The application of population pyramid information for COVID-19 cases and deaths mortality

Yasiru Ranasinghe[a,*], Harshana Weligampola[a], Gayanthi Ilangarathna[a], Roshan Godaliyadda[a], Vijitha Herath[a], Parakrama Ekanayake[a], Janaka Ekanayake[a], Samath Dharmaratne[b], Anuruddhika Rathnayake[c], Mallika Pinnawala[d], Sakunthala Yatigammana[d], Ganga Tilakaratna[e]

[a]*Faculty of Engineering, University of Peradeniya, Peradeniya, Sri Lanka, 20400*
[b]*Faculty of Medicine, University of Peradeniya, Peradeniya, Sri Lanka, 20400*
[c]*Ministry of Health, Colombo, Sri Lanka, 01000*
[d]*Faculty of Arts, University of Peradeniya, Peradeniya, Sri Lanka, 20400*
[e]*Institute of Policy Studies, Colombo, Sri Lanka, 00700*

Abstract

Background

In modern times, the COVID-19 pandemic presented the world with an unprecedented health crisis, yet certain countries have had far more success in limiting COVID-19. This disproportionate impact of COVID-19 may be attributed to socioeconomic standards, healthcare conditions, and cultural and political factors. Inordinately, these factors contribute to the population pyramid of a country. Hence, we examine how a novel statistic defined using the population pyramid affects the COVID-19 severity in the country.

Methods

We derived a scalar variable, named $\text{PoPStat}_{\text{COVID-19}}$, from the information-rich population pyramid correlated with COVID-19 fatality statistics. Then, we estimated the relationship between COVID-19 cases and deaths per million with the proposed $\text{PoPStat}_{\text{COVID-19}}$ as of April 08, 2022.

Findings

The proposed mechanism processed the population pyramid shape to generate the $\text{PoPStat}_{\text{COVID-19}}$ parameter, which has a better correlation to cases and deaths than standard statistical parameters, which are not generated with any specific optimization criterion. The results indicated that regressive population structures had more significant cases and deaths per million than progressive population pyramids. $\text{PoPStat}_{\text{COVID-19}}$ enables visualizing transformation of the population pyramid along the direction of COVID-19 severity increase.

Interpretation

The high positive correlation with COVID-19 exhibited by the proposed $\text{PoPStat}_{\text{COVID-19}}$ can be attributed to its ability to better gauge the underpinning demographic information compared to other standard statistical measures. Countries with regressive population pyramids were more severely affected by COVID-19, wherein socioeconomic standards tend to be high because the pandemic severity becomes fatal for the vulnerable groups of the infected population.

Funding

International Development Research Centre, Canada.

*Corresponding author

Email addresses: e14273@eng.pdn.ac.lk (Yasiru Ranasinghe), harshana.w@eng.pdn.ac.lk (Harshana Weligampola), ilangarathna.g@eng.pdn.ac.lk (Gayanthi Ilangarathna), roshangodd@ee.pdn.ac.lk (Roshan Godaliyadda), vijitha@eng.pdn.ac.lk (Vijitha Herath), mpb.ekanayake@ee.pdn.ac.lk (Parakrama Ekanayake), ekanayakej@eng.pdn.ac.lk (Janaka Ekanayake), samathd@pdn.ac.lk (Samath Dharmaratne), m29782@pgim.cmb.ac.lk (Anuruddhika Rathnayake), mallikap@pdn.ac.lk (Mallika Pinnawala), sakuyatigammana@arts.pdn.ac.lk (Sakunthala Yatigammana), ganga@ips.lk (Ganga Tilakaratna)

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first declared a public health emergency of international concern and then a pandemic on 30th January and 11th March 2020, respectively. By the time this paper was written, nearly five hundred million cases with more than six million fatalities were reported globally. The impact of the disease, in terms of cases and fatalities, varied from country to country.

Research in context

Evidence before this study

There are few studies about the impact of demographic and economic variables: median age, population density, and GDP per capita to COVID-19. Besides that, the cultural impact on COVID-19 has been analyzed in a study among the other statistical studies. We searched PubMed, Scopus, and Google Scholar on July 1, 2022, with no date restrictions, for publications in English, using the keywords, "population pyramid", "COVID-19" and "correlation" to identify correlation analysis performed using population demographics and COVID-19 fatality records. We only identified one study, published in 2020, which considered a systematic review of the population pyramids of 20 countries with the most and least COVID-19 mortality.

Added value of this study

To our best knowledge, our work is the first analysis to correlate the population pyramid shape to COVID-19 cases and deaths. We devised a mechanism to generate the maximally correlated scalar parameter derivable from the population pyramid to COVID-19 severity for 163 countries. The proposed scalar variable recorded a strong correlation with cases (r=0.81, p<0.0001, 95% CI=[0.7457, 0.8546]) and deaths (r=0.71, p<0.0001, 95% CI=[0.6205, 0.7762]) per million, respectively, as of April 2022. Nations with a regressive pyramid structure recorded a higher severity than countries with progressive pyramid shapes. The positive correlation suggests that the countries with more elderly populations have been vulnerable to COVID-19 though these countries generally prosper in economy and healthcare. With repeated infections from multiple variants, a significant proportion of the population is bound to get infected.

Implications of all the available evidence

Our results suggest that nations with higher life expectancy, hence more older people, were more vulnerable to COVID-19 despite having better socioeconomic standards as of April 2022. Research in socioeconomic, public health, and political science has shown the viability of the population pyramid as an instrument for policymaking. Policymakers could use the proposed mechanism to benchmark against countries with similar socioeconomic standards in initiating policies for COVID-19 or other similar deceases. The proposed mechanism visually traces how the shape of the population pyramid changes as vulnerability to COVID-19 increases allowing for more in-depth exploration. Finally, the mechanism is easily extendable for any cause of death; hence, the contributions are not limited to COVID-19.

The differences in case and death fatality rates of COVID-19 have been studied against the age statistics[1,2,3,4] of the global population. In the United States, the highest case fatality rate of 10% - 27% was recorded for patients with age $\geq$ 85 years, followed by 65-84 years (3% to 11%) and 20-54 years (<1%) age groups according to morbidity and mortality weekly report.[2] In Italy, the reported case rates of fatality by Livingston and Bucher[5] were 22.7% for those aged $\geq$ 90 years, 19.7% for those aged 80-89 years, 12.5% for those aged 70-79 years and 3.5% for those aged 60-69 years. Furthermore, Wu C et al.[4] found that old age ($\geq$ 65 years) was associated with a greater risk of death (hazard ratio 6.17, 95% confidence interval 3.26-11.67). Similarly, Fu et al.[6] revealed that old age (> 70 years) was significantly at a higher risk of death from COVID-19 (relative risk 10.67).

In literature, several studies are available that attempt to deduce an explanatory variable to examine how different demographics and populations are affected by and how have they responded to the threat of COVID-19[7] and other diseases.[8,9,10] Census data of the population, economic indicators, and socio-cultural variances in the population of different countries have been employed to perceive significant descriptors. For example, Gelfand et al.[11] propounded that the differences in cultural tightness-looseness can be matched to the variation in cases and deaths during COVID-19 and that social norms impacted effective COVID-19 mitigation since cultures with tight norms were able to respond to the pandemic successfully. Then, Tan et al.[12] studied the association between the country-level income and the COVID-19 cases and deaths to understand the impact of the economic activity level on the prevention of COVID-19 cases and deaths.

However, these parameters are skewed towards specific groups in the population. For example, even though remittances influence income level, there is no causation between the immigrant workers or population and the pandemic progression of the home country. Further, recent surges in countries that were less affected by earlier variants due to the high transmission capability of omicron had high case counts. Therefore a shift in focus from a high level of containment to optimal containment with minimal socioeconomic impact might be a focus. More importantly, it[11] has indicated that the death count is a more reliable indicator of the pandemic severity than the case count.

Moreover, Bhadra et al.[13] conducted a study on the impact of population density on the COVID-19 progression exclusive to India. However, population density could be misrepresented as the populace's distribution over the country's land coverage will dilute the population density due to unpopulated areas. On the other hand, Wang et al.[14] analyzed the correlation of the ageing population, median age, and life expectancy to the mortality cases of COVID-19. However, despite the frequent application of median age, the scalar parameter does not represent the differences in the population distribution among countries because it condenses the distribution's information. Further, from a statistical perspective, the median does not reflect certain types of statistics in the distribution. Most importantly, the shape-based parameters of the distribution like skewness, which carry more information about the regressive vs progressive nature of the population pyramid, are not reflected in median age.

This work proposes using the differences between the overall age distributions among countries to derive an explanatory variable for nationwide COVID-19 cases and deaths. The proposed mechanism yields a scalar distance statistic that exemplifies the correlation with COVID-19 severity. We tested the capability of the proposed variable to explain the recorded COVID-19 cases and deaths per million as of April 2022. The proposed mechanism as a statistical method could be used in policy-making[15,16] specific to the mortality cause. Though the proposed mechanism was constructed using census data at a country resolution, the method applies to lower administrative levels such as states and provinces of individual countries. As the proposed method provides a comprehensive understanding of the connection between age demographics and disease severity, this work could be used by both global and national policymakers. Furthermore, it can be easily extended to study similar impacts and vulnerabilities to other causes of death and interpret them through a socioeconomic lens due to the proposed method's ability to track the shape changes of the population pyramid as severity increases.

2. Methods

2.1. COVID-19 cases and deaths

We retrieved data on COVID-19 as of April 08 2022, from "Our World in Data", which provides daily updates on the number of COVID-19 cases and deaths globally, starting from the first documented case using data from the European Centre for Disease Prevention and Control. We log transformed cases and deaths per million to distribute normally. The data are from 163 countries for which COVID-19 data were downloaded on April 09 2022, and are available online.

2.2. Implications of the population pyramid

A population pyramid is a graphical presentation of the age and gender of a population. Economic variations, public health standards,[17] natural events, conflicts, and political changes can impact birth and death rates, influencing the shape of the population pyramid. For example, China's pyramid shrank inwards for the age cohorts around 50, showing the slowing of population growth during the Great Chinese famine. After the famine, the population started increasing, but about 30 years ago, the pyramid started thinner again with the "one-child policy"[18,19] enacted in 1979. Hence, the population pyramid reflects national policies[20] and how said policies affect the country's living conditions.

Apart from the derivative variables of the population pyramid, such as median age, a correlation exists between the economic variables and the pyramid shape.[21] For example, a macro-economist may study the relationship between population growth and growth in GDP per capita. In contrast, a health economist may study how local population composition impacts the demand for long-term care facilities. Besides, countries with different population compositions (but similar in other respects) will allocate their resources differently; therefore, the composition of their GDPs will also differ. For example, an inverted pyramid implies that the government could negotiate more health service payments. Therefore, the proposed mechanic allows for constructing such parameters that can tie these factors: economic or educational, to $\text{PoPStat}_{\text{COVID-19}}$.

Besides that, the socioeconomic and healthcare standards immediately impacted[22] the COVID-19 infections and deaths. It was observed that counties with regressive population pyramids had higher COVID-19 fatality rates as these countries had a more significant percentage of vulnerable groups in the population. Even though most communicable diseases are more severe in countries with lower healthcare standards, COVID-19 has exhausted countries with better economic standards. Besides, with multiple waves and variants, the pandemic would most likely ultimately infect a majority of the population means that the percentage of vulnerable groups exposed will be high. Therefore countries with a high proportion of vulnerable demographics are bound to be affected more as time progresses. Therefore, basic tools such as the population pyramid are conducive to understanding the characteristics of the progression of the disease[23,24] even though COVID-19 is a recent pandemic.

2.3. Significance of the demographic distribution

The population pyramid represents the population breakdown by gender and age at a given time. The most prominent population pyramid structures are progressive and regressive. Progressive structures are considered youthful populations, and regressive structures are considered aged populations. Nevertheless, for brevity, these quantitative data of each country will be referred to as the population pyramid. In this study, we considered the population data of each country of the year before the pandemic, and necessary data were retrieved from the World Bank, 2018. Moreover, similar to COVID-19 data, we standardized the downloaded data by the respective population size. With this modification, the standardized population pyramid fits the probability distribution definition, and this form is more pertinent for statistical analysis.

In Fig. 1, a few examples illustrate the significance of the population pyramid from the other commonly used variables: median age, population density, and GDP per capita for statistical analysis. We have provided the population pyramids of the countries with similar indices for the other variables but with disparate demographic distributions and vice versa. The difference suggests that the alternative variables are affected by factors other than the population demographics. Consider two countries with a similar population pyramid specific to the median age. Then, it can be inferred that their median ages would also be the same. Nonetheless, the converse of that statement does not hold water (i.e. countries with similar median ages might have dramatically different population pyramids as per Fig. 1a) because the median age is indifferent to any symmetrical changes in the population pyramid. Accordingly, it is

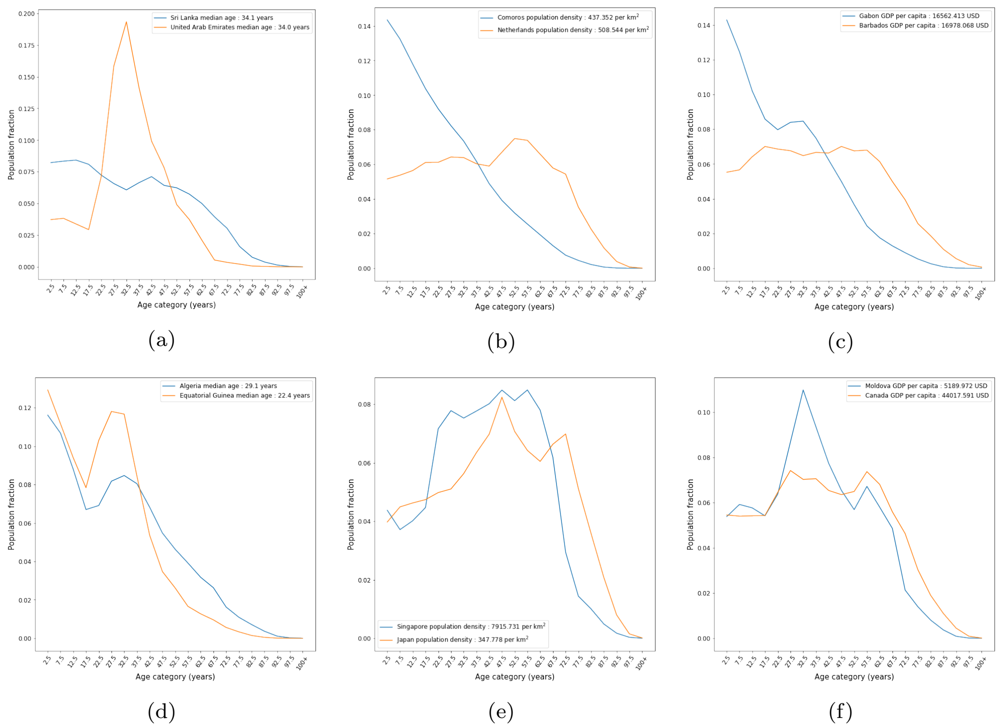

Fig. 1. Difference in population distribution for similar (a) median ages, (b) population density, and (c) GDP per capita and Similarity in population distribution for different (d) median ages, (e) population density, and (f) GDP per capita.

conclusive that countries with close population pyramids have similar demographic profiles, while such a claim is not necessarily valid with the median age as a parameter.

2.4. Accounting for under-reporting

Under-reporting is a concern when analyzing COVID-19 case rates in that the countries with higher testing rates are likely to report as suggested by multiple sources accurately.[25,26] Countries that adopted widespread testing have lower under-reporting probabilities than those countries that slowly adopted widespread testing (e.g., the USA). Accordingly, we used the ratio of tests to cases as our primary proxy for under-reporting following Gelfand et al.[11]

2.5. Constructing the PoPStat$_{COVID\text{-}19}$

Population pyramids are multi-dimensional objects, and to define a scalar parameter that can be correlated with metrics such as death rate, the distance between two population pyramids is a good candidate. Hence given a "pre-selected" reference, the distance or disparity from another population pyramid can be proposed. The distance can be more reliably estimated by the extent to which one population pyramid can be inferred from the pre-selected reference. i.e. the metric weighs the deformation between the shapes of the two population pyramids starting from the reference country. Therefore, the metric representing the relative difference between two distributions reflects this asymmetry unless the two probability distributions are identical. Therefore, we measure the distance between the two probability distributions using Kullback-Leibler (KL) divergence as

$$D_{KL}(P \parallel Q) = \sum_{i \in I} P_{(i)} \log \left(\frac{P_{(i)}}{Q_{(i)}} \right), \tag{1}$$

where i is the class index of each age interval of the population distributions P and Q and $P_{(i)}$ and $Q_{(i)}$ are the individual probabilities of the corresponding i^{th} age interval. $D_{KL}(P \parallel Q)$ is called the relative entropy between the two random vectors with respect to Q. The construction of the PoPStat$_{\text{COVID-19}}$ has been illustrated as a flow chart in the appendix pp 2.

2.6. Relationship between COVID-19 data and the population pyramid

We consider the $\text{PoPStat}_{\text{COVID-19}}$ described in section 2.5 as the explanatory for COVID-19 data to compute the correlation. However, as presented earlier, the distance computation is not symmetric with KL divergence and therefore, it is essential to construct a reference population pyramid. To establish a reference distribution, we formulated an objective function as follows.

$$\underset{\omega \in \Omega}{\arg\max} \quad \text{Cor}\big[S, D_{KL}(\omega)\big], \tag{2}$$

where Ω is the set of all countries included in the study and ω is the country which, when selected as the reference probability distribution for the $\text{PoPStat}_{\text{COVID-19}}$ calculation, yields the highest correlation between the two random variables (KL divergence based distance from the target country's population pyramid to reference population pyramid and COVID-19 cases or death). Therefore, S denotes the random variable representing COVID-19 cases or deaths information, and $D_{KL}(\omega)$ is the random variable constructed for the distance between two population pyramids with ω as the reference country. Then, we used brute force searching to establish a reference population pyramid because there is no closed-form solution to the above optimization problem. It is computationally economical given the discrete nature of all the variables involved in the study. It should be highlighted that this is a highly extendable optimization mechanism that can be used to find the optimum origin to have the highest correlation with population pyramid statistics and any other metric, such as a cause of death or an economic factor.

2.7. Role of the funding body

The funding body of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

3. Results

In Fig. 2, the transformation of the population pyramid shape from progressive to regressive as $\text{PoPStat}_{\text{COVID-19}}$ is increased is illustrated. Hence, the proposed mechanism visually quantifies the shape transformation of the population pyramid in the direction of increase or decrease of severity for a given parameter, in this case, COVID-19. This enables a more expansive analysis of underlying socioeconomic factors and their relationship to disease severity (as a policy-maker can map the influence of socioeconomic conditions or policy decisions the shape changes of the pyramid using this tool), in this study COVID-19 for the sake of demonstration of capability.

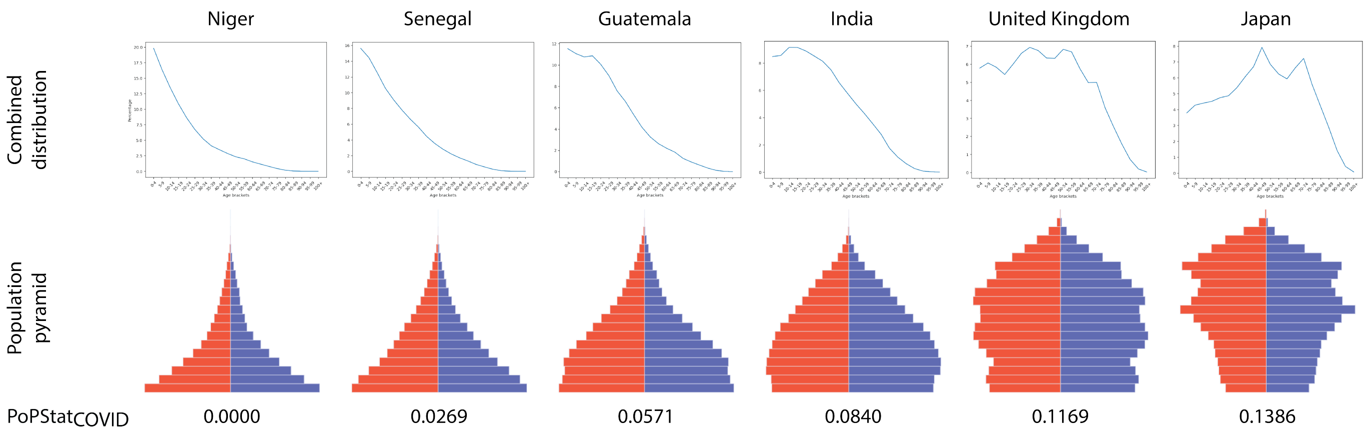

Fig. 2. Transformation of the population pyramid shape with $\text{PoPStat}_{\text{COVID-19}}$

In Fig. 3, the variation of COVID-19 cases and deaths (per million) with the proposed $\text{PoPStat}_{\text{COVID-19}}$ and the comparative variables are illustrated. According to Fig. 3a and Fig. 3e, there is a better linear variation for the cases and deaths with the proposed $\text{PoPStat}_{\text{COVID-19}}$ and the median age Fig. 3b and Fig. 3f than the GDP per capita Fig. 3c and Fig. 3g and the population density Fig. 3d and Fig. 3h as variables. The better linear variation translates as a strong correlation with the first two variables, especially with the proposed metric based on the population distribution and vice versa. The $\text{PoPStat}_{\text{COVID-19}}$ distances and the respective comparative variables of the countries are provided in the appendix pp 3 - pp 7 with the cases and deaths per million.

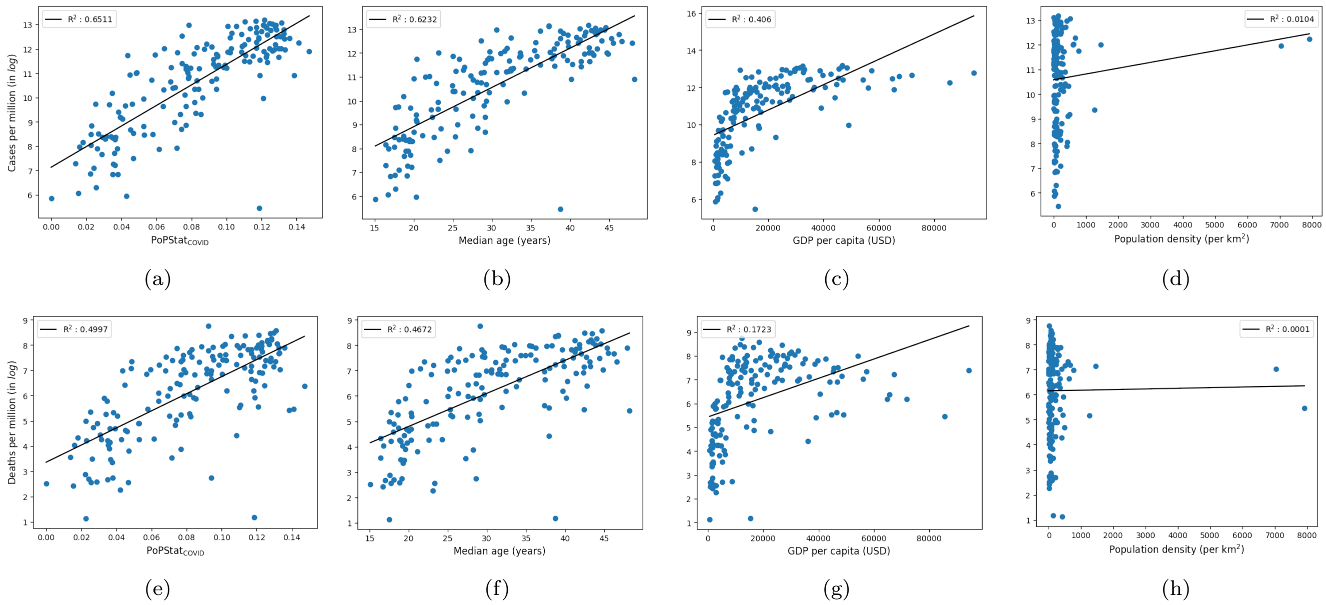

Fig. 3. Variation between the cases per million with (a) proposed $\text{PoPStat}_{\text{COVID-19}}$ (b) median age (c) GDP per capita (d) population density and deaths per million with (e) proposed $\text{PoPStat}_{\text{COVID-19}}$ (f) median age (g) GDP per capita (h) population density as at April 08,2022.

In this work, the reference country resulted to compute the $\text{PoPStat}_{\text{COVID-19}}$ from the optimization criterion was Niger. With the proposed $\text{PoPStat}_{\text{COVID-19}}$, a correlation coefficient of 0.81($p<0.0001$, 95% CI = [0.7457, 0.8546]) and 0.71($p<0.0001$, 95% CI = [0.6205, 0.7762]) was recorded for COVID-19 cases per million and deaths per million, as depicted in Fig. 3a and Fig. 3e from COVID-19. Therefore, the proposed $\text{PoPStat}_{\text{COVID-19}}$ correlates strongly with COVID-19 cases per million and the COVID-19 deaths per million. Next, it is necessary to compare the correlation results for other demographic variables: median age and population density,[13,27,28,29,30] to signify the superiority of the population pyramid-based distance metric. With median age as the independent variable, a correlation coefficient of 0.78($p<0.0001$, 95% CI = [0.7235, 0.8411]) and 0.68($p<0.0001$, 95% CI = [0.592, 0.7576]) was recorded per million cases and deaths. Then, a correlation coefficient of 0.10($p=0.1946$, 95% CI = [-0.0524, 0.2519]) and 0.01($p=0.8793$, 95% CI = [-0.142, 0.1654]) was recorded with population density and 0.64($p<0.0001$, 95% CI = [0.536, 0.7204]) and 0.42($p<0.0001$, 95% CI = [0.2792, 0.5347]) with GDP per capita per million cases and deaths, respectively. Thus, when we compare the correlation results of the proposed $\text{PoPStat}_{\text{COVID-19}}$ with other demographic variables such as median age and population density, the superiority of the $\text{PoPStat}_{\text{COVID-19}}$ metric is evident.

Moreover, when the $\text{PoPStat}_{\text{COVID-19}}$ was applied to test the correlation to all-cause mortality, a weak correlation coefficient of 0.34 was significantly reduced compared to the correlation to COVID-19 mortality and severity. The correlation dichotomy conveys that the observed correlation is statistically significant for COVID-19 cases and deaths. Furthermore, in Fig. 4, the effect of different aspects of the population pyramid on COVID-19 has been illustrated. In Fig. 4, the origin was centred around the normalized mean of the COVID-19 deaths per million and $\text{PoPStat}_{\text{COVID-19}}$. For example, COVID-19 had a higher impact in countries with high life expectancy, shown in the upper right quadrant. Consequently, COVID-19 was severe in countries with regressive population pyramids with a higher percentage of elderly populations vulnerable to COVID-19 and had higher fatalities, as opposed to countries with progressive population pyramids with fewer percentages in those ages, which had fewer percentages of vulnerable groups. The positive correlation recorded with the proposed PoPStat$_{\text{COVID-19}}$ and GDP per capita validates this assertion as there is a link between the population distribution and the economic standard of the countries, as this reflects the better conditions of living as observed by the upper quadrant's regressive population pyramids and higher GDP per capita.

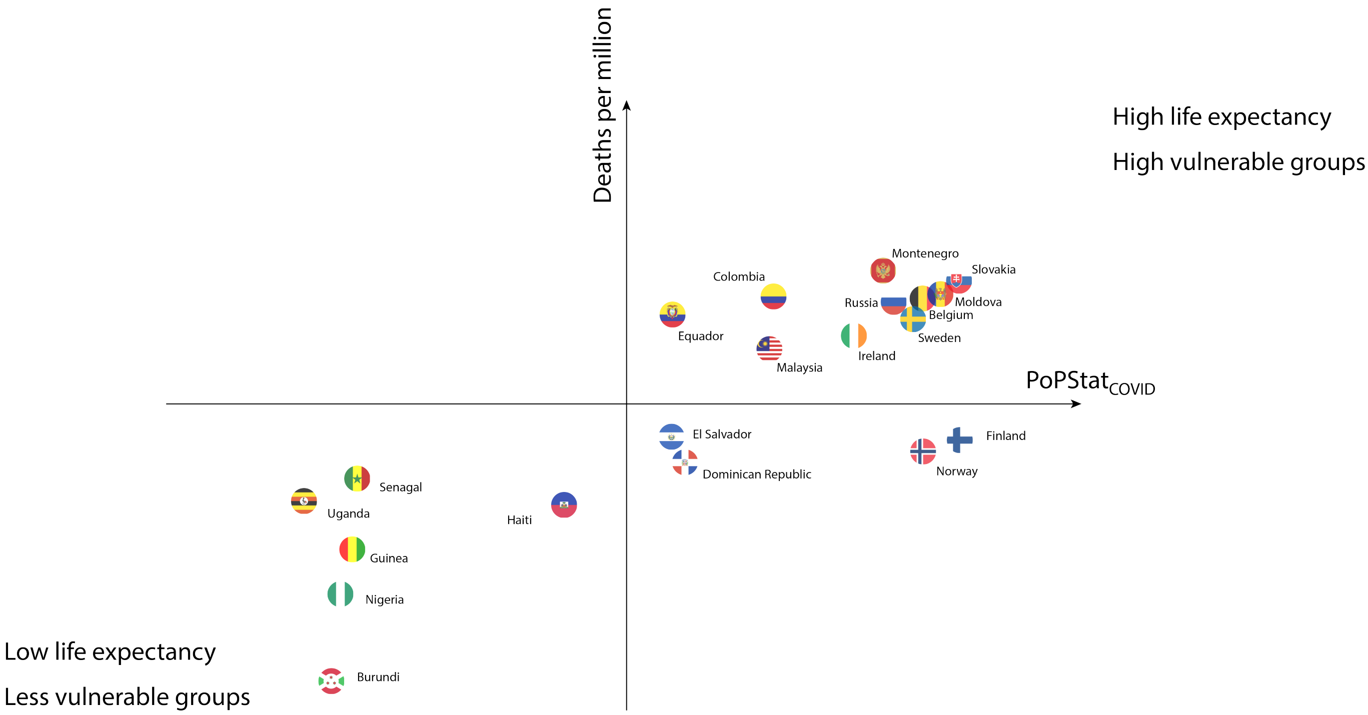

Fig. 4. Quadrant placement for selected countries according to the variation with the PoPStat$_{\text{COVID-19}}$ for COVID-19.

4. Discussion

The population pyramid of a country is a descriptive demographic statistic. It is an agglomeration of various socioeconomic factors of a country. Therefore, it is plausible that the population pyramid may be correlated with the impact of the COVID-19 pandemic, which is dependent on numerous socioeconomic factors. This work proposes a mechanism to generate a statistical measure highly correlated with COVID-19 severity through the population pyramid. Thereby, PoPStat$_{\text{COVID-19}}$ latches on to those attributes of the population pyramid that influence disease severity.

Therefore, this work provides yet another quantitative justification that the socioeconomic condition is linked to the impact of COVID-19. This methodology will be vital to earmark public health initiatives and other strategies and optimally plan the roll-out of such strategies, and valuable when dealing with a future public health emergency by providing a convenient tool for orienting policies and strategies, particularly with constrained resources. The same analysis can be performed at the state or provincial level but also for dynamic response to pandemics: attention or funding allocation for control and mitigation.

The population pyramid shape embodies age statistics and other influencing statistics such as quality of life, economic and healthcare, and education system. The PoPStat$_{\text{COVID-19}}$ quantifies the shape of the pyramid to distill the effect of these factors to a scalar variable given a parameter such as COVID-19 mortality. The transformation of the multidimensional population pyramid to a scalar parameter optimally distills the most influential aspects of the population pyramid to the disease severity. Consequently, a much higher correlation compared to existing metrics can be observed. When considering the population pyramid of Niger, the country has a strict progressive structure where the frequency of each age interval in the population histogram decreases monotonically. Since Niger has the lowest median age and is also the reference point, the proposed $\text{PoPStat}_{\text{COVID-19}}$ can reproduce inferences from 'median age'. Furthermore, as it is possible to analyze the shape transformation of the pyramid as severity increases, it enables the identification of the underlying socioeconomic factors that influence increased vulnerability in population groups.

When considering median age and the proposed $\text{PoPStat}_{\text{COVID-19}}$, there is a direct correlation between these variables as the two parameters are different statistical measures of the same probability distribution. Though the median is an absolute measure and has the luxury of not requiring a reference, the median could lead to a false impression, as explained in section 2.3. In most studies, the median age has been used as the independent variable since it is already in a scalar form. However, the median is a standard statistical measure which does not incorporate the shape of the population pyramid and is not specifically optimized to latch on to those components in the population pyramid that are maximally correlated with the COVID-19 or any such specific cause. This specificity is the reason for the high correlation and enables the identification of specific shape changes that affect the severity.

The positive correlation recorded with COVID-19 cases and deaths is primarily ascribed to high percentages of vulnerable groups in the population, especially the elderly, resulting from higher life expectancy from better living standards. However, higher occupancy of the working-age population as a characteristic of regressive population pyramid shapes contributes to the transmission of COVID-19. Because the population pyramid has a significant relationship with the pandemic, policymakers could benchmark containment strategies of other countries with similar age demographics. Moreover, the proposed $\text{PoPStat}_{\text{COVID-19}}$ provides a mathematical foundation to identify countries with similar factors that influence the severity of COVID-19 subsumed in the population pyramid shape.

Though $\text{PoPStat}_{\text{COVID-19}}$ was applied to study the correlation between COVID-19 and the population pyramid, applying $\text{PoPStat}_{\text{COVID-19}}$ to other mortality causes is warranted and only requires appropriate information about the population pyramid and the disease mortality data. This has been studied in our successive work. Also, further optimization could be introduced to the population pyramid to find an optimal weight vector that generates the relative impact from each age group.

4.1. Contributors

YR conceptualized and designed the mechanism with support from HW, GI, RG, VH, and PE. YR and GI curated the data and carried the analysis with support from JE, SD, and AR. YR, RG, MP, SY, and GT visualised the results for the manuscript. YR compiled the manuscript with help and revisions from RG, PE, VH, MP, SY, and GT. All authors commented on the manuscript and approved the final version. All authors had full access to all the data in the study and accept responsibility to submit for publication.

5. Declaration of interests

We declare no competing interests.

References

[1] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al., "Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study," *The lancet*, vol. 395, no. 10229, pp. 1054–1062, 2020.

[2] CDCx COVID, Response Team, CDC COVID, Response Team, CDC COVID, Response Team, Stephanie Bialek, Ellen Boundy, Virginia Bowen, Nancy Chow, et al., "Severe outcomes among patients with coronavirus disease 2019 (covid-19)—united states, february 12–march 16, 2020," *Morbidity and mortality weekly report*, vol. 69, no. 12, pp. 343, 2020.

[3] Chaomin Wu, Xiaoyan Chen, Yanping Cai, Xing Zhou, Sha Xu, Hanping Huang, Li Zhang, Xia Zhou, Chunling Du, Yuye Zhang, et al., "Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china," *JAMA internal medicine*, vol. 180, no. 7, pp. 934–943, 2020.

[4] Ruchong Chen, Wenhua Liang, Mei Jiang, Weijie Guan, Chen Zhan, Tao Wang, Chunli Tang, Ling Sang, Jiaxing Liu, Zhengyi Ni, et al., “Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in china,” *Chest*, vol. 158, no. 1, pp. 97–105, 2020.

[5] Edward Livingston and Karen Bucher, “Coronavirus disease 2019 (covid-19) in italy,” *Jama*, vol. 323, no. 14, pp. 1335–1335, 2020.

[6] Lin Fu, Jun Fei, Hui-Xian Xiang, Ying Xiang, Zhu-Xia Tan, Meng-Die Li, Fang-Fang Liu, Hong-Yan Liu, Ling Zheng, Ying Li, et al., “Influence factors of death risk among covid-19 patients in wuhan, china: a hospital-based case-cohort study,” *MedRxiv*, 2020.

[7] Dimple D Rajgor, Meng Har Lee, Sophia Archuleta, Natasha Bagdasarian, and Swee Chye Quek, “The many estimates of the covid-19 case fatality rate,” *The Lancet Infectious Diseases*, vol. 20, no. 7, pp. 776–777, 2020.

[8] Gerard Joseph Abou Jaoude, Ines Garcia Baena, Peter Nguhiu, Andrew Siroka, Tom Palmer, Lara Goscé, Kasim Allel, Edina Sinanovic, Jolene Skordis, and Hassan Haghparast-Bidgoli, “National tuberculosis spending efficiency and its associated factors in 121 low-income and middle-income countries, 2010–19: a data envelopment and stochastic frontier analysis,” *The Lancet Global Health*, vol. 10, no. 5, pp. e649–e660, 2022.

[9] Neff Walker, Jennifer Bryce, and Robert E Black, “Interpreting health statistics for policymaking: the story behind the headlines,” *The Lancet*, vol. 369, no. 9565, pp. 956–963, 2007.

[10] Carla AbouZahr, Sam Adjei, and Churnrurtai Kanchanachitra, “From data to policy: good practices and cautionary tales,” *The Lancet*, vol. 369, no. 9566, pp. 1039–1046, 2007.

[11] Michele J Gelfand, Joshua Conrad Jackson, Xinyue Pan, Dana Nau, Dylan Pieper, Emmy Denison, Munqith Dagher, Paul AM Van Lange, Chi-Yue Chiu, and Mo Wang, “The relationship between cultural tightness–looseness and covid-19 cases and deaths: a global analysis,” *The Lancet Planetary Health*, vol. 5, no. 3, pp. e135–e144, 2021.

[12] Annabel X Tan, Jessica A Hinman, Hoda S Abdel Magid, Lorene M Nelson, and Michelle C Odden, “Association between income inequality and county-level covid-19 cases and deaths in the us,” *JAMA Network Open*, vol. 4, no. 5, pp. e218799–e218799, 2021.

[13] Arunava Bhadra, Arindam Mukherjee, and Kabita Sarkar, “Impact of population density on covid-19 infected and mortality rate in india,” *Modeling Earth Systems and Environment*, vol. 7, no. 1, pp. 623–629, 2021.

[14] Xue-Qiang Wang, Ge Song, Zheng Yang, Ren-Jie Chen, Yi-Li Zheng, Hao-Yu Hu, Xuan Su, and Pei-Jie Chen, “Association between ageing population, median age, life expectancy and mortality in coronavirus disease (covid-19),” *Aging (Albany NY)*, vol. 12, no. 24, pp. 24570, 2020.

[15] Mariachiara Di Cesare, Young-Ho Khang, Perviz Asaria, Tony Blakely, Melanie J Cowan, Farshad Farzadfar, Ramiro Guerrero, Nayu Ikeda, Catherine Kyobutungi, Kelias P Msyamboza, et al., “Inequalities in non-communicable diseases and effective responses,” *The Lancet*, vol. 381, no. 9866, pp. 585–597, 2013.

[16] J Ties Boerma and Sally K Stansfield, “Health statistics now: are we making the right investments?,” *The Lancet*, vol. 369, no. 9563, pp. 779–786, 2007.

[17] Stuart Gietel-Basten and Tomas Sobotka, “Trends in population health and demography,” *The Lancet*, vol. 398, no. 10300, pp. 580–581, 2021.

[18] Jessica Bernman, “beijing china attempts to soften its one-child policy,” *The Lancet*, vol. 353, no. 9152, pp. 567, 1999.

[19] Megan Tatum, “China’s population peak,” *The Lancet*, vol. 399, no. 10324, pp. 509, 2022.

[20] Yi Zeng and Therese Hesketh, “The effects of china’s universal two-child policy,” *The Lancet*, vol. 388, no. 10054, pp. 1930–1938, 2016.

[21] Ibrahim Abubakar, Sarah L Dalglish, Blake Angell, Olutobi Sanuade, Seye Abimbola, Aishatu Lawal Adamu, Ifedayo MO Adetifa, Tim Colbourn, Afolabi Olaniyi Ogunlesi, Obinna Onwujekwe, et al., “The lancet nigeria commission: Investing in health and the future of the nation,” *The Lancet*, vol. 399, no. 10330, pp. 1155–1200, 2022.

[22] Zaki Hassan-Smith, Wasim Hanif, and Kamlesh Khunti, “Who should be prioritised for covid-19 vaccines?,” *The Lancet*, vol. 396, no. 10264, pp. 1732–1733, 2020.

[23] The Lancet, “Global health: time for radical change?,” *Lancet (London, England)*, vol. 396, no. 10258, pp. 1129, 2020.

[24] The Lancet Healthy Longevity, “The lancet healthy longevity: Health for all, for longer,” *The Lancet Healthy Longevity*, vol. 1, no. 1, pp. e1, 2020.

[25] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Unwin, Helen Coupland, T Mellan, Harisson Zhu, Tresnia Berah, J Eaton, P Perez Guzman, et al., “Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries,” 2020.

[26] Graziano Onder, Giovanni Rezza, and Silvio Brusaferro, “Case-fatality rate and characteristics of patients dying in relation to covid-19 in italy,” *Jama*, vol. 323, no. 18, pp. 1775–1776, 2020.

[27] Zhibin Sun, Hui Zhang, Yifei Yang, Hua Wan, and Yixiang Wang, "Impacts of geographic factors and population density on the covid-19 spreading under the lockdown policies of china," *Science of The Total Environment*, vol. 746, pp. 141347, 2020.

[28] Hamit Coşkun, Nazmiye Yıldırım, and Samettin Gündüz, "The spread of covid-19 virus through population density and wind in turkey cities," *Science of the Total Environment*, vol. 751, pp. 141663, 2021.

[29] Nadjat Kadi and Mounia Khelfaoui, "Population density, a factor in the spread of covid-19 in algeria: statistic study," *Bulletin of the National Research Centre*, vol. 44, no. 1, pp. 1–7, 2020.

[30] Paulo R Martins-Filho, "Relationship between population density and covid-19 incidence and mortality estimates: A county-level analysis," *Journal of Infection and Public Health*, vol. 14, no. 8, pp. 1087, 2021.