Mortality due to COVID-19 in different countries is associated with their demographic character and prevalence of autoimmunity

Bithika Chatterjee[1], Rajeeva Laxman Karandikar[2] and Shekhar C. Mande[1,3,*]

[1]National Centre for Cell Science, NCCS Complex, Ganeshkhind, Pune 411 007, India
[2]Chennai Mathematical Institute, H1 SIPCOT IT Park, Siruseri 603 103, India
[3]Council of Scientific and Industrial Research, 2, Rafi Marg, Anusandhan Bhavan, New Delhi 110 001, India

In the first few months of its deadly spread across the world, COVID-19 mortality has exhibited a wide range of variability across different nations. In order to explain this phenomenon empirically, we have taken into consideration all publicly available data for 106 countries on parameters like demography, prevalence of communicable and non-communicable diseases, BCG vaccination status, sanitation parameters, etc. We used multivariate linear regression models and found that the incidence of communicable diseases correlated negatively. Demography, improved hygiene and higher incidence of autoimmune disorders correlated positively with COVID-19 mortality and they were among the most plausible factors to explain COVID-19 mortality compared to GDP of the nations.

Keywords: Autoimmunity, epidemiology, hygiene, mortality, risk factors, COVID-19.

HIGHER GDP and improved human development index (HDI) have led to improved sanitation and consequently reduction in the load of communicable disease burden in many countries[1]. Interestingly, the non-communicable diseases burden now occupies areas of major concern in the higher HDI countries. Thus, a distinct correlation between HDI status of a country and the prevalence of specific diseases has emerged in recent history[2].

About half of the world's population lives in low- and low–middle-income countries[3]. Typically access to healthcare facilities, hygiene and sanitation is poor in these countries and is often believed to be the contributing factor of higher incidence of communicable diseases. Thus, it is not unexpected if infectious disease pandemics, such as that due to SARS-CoV-2, have catastrophic consequences in the low and low–middle-income countries. On the contrary, disease prevalence and case fatality ratio (CFR) during the COVID-19 pandemic show a contrasting opposite trend in the low and low–middle income countries when compared to that of the high-income countries. It is fascinating to explore reasons that would explain higher prevalence and deaths due to COVID-19 among richer nations.

An interesting relationship between severity of COVID-19 outcome and several non-communicable disorders such as diabetes, hypertension and cardiovascular disorders has been noted[4–6]. However, the non-communicable disease burden of a country and its apparent relation to CFR due to COVID-19 has not been explored in detail yet. As a large population with these disorders lives in the high HDI countries, co-morbidities with diabetes, hypertension, cardiovascular disorders and respiratory disorders might have emerged as important determinants of CFR due to COVID-19 in these countries. Similarly, people above the age of 65 years are believed to be at a greater risk[7], with the percentage of such people being significantly more in the higher HDI countries[8]. Thus, co-morbidities with non-communicable diseases and the fraction of people living above the age of 65 years being skewed towards the high-income countries, offer possible explanations to the perplexing observation of CFR dichotomy among nations.

Among the many parameters that lead to non-communicable diseases, autoimmunity occupies an important place. Interestingly, a correlation between autoimmune disorders and HDI has been proposed[9]. Since one of the primary manifestations of COVID-19 has been a severe autoimmune reaction in the later phase of the disease[10], susceptibility to autoimmunity in SARS-CoV-2 infection is a possibility to be considered. One of the reasons for rising prevalence in autoimmune disorders in the higher income countries is considered to be related to the 'hygiene hypothesis'[11]. This hypothesis postulates that exposure to pathogens early in life protects people from allergic diseases later on. Moreover, improvement in hygiene practices such as better sanitation, availability of safe drinking water, hand-washing facilities, etc. reduces the impact of communicable diseases. On the contrary, such a reduction in the exposure to infectious agents might be related to higher prevalence of autoimmune disorders[12]. We wanted to explore if epidemiological data supports correlation between prevalence of autoimmune diseases and sanitation parameters, and how much of the

---

*For correspondence. (e-mail: shekhar@csir.res.in)

variation in deaths per million due to COVID-19 is explained by the same. We have chosen more than 25 parameters to find the correlation between these and CFR, if any. These include GDP, HDI, prevalence of various diseases, demographic parameters, various sanitation parameters, etc. Our findings reveal that the apparent correlation between GDP of a nation and COVID-19-related deaths, or that between GDP and CFR, is a manifestation of other parameters.

Methods

The country-wise COVID-19 deaths per million data was collected from *Our World in Data*[13] with the original source: European Centre for Disease Prevention and Control (ECDC) on 29 June 2020. We chose deaths per million compared to the confirmed reported cases, as it gives a more reliable parameter to assess the current pandemic. The parameters on obesity (BMI > 30, age and sex-standardized) were sourced from the World Health Organization[14], whereas parameters on GDP per capita, population density, age profile, gender ratio, HDI and urban population percentage were sourced from World Bank[15].

We also collected all possible variables that quantify the cleanliness parameters in countries. This included the percentage of access to handwashing facilities, access to basic drinking water, availability of basic sanitation, prevalence of open defecation, safe drinking and safe sanitation. Most of these variables were available through WHO and UNICEF water and sanitation surveillance projects[16]. The prevalence percentages of different autoimmune disorders such as multiple sclerosis, type-1 diabetes mellitus, psoriasis, rheumatoid arthritis, asthma, and communicable diseases such as schistosomiasis, onchocerciasis, lymphatic filariasis, ascaris, hook worm, malaria, tuberculosis, dengue, upper respiratory infections, lower respiratory infections and H influenzae type-B meningitis for the year 2017 (age- and sex-standardized) were downloaded from the Global Burden of Disease[17]. The diseases were classified into two broad categories as communicable and non-communicable.

The mean BCG vaccination coverage in percentage was downloaded from WHO[18]. The average mean coverage was calculated for all years. The 2017 GDP gross domestic products per capita based on purchasing power parity converted to 2011 international dollars and HDI 2018 was used for analysis. For details of sources of each variable see Supplementary Material 1.

We further filtered countries that had at least 10 deaths in total and population size of at least 100,000. Moreover, we restricted consideration of the countries to have at least four deaths/million. For countries missing corresponding values for our variables, we either imputed them with values smaller than the smallest values in the other countries, or discarded them from further analyses. For

the missing values in variables related to drinking water, sanitation and wash, we predicted them using a regression model as there was a high correlation amongst each of them. For missing values of HDI, we used GDP data to impute them. After filtration and imputation of missing values, there were 106 countries with values corresponding to more than 25 variables. For the same countries we downloaded the deaths per million up to 60 days from the onset of the first death occurring in each country on 7 August 2020. Subsequently, we also added the 90 and 120 days interval data collected on 16 September 2020 to account for the difference in the stage of the epidemic and lockdown restrictions of each country. It is customary to consider log of GDP instead of GDP for analyses in many disciplines to bring it closer to normal distribution. We also considered log of deaths per million (LDM) instead of deaths per million for all our calculations.

We compared individual (Pearson) correlation coefficients of different variables with the LDM on 29 June 2020 as well as those for 60, 90 and 120 days interval. The variables were compared with log (GDP) and HDI as well. This helped in assessing the important variables to be used in the linear regression model. The country-wise data on various parameters: GDP, HDI, sanitation parameters, prevalence of various communicable and non-communicable diseases were clubbed into development variables (GDP, HDI), demographics (percentage of older population, urban population and obesity among adults), sanitation (percentage of access to handwashing facilities, basic drinking, basic sanitation, no open defecation, safe drinking and safe sanitation), communicable diseases (schistosomiasis, onchocerciasis, lymphatic filariasis, ascaris, hook worm, malaria, tuberculosis, dengue, upper respiratory infections, lower respiratory infections and H influenzae type-B meningitis) and non-communicable diseases (multiple sclerosis, type-1 diabetes mellitus, psoriasis, rheumatoid arthritis and asthma).

Statistical analysis

To explore which of these combined variables are important, we employed multivariate linear regression model using R with several combinations of variables for a better understanding of the epidemiology of COVID-19. LDM of both scenarios was assessed for its association with the predictor/explanatory variables grouped as described above into development variables, demographic variables, sanitation, tropical diseases and autoimmune disorders. Our final objective was to select the best combination of these groups of variables that would yield the largest adjusted-R^2 value.

Results

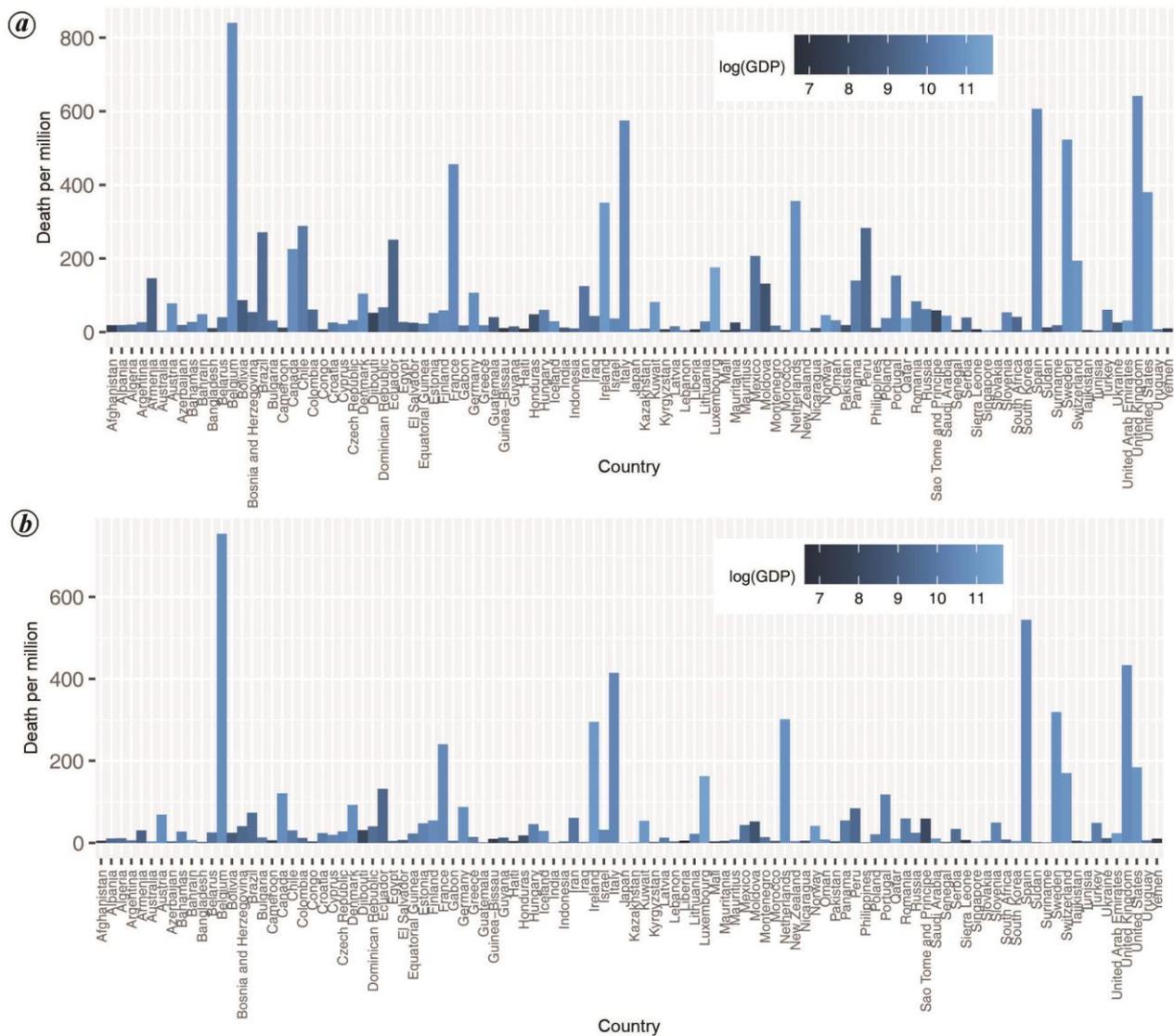Analysis was carried out on total 10,112,754 cases 501,562 deaths as on 29 June 2020, as well as with

Figure 1. Number of deaths per million inhabitants in each country due to COVID-19 (*a*) up to 29 June 2020 and (*b*) 60 days after the onset of the first death. Colour gradient indicates the country's log (GDP).

60-day interval data as on the 7 August 2020. Further, 90 and 120 days interval data as on 16 September 2020 were also included. Data from well-known corona virus global deaths tracking sources[19–21] shows that more than 70% COVID-19 mortality has occurred in high-income countries like Italy, Spain, UK, France and USA[22]. The present analysis also reveals a similar trend (Figure 1 *a* and *b*). To explore the dichotomy of rich and poor countries in relation to COVID-19 deaths, we considered several potential explanatory parameters. These included percentage of older population, population density, incidence of communicable and non-communicable diseases, etc. The data shows a positive correlation between log GDP and LDM (Figure 2, Supplementary Table 1 and Figure 1 *b*), which reinforces the fact that richer countries are being most burdened by the disease.

Demographics

Among the many distinctive features of higher GDP countries are its ageing population and a larger fraction of urbanization. We observed that while proportion of population above 65 years, obesity and percentage of urban population are positively correlated with COVID-19 deaths (Figure 2, Supplementary Table 1 and Figure 1 *c*, *e* and *g*), the gender ratio and population density have negligible negative correlation (Figure 2, Supplementary Table 1, Figure 1 *d* and *f*). Interestingly, while at an individual level gender does seem to have an impact on COVID-19 related death as shown in numerous studies[23,24], at the aggregate level the gender ratio varies little across countries. Likewise, urbanization has an impact while population density does not seem to have an
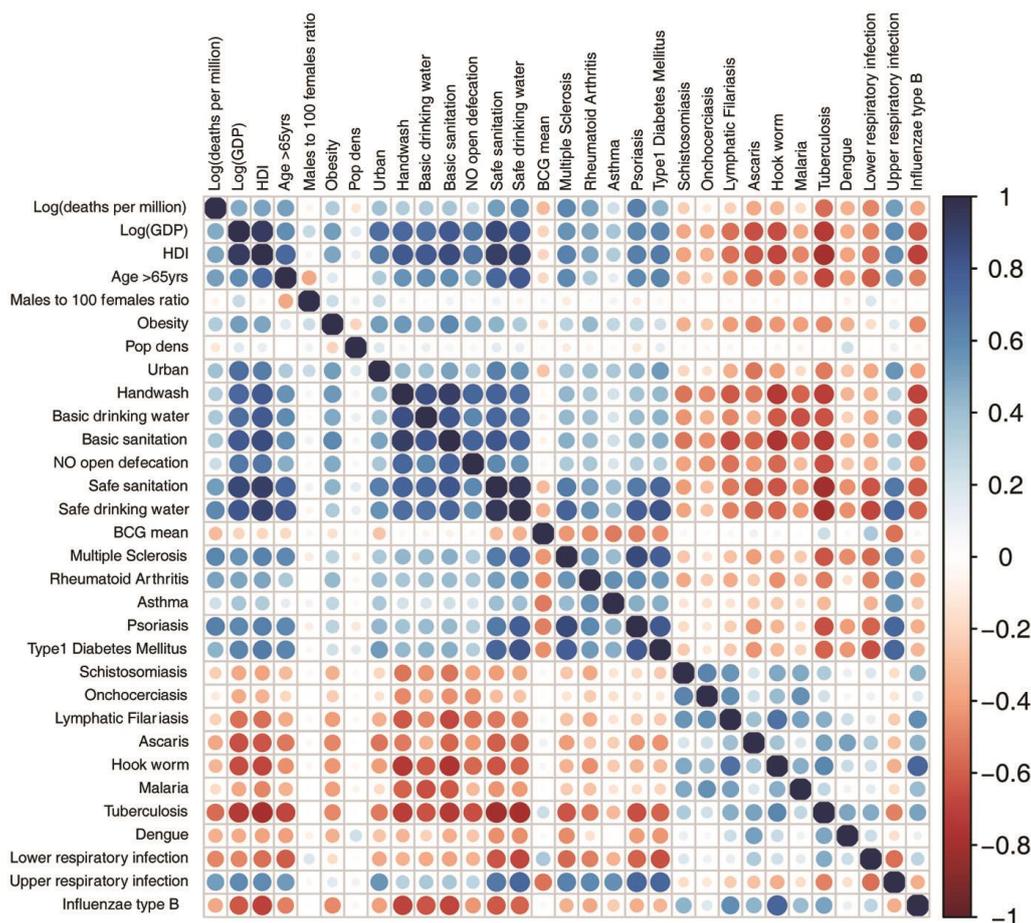
Figure 2. Plot showing correlation matrix of all variables with each other. The colour key is depicted as –1 (red) showing inverse correlation, 0 showing no correlation, and +1 (blue) showing maximum positive correlation. The size of the circles ranges from small (low correlation) to large (high correlation). Here log (deaths per million) is from the 60 days interval data; $N = 106$.

impact. We observed a positive correlation between LDM and percentage of population over 65 years, which is expected from the reports showing that the older population is more susceptible to the virus[7]. Thus, the overall demographic parameters yield a positive correlation with COVID-19 deaths per million (Figure 3 a).

Safe sanitation

Lack of sanitation and poor hygiene practices are known to be responsible for higher communicable disease burden in the low GDP countries[1]. It is therefore reasonable to expect that parameters describing safe sanitation and safe drinking water will be negatively correlated with COVID-19 deaths. Surprisingly, on the contrary we found that different sanitation parameters were positively correlated with the COVID-19 outcome (Figure 3 b). It is therefore perplexing to note positive correlation of sanitation parameters with the COVID-19 CFR (Figure 2, Supplementary Table 1 and Figure 1 h–m).

Communicable diseases

Several studies have reported the protective role of parasitic/bacterial infections in bolstering human immunity, also referred to as immune training[25,26]. We observed that the prevalence of communicable diseases such as malaria and tuberculosis as well as parasitic diseases such as schistosomiasis, onchocerciasis had a weak negative correlation with LDM (Figure 2; Supplementary Table 1 and Figure 1 t–z). Upper respiratory diseases, showed positive correlation, which may be indicative of hyperactive, immune response in individuals suffering from such disorders (Figure 2; Supplementary Table 1). These intuitive observations related to communicable diseases with respect to COVID-19 deaths are suggestive of the role of 'immune training' on the final disease outcome.

BCG vaccination

Recent studies have suggested that the countries in which majority population was immunized by BCG vaccine had
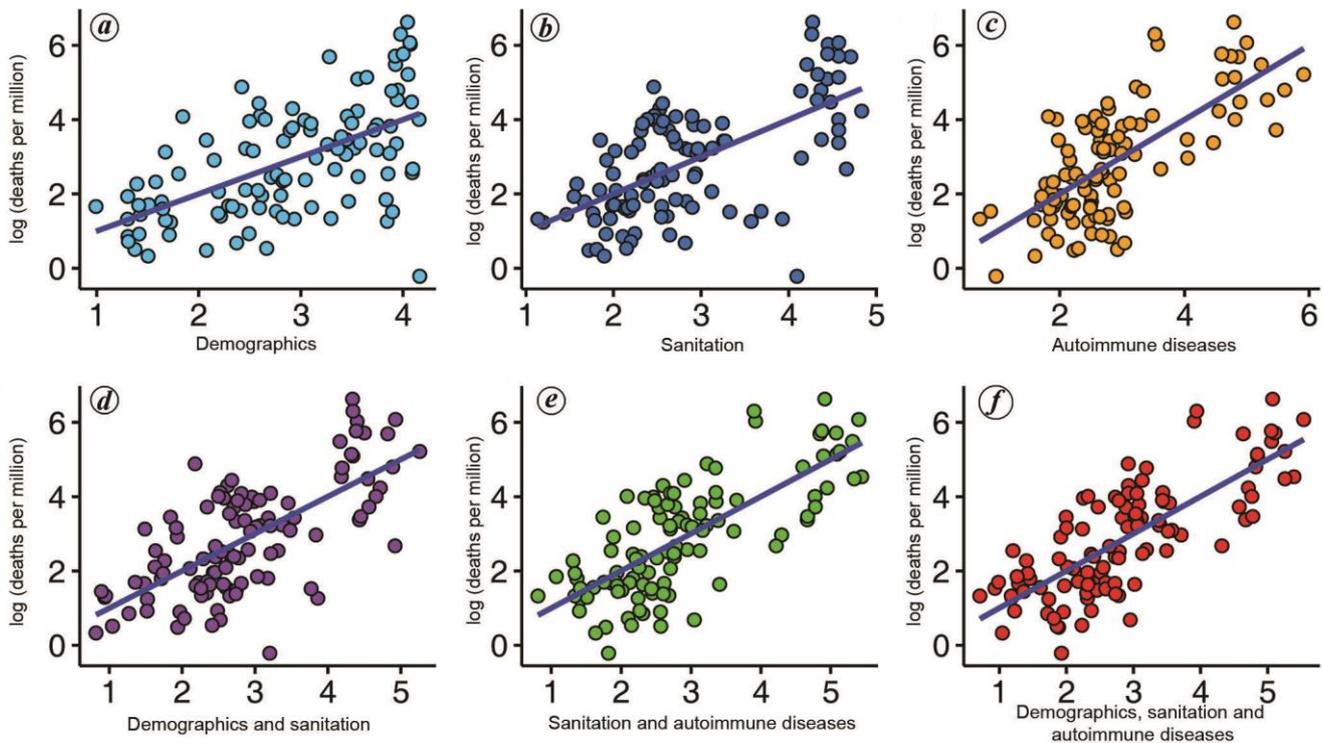
Figure 3. Actual values of log (deaths per million) 60 days after the onset of the first death or LDM plotted against their predicted values. The explanatory variables have been grouped, as described in the text, into (*a*) demographics, (*b*) sanitation, (*c*) autoimmune diseases, (*d*) demographics and sanitation, (*e*) sanitation and autoimmune diseases, (*f*) demographics, sanitation and autoimmune diseases. Regression analysis was done with LDM as the dependent variable and combinations of different variables; $N = 106$. Each dot represents a country. The R^2 of each graph is depicted in Table 1.

lower LDM due to trained immunity[27,28]. We found negligible correlation between BCG vaccination in different countries and LDM (Figure 2, Supplementary Table 1 and Figure 1 *n*). A possible explanation for this difference could be due to the fact that comparisons in previous studies were made with socially similar nations, while we considered a diverse group of countries. Moreover, the distribution showing two distinct clusters for BCG vaccination (Supplementary Figure 1 *n*), and with no apparent correlation within the clusters also makes it difficult to draw any conclusions on the effect of BCG vaccination on CFR at the population level. It is also pertinent to note that we found significant negative correlation between mean BCG vaccination of a country and prevalence of autoimmune disorders (Figure 2).

Autoimmunity

One of the manifestations of better hygiene and safe sanitation practices in high GDP countries is the increased incidence in autoimmune disorders[29]. It has been postulated that better hygiene practices could lower a person's immunity and make him/her susceptible to autoimmune diseases. We therefore tested if any of the autoimmune diseases showed an association with the LDM variable.

Interestingly, we found a strong positive correlation with multiple sclerosis, type-1 diabetes mellitus, rheumatoid arthritis and psoriasis, and a weaker correlation of LDM with asthma (Figures 2 and 3 *c*, Supplementary Table 1 and Figure 1 *o–s*).

Regression analysis

We combined many related factors such as demographic parameters, sanitation, communicable and non-communicable diseases. These combinations were arrived at after calculating correlations among all the parameters considered (Figure 2; Supplementary Table 1). In order to find whether development indices like HDI and GDP of countries had an impact on LDM, we employed regression models using combinations of explanatory factors outlined above and checked the correlation between the residues to GDP and HDI. The residues are the difference in actual values of the dependent variable (LDM) and its predicted values predicted from predictor variables.

After testing various combinations of these groups, we found that demographic variables, sanitation parameters and incidence of autoimmune disorders together explained most of the observed variations in LDM and development variables become insignificant once the

Table 1. R – the correlation coefficient of combinations of variables with log (deaths/million) LDM for 60, 90 and 120 days interval and 29 June based on multiple regression analysis as well as correlation of the residues with log (GDP) and HDI. Also shown is the adjusted R^2; $N = 106$

Broad category variables	Explanatory variables	Duration	R with LDM as dependent variable	Adjusted R^2	R between residue and (log) GDP	R between residue and HDI
Demographics	Age >65 yrs, obesity, urban	29 June	0.49	0.22	0.04	0.01
		60 days	0.58	0.32	0.02	0.00
		90 days	0.50	0.22	0.03	0.00
		120 days	0.40	0.14	0.02	0.01
Sanitation	Handwash, basic drinking water, basic sanitation, NO open defecation, safe sanitation, safe drinking water	29 June	0.54	0.25	0.11	0.03
		60 days	0.63	0.36	0.05	0.02
		90 days	0.56	0.27	0.10	0.03
		120 days	0.47	0.18	0.08	0.01
Autoimmune diseases	Multiple sclerosis, rheumatoid arthritis, asthma, psoriasis, type-1 diabetes mellitus	29 June	0.64	0.39	0.15	0.11
		60 days	0.70	0.46	0.12	0.12
		90 days	0.66	0.40	0.12	0.08
		120 days	0.61	0.34	0.10	0.07
Demographics and sanitation	Age >65 yrs, obesity, urban population, handwash, basic drinking water, basic sanitation, NO open defecation, safe sanitation, safe drinking water	29 June	0.63	0.35	0.06	0.00
		60 days	0.69	0.42	0.03	–0.01
		90 days	0.64	0.35	0.05	–0.01
		120 days	0.57	0.27	0.02	–0.01
Sanitation and autoimmune diseases	Handwash, basic drinking water, basic sanitation, NO open defecation, safe sanitation, safe drinking water, multiple sclerosis, rheumatoid arthritis, asthma, psoriasis, type-1 diabetes mellitus	29 June	0.68	0.40	0.10	0.02
		60 days	0.75	0.50	0.03	0.00
		90 days	0.69	0.42	0.09	0.01
		120 days	0.64	0.35	0.07	0.01
Demographics, sanitation and autoimmune diseases	Age >65 yrs, obesity, urban, handwash, basic drinking water, basic sanitation, NO open defecation, safe sanitation, safe drinking water, multiple sclerosis, rheumatoid arthritis, asthma, psoriasis, type-1 diabetes mellitus	29 June	0.74	0.48	0.04	–0.03
		60 days	0.77	0.53	0.00	–0.04
		90 days	0.74	0.48	0.03	–0.03
		120 days	0.70	0.42	0.01	–0.03

effect of these explanatory variables was accounted for (Supplementary Table 2). More precisely, if one uses multiple regression to predict death rates, then the residues are uncorrelated with GDP and HDI. The R value for our final best model for 60 days interval was 0.77 and the adjusted-R^2 value was 0.53 (P value < 0.05). The correlation of the residue of this model with GDP and HDI 0.00 and –0.04 respectively (Table 1). Since during this time there was strict lockdown imposed in some countries, we also did the same analysis on 90 and 120 days interval data. The R value for 90 days interval was 0.74 and adjusted-R^2 was 0.48. Analysis on 120 days interval reduced R value to 0.70 and adjusted-R^2 value to 0.42 (Table 1). Although still significant, the reduced R^2 and R values as the number of days increases can be attributed to the lowering of deaths in countries whose first wave of

infection was completed as well as their better preparedness to fight the infection.

Discussion

Various corona virus global death tracking sources point towards an uneven distribution of COVID-19 deaths with richer countries having higher CFR[19–21,30]. In order to understand the reasons behind this paradoxical observation, we analysed publicly available data on various parameters. In this context, it is well known that when we have multiple factors correlated with the dependent variables and these factors are themselves mutually correlated, it is not possible to determine as to which of these may be the causal variables using purely statistical techniques. To address this, we could collect data from a

controlled experiment or look for data on individuals rather than countrywide aggregated data. However, in this instance many variables make sense only at the aggregate level, ruling out both these options. Moreover, trends observed at individual level can be opposite to that in the ecological or population level owing to the Simpson's paradox. In such challenging situations, it is desirable to use domain knowledge and identify the variables whose correlation with the dependent variables has a logical explanation. In the present context, older population having higher death rate can be explained biologically. Similarly, since COVID-19 spreads rapidly through person-to-person contact, population density is likely to have a big impact.

Our observation of the weak negative correlation of COVID-19 LDM with communicable diseases, and its positive correlation with incidence of autoimmune disorders in the high GDP countries is an indication of the interplay of host immunity and viral infection. As the parasite and bacterial disease burden is high in low- and low–middle-income countries, this can best be inferred by 'immune training' in the population of these countries due to chronic exposure to communicable diseases. It has been observed recently that patients susceptible to the corona virus exhibited impaired type-1 interferon activity[31]. Since initially the innate immunity acts upon the virus mediated by interferon activity, individuals with prior infections tend to cope with the virus better than those without exposure to pathogens[32]. The evidence for innate immunity playing a major role in susceptibility to the virus is available from the fact that asymptotic patients or patients with mild infection produced poor neutralizing antibodies[33]. Hence it can be speculated that viral susceptibility is established at the first line of defence known as innate immunity, much before the adaptive immunity. As mentioned earlier, the observed correlation between incidence of autoimmune disorders and LDM can be explained as COVID-19 seems to lead to severe autoimmune reaction. Paradoxically, better sanitation leads to poorer 'immune training' and thus could be leading to higher LDM. Thus, the observed correlation of demographics, autoimmune disorders and sanitation with LDM derives from the domain. The transfer rate for COVID-19 being high, we found a lower correlation between incidence of communicable diseases and LDM due to COVID-19 as compared to autoimmune disorders.

The correlation of GDP and HDI with the residue of each model was almost negligible (Supplementary Table 2). Both GDP and HDI seem to have a dominant role in determining the deaths per million. Yet both become insignificant when the linear regression models considered other notable and explainable confounding factors such as demography, sanitation and pre-COVID prevalence of autoimmune and tropical diseases. The parameters of sanitation and autoimmune diseases consistently contributed significantly to R^2, no matter which combination of variables was chosen. It has not escaped our attention that there could be other factors such as stage of the epidemic in a country, and lower reporting/testing in less-developed countries that could also affect the mortality numbers. However, statistical evidence provided in this study should give a head start to study the role of hyper immune reaction on COVID-19 susceptibility at individual patient's level.

We note that statistical analyses such as our work attempts to provide evidence of practices at a macroscopic scale. Individual-level variations among people or communities typically get masked in such analyses. Thus, although we provide a possible explanation based on sanitation practices for the CFR differences among economically stronger and weaker countries, this should not be inferred as our advocating a move towards weaker hygiene practices for handling future pandemics. Rather, this analysis opens up avenues to consider 'immune training' with possibilities of microbiome therapies to supplement improved hygiene and sanitation practices.

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*, 2018, **392**, 1789–1858.
2. Pan, H.-Y. *et al.*, National incidence of autoimmune liver diseases and its relationship with the human development index. *Oncotarget*, 2016, **7**, 46273–46282.
3. https://datahelpdesk.worldbank.org/knowledgebase/articles/90651-9-world-bank-country-and-lending-groups (accessed on 29 June 2020).
4. Li, B. *et al.*, Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin. Res. Cardiol.*, 2020, **109**, 531–538.
5. Yang, J. *et al.*, Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *Int. J. Infect. Dis.*, 2020, **94**, 91–95.
6. Emami, A., Javanmardi, F., Pirbonyeh, N. and Akbari, A., Prevalence of underlying diseases in hospitalized patients with COVID-19: a systematic review and meta-analysis. *Arch. Acad. Emerg. Med.*, 2020, **8**(1), e35.
7. UN, COVID-19 and older persons: a defining moment for an informed, inclusive and targeted response. United Nations for Ageing; https://www.un.org/development/desa/ageing/news/2020/05/covid19/ (accessed on 29 June 2020).
8. Roser, M., Human development index (HDI). *Our World in Data*, 2014; https://ourworldindata.org/human-developmentindex#:~:text=The%20Human%20Development%20Index%20(HDI)%20provides%20a%20single%20index%20measure,a%20decent%20standard%20of%20living (accessed on 29 June 2020).
9. Grosse, J., Hornstein, H., Manuwald, U., Kugler, J., Glauche, I. and Rothe, U., Incidence of diabetic ketoacidosis of new-onset type 1 diabetes in children and adolescents in different countries correlates with human development index (HDI): an updated systematic review, meta-analysis, and meta-regression. *Hormone Metab. Res.*, 2018, **50**, 209–222.

10. Galeotti, C. and Bayry, J., Autoimmune and inflammatory diseases following COVID-19. *Nature Rev. Rheumatol.*, 2020, **16**, 413–414.

11. Okada, H., Kuhn, C., Feillet, H. and Bach, J., The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clin. Exp. Immunol.*, 2010, **160**, 1–9.

12. Bach, J.-F., The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. *Nature Rev. Immunol.*, 2018, **18**, 105.

13. Roser, M., Ritchie, H., Ortiz-Ospina, E. and Hasell, J., Coronavirus pandemic (COVID-19). *Our World in Data*, 2020; https://ourworldindata.org/coronavirus (accessed on 29 June 2020).

14. https://www.who.int/data/gho (accessed on 29 June 2020).

15. https://data.worldbank.org/ (accessed on 29 June 2020).

16. https://www.who.int/data/gho/data/themes/water-sanitation-and-hygiene (accessed on 29 June 2020).

17. http://ghdx.healthdata.org/gbd-results-tool (accessed on 29 June 2020).

18. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/bcg (accessed on 29 June 2020).

19. WHO coronavirus disease (COVID-19) dashboard; https://covid19.who.int (accessed on 29 June 2020).

20. COVID-19 map. Johns Hopkins Coronavirus Resource Center, USA, https://coronavirus.jhu.edu/map.html (accessed on 29 June 2020).

21. COVID-19 situation update worldwide, as of 18 July 2020. European Centre for Disease Prevention and Control, https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases (accessed on 29 June 2020).

22. Schellekens, P. and Sourrouille, D., Tracking COVID-19 as cause of death: global estimates of relative severity, 2020.

23. Walter, L. A. and McGregor, A. J., Sex-and gender-specific observations and implications for COVID-19. *West. J. Emerg. Med.*, 2020, **21**, 507.

24. Jin, J.-M. *et al.*, Gender differences in patients with COVID-19: focus on severity and mortality. *Front. Public Health*, 2020, **8**, 152.

25. Netea, M. G., Quintin, J. and van der Meer, J. W. M., Trained immunity: a memory for innate host defense. *Cell Host Microbe*, 2011, **9**, 355–361.

26. Quinn, S. M. *et al.*, Anti-inflammatory trained immunity mediated by helminth products attenuates the induction of T cell-mediated autoimmune disease. *Front. Immunol.*, 2019, **10**.

27. Escobar, L. E., Molina-Cruz, A., and Barillas-Mury, C., BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). *Proc. Natl. Acad. Sci., USA*, 2020, **117**(30), 17720–17726.

28. Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19. *Science Advances*; https://advances.sciencemag.org/content/6/32/eabc1463 (accessed on 29 June 2020).

29. Bach, J.-F. and Chatenoud, L., The hygiene hypothesis: an explanation for the increased frequency of insulin-dependent diabetes. *Cold Spring Harbor Perspect. Med.*, 2012, **2**, a007799.

30. Schellekens, P. and Sourrouille, D. M., Covid-19 mortality in rich and poor countries: a tale of two pandemics? SSRN Scholarly Paper, Social Science Research Network, Rochester, NY, USA, 2020.

31. Hadjadj, J. *et al.*, Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*, 2020, **369**, 718–724.

32. Rouse, B. T. and Sehrawat, S., Immunity and immunopathology to viruses: what decides the outcome? *Nature Rev. Immunol.*, 2010, **10**, 514–526.

33. Long, Q.-X. *et al.*, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature Med.*, 2020, **26**, 1200–1204.