

1 **Title:** The mortality due to COVID-19 in different nations is associated with the demographic  
2 character of nations and the prevalence of autoimmunity.

3 **Authors:** Bithika Chatterjee<sup>1</sup>, Rajeeva Laxman Karandikar<sup>2, \*</sup> and Shekhar C. Mande<sup>1, 3, \*</sup>

4 1. National Centre for Cell Science, NCCS Complex, Ganeshkhind, Pune- 411007

5 2. Chennai Mathematical Institute, H1 SIPCOT IT Park, Siruseri- 603103

6 3. Council of Scientific and Industrial Research, 2, Rafi Marg, Anusandhan Bhavan, New  
7 Delhi- 110001

8

9 **\*Corresponding Author:**

10 <sup>3</sup>Shekhar C. Mande

11 Director General, Council of Scientific and Industrial Research (CSIR)

12 2, Rafi Marg, Anusandhan Bhavan

13 New Delhi 110001, INDIA

14 Email: [shekhar@csir.res.in](mailto:shekhar@csir.res.in)

15 And

16 <sup>2</sup>Rajeeva Laxman Karandikar

17 Chennai Mathematical Institute

18 H1, SIPCOT IT park

19 Padur PO, Siruseri 603103, INDIA

20 Email: [rlk@cmi.ac.in](mailto:rlk@cmi.ac.in)

21

22

23

24 **Abstract**

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

34

35 **Keywords:** SARS-CoV-2 ,Risk factors, Mortality, Epidemiology, Autoimmunity, Hygiene.

36

## 37 **Introduction**

38 Higher GDP and improved Human Development Index (HDI) has led to improved  
39 sanitation and consequently reduction in the load of communicable disease burden in many  
40 countries<sup>1</sup>. Interestingly, disease burden of non-communicable diseases now occupies areas of  
41 major concern in the higher HDI countries. Thus, a distinct correlation between HDI status of a  
42 country, and the prevalence of specific diseases has emerged in recent history<sup>2</sup>.

43 About half of the world's population lives in low and low middle income countries<sup>3</sup>.  
44 Typically access to healthcare facilities, hygiene and sanitation is poorer in these countries and is  
45 often believed to be the contributing factor of higher incidence of communicable diseases in  
46 these countries. Thus, it is not unexpected if infectious disease pandemics, such as that due to  
47 SARS-CoV-2, have catastrophic consequences in the low and low-middle income countries. Yet  
48 on the contrary, the disease prevalence and the Case Fatality Ratio (CFR) during the Covid-19  
49 pandemic shows a contrasting opposite trend in the low and low-middle income countries when  
50 compared to that of the high income countries. It is fascinating to explore reasons that would  
51 explain higher prevalence and deaths due to Covid-19 among richer nations.

52 An interesting relationship between severity of Covid-19 outcome and several non-  
53 communicable disorders such as diabetes, hypertension, cardiovascular disorders has been  
54 noted<sup>4-6</sup>. However, the non-communicable disease burden of a country and its apparent relation  
55 to CFR due to Covid-19 has not been explored in detail yet. As a large population with these  
56 disorders lives in the high HDI countries, co-morbidities with diabetes, hypertension,  
57 cardiovascular disorders and respiratory disorders might have emerged as important determinants  
58 of CFR due to Covid-19 in these countries. Similarly, people above the age of 65 are also  
59 believed to be at a greater risk<sup>7</sup>, with the percentage of such people being significantly more in

60 the higher HDI countries<sup>8</sup>. Thus, co-morbidities with non-communicable diseases and the  
61 fraction of people living above the age of 65, being skewed towards the high-income countries,  
62 offers possible explanations to the perplexing observation of CFR dichotomy among nations.

63 Among the many parameters that lead to non-communicable diseases, autoimmunity  
64 occupies an important place. Interestingly, a correlation between autoimmune disorders and HDI  
65 has been proposed<sup>9</sup>. Since, one of the primary manifestations of Covid-19 has been a severe  
66 autoimmune reaction in the later phase of the disease<sup>10</sup>, susceptibility to autoimmunity in SARS-  
67 CoV-2 infection is a possibility to consider. One of the reasons of rising prevalence in  
68 autoimmune disorders in the western countries has been proposed to be that related to the  
69 “hygiene hypothesis”<sup>11</sup>. The hypothesis postulates that exposure to pathogens early in life  
70 protects people from allergic diseases later in their lives. Moreover, improvement in hygiene  
71 practices such as better sanitation, availability of safe drinking water, hand washing facilities etc.  
72 reduces the impact of communicable diseases. On the contrary such a reduction to the exposure  
73 to infectious agents might be related to higher prevalence of autoimmune disorders<sup>12</sup>. We wished  
74 to explore if epidemiological data supports correlation between autoimmune diseases’  
75 prevalence, sanitation parameters and how much of the variation in death per million due to  
76 Covid-19 is explained by the same. We have chosen more than 25 parameters and attempted to  
77 find correlation between these and CFR, if any. These parameters include GDP, HDI,  
78 prevalence of various diseases, demographic parameters, various sanitation parameters etc. Our  
79 findings reveal that the apparent correlation between GDP of a nation and Covid-19 related  
80 deaths, or that between GDP and CFR, is a manifestation of other parameters.

81

82 **Methods**

83 The country-wise Covid-19 deaths per million data was collected from  
84 <https://ourworldindata.org/coronavirus><sup>13</sup> with the original source: European Centre for Disease  
85 Prevention and Control (ECDC) on the 29<sup>th</sup> June 2020. We chose deaths per million as compared  
86 to the confirmed reported cases as it gives a more reliable parameter to assess the current  
87 pandemic. The parameters on obesity (BMI>30, Age and sex standardized) was sourced from the  
88 WHO<sup>14</sup> while the parameters on GDP per capita, population density, age profile, gender ratio,  
89 HDI and the urban population percentage was sourced from the World Bank organization<sup>15</sup>.

90 We also collected all possible variables that quantify the cleanliness parameters in  
91 countries. This included the percentage of access to handwashing facilities, access to basic  
92 drinking water, availability of basic sanitation, prevalence of open defecation, safe drinking and  
93 safe sanitation. Most of these variables were available through WHO and UNICEF water and  
94 sanitation surveillance projects<sup>16</sup>. The prevalence percentages of different autoimmune disorders  
95 such as Multiple Sclerosis, Type1 Diabetes Mellitus, Psoriasis, Rheumatoid Arthritis, Asthma,  
96 and communicable diseases such as Schistosomiasis, Onchocerciasis, Lymphatic Filariasis,  
97 Ascaris, Hook worm, Malaria, Tuberculosis, Dengue, Upper respiratory infections, Lower  
98 respiratory infections and H influenzae type B meningitis for the year 2017(Age and sex  
99 standardized) were downloaded from the <http://ghdx.healthdata.org/gbd-results-tool><sup>17</sup>. The  
100 diseases were classified into two broad categories as communicable and non-communicable  
101 diseases.

102 The Mean BCG vaccination coverage in percentage was downloaded from WHO<sup>18</sup>. The  
103 average mean coverage was calculated for all years. The 2017 GDP gross domestic products per  
104 capita based on purchasing power parity converted to 2011 international dollars and Human

105 Development Index 2018 was used for analysis. For details of sources of each variable see  
106 Supplementary Dataset S1.

107 We further filtered countries that had at least 10 deaths in total and population size of at  
108 least 100,000. Moreover, we restricted consideration of the countries to have at least 4  
109 deaths/million. For the countries missing corresponding values for our variables, we either  
110 imputed them with values smaller than the smallest values in the other countries or dropped the  
111 countries from further analyses. For the missing values in variables related to drinking water,  
112 sanitation and wash we predicted them using a regression model as we noticed a high correlation  
113 amongst each other. For missing values of HDI, we used GDP data to impute values for HDI.  
114 After filtration and imputation of missing values we were left with 106 countries with values  
115 corresponding to more than 25 variables. For the same countries we downloaded the deaths per  
116 million up to 60 days from the onset of the first death occurring in each country on 7<sup>th</sup> August  
117 2020. Subsequently we also added the 90 days and 120 days interval data collected on 16<sup>th</sup>  
118 September 2020 to account for the difference in the stage of the epidemic and the lockdown  
119 restrictions of each country. It is customary to consider log of GDP instead of GDP for analyses  
120 in many disciplines to bring it closer to normal distribution. We also considered log of deaths  
121 per million instead of deaths per million for all our calculations.

122 We compared individual (Pearson) correlation coefficients of different variables with the  
123 log of deaths per million (LDM) on the 29<sup>th</sup> June as well as with the 60, 90 and 120 days interval  
124 LDM. The variables were compared with log (GDP) and HDI as well. This helped in assessing  
125 the important variables to be used in our linear regression model. The country-wise data on  
126 various parameters: GDP, HDI, sanitation parameters, prevalence of various communicable and  
127 non-communicable diseases were clubbed into development variables (GDP, HDI),

128 demographics (percentages of older population, urban population and obesity among adults),  
129 Sanitation (percentage of access to handwashing facilities, basic drinking, basic sanitation, no  
130 open defecation, safe drinking and safe sanitation), communicable diseases (Schistosomiasis,  
131 Onchocerciasis, Lymphatic Filariasis, Ascaris, Hook worm, Malaria, Tuberculosis, Dengue,  
132 Upper respiratory infections, Lower respiratory infections and H influenzae type B meningitis)  
133 and non-communicable diseases (Multiple Sclerosis, Type1 Diabetes Mellitus, Psoriasis,  
134 Rheumatoid Arthritis and Asthma).

### 135 *Statistical Analysis used*

136 To explore which of these combined variables are important, we tried multivariate linear  
137 regression model using R with several combinations of variables to obtain insights into our  
138 understanding of epidemiology of Covid-19. LDM of both scenarios was assessed for its  
139 association with our predictor/explanatory variables grouped as described above into  
140 development variables, demographic variables, sanitation, tropical diseases and autoimmune  
141 disorders. Our final objective was to select the best combination of these groups of variables that  
142 would yield the largest adjusted-R<sup>2</sup> value.

143

## 144 **Results**

145 Analysis was carried on reported numbers of total 10112754 cases 501562 deaths as on  
146 the 29<sup>th</sup> June 2020 as well as with 60-day interval data as on the 7<sup>th</sup> August 2020. Further 90 days  
147 and 120 days interval data as on 16<sup>th</sup> September 2020 was also included. Data from well-known  
148 corona virus global deaths tracking sources<sup>19-21</sup> show that more than 70 percent Covid-19  
149 mortality has occurred in high income countries like Italy, Spain, UK, France and USA<sup>22</sup>. Our  
150 analyses also reveal a similar trend (Fig. 1a, b). To explore the dichotomy of rich and poor

151 countries in relation to Covid-19 deaths, we considered several potential explanatory parameters.  
152 The factors included percentage of older population, population density, incidence of  
153 communicable and non-communicable diseases etc. The data shows a positive correlation  
154 between log (GDP per capita) and log (deaths per million), abbreviated as LDM (Fig. 3,  
155 Supplementary, Tables S1, Fig. S1b), which reinforces the information of richer countries being  
156 most burdened by the disease.

### 157 *Demographics*

158 Among the many distinctive features of higher GDP countries, are its ageing population  
159 and a larger fraction of urbanization. We observe that while proportion of population above 65,  
160 obesity and the percentage of urban population are positively correlated with the Covid-19  
161 deaths (Fig. 3, Supplementary, Tables S1, Fig.S1c, Fig. S1e, Fig. S1g) the gender ratio and  
162 population density have negligible negative correlation (Fig. 3, Supplementary, Tables S1,  
163 Fig.S1d, Fig. S1f). Interestingly, whereas at an individual level gender does seem to have an  
164 impact on Covid-19 related death as numerous studies have shown<sup>23,24</sup>, at the aggregate level the  
165 gender ratio varies little across countries. Likewise, urbanization has an impact while population  
166 density does not seem to have an impact. We observe a positive correlation between LDM and  
167 percentage of population with age over 65, which is expected from the reports showing that older  
168 population is more susceptible to the virus<sup>7</sup>. Thus, the overall demographic parameters yield a  
169 positive correlation with Covid-19 deaths per million (Fig. 2a).

### 170 *Safe Sanitation*

171 Lack of sanitation and poorer hygiene practices are known to be responsible for the  
172 higher communicable disease burden in the low GDP countries<sup>1</sup>. It is therefore reasonable to  
173 expect that parameters describing safe sanitation and safe drinking water to be correlated



174 negatively with the Covid-19 deaths. Surprisingly we find a contrary observation, where  
175 different sanitation parameters are correlated positively with the Covid-19 outcome (Fig. 2b). It  
176 is therefore perplexing to note positive correlation of sanitation parameters, as described in  
177 methods, with the Covid-19 CFR (Fig. 3, Supplementary, Tables S1, Fig. S1h-m).

### 178 *Communicable Diseases*

179 Several studies have reported the protective role of parasitic/bacterial infections in  
180 bolstering human immunity, also referred to as immune training<sup>25,26</sup>. We observe that the  
181 prevalence of communicable diseases such as Malaria and Tuberculosis as well as parasitic  
182 diseases such as schistosomiasis, Onchocerciasis have a weak negative correlation with LDM  
183 (Fig. 3, Supplementary, Tables S1, Fig. S1t-z1). Upper respiratory diseases however show  
184 positive correlation which may be indicative of hyper active immune response in individual  
185 suffering from such disorder(Fig. 3, Supplementary, Tables S1). These intuitive observations  
186 related to communicable diseases with respect to Covid-19 deaths are suggestive of the role of  
187 “immune training” and the final disease outcome.

### 188 *BCG vaccination*

189 Recent studies have suggested that the countries in which the majority population was  
190 immunized by BCG vaccine had lower LDM due to trained immunity<sup>27,28</sup>. We find negligible of  
191 correlation between BCG vaccination in different countries with LDM (Fig. 3, Supplementary,  
192 Tables S1, Fig. S1n). A possible explanation for this difference could be due to the fact that  
193 comparisons in the previous studies were made with socially similar nations while we consider a  
194 diverse group of countries. Moreover, the distribution showing two distinct clusters for BCG  
195 vaccination (Supplementary, Fig. S1n), and with no apparent correlation within the clusters also  
196 makes it difficult to draw any conclusions on the effect of BCG vaccination on CFR at the

197 population level. It is also pertinent to note that we find significant negative correlation of mean  
198 BCG vaccination of a country and prevalence of autoimmune disorders (Fig. 3).

### 199 *Autoimmunity*

200 One of the manifestations of better hygiene and safe sanitation practices in high GDP  
201 countries is the increased incidence in autoimmune disorders<sup>29</sup>. It has been postulated that better  
202 hygiene practices could lower a person's immunity and make the person susceptible to  
203 autoimmune diseases. We therefore tested if any of the autoimmune diseases showed an  
204 association with the LDM variable. We interestingly find a strong positive correlation with  
205 Multiple Sclerosis, Type 1 Diabetes Mellitus, Rheumatoid Arthritis and Psoriasis and a weaker  
206 correlation with Asthma (Fig. 2c, Fig. 3, Supplementary, Tables S1, Fig. S10-s).

### 207 *Regression Analysis*

208 We now combined many related factors as described in methods such as demographic  
209 parameters, sanitation, communicable and non-communicable diseases. These combinations  
210 were arrived at after calculating correlations among all different parameters under consideration  
211 (Fig3, Supplementary, Tables S1). In order to inspect whether development indices like HDI  
212 and GDP of countries had an actual impact on our variable of LDM, we ran regression models  
213 using combinations of explanatory factors outlined above and checked the correlation between  
214 the residues, i.e. difference in actual values of the dependent variable and its predicted values  
215 from predictor variables, to GDP and HDI.

216 After testing various combinations of these groups, we observe that demographic  
217 variables, sanitation parameters and incidence of autoimmune disorders together explain most of  
218 the observed variation in LDM and development variables do not seem to count once the effect  
219 of these explanatory variables is accounted for (Supplementary, Tables S2). More precisely, if

220 one uses multiple regression to predict death rates, then the residues are uncorrelated with GDP  
221 and HDI. The R value for our final best model for 60 days interval is 0.77 and the adjusted R<sup>2</sup>  
222 value is 0.53(P-value<0.05). The correlation of this model's residue with GDP and HDI came to  
223 0.00 and -0.04 respectively (Supplementary, Tables S2). Since in this duration there was strict  
224 lockdown imposed upon some of the countries, we also did the same analysis on 90 and 120 days  
225 interval data. The R value for 90 days interval came to 0.74 and adjusted R<sup>2</sup> came to 0.48. The  
226 analysis on the 120 days interval reduced R values to 0.70 and adjusted R<sup>2</sup> value to 0.42.  
227 Although still significant the reduced R<sup>2</sup> and R values as the number of days increases can be  
228 attributed to the lowering of deaths in countries whose first wave of infection was completed as  
229 well as better preparedness to fight the infection.

230

## 231 **Discussion**

232 Various corona virus global deaths tracking sources<sup>19-21,30</sup> point towards uneven  
233 distribution of Covid-19 deaths with richer countries having higher CFR. In order to understand  
234 the reasons behind this paradoxical observation we analyzed publicly available data on various  
235 parameters. In this context, it is well known that when we have multiple factors correlated with  
236 dependent variable and these factors are themselves mutually correlated, it is not possible to  
237 determine as to which of these may be the causal variables using purely statistical techniques. To  
238 address this, we could collect data from a controlled experiment or look for data on individuals  
239 rather than countrywide aggregated data. However, in this instance many variables make sense  
240 only at the aggregate level ruling out both these options. Moreover, trends observed at individual  
241 level can be opposite to ecological or population level owing to the Simpson's paradox. In such  
242 challenging situations, it is desirable to use domain knowledge and identify the variables whose

243 correlation with the dependent variables has a logical explanation. In the present context, older  
244 population having higher death rate can be explained biologically. Similarly, since Covid-19  
245 spreads rapidly via person-to-person contact, population density is likely to have a big impact.

246 Our observation of the weak negative correlation of Covid-19 LDM with communicable  
247 diseases, and its positive correlation with incidence of autoimmune disorders in the high GDP  
248 countries, is indication of the interplay of host immunity and viral infection. As the parasite and  
249 bacterial disease burden is high in low and low-middle income countries, this can best be  
250 inferred upon by the “immune training” in the population of these countries due to chronic  
251 exposure to communicable diseases. It has been observed recently that patients susceptible to the  
252 Corona virus exhibited impaired Type 1 interferon activity<sup>31</sup>. Since initially the innate immunity  
253 acts upon the virus mediated by the interferon activity, individual with prior infections tend to  
254 coup with the virus better than the ones without exposure to pathogens<sup>32</sup>. The evidence for the  
255 innate immunity playing major role in susceptibility to the virus is available from the fact that  
256 asymptomatic patients or patients with mild infection produced poor neutralizing antibodies<sup>33</sup>.  
257 Hence it can be speculated that viral susceptibility is established at the first line of defense  
258 known as innate immunity much before the adaptive immunity. As mentioned earlier, the  
259 observed correlation between incidence of autoimmune disorders and LDM can be explained as  
260 Covid-19 seems to lead to severe autoimmune reaction. Paradoxically, better sanitation leads to  
261 poorer "immune training" and thus could be leading to higher LDM. Thus, the observed  
262 correlation of demographics, autoimmune disorders and sanitation with LDM appears to have  
263 explanation coming from the domain. The transfer rate for COVID-19 being high, we see a  
264 lower correlation between incidence of communicable diseases and LDM due to COVID-19 as  
265 compared to autoimmune disorders.

266 As can be seen in (Supplementary, Tables S2), the correlation of GDP and HDI with each  
267 model's residue is almost negligible. Both GDP and HDI seem to have a dominant role in  
268 determining the deaths per million yet both become insignificant when our linear regression  
269 models considered other notable and explainable confounding factors such as the demography,  
270 sanitation and the pre-Covid prevalence's of autoimmune and tropical diseases. The parameters  
271 of sanitation and autoimmune diseases consistently contributed significantly to the  $R^2$  no matter  
272 which combination of variables were chosen. It has not escaped our attention that there could be  
273 other factors such as the country's stage of the epidemic and lower reporting/testing in less  
274 developed countries that could also affect the mortality numbers. However, the statistical  
275 evidence given by us should give a head start to investigate the role of hyper immune reaction on  
276 Covid-19 susceptibility at individual patient's level.

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

284

## 285 **Acknowledgements**

286 BC gratefully acknowledges financial support by the National Centre for Cell Science, Pune.

287 Authors declare no competing financial interests.

288

289 **References**

- 290 1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, Global, regional, and national  
291 incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries  
292 and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.  
293 *Lancet*, 2018, **392**, 1789–1858.
- 294 2. Pan, H.-Y., Dai, Y.-N., Zheng, J.-N., et al., National incidence of autoimmune liver diseases and its  
295 relationship with the human development index. *Oncotarget*, 2016, **7**, 46273–46282.
- 296 3. World Bank Country and Lending Groups – World Bank Data Help Desk.  
297 [https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups)  
298 [lending-groups](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups).
- 299 4. Li, B., Yang, J., Zhao, F., et al., Prevalence and impact of cardiovascular metabolic diseases on  
300 COVID-19 in China. *Clin Res Cardiol*, 2020, **109**, 531–538.
- 301 5. Yang, J., Zheng, Y., Gou, X., et al., Prevalence of comorbidities in the novel Wuhan coronavirus  
302 (COVID-19) infection: a systematic review and meta-analysis. *International journal of infectious*  
303 *diseases*, 2020.
- 304 6. Emami, A., Javanmardi, F., Pirbonyeh, N., and Akbari, A., Prevalence of underlying diseases in  
305 hospitalized patients with COVID-19: a systematic review and meta-analysis. *Archives of academic*  
306 *emergency medicine*, 2020, **8**.
- 307 7. COVID-19 and Older Persons: A Defining Moment for an Informed, Inclusive and Targeted  
308 ResponseUnited Nations For Ageing | United Nations For Ageing.  
309 <https://www.un.org/development/desa/ageing/news/2020/05/covid19/>  
310
- 311 8. Roser, M., Human development index (HDI). *Our World in Data*, 2014.

- 312 [https://ourworldindata.org/human-](https://ourworldindata.org/human-developmentindex#:~:text=The%20Human%20Development%20Index%20(HDI)%20provides%20a%20single%20index%20measure,a%20decent%20standard%20of%20living)
- 313 [developmentindex#:~:text=The%20Human%20Development%20Index%20\(HDI\)%20provides](https://ourworldindata.org/human-developmentindex#:~:text=The%20Human%20Development%20Index%20(HDI)%20provides%20a%20single%20index%20measure,a%20decent%20standard%20of%20living)
- 314 [%20a%20single%20index%20measure,a%20decent%20standard%20of%20living](https://ourworldindata.org/human-developmentindex#:~:text=The%20Human%20Development%20Index%20(HDI)%20provides%20a%20single%20index%20measure,a%20decent%20standard%20of%20living).
- 315 9. Grosse, J., Hornstein, H., Manuwald, U., Kugler, J., Glauche, I., and Rothe, U., Incidence of diabetic
- 316 ketoacidosis of new-onset type 1 diabetes in children and adolescents in different countries
- 317 correlates with human development index (HDI): an updated systematic review, meta-analysis, and
- 318 meta-regression. *Hormone and Metabolic Research*, 2018, **50**, 209–222.
- 319 10. Galeotti, C. and Bayry, J., Autoimmune and inflammatory diseases following COVID-19. *Nature*
- 320 *Reviews Rheumatology*, 2020, **16**, 413–414.
- 321 11. Okada, H., Kuhn, C., Feillet, H., and Bach, J., The ‘hygiene hypothesis’ for autoimmune and allergic
- 322 diseases: an update. *Clinical & Experimental Immunology*, 2010, **160**, 1–9.
- 323 12. Bach, J.-F., The hygiene hypothesis in autoimmunity: the role of pathogens and commensals. *Nature*
- 324 *Reviews Immunology*, 2018, **18**, 105.
- 325 13. Roser, M., Ritchie, H., Ortiz-Ospina, E., and Hasell, J., Coronavirus pandemic (COVID-19). *Our*
- 326 *World in Data*, 2020.
- 327 *Our World in Data*, 2020. <https://ourworldindata.org/coronavirus>
- 328 14. GHO. <https://www.who.int/data/gho>. Accessed 29 June 2020.
- 329 15. World Bank Open Data | Data. <https://data.worldbank.org/>.
- 330 16. Water, sanitation and hygiene - exposure. [https://www.who.int/data/gho/data/themes/water-](https://www.who.int/data/gho/data/themes/water-sanitation-and-hygiene)
- 331 [sanitation-and-hygiene](https://www.who.int/data/gho/data/themes/water-sanitation-and-hygiene).
- 332 17. GBD Results Tool | GHDx. <http://ghdx.healthdata.org/gbd-results-tool>.
- 333 18. BCG. <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/bcg>.
- 334 19. WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int>.
- 335
- 336 20. COVID-19 Map. *Johns Hopkins Coronavirus Resource Center*.

337 <https://coronavirus.jhu.edu/map.html>.

338 21. COVID-19 situation update worldwide, as of 18 July 2020. *European Centre for Disease Prevention*  
339 *and Control*. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.

340 22. Schellekens, P. and Sourrouille, D., Tracking COVID-19 as Cause of Death: Global Estimates of  
341 Relative Severity. 2020.

342 23. Walter, L. A. and McGregor, A. J., Sex-and Gender-specific Observations and Implications for  
343 COVID-19. *Western Journal of Emergency Medicine*, 2020, **21**, 507.

344 24. Jin, J.-M., Bai, P., He, W., et al., Gender Differences in Patients With COVID-19: Focus on Severity  
345 and Mortality. *Front. Public Health*, 2020, **8**.

346 25. Netea, M. G., Quintin, J., and van der Meer, J. W. M., Trained Immunity: A Memory for Innate Host  
347 Defense. *Cell Host & Microbe*, 2011, **9**, 355–361.

348 26. Quinn, S. M., Cunningham, K., Raverdeau, M., et al., Anti-inflammatory Trained Immunity  
349 Mediated by Helminth Products Attenuates the Induction of T Cell-Mediated Autoimmune Disease.  
350 *Front. Immunol.*, 2019, **10**.

351 27. Escobar, L. E., Molina-Cruz, A., and Barillas-Mury, C., BCG vaccine protection from severe  
352 coronavirus disease 2019 (COVID-19). *PNAS*, 2020.

353 28. Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of  
354 COVID-19 | Science Advances. <https://advances.sciencemag.org/content/6/32/eabc1463>

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

357 30. Schellekens, P. and Sourrouille, D. M., *Covid-19 Mortality in Rich and Poor Countries: A Tale of*  
358 *Two Pandemics?* SSRN Scholarly Paper, Social Science Research Network, Rochester, NY, 2020.

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



- 361 32. Rouse, B. T. and Sehrawat, S., Immunity and immunopathology to viruses: what decides the  
362 outcome? *Nat Rev Immunol*, 2010, **10**, 514–526.
- 363 33. Long, Q.-X., Tang, X.-J., Shi, Q.-L., et al., Clinical and immunological assessment of asymptomatic  
364 SARS-CoV-2 infections. *Nature Medicine*, 2020, **26**, 1200–1204.

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384 **Table 1.** R- the correlation coefficient of combinations of variables with log (Deaths/Million)  
 385 LDM for 60,90 and 120 days interval and 29<sup>th</sup> June based on multiple regression analysis as well  
 386 as correlation of the residues with log (GDP) and HDI. Also showing the adjusted R<sup>2</sup>; N=106.  
 387

<b>Broad category variables</b>	<b>Explanatory variables</b>	<b>Duration</b>	<b>R with LDM as dependent variable</b>	<b>Adjusted R<sup>2</sup></b>	<b>R between residue and (log) GDP</b>	<b>R between residue and HDI</b>
		29 <sup>th</sup> June	0.49	0.22	0.04	0.01
<b>Demographics</b>	Age>65yrs, Obesity, Urban	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
<b>Sanitation</b>	Handwash, Basic drinking water, Basic sanitation, NO open defecation, Safe sanitation,	29 <sup>th</sup> 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

	Safe drinking						
	water						
	<hr/>						
<b>Autoimmune Diseases</b>	Multiple Sclerosis, Rheumatoid Arthritis, Asthma, Psoriasis, Type1 Diabetes Mellitus	29 <sup>th</sup> 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	
	<hr/>						
	<b>Demographics &amp; Sanitation</b>	Age >65yrs, Obesity, Urban population, Handwash, Basic drinking water, Basic sanitation, NO open defecation, Safe sanitation, Safe drinking water	29 <sup>th</sup> 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	
<hr/>							

	Handwash,					
	Basic drinking	29 <sup>th</sup> June	0.68	0.40	0.10	0.02
	water, Basic					
	sanitation, NO					
	open					
	defecation,					
	Safe sanitation,					
<b>Sanitation &amp;</b>	Safe drinking					
<b>Autoimmune</b>	water, Multiple					
<b>Diseases</b>	Sclerosis,	60 days	0.75	0.50	0.03	0.00
	Rheumatoid	90 days	0.69	0.42	0.09	0.01
	Arthritis,	120 days	0.64	0.35	0.07	0.01
	Asthma,					
	Psoriasis,					
	Type1					
	Diabetes					
	Mellitus					
<b>Demographics</b>	Age >65yrs,					
<b>, Sanitation &amp;</b>	Obesity,					
<b>Autoimmune</b>	Urban,	29 <sup>th</sup> June	0.74	0.48	0.04	-0.03
<b>Diseases</b>	Handwash,					
	Basic drinking					

water, Basic

sanitation, NO

open

defecation,

Safe sanitation,

Safe drinking

water, Multiple	60 days	0.77	0.53	0.00	-0.04
-----------------	---------	------	------	------	-------

Sclerosis,	90 days	0.74	0.48	0.03	-0.03
------------	---------	------	------	------	-------

Rheumatoid	120 days	0.70	0.42	0.01	-0.03
------------	----------	------	------	------	-------

Arthritis,

Asthma,

Psoriasis,

Type1

Diabetes

Mellitus

---

388

389

390

391

392

393

394

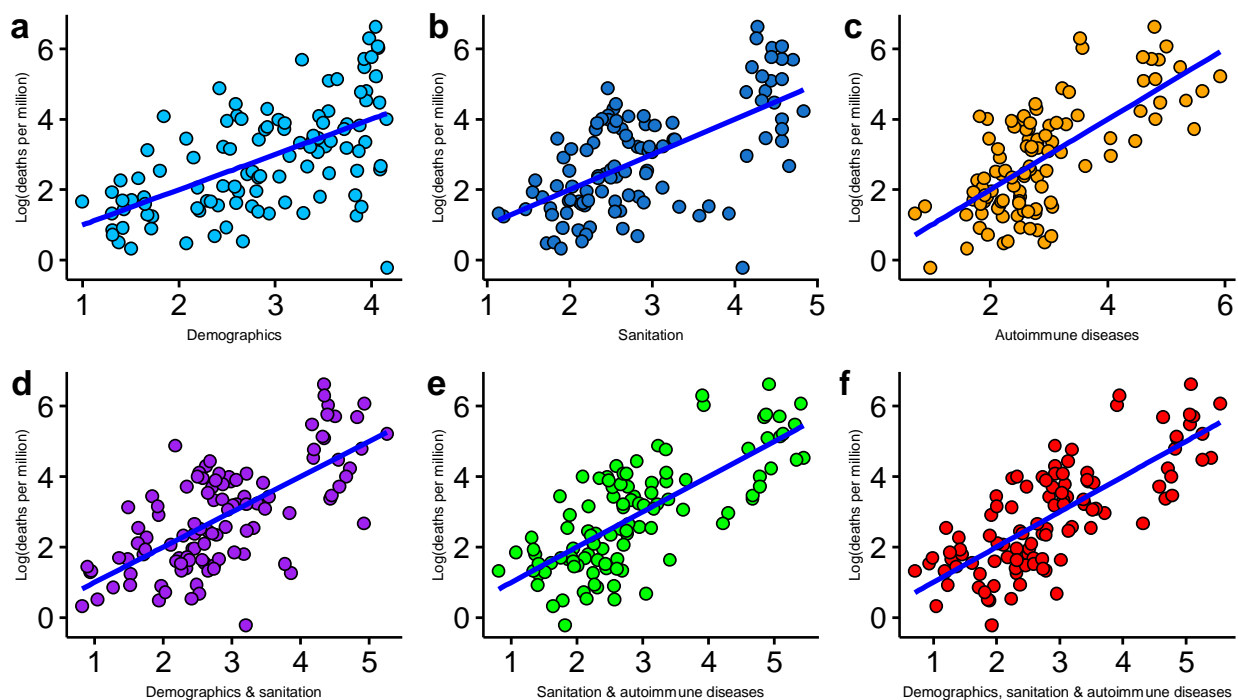
395





416 **Figure 2.** The actual values of log (deaths per million) 60 days after onset of 1<sup>st</sup> death or LDM  
417 plotted against their predicted values. The explanatory variables have been grouped, as described  
418 in the text, into a) demographics, b) sanitation, c) autoimmune diseases, d) demographics &  
419 sanitation, e) sanitation & autoimmune diseases and f) demographics, sanitation and autoimmune  
420 diseases. Regression analysis were done with LDM as dependent variable and the combinations  
421 of different variables; N=106. Each dot represents a country. The R<sup>2</sup> of each graph is depicted in  
422 Table 1.

423  
424

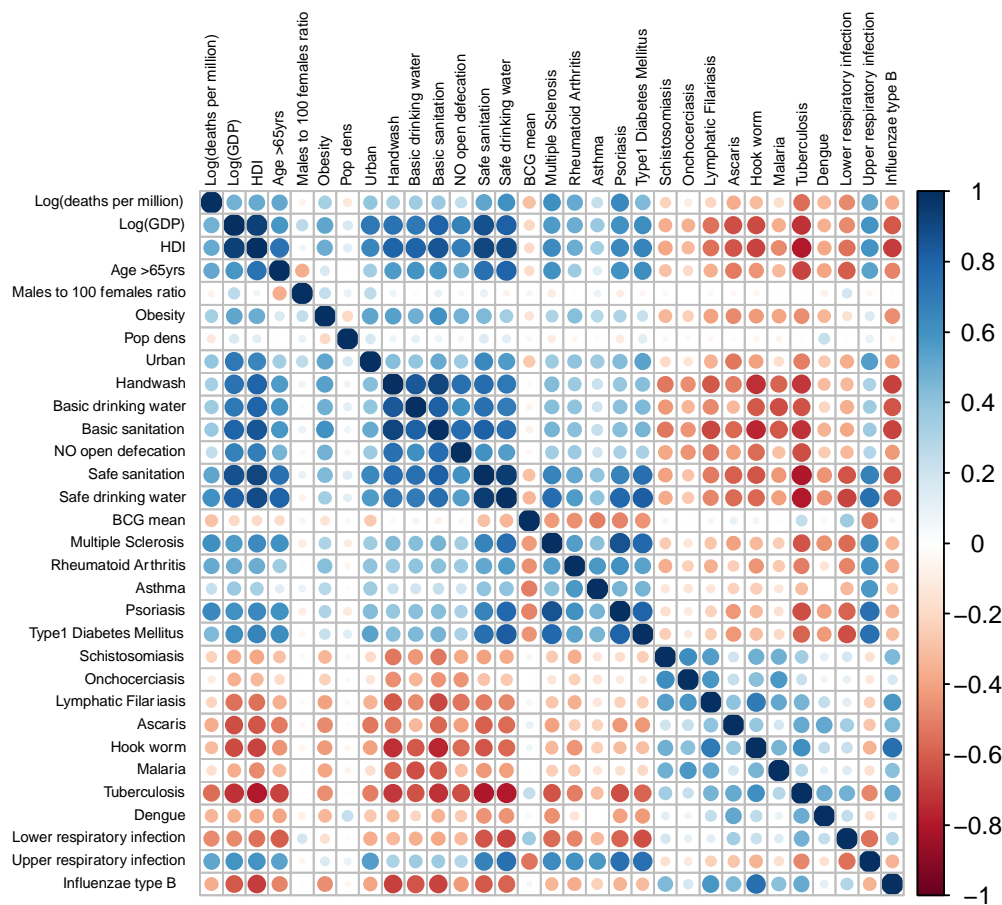


425



426 **Figure 3.** Plot showing correlation matrix of every variable with each other. The color key is  
427 depicted as -1(red) showing inverse correlation, 0 showing no correlation, and +1(blue) showing  
428 maximum positive correlation. The sizes of the circles range from small (low correlation) to  
429 larger (high correlation). Here the Log (deaths per million) is from the 60 days interval data;  
430 N=106.

431



432

433