. (2020). Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. *Proceedings of the National Academy of Sciences*, 117(26), 14642–14644.
- Bernardes, P., Ekel, P. I., Rezende, S. F. L., Pereira Júnior, J. G., dos Santos, A. C. G., da Costa, M. A. R., & Libório, M. P. (2021). Cost of doing business index in Latin America. *Quality and Quantity*. <https://doi.org/10.1007/s11135-021-01221-8>
- Boyd, M., & Wilson, N. (2021). Optimizing island refuges against global catastrophic and existential biological threats: Priorities and preparations. *Risk Analysis*. <https://doi.org/10.1111/risa.13735>
- Bukari, C., Aning-Agyei, M. A., Kyeremeh, C., Essilfie, G., Amuquandoh, K. F., Owusu, A. A., & Bukari, K. I. (2021). Effect of COVID-19 on household food insecurity and poverty: Evidence from Ghana. *Social Indicators Research*. <https://doi.org/10.1007/s11205-021-02766-9>
- Cinelli, M., Spada, M., Kim, W., Zhang, Y., & Burgherr, P. (2021). MCDA Index Tool: An interactive software to develop indices and rankings. *Environment Systems and Decisions*, 41(1), 82–109.
- Cristofaro, L., Gil-Alana, L. A., Chen, Z., & Wanke, P. (2021). Modelling stock market data in China: Crisis and Coronavirus. *Finance Research Letters*, 41, 101865.
- Dialga, I., & Le Giang, T. H. (2017). Highlighting methodological limitations in the steps of composite indicators construction. *Social Indicators Research*, 131(2), 441–465.
- Dlamini, W. M., Dlamini, S. N., Mabaso, S. D., & Simelane, S. P. (2020). Spatial risk assessment of an emerging pandemic under data scarcity: A case of COVID-19 in Eswatini. *Applied Geography*, 125, 102358.
- Dowd, J. B., Andriano, L., Brazel, D. M., Rotondi, V., Block, P., Ding, X., & Mills, M. C. (2020). Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proceedings of the National Academy of Sciences*, 117(18), 9696–9698.
- Drefahl, S., Wallace, M., Mussino, E., Aradhya, S., Kolk, M., Brandén, M., & Andersson, G. (2020). A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nature Communications*, 11(1), 1–7.
- Dzúrová, D., & Květoň, V. (2021). How health capabilities and government restrictions affect the COVID-19 pandemic: Cross-country differences in Europe. *Applied Geography*, 135, 102551.
- El Gibari, S., Gómez, T., & Ruiz, F. (2019). Building composite indicators using multicriteria methods: A review. *Journal of Business Economics*, 89(1), 1–24.
- Erić, O., Popović, G., & Bjelić, J. (2021). Economic response of the European countries to the first wave of COVID-19. *EMC REVIEW-ČASOPIS ZA EKONOMIJU*, 21(1), 63–78.
- Ferraz, D., Mariano, E. B., Manzone, P. R., Morales, H. F., Morceiro, P. C., Torres, B. G., & Do Nascimento Rebelatto, D. A. (2021). COVID health structure index: The vulnerability of Brazilian microregions. *Social Indicators Research*, 158, 197–215. <https://doi.org/10.1007/s11205-021-02699-3>
- Forati, A. M., & Ghose, R. (2021). Geospatial analysis of misinformation in COVID-19 related tweets. *Applied Geography*, 133, 102473.
- Funke, M., & Tsang, A. (2020). The People's bank of China's response to the coronavirus pandemic: A quantitative assessment. *Economic Modelling*, 93, 465–473.
- Goh, T. S., Henry, H., & Albert, A. (2021). Determinants and prediction of the stock market during COVID-19: Evidence from Indonesia. *The Journal of Asian Finance, Economics, and Business*, 8(1), 1–6.
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(1), 61–94.
- Gupta, D., Biswas, D., & Kabiraj, P. (2021). COVID-19 outbreak and Urban dynamics: Regional variations in India. *Geojournal*. <https://doi.org/10.1007/s10708-021-10394-6>
- Hryhoruk, P., Khrushch, N., Grygoruk, S., Grygoruk, S., Prystup, L., & Gorbatiuk, K. (2021). Assessing the impact of COVID-19 pandemic on the regions' socio-economic development: The case of Ukraine. *European Journal of Sustainable Development*, 10(1), 63.
- Ignat, R., & Constantin, M. (2020). Multidimensional facets of entrepreneurial resilience during the COVID-19 crisis through the lens of the wealthiest Romanian counties. *Sustainability*, 12(23), 10220.
- Jolliffe, I. (2005). Principal component analysis. *Encyclopedia of statistics in behavioral science*. Wiley.
- Joshua, V., Grace, J. S., Emmanuel, J. G., Satish, S., & Elangovan, A. (2021). Spatial mapping of COVID-19 for Indian states using principal component analysis. *Clinical Epidemiology and Global Health*, 10, 100690.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
- Kaiser, M., Chen, A. T. Y., & Gluckman, P. (2021). Should policy makers trust composite indices? A commentary on

- the pitfalls of inappropriate indices for policy formation. *Health Research Policy and Systems*, 19(1), 1–11.
- Kalla, M. I., Lahmar, B., Geullouh, S., & Kalla, M. (2021). Health geo-governance to assess the vulnerability of Batna, Algeria to COVID-19: The role of GIS in the fight against a pandemic. *GeoJournal*. <https://doi.org/10.1007/s10708-021-10449-8>
- Karácsonyi, D., Dyrting, S., & Taylor, A. (2021). A spatial interpretation of Australia's COVID-vulnerability. *International Journal of Disaster Risk Reduction*, 61, 102299.
- Karmakar, M., Lantz, P. M., & Tipirineni, R. (2021). Association of social and demographic factors with COVID-19 incidence and death rates in the US. *JAMA Network Open*, 4(1), e2036462–e2036462.
- Khan, K., Zhao, H., Zhang, H., Yang, H., Shah, M. H., & Jahanger, A. (2020). The impact of COVID-19 pandemic on stock markets: An empirical analysis of world major stock indices. *The Journal of Asian Finance, Economics, and Business*, 7(7), 463–474.
- Kinnunen, J., Georgescu, I., Hosseini, Z., & Androniceanu, A. M. (2021). Dynamic indexing and clustering of government strategies to mitigate Covid-19. *Entrepreneurial Business and Economics Review*, 9(2), 7–20.
- Kitrar, L. (2021). The relationship of economic sentiment and GDP growth in Russia in light of the Covid-19 crisis. *Entrepreneurial Business and Economics Review*, 9(1), 7–29.
- Knaul, F., Arreola-Ornelas, H., Porteny, T., Touchton, M., Sánchez-Talanquer, M., Méndez, Ó., Group from the Observatory for the Containment of COVID-19 in the Americas. (2021). Not far enough: Public health policies to combat COVID-19 in Mexico's states. *Plos one*, 16(6), e0251722.
- Krishna, T. S., Babu, A. Y., & Kumar, R. K. (2018). Determination of optimal clusters for a non-hierarchical clustering paradigm K-means algorithm. In *Proceedings of International Conference on Computational Intelligence and Data Engineering* (pp. 301–316). Springer, Singapore.
- Kuc-Czarnecka, M., Piano, S. L., & Saltelli, A. (2020). Quantitative storytelling in the making of a composite indicator. *Social Indicators Research*. <https://doi.org/10.1007/s11205-020-02276-0>
- Kujur, S. K., & Goswami, D. (2021). Do Covid-19 induced NHRD policies have a dampening effect on employment? *Human Resource Development International*, 24(4), 454–464.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lasdon, L. S., Fox, R. L., & Ratner, M. W. (1974). Nonlinear optimization using the generalized reduced gradient method. *Revue Française d'automatique, Informatique, Recherche Opérationnelle. Recherche Opérationnelle*, 8(3), 73–103.
- Lee, K. Y. M., Jais, M., & Chan, C. W. (2020). Impact of covid-19: Evidence from Malaysian stock market. *International Journal of Business and Society*, 21(2), 607–628.
- Li, K., Liang, Y., Li, J., Liu, M., Feng, Y., & Shao, Y. (2020). Internet search data could Be used as novel indicator for assessing COVID-19 epidemic. *Infectious Disease Modelling*, 5, 848–854.
- Li, S., Ning, K., & Zhang, T. (2021). Sentiment-aware jump forecasting. *Knowledge-Based Systems*, 228, 107292.
- Libório, M. P., da Silva Martinuci, O., Machado, A. M. C., Machado-Coelho, T. M., Laudares, S., & Bernardes, P. (2020). Principal component analysis applied to multidimensional social indicators longitudinal studies: Limitations and possibilities. *GeoJournal*. <https://doi.org/10.1007/s10708-020-10322-0>
- Libório, M. P., da Silva Martinuci, O., Machado, A. M. C., de Mello Lyrio, R., & Bernardes, P. (2021). Time-space analysis of multidimensional phenomena: A composite indicator of social exclusion through k-means. *Social Indicators Research*. <https://doi.org/10.1007/s11205-021-02763-y>
- Libório, M. P., Martinuci, O. D. S., Machado, A. M. C., Hadad, R. M., Bernardes, P., & Camacho, V. A. L. (2021b). Adequacy and consistency of an intraurban inequality indicator constructed through principal component analysis. *The Professional Geographer*, 73(2), 282–296.
- Liu, K. (2021). The effects of COVID-19 on Chinese stock markets: an EGARCH approach. *Economic and Political Studies*. <https://doi.org/10.1080/20954816.2021.1914415>
- Liu, A., Vici, L., Ramos, V., Giannoni, S., & Blake, A. (2021). Visitor arrivals forecasts amid COVID-19: A perspective from the Europe team. *Annals of Tourism Research*, 88, 103182.
- Luca, M. M., Mustea, L., Taran, A., Stefea, P., & Vatavu, S. (2021). Challenges on radical health redesign to reconfigure the level of e-health adoption in EU countries. *Frontiers in Public Health*. <https://doi.org/10.3389/fpubh.2021.728287>
- Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., & Pilla, F. (2021). Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. *Sustainable Cities and Society*, 68, 102784.
- Marković, R., Šterk, M., Marhl, M., Perc, M., & Gosak, M. (2021). Socio-demographic and health factors drive the epidemic progression and should guide vaccination strategies for best COVID-19 containment. *Results in Physics*, 26, 104433.
- Martines, M. R., Ferreira, R. V., Toppa, R. H., Assunção, L. M., Desjardins, M. R., & Delmelle, E. M. (2021). Detecting space–time clusters of COVID-19 in Brazil: Mortality, inequality, socioeconomic vulnerability, and the relative risk of the disease in Brazilian municipalities. *Journal of Geographical Systems*, 23(1), 7–36.
- Mazziotta, M., & Pareto, A. (2017). Synthesis of indicators: The composite indicators approach. *Complexity in society: From indicators construction to their synthesis* (pp. 159–191). Cham: Springer.
- Mitze, T., & Kosfeld, R. (2021). The propagation effect of commuting to work in the spatial transmission of COVID-19. *Journal of Geographical Systems*. <https://doi.org/10.1007/s10109-021-00349-3>
- Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. *European Commission, Ispra*, 15(1), 19–20.
- Nasiri, R., Akbarpour, S., Zali, A. R., Khodakarami, N., Boochani, M. H., Noory, A. R., & Soori, H. (2021). Spatio-temporal analysis of COVID-19 incidence rate using GIS:

- A case study—Tehran metropolitan Iran. *Geojournal*. <https://doi.org/10.1007/s10708-021-10438-x>
- Nicodemo, C., Barzin, S., Lasserson, D., Moscone, F., Redding, S., & Shaikh, M. (2020). Measuring geographical disparities in England at the time of COVID-19: Results using a composite indicator of population vulnerability. *BMJ Open*, 10(9), e039749.
- Olusola, A., Olusola, B., Onafeso, O., Ajiola, F., & Adelabu, S. (2020). Early geography of the coronavirus disease outbreak in Nigeria. *GeoJournal*. <https://doi.org/10.1007/s10708-020-10278-1>
- Onafeso, O. D., Onafeso, T. E., Olumuyiwa-Oluwabiye, G. T., Faniyi, M. O., Olusola, A. O., Dina, A. O., & Adagbasa, E. (2021). Geographical trend analysis of COVID-19 pandemic onset in Africa. *Social Sciences and Humanities Open*, 4(1), 100137.
- Otoiu, A., Pareto, A., Grimaldi, E., Mazziotto, M., & Terzi, S. (2021). *Open issues in composite indicators. A starting point and a reference on some state-of-the-art issues (Vol 3)*. Roma TrE-Press.
- Oyedotun, T. D. T., & Moonsammy, S. (2021). Spatiotemporal variation of COVID-19 and its spread in South America: A rapid assessment. *Annals of the American Association of Geographers*, 111(6), 1868–1879.
- Paez, A., Lopez, F. A., Menezes, T., Cavalcanti, R., & Pitta, M. G. D. R. (2020). A spatio-temporal analysis of the environmental correlates of COVID-19 incidence in Spain. *Geographical Analysis*, 53(3), 397–421. <https://doi.org/10.1111/gean.12241>
- Pandey, A., Prakash, A., Agur, R., & Maruvada, G. (2021). Determinants of COVID-19 pandemic in India: An exploratory study of Indian states and districts. *Journal of Social and Economic Development*. <https://doi.org/10.1007/s40847-021-00154-0>
- Pang, S., Hu, X., & Wen, Z. (2021). Environmental risk assessment and comprehensive index model of disaster loss for COVID-19 transmission. *Environmental Technology and Innovation*, 23, 101597.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.
- Powell, S. G., & Batt, R. J. (2008). *Modeling for insight*. Hoboken, NJ: John Wiley & Sons Inc.
- Quirk, T. J. (2012). One-way analysis of variance (ANOVA). *Excel 2007 for educational and psychological Statistics* (pp. 163–179). New York, NY: Springer.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Salisu, A. A., Ogbonna, A. E., Oloko, T. F., & Adediran, I. A. (2021). A new index for measuring uncertainty due to the COVID-19 pandemic. *Sustainability*, 13(6), 3212.
- Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*, 62, 102418.
- Sarkar, A., & Chouhan, P. (2021). COVID-19: District level vulnerability assessment in India. *Clinical Epidemiology and Global Health*, 9, 204–215.
- Sarmadi, M., Marufi, N., & Moghaddam, V. K. (2020). Association of COVID-19 global distribution and environmental and demographic factors: An updated three-month study. *Environmental Research*, 188, 109748.
- Shabani, Z. D., & Shahnazi, R. (2020). Spatial distribution dynamics and prediction of COVID-19 in Asian countries: Spatial Markov chain approach. *Regional Science Policy and Practice*, 12(6), 1005–1025.
- Shehzad, K., Xiaoxing, L., & Kazouz, H. (2020). COVID-19's disasters are perilous than global financial crisis: A rumor or fact? *Finance Research Letters*, 36, 101669.
- Shehzad, K., Xiaoxing, L., Arif, M., Rehman, K. U., & Ilyas, M. (2020b). Investigating the psychology of financial markets during covid-19 era: A case study of the us and european markets. *Frontiers in Psychology*, 11, 1924.
- Shehzad, K., Xiaoxing, L., Kazouz, H., Balsalobre-Lorente, D., Zeraibi, A., & Rauf, A. (2021). An asymmetric spillover between China and Pakistan' stock markets: A comparative analysis before and during COVID-19 crisis. *Journal of Sustainable Finance and Investment*. <https://doi.org/10.1080/20430795.2020.1870203>
- Suh, E., & Alhaery, M. (2021). Measuring reopening readiness: A universal COVID-19 index for US states. *Library Hi Tech*. <https://doi.org/10.1108/LHT-03-2021-0115>
- Tang, X., Li, Z., Hu, X., Xu, Z., & Peng, L. (2021). Self-correcting error-based prediction model for the COVID-19 pandemic and analysis of economic impacts. *Sustainable Cities and Society*, 74, 103219.
- Tribby, C. P., & Hartmann, C. (2021). COVID-19 cases and the built environment: Initial evidence from New York City. *The Professional Geographer*, 73(3), 365–376. <https://doi.org/10.1080/00330124.2021.1895851>
- Tripathi, S., Gist, K. M., Bjornstad, E. C., Kashyap, R., Boman, K., Chiotos, K., & Bhalala, U. S. (2021). Coronavirus disease 2019–associated PICU admissions: A report from the society of critical care medicine discovery network viral infection and respiratory illness universal study registry. *Pediatric Critical Care Medicine*, 22(7), 603.
- Vukovic, D., Maiti, M., Grubisic, Z., Grigorieva, E. M., & Frömmel, M. (2021). COVID-19 pandemic: Is the crypto market a safe haven? The impact of the first wave. *Sustainability*, 13(15), 8578.
- Weinstein, B., da Silva, A. R., Kouzoukas, D. E., Bose, T., Kim, G. J., Correa, P. A., & Carpenter, D. O. (2021). Precision mapping of COVID-19 vulnerable locales by epidemiological and socioeconomic risk factors, developed using South Korean data. *International Journal of Environmental Research and Public Health*, 18(2), 604.
- Wen, F., Cao, J., Liu, Z., & Wang, X. (2021). Dynamic volatility spillovers and investment strategies between the Chinese stock market and commodity markets. *International Review of Financial Analysis*, 76, 101772.
- Wyper, G. M., Assunção, R., Cuschieri, S., Devleeschauwer, B., Fletcher, E., Haagsma, J. A., & Grant, I. (2020). Population vulnerability to COVID-19 in Europe: A burden of disease analysis. *Archives of Public Health*, 78, 1–8.
- Yang, T., Kim, S., & Matthews, S. A. (2021). Face masking violations, policing, and COVID-19 death rates: A spatial analysis in New York city zip codes. *The Professional Geographer*. <https://doi.org/10.1080/00330124.2021.1933552>

- Yigitcanlar, T., Kankanamge, N., Inkinen, T., Butler, L., Preston, A., Rezayee, M., & Senevirathne, M. (2021). Pandemic vulnerability knowledge visualisation for strategic decision-making: A COVID-19 index for government response in Australia. *Management Decision*. <https://doi.org/10.1108/MD-11-2020-1527>
- Zhou, S. J., Zhang, L. G., Wang, L. L., Guo, Z. C., Wang, J. Q., Chen, J. C., & Chen, J. X. (2020). Prevalence and socio-demographic correlates of psychological health problems in Chinese adolescents during the outbreak of COVID-19. *European Child and Adolescent Psychiatry*, 29(6), 749–758.

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