



OPEN

## Analyzing the GHSI puzzle of whether highly developed countries fared worse in COVID-19

Sofija Markovic<sup>1</sup>, Igor Salom<sup>2</sup>, Andjela Rodic<sup>1</sup> & Marko Djordjevic<sup>1✉</sup>

Global Health Security Index (GHSI) categories are formulated to assess the capacity of world countries to deal with infectious disease risks. Thus, higher values of these indices were expected to translate to lower COVID-19 severity. However, it turned out to be the opposite, surprisingly suggesting that higher estimated country preparedness to epidemics may lead to higher disease mortality. To address this puzzle, we: (i) use a model-derived measure of COVID-19 severity; (ii) employ a range of statistical learning approaches, including non-parametric machine learning methods; (iii) consider the overall excess mortality, in addition to official COVID-19 fatality counts. Our results suggest that the puzzle is, to a large extent, an artifact of oversimplified data analysis and a consequence of misclassified COVID-19 deaths, combined with the higher median age of the population and earlier epidemics onset in countries with high GHSI scores.

It is natural to assume that the most prosperous world countries, well organized and with the most advanced medical institutions, would soon gain an advantage in the global race to minimize the harm from a new viral infection. This expectation is quantified in the form of the Global Health Security Index (GHSI)—a measure of preparedness for various health emergencies, precisely of the kind of the present COVID-19 pandemic<sup>1</sup>. The categories included in this index were carefully designed to help countries identify weaknesses primarily in their healthcare systems but also in their political and socio-economic structure, as well as to guide the process of reinforcing their health security capacities<sup>2–4</sup>. To this end, capacities to prevent medical hazards, to timely detect and report them, to rapidly respond in these situations, to comply in all these with international norms, as well as the health system capacity, and the overall vulnerability to biological threats—are all separately assessed and assigned a specific GHSI category: Prevent, Detect, Respond, Norms, Health and Risk, respectively.

With the COVID-19 pandemic, the GHSI scores were put for the first time to a real-life global test – which they failed, according to most of the published attempts to assess their correlation with the epidemic scale<sup>5–18</sup>. An unexpected positive correlation between the GHSI categories and differently calculated mortality<sup>5–7,12,14,17</sup> and infection spread<sup>5,7,8,10,12–14</sup> – deserving to be called the "GHSI puzzle" – suggested that the GHSI cannot be used to describe the national pandemic preparedness<sup>19</sup>. As this result seemed highly convincing, the focus was shifted to the search for explanations of the GHSI inappropriateness<sup>19</sup>. Thereby, the observed large COVID-19 infection counts in highly developed (high GHSI countries) motivated the firm opinion that their responses were "sclerotic and delayed at best"<sup>20</sup>. The first-ranked country in the GHSI 2019 list, the USA, provided a striking example. Despite its top overall score of 76/100—equaling roughly twice the average score of 38.9—COVID-19 has taken a devastating toll in the US, especially in the first wave of the pandemic<sup>21,22</sup>, introducing a two-year drop in life expectancy<sup>23</sup>. It was pointed out that GHSI scores cannot anticipate how a country decides to respond to epidemics, so these scores cannot be expected to necessarily be predictive of reported infections and fatalities<sup>24,25</sup>.

However, it is also necessary to revisit the methodological approaches that lead to the unintuitive relation of GHSI scores to the observed pandemic effects. First, the outcome variable, assumed to depend on the GHSI, should measure the relevant, effective epidemic burden to reasonably represent the country's overall success in minimizing the harm from the epidemic. Significantly, it should predominantly depend on the aspects of preparedness rather than other prevailing factors. As noted above, the GHSI predictivity was analyzed for different mortality and infection rates, yielding mostly positive correlations, and these variables depend on factors that directly or indirectly increase the frequency and range of personal contacts<sup>26</sup> and thus promote the disease spread<sup>27</sup>. Developed countries, typically carrying high GHSI scores, are characterized by intense population mobility and mixing due to ease of travel, higher local concentrations of people, and business and social activities. So,

<sup>1</sup>Quantitative Biology Group, Faculty of Biology, University of Belgrade, Belgrade, Serbia. <sup>2</sup>Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia. ✉email: dmarko@bio.bg.ac.rs

the consequential rapid disease spread may cancel and overshadow the ameliorating effects of large preparedness capacities, explaining the lack of a negative correlation of the GHSI with COVID-19 growth numbers. In addition, these outcome variables depend on the calculation time frame, and there are even examples of incidence rates with negative dependence on the GHSI<sup>18</sup>. Therefore, it is quite nontrivial to select the appropriate outcome variable (among mortality and incidence rates) for describing the relationship between the GHSI and the epidemic burden.

The relation of the GHSI with clinical severity measured by the Case Fatality Ratio (CFR), as a more obvious indicator of the burden magnitude, was considered in a few papers<sup>8,13,16</sup>, with the most extensive study (for a large number of countries and with including numerous other potential predictors) carried out in<sup>13</sup>—and the connection turned out to be positive. This is genuinely perplexing since CFR is an objective measure of disease severity, thus expected to be causally independent of the disease transmissibility. Consequently, there is no obvious explanation why higher GHSI would be linked to a worse case fatality ratio.

While CFR is a specific measure of severity, its shortcoming is being time-dependent and affected by the varying lag between case detections and fatal outcomes (in addition to lacking a direct mechanistic interpretation in terms of the disease dynamics)<sup>28</sup>. Instead, we choose a timing-independent measure of severity:  $m/r$  value that we previously derived from an epidemic model<sup>29</sup>, which represents the ratio of population-averaged mortality to recovery rates (so that many fatalities and/or slow recovery correspond to higher epidemic severity). It can be shown that this measure is independent of the SARS-CoV-2 transmissibility and directly relates to the asymptotic CFR value calculated at the end of the epidemic wave, so it does not rely on an arbitrary time frame. Thus, we could estimate it straightforwardly using the data from the end of the first pandemic's peak, avoiding interpretation of complex effects of vaccination and different SARS-CoV-2 strains on the disease severity.

Furthermore, we aimed to check if the observed positive GHSI correlations with various measures of epidemics intensity may be an artifact of the applied methods that are not well adapted for the problem complexity – i.e., cannot extract the impacts specific to individual factors entangled in mutual correlations. To that end, we searched for the  $m/r$  predictors among a number of variables on 85 countries using methods such as the regularization-based linear regressions (Lasso and Elastic Net)<sup>30</sup>, and non-parametric machine learning techniques (Random Forest and Gradient Boost)<sup>30</sup>. Together with different GHS Indices, we also included a number of sociodemographic and health-related variables in the analysis, in line with a recent result that factors external to GHSI may aid in characterizing countries' ability to respond to epidemic outbursts<sup>31</sup>. Finally, we investigated the relationship between the GHSI and the total excess deaths, which was not performed before; the main advantage of such analysis is that, unlike official COVID-19 figures, total excess deaths counts do not depend on diagnosis and reporting policies<sup>17,18</sup>. Combined together, our results, to a large extent, solve—or more precisely dissolve—the puzzle: it is not justified to relate high GHSI values (i.e., high estimated country preparedness) with an unsatisfactory pandemic response and high severity/mortality burden.

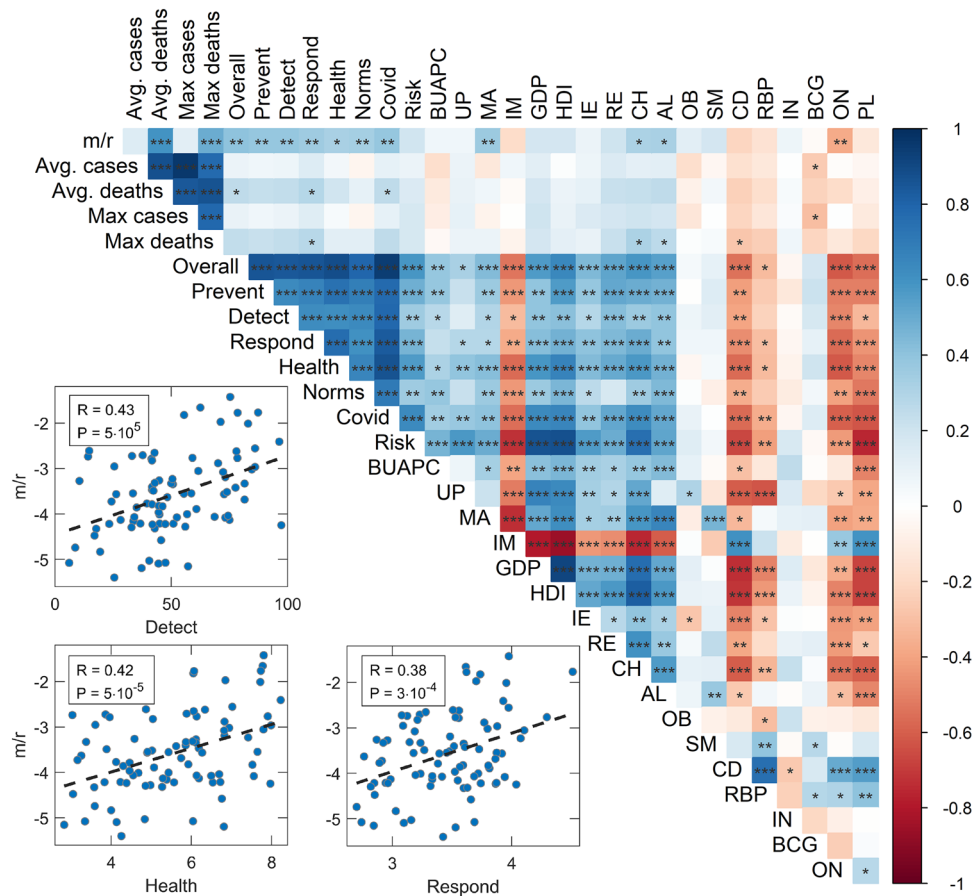
## Results

**Analysis based on COVID-19 case counts.** Univariate analysis (Fig. 1) reveals a statistically significant positive correlation of all selected GHSI categories with our severity measure ( $m/r$ ), demonstrating the puzzle. Strikingly, GHSI categories show the highest positive correlations compared to all other variables in the analysis. Also, they are highly correlated with measures of a country's development (e.g., GDP per capita and HDI), so that, on average, more developed countries have higher GHSI.

However, the univariate analysis does not control for simultaneous effects of multiple factors that can influence the disease severity. Additionally, many factors that we initially included as potentially important (and thus considered in the analysis) may not significantly influence the severity, so retaining them in the final model would introduce large noise. Another problem is that many factors are mutually highly correlated. Consequently, we next applied linear regressions with regularizations and feature selection (Lasso and Elastic Net, which can eliminate non-significant predictors), combined with PCA on groups of related variables (to partially decorrelate the variable set while retaining its interpretability).

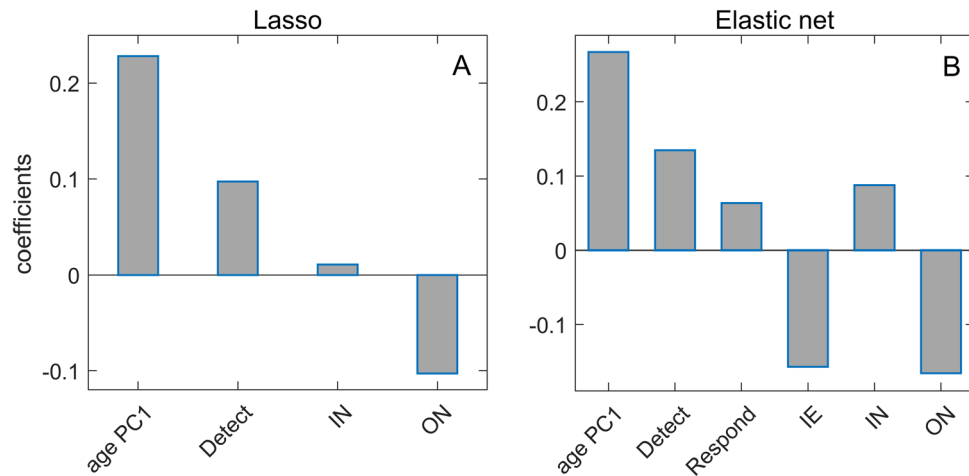
Mutually related variables were thus grouped in three categories (age, chronic disease, prosperity) on which PCA was performed (Supplement Fig. 1, Supplement Table 2), while the other variables remained unchanged (Supplement Table 1). The data were then independently analyzed using Lasso (Fig. 2A) and Elastic net regressions (Fig. 2B), which were implemented in the so-called "relaxed"<sup>30</sup> way to reduce noise from the multidimensional data.

From Fig. 2, we see that age PC1 is the dominant predictor in both methods, unsurprisingly suggesting that the older population is more severely affected by COVID-19. Furthermore, epidemic onset appears to be an important factor negatively associated with the disease severity, meaning that earlier epidemic onset is associated with higher severity of the disease. Of all four analyzed GHSI categories, the disease detection index is selected as the most important, with an (unintuitive) positive effect on  $m/r$ . While positive contribution was also obtained through univariate analysis, its magnitude diminishes when the more advanced analysis is used, i.e., in Fig. 2, PCA followed by multivariate regressions with regularization. Consequently, while the paradox is still there (positive contribution of one of GHSI categories), the regression coefficient of this predictor significantly decreases (e.g., compared to age) as the analysis progresses from univariate to multivariate (with feature selection and correlated variables accounted for). Also, in Elastic net regression (Fig. 2B), net immigration appears to contribute negatively to the disease severity, which we will further discuss below. The obtained results remain robust when the regression methods are applied to the initial data without PCA, when besides 18 demographic parameters, Detect, Respond, Health and Risk GHSI categories (Supplement Fig. 2), or Covid index (Supplement Fig. 3) are used.



**Figure 1.** Graphical representation of the predictor correlation matrix (as the matrix is symmetric, only its upper half is shown). Colors correspond to the magnitude of Person correlation coefficients, as indicated by the color bar on the right, while asterisks correspond to the statistical significance of the results (‘ ’ –  $P > 0.05$ , ‘ \* ’ –  $0.05 > P > 0.01$ , ‘ \*\* ’ –  $0.01 > P > 0.001$ , ‘ \*\*\* ’ –  $P < 0.001$  ). Correlations of  $m/r$  and selected three Global Health Security Index categories, relevant for the pandemic, are also represented by scatterplots in the inset.  $m/r$  – disease severity measure, Overall – Overall GHSI, Prevent – GHSI Prevent category, Detect – GHSI Detect category, Respond – GHSI Respond category, Health – GHSI Health category, Norms – GHSI Norms category, Risk – GHSI Risk category, Covid – a combination of COVID-related GHSI indicators, BUAPC – built-up area per capita, UP – urban population, MA – median age, IM – infant mortality, GDP – gross domestic product per capita, HDI – human development index, IE – net immigration, RE – refugees, CH – blood cholesterol level, AL – alcohol consumption, OB – prevalence of obesity, SM – prevalence of smoking, CD – prevalence of cardiovascular diseases, RBP – raised blood pressure, IN – physical inactivity, BCG – BCG vaccination coverage, ON – the onset of the epidemic, PL – air pollution. The figure was created using R<sup>55</sup> corrplot package<sup>56</sup>, in RStudio integrated development environment for R<sup>57</sup>.

In the analysis above, we grouped variables in subsets according to their natural interpretation (e.g., age-related, chronic diseases, country prosperity) for PCA. While this approach facilitates the interpretation of the regression results, there is also some arbitrariness in the grouping criteria. This motivated us to perform supervised PCA—a mathematically well-defined procedure that chooses the variables for PCA solely based on their numerical correlation with the response variable. The optimal hyperparameter values, selected through cross-validation, were  $\theta = 0.05$  for correlation coefficient cutoff and  $n = 6$  for the number of retained principal components. The selected six principal components were used as predictors in a linear regression model with  $m/r$  as the response variable. As only two principal components, PC1 and PC4, turned out to be significant predictors (with the standard significance cutoff of  $P < 0.05$ ) in the linear regression, we proceeded to further analyze these principal components. The results in Fig. 3 show scatterplots of PC1 and PC4 vs.  $m/r$  and variables entering PCA. One can see a high correlation of  $m/r$  with PC1, which technically makes it a good risk index for COVID-19 severity – though some of its constituents, such as GHSI, have an unintuitive contribution to  $m/r$ . More importantly, PC1 shows that the composite variable, which can explain much of the variability in  $m/r$ , contains GHSI categories intertwined with HDI and GDP in terms of their effect on  $m/r$ . Consequently, the unintuitive positive contribution of GHSI to  $m/r$  (as well as the absence of negative contribution in more complex models) may be an indirect consequence of some effect that increases  $m/r$  for more developed countries (for which both HDI/GDP and GHSI are high). We will further explore this possibility through the analysis of excess deaths.



**Figure 2.** Lasso and Elastic Net regressions with  $m/r$  as the response variable. (A) Regression coefficients of variables selected by Relaxed Lasso regression, (B) Regression coefficients of variables selected by Relaxed Elastic Net regression, age PC1 – age principal component 1, Detect – GHSI Detect category, Respond – GHSI Respond category, IE – net immigration, IN – physical inactivity, ON – the onset of the epidemic.

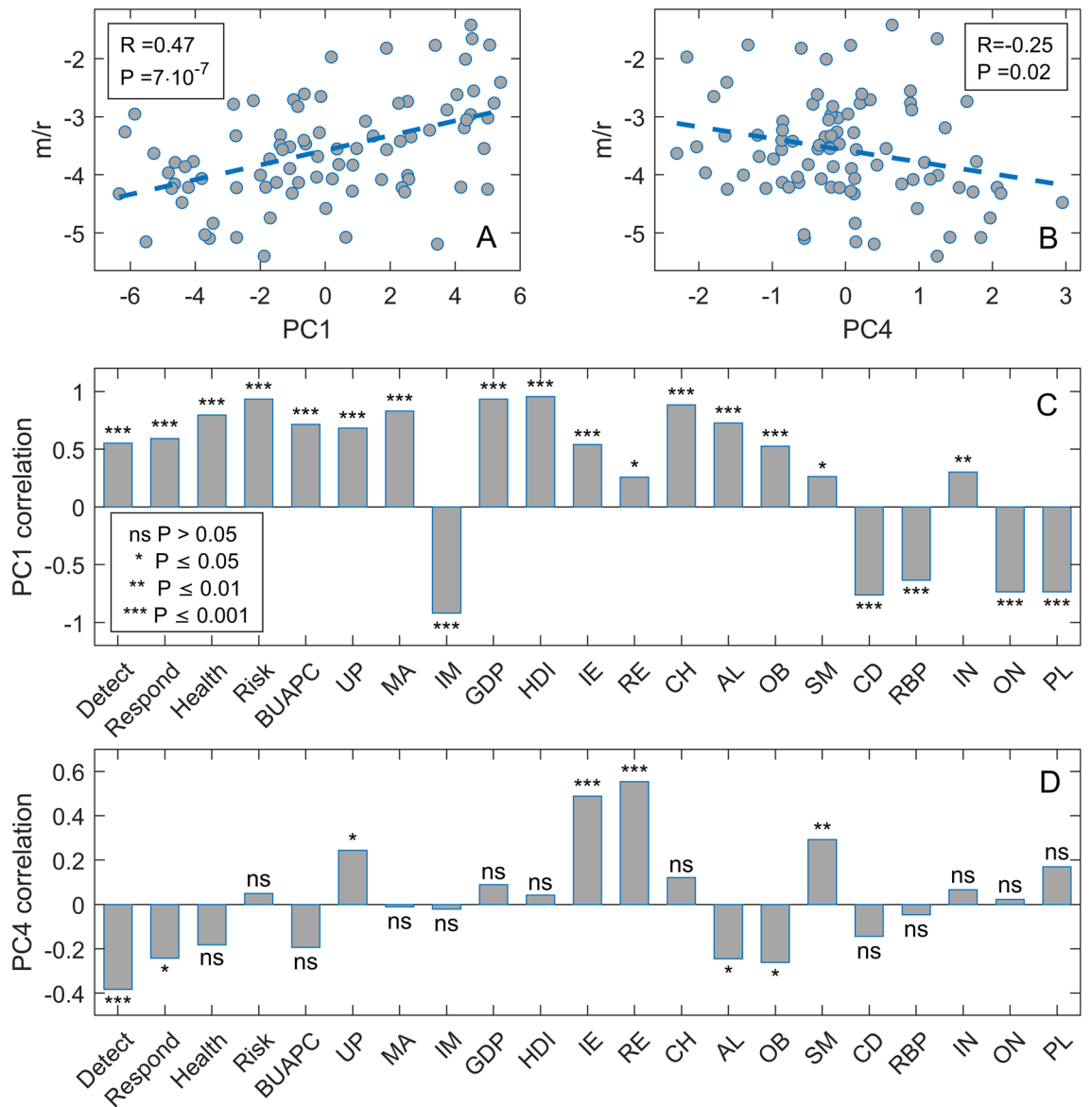
The regression methods used so far do not account for nonlinearities or possible interactions between predictor variables. For this reason, we applied machine learning methods based on ensembles of weak learners (decision trees), Random Forest, and Gradient Boost, on the same set of predictors and PCs used in the linear regression models above. Figure 4 shows the estimate of variable importance (A and D) and Partial Dependence (PD) plots (B, C, E, F) for the variables with the highest importance in explaining disease severity. PD plots also provide information about the direction (positive or negative) in which the predictors affect  $m/r$ . We obtain age PC1 as by far the most important variable, where older age promotes the disease severity. Detect GHSI and epidemic onset also appear above the standard importance threshold. However, the estimated effect of Detect GHSI is now small, based on both the variable importance and PD plots that show a much larger dependence of  $m/r$  on age PC1.

**Analysis of excess deaths.** We next focus on excess deaths (see the definition in Methods), where we again use several univariate and multivariate approaches. As predictors, we here use COVID-19 counts, GHSI, and demographic variables, which we separated into mutually related subsets on which PCA was performed (see Supplement Table 3 and Supplement Fig. 4). In Fig. 5, which shows Random Forest results, we see that Counts PC1, which equally well reflect COVID-19 infection and death case counts, are by far the most dominant predictor of excess deaths. While expected, this is also a highly nontrivial result, given that the two quantities (excess deaths and classified COVID-19 deaths) are inferred in an entirely independent manner, and that in-principle ununiform policies of COVID-19 related deaths classification are applied in different countries<sup>32</sup>. This result is robust, i.e., consistently obtained through different methods, ranging from simple univariate regressions to linear regressions (with variable selection and regularization), see Supplement Figs. 5 and 6 and Supplement Table 4. The only other predictor with importance above the threshold in Fig. 5 is CFR (which we also calculate in saturation, i.e., at the end of the first wave), which has a small to moderate positive effect on  $m/r$ , though much smaller than the COVID-19 case counts.

We then analyzed the unexplained deaths with the same approach described above for the excess deaths. The unexplained deaths are the difference between the excess deaths and COVID-19 deaths, i.e., those excess deaths not attributed to COVID-19 (for the definition, see Methods). From Fig. 6, we see that the most important predictor is demo PC1, which has a clear *negative* association with unexplained deaths (see the corresponding PD plot). These results are robust, i.e., independently obtained by several different univariate and multivariate analyses (Supplement Figs. 7 and 8 and Supplement Table 5). As demo PC1 is strongly related to country GDP/HDI (see Supplement Fig. 4), more developed countries are associated with a smaller number of unexplained deaths. The other demographic-related PC components above the importance threshold (demo PC 4 and 6) have a smaller influence on unexplained deaths and can also be related to the negative influence of HDI/GDP (see Supplement Fig. 4). Interestingly, GHSI PC1 (highly correlated to all GHSI categories) is also selected as a significant predictor, negatively associated with unexplained deaths. This implies that higher GHSI (significantly correlated to HDI/GDP) is associated with fewer unexplained COVID-19 deaths.

## Discussion

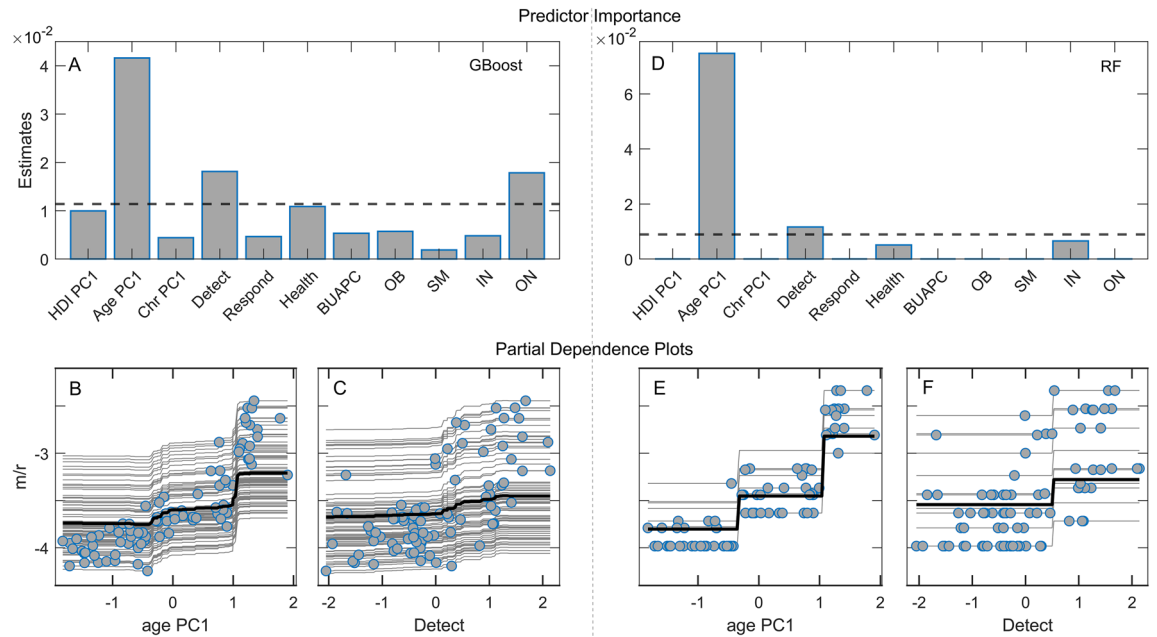
Although the  $m/r$  quantity does not depend on the epidemic growth rate, which is boosted by the large contact rate in countries with high GDP, HDI, and GHSI, the univariate analysis shows that GHSI categories significantly correlate with this measure of severity – having the highest correlation of all considered variables. This could be naively interpreted as a higher likelihood of dying from COVID-19 in well-equipped medical facilities than in settings ranked by far lower GHSI. In the subsequent two parallel analyses—Lasso and Elastic Net on PCs



**Figure 3.** Supervised Principal Component Analysis. Scatterplot of (A) PC1 and  $m/r$ , (B) PC4 and  $m/r$ , (C) PC1 and variables entering PCA, (D) PC4, and variables entering PCA.  $m/r$  – mortality over recovery rate, Detect – GHS Detect category, Respond – GHSI Respond category, Health – GHSI Health category, Risk – GHSI Risk category, BUAPC – built-up area per capita, UP – urban population, MA – median age, IM – infant mortality, GDP – gross domestic product per capita, HDI – human development index, IE – net immigration, RE – refugees, CH – blood cholesterol level, AL – alcohol consumption, OB – prevalence of obesity, SM – prevalence of smoking, CD – prevalence of cardiovascular diseases, RBP – raised blood pressure, IN – physical inactivity, BCG – BCG vaccination coverage, ON – the onset of the epidemic, PL – air pollution.

obtained from three predefined groups of variables, on one side, and the Supervised PCA, on the other – the positive effect of GHSI is acknowledged, primarily through the Detect category. However, it is smaller than the effects of PCs brought to the fore, determined by factors like the population median age, the epidemic onset, and the net immigration. Thereby, the diversity of the variables constituting PC1 in the supervised PCA, together with multiple GHSI categories, additionally questions the GHSI significance, as it makes it hard to separate the influence of the GHSI from the influences of other variables related to the country development level (HDI/GDP). Finally, the variable importance analysis points to a robust and dominant influence of the age PC1, while a positive influence of the GHSI Detect category is present but almost negligible compared to the age PC1 effect.

Moreover, the Random Forest method selects the Onset as an additional variable with higher significance than the GHSI Detect. Therefore, our results show that oversimplified analytical methods play a major role in the puzzle: While simple pairwise correlations indeed strongly suggest that "better healthcare leads to higher COVID-19 mortality", there is hardly any hint of such paradox with advanced machine-learning analyses. Instead, old age and an earlier epidemic onset are likely behind the increased disease severity—and simple univariate analysis fails to identify these confounding variables, both significantly correlated with the GHSI.

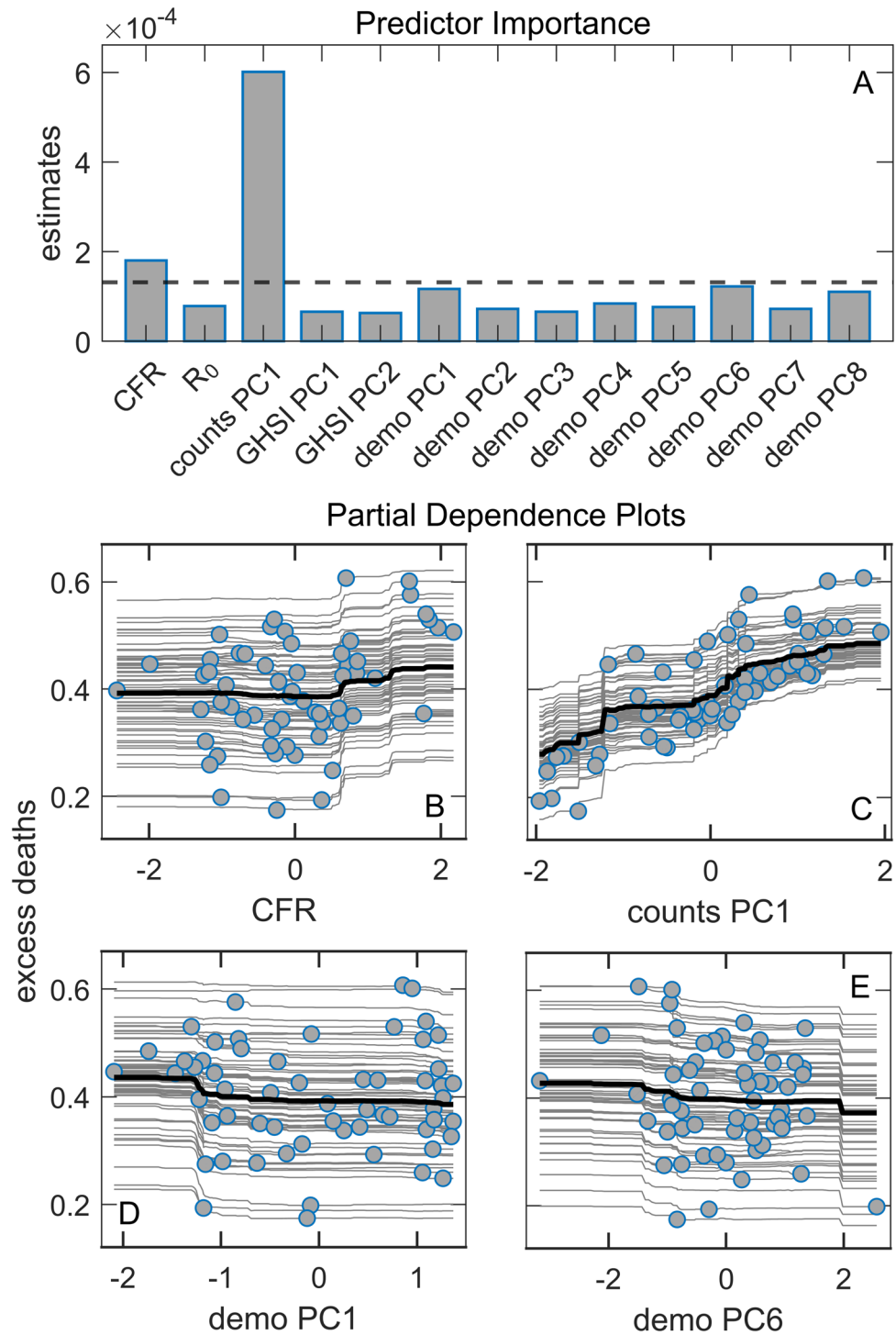


**Figure 4.** Machine learning algorithms and Partial dependence (PD) plots with disease severity as the response variable. **(A)** Variable importance estimates for Random Forest regression, **(B–C)** PD plots corresponding to Random Forest, **(D)** Variable importance estimates from Gradient Boost regression, **(E–F)** PD plots corresponding to Gradient Boost. The dashed lines in A and D correspond to the mean variable importance in the given model. HDI PC1 – human development index PC1, Chr PC1 – chronic diseases PC1, Detect – GHSI Detect category, Respond – GHSI Respond category, Health – GHSI Health category, BUAPC – built-up area per capita, OB – obesity, SM – smoking, IN – physical inactivity, ON – epidemic onset, m/r – disease severity measure.

Three severity predictors robustly selected by different methods – the median age, the epidemic onset, and the net immigrations – are generally supported by other studies' results, obtained in combinations with other predictors<sup>6,8,11,13,14</sup>. The fact that our analysis singled them out argues about their dominant roles. Old age is undoubtedly a dominant risk factor for developing severe illness and dying from COVID-19<sup>6,8,13,14</sup>. The onset variable, not acknowledged only by the Random Forest, is obtained with a negative association to severity<sup>11,33</sup>. The time in which the pandemic reached a country may have been a significant factor in the first wave, as countries could use it to organize medical resources and adjust treatment protocols, learning from the experiences of those affected earlier. The Elastic net and the supervised PCA results identify the net immigration as a negative contributor to the COVID-19 severity. The supervised PCA groups this variable inside the same PC with the percent of refugees. This negative effect of immigration in the broader sense on the severity should not be interpreted through a hypothetical mitigating effect of the full lockdowns on the virus spread, rapidly imposed in those countries expecting a massive inflow of people<sup>13,34</sup>, because m/r is independent of the disease transmissibility. Rather than being more efficient in treating COVID-19, countries leading in net immigration probably had an additional under-privileged population<sup>35</sup> whose deaths from COVID-19 were more frequently misattributed to other causes, e.g., a lack of medical insurance or a general reluctance to go to hospitals<sup>36,37</sup>. If so, the presence of immigrants did not decrease the actual COVID-19 mortality, but only the officially recognized toll (further discussed below).

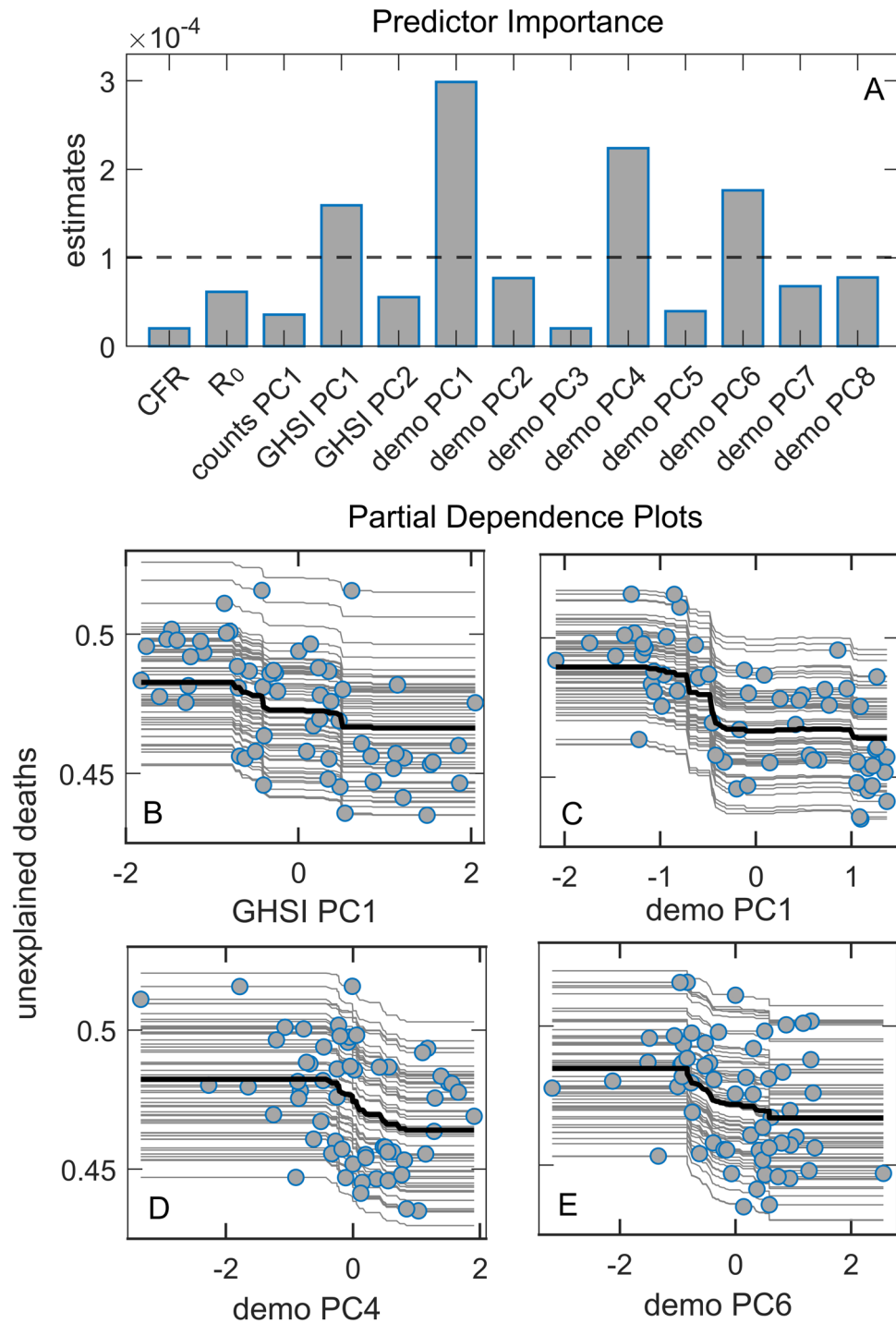
We further demonstrate that the reliability of COVID-19 data also plays an important role. Many factors can lead to underreporting, or even overreporting, of COVID-19 deaths: testing capabilities, protocols for classifying causes of death, to outright political influences<sup>38</sup>. Since these factors can be highly variable between countries, we considered the overall excess deaths as an alternative measure to official COVID-19 mortality statistics. We obtain that, similarly to<sup>39</sup>, the excess deaths variable depends almost only on the sheer number of COVID-19 cases (itself not significantly correlated with any other of our predictor variables, see Fig. 1), together with some influence of the CFR value (where CFR itself is dominantly correlated with median age, and with epidemic onset—possibly due to the accumulated experience in medical treatments). The counterintuitive connection of COVID-19 severity with GHSI categories (and country's prosperity) does not appear when considering the overall excess mortality. Another perspective on the phenomenon is gained by considering the "unexplained deaths" variable, quantifying the part of the overall excess deaths not officially attributed to COVID-19 mortality. We obtained that unexplained deaths are smaller in developed countries with high GHSI scores. Other studies also found a similar relationship between unaccounted excess deaths and the lower average income/fewer physicians per capita<sup>6,8,14,37,38,40</sup>. This strongly suggests that underreporting COVID-19 deaths plays another crucial role in the GHSI puzzle, i.e., that the actual COVID-19 toll in low GHSI countries is higher than the official figures reveal.

For our present purposes, it is not crucial whether the excess deaths represent direct COVID-19 victims or indirect pandemic casualties of reduced access to healthcare services (i.e., due to overwhelmed medical facilities):



**Figure 5.** Random Forest regression with excess deaths as the response. (A) Predictor importance estimates from Random Forest regression, where dashed line represents the importance threshold, i.e., the mean importance value, (B–E) Partial dependence plots of predictors with the highest importance estimates from Random Forest regression model, CFR – case fatality rate,  $R_0$  – basic reproductive number of the virus.

GHSI scores should reflect capabilities to cope with both aspects of the pandemic consequences. Nevertheless, most of the unaccounted deaths likely represent direct consequences of SARS-CoV-2 infection, especially in countries where the difference between the official COVID-19 mortality and the excess of deaths is substantial. Namely, our results are based on the excess of deaths that occurred during a relatively short time window of the first pandemic wave (a few months), where the negligence of other/chronic medical conditions (e.g., postponed oncological or cardiological check-ups and treatments) will unlikely have such immediate effects on mortality.



**Figure 6.** Random Forest regression with unexplained deaths as the response. (A) Predictor importance estimates from Random Forest regression, where dashed line represents the importance threshold (the mean value of the variables), (B–E) Partial dependence plots of predictor with the highest importance estimates from Random Forest regression model, CFR – case fatality rate,  $R_0$  – basic reproductive number of the virus.

Besides, if the unexplained deaths were primarily an indirect consequence of saturation of medical resources, COVID-19 case counts should reflect this saturation and thus be expected to appear as a relevant predictor of unexplained deaths—but this does not happen. On similar grounds, the authors of<sup>39</sup> support the same interpretation of excess mortality in regions of Italy. As an additional argument, we note that, of all GHS indices, both the linear regressions with feature selection (Fig. 2) and the machine learning techniques (Fig. 4) consistently single out the Detect score (quantifying testing capabilities) as the one most affecting m/r. This is consistent with the interpretation that high GHSI scores—and the Detect index category in particular—are not connected to higher



COVID-19 mortality but rather reflect better capacity to recognize (i.e., "detect") COVID-19 deaths as such, and thus lead to less unexplained deaths. On the other hand, a lower Detect score relates to a larger number of unrecognized COVID-19 deaths and falsely lower mortality, adding to the apparent paradox.

We have provided ample arguments that the puzzling connection of high GHSI scores with severe COVID-19 epidemic outcomes is an artifact of oversimplified analyses and low-quality data, but neither of our methods could directly establish the anticipated negative correlation. Notably, the negative correlation with the scale of under-reporting of COVID-19 deaths is suggested by our results, demonstrating the effectiveness of better detection capacity of high-GHSI countries. However, it seems that the anticipated benefits of the elaborated preparedness capacities for minimizing the overall harm did not manifest, at least not directly. Note that our results clearly show the strong influence of population age and, in the second place, the epidemic onset on the m/r measure of severity. The latter predictor emphasizes the particular vulnerability of the initial pandemic period<sup>11,12</sup>. Not only did a later onset provide more time to organize a response, but the large unpredictability effect when the world encountered an unknown virus, obstructing any organizational efforts<sup>1,41,42</sup>, put the first-hit countries in an additionally vulnerable position. The lack of knowledge and experience could have primarily determined the reach of the initial efforts to minimize severity regardless of the pre-established capacities, and this factor would be hard to incorporate in a static preparedness measure such as the GHSI. Nevertheless, the GHSI might show a better, anticipated agreement later during the pandemic, when the advantages of the prepared system start to reveal. Previous studies, like ours, assessed the GHSI predictivity focusing on the pre-vaccination period, so its performance separated from the sensitive, first outbreak period remains largely unexplored.

On the other hand, GHSI indices are positively correlated with the population age and negatively correlated with the onset time. In turn, if there was no direct influence of GHSI scores on mortality (nor additional confounding correlations), we should expect more excess deaths in countries with higher GHSI indices—simply because the population is, on average, older, while the epidemic came earlier, in these countries. However, since the plot in Fig. 6 does not reveal such a trend, and overall excess mortality seems to depend almost solely on the number of infected individuals (with neither age nor GHSI playing any role in Fig. 5), there must be a factor that compensates for this age and early epidemic onset disadvantage. While we cannot rule out possible confounders, it is plausible that the difference in levels of medical care—reflected via GHSI scores—counterbalances these two effects.

In summary, the main goal of this study was to reinvestigate the "GHSI puzzle," giving attention to three essential aspects (i) the choice of the effective epidemic burden measure, (ii) the advanced multivariate data analysis, suiting the data complexity, and (iii) the analysis of excess deaths data, in addition to more questionable official COVID-19 infection and mortality figures. Our results point to a major weak point in the previous efforts to assess the predictivity of the GHSI index—namely, using the oversimplified methodology, which emphasizes the importance of implementing the three essential steps in our analysis mentioned above. We show that high GHSI scores cannot be associated with low efficacy in combating epidemics, contrary to the dominant claim in academic literature. On the contrary, more developed countries with high GHSI values likely provided better quality data due to more reliable reporting of COVID-19 fatality counts. This becomes particularly significant in light of the notion strongly supported by our results: that a majority of excess deaths after the first wave should be attributed to direct COVID-19 fatalities instead of indirect consequences of the pandemic or other concurrent causes.

GHSI categories, therefore, do not contradict the observed data and can still be a valuable instrument in ameliorating the consequences of future global health-related crises, especially with some refinement to account for the country's vulnerability to a high level of initial unpredictability when it faces an unknown, fast-transmitting virus. Refining this and developing similar measures can aid the prevention of future epidemic catastrophes, where our study may provide some guidelines on how to perform an accurate assessment of their appropriateness. Some of the GHSI shortcomings have already been reported<sup>31</sup> and addressed in the revised GHSI Index 2021<sup>43</sup>, where the authors have attempted to improve the index by adding several new subcategories, which can certainly be viewed as a step in the right direction. While the lack of a positive association between higher GHSI scores and lower disease severity cannot be fully explained by our results, it is possible that including more variables in the research and obtaining excess and unexplained deaths data for a higher number of countries (which would allow for introducing more predictor variables without the danger of overfitting the data) would help bring to light some of the dependencies we were unable to capture in this research. One such variable could be countries' corruption index, as it is reasonable to assume that countries with a higher corruption prevalence, would be less transparent in reporting COVID-19-related deaths. Less reported deaths would lead to lower disease severity estimates in countries with a higher corruption index, contributing to the paradox. Higher perceptions of corruption and social polarization have already been identified as contributing to higher excess mortality<sup>31</sup>.

Also, the results from later COVID-19 peaks might differ from the first peak, as it is reasonable to anticipate that more developed countries, with higher GHSI and, on average, higher healthcare budgets, given enough time and resources, would be better prepared for the later COVID-19 waves<sup>44,45</sup>. Additionally, a quicker vaccine rollout in developed countries could lead to quicker drop in disease severity, regardless of the increasing transmissibility of new viral strains<sup>46</sup>.

## Methods

**Data collection and processing.** The majority of the data used in this research, including 18 demographic and health parameters, and the basic reproduction number of the virus was taken from the dataset previously assembled in<sup>27</sup>. Additionally, Global Health Security Index (GHSI) data were taken from<sup>1</sup>. In addition to the main (Overall) GHSI and six index categories (Prevent, Detect, Respond, Health, Norms, and Risk), the Covid index was calculated from the averaged values of COVID-19 related GHSI indicators, according to<sup>17</sup>.

COVID-19 counts (average and maximal daily cases and deaths) were taken from<sup>48</sup>. At the end of the first peak, the case fatality rate and severity measure (m/r) for each country were calculated according to the methodology described in<sup>29</sup>. To determine the numbers of total and excess deaths during the first peak of the pandemic, data from<sup>49</sup> was used. Relative excess deaths and relative unexplained deaths (excess and unexplained deaths in further text and figures) were calculated as follows:

$$\text{relative excess deaths} = \frac{\text{excess deaths}}{\text{total deaths} - \text{excess deaths}}$$

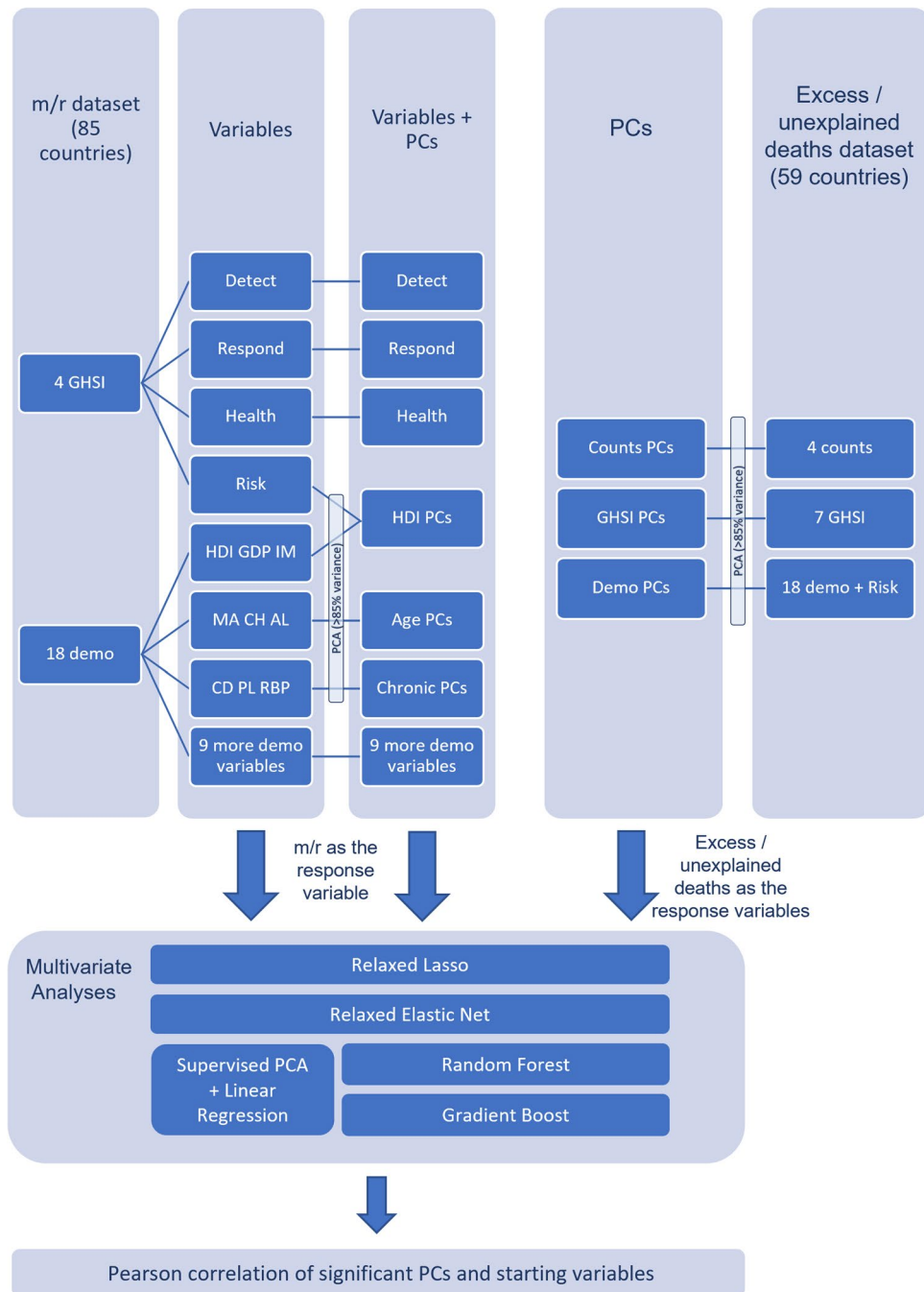
$$\text{relative unexplained deaths} = \frac{\text{excess deaths} - \text{COVID19 deaths}}{\text{total deaths} - \text{excess deaths}}.$$

For the research two datasets were constructed: (i) m/r dataset—a subset of the dataset used in<sup>27</sup>, consisting of the data from 85 countries for which both GHSI were available and m/r was inferred; (ii) excess deaths dataset—a subset of (i) for which total and excess deaths data was available (59 countries). Since the distribution of most variables initially deviated from normality, data was transformed, so that the skewness of the variables was as close to zero as possible, with the minimal number of remaining outliers. To obtain the best possible results, transformations were applied separately for variables in each dataset. Interpretation of all the variables and the used transformation for both datasets are presented in Supplementary Table 1. For easier tracking of the course of the research, analyses performed on each dataset are graphically represented by a flowchart in Fig. 7.

**Principal component analysis.** To partially decorrelate data and reduce dimensionality while retaining relatively simple interpretation of principal components, some of the variables were grouped into mutually related groups, on which Principal Component Analysis (PCA) was performed. The number of principal components retained from each group after the PCA was determined to explain over 85% of the variance of the data. PCA grouping was performed independently for both datasets and variables entering PCA; retained principal components, and the percentage of variance they account for are presented in Supplementary Table 2 (m/r dataset) and 3 (excess deaths dataset). New datasets, now consisting of both retained principal components and remaining variables that did not enter PCA, were next used as input in univariate and multivariate analyses. For easier interpretation of the most relevant principal components, their correlation with the variables entering PCA is given in Supplementary Figs. 1 and 4. Two now obtained sets of predictors were: (i) variables used in m/r regression analysis, consisting of: Detect, Respond, and Health GHSI categories, five selected PCs from Supplementary Table 2, and remaining demographic and health parameters that did not enter PCA (BUAPC, UP, IE, RE, OB, SM, IN, BCG, ON) (ii) variables used in excess and unexplained deaths analysis with principal components consisting of  $R_0$ , CFR and 11 principal components from Supplementary Table 3.

**Linear regression models.** LASSO and Elastic net regressions were used as the implementations of L1 (LASSO) and L1 and L2 norms (Elastic net) on three variations of the m/r dataset and on excess deaths dataset with principal components (regressions were implemented separately for excess and unexplained deaths as response variables). The variables (and PCs) were standardized before the regressions, so the regression coefficients obtained by these methods can be interpreted as the relative importance of predictors in explaining the response variables. Values of the hyperparameters  $\lambda$  in LASSO and  $\alpha$  and  $\lambda$  in Elastic Net were determined through fivefold cross-validation, with the data repartitioned 40 times. In m/r dataset regressions, the sparsest model—with mean squared error (MSE) within 1 standard error (SE) from the minimal MSE—model was chosen. To further reduce the noise in m/r dataset regressions (where the number of predictors retained after the first round of regression was relatively high), LASSO and Elastic Net were implemented as “Relaxed”<sup>50</sup>, meaning that the regression was performed in two rounds so that only the variables selected in the first round were used as the input for the second round of regression. Regressions were performed on the initial m/r dataset, without grouping and PCA and with Detect, Respond, Health and Risk GHSI categories (Supplementary Fig. 2); on a dataset containing Covid index and 18 demographic and health variables, without any other GHSI categories (Supplementary Fig. 3); and on m/r dataset with principal components (Fig. 2). Since the number of selected variables after the first round of regression with both excess and unexplained deaths data was relatively small (even though this time minimal MSE model was selected instead of the sparsest), the second round of regression was not performed. As an addition to LASSO and Elastic Net for the excess and unexplained deaths analysis, Forward stepwise linear regression<sup>30</sup> was performed. In this method, starting from the constant model, a predictor is added to the regression if it improves the fit significantly ( $P < 0.05$  in F-statistics). The process is repeated sequentially, adding significant and removing non-significant terms until the best fit is obtained.

**Random forest and gradient boost regressions.** Two non-parametric, decision tree-based methods – Random Forest<sup>30,51</sup> and Gradient Boost<sup>30,52</sup>—were implemented on both m/r and excess deaths datasets with principal components. These methods are of particular importance, as they can accommodate potential interaction between the variables and highly non-linear relations between the predictors and the response variable. In the analysis of disease severity, the m/r dataset with principal components was used, with a constraint that only variables correlated to m/r with  $P < 0.1$  in either Pierson, Kendall, or Spearman correlation are used as the input in the model. This way, overfitting, which would present a problem for these algorithms, is avoided. As the number of input variables was smaller in the excess/unexplained deaths dataset, preselection was not required before those regressions. Values of the hyperparameters in Random Forest and Gradient Boost methods were selected through extensive grid search, by the same method of cross-validation used for LASSO and Elastic Net regres-



**Figure 7.** Flowchart of the methods applied on each of the datasets. m/r – disease severity measure, GHSI – Global Health Security Index, demo – demographic, counts – COVID-19 counts, PCs – principal components, Detect – GHSI Detect category, Respond – GHSI Respond category, Health – GHSI Health category, Risk – GHSI Risk category, MA – median age, IM – infant mortality, GDP – gross domestic product per capita, HDI – human development index, CH – blood cholesterol level, AL – alcohol consumption, CD – prevalence of cardiovascular diseases, RBP – raised blood pressure, PL – air pollution.

sions, and a model with the minimal MSE was chosen to be retained on the entire (reduced) dataset. Unlike linear regression models, where regression coefficients measure the impact of each feature on the outcome variable, in these two methods, the contribution of each of the features is measured by feature importance. For each decision tree, the feature importance is calculated as the sum of changes in the node risk due to splits for each feature, divided by the total number of branch nodes<sup>53</sup>. The feature importance for the ensemble is obtained by averaging over all the trees<sup>54</sup>. Note that, unlike regression coefficients, feature importance does not indicate the direction of the association (that is, whether the feature affects the outcome positively or negatively). Therefore, the contribution of the relevant variables in Random Forest and Gradient Boost was estimated through Partial

Dependence (PD) plots<sup>30</sup>, which estimate the effect of a single predictor by considering the average effects of all other predictors.

**Supervised PCA.** Supervised Principle Component Analysis<sup>30</sup> was performed. In this method, PCA is done only on those variables that show sufficiently high correlation with the response variable (m/r), where a certain number of PCs is then used in multiple regression with m/r as the response. The univariate correlation constant cutoff (i.e., the number of variables retained for PCA), and the number of retained PCs, are treated as hyper-parameters, i.e., determined through cross-validation. Principal components that appeared significant in the multivariable regressions were further analyzed.

### Data availability

We confirm that all methods were carried out in accordance with relevant guidelines and regulations as publicly available data set is used. All data generated or analyzed during this study are included in this published article and its Supplementary material.

### Code availability

All source codes are available at [https://github.com/Q-bio-Belgrade/GHSI\\_puzzle](https://github.com/Q-bio-Belgrade/GHSI_puzzle).

Received: 15 March 2022; Accepted: 17 October 2022

Published online: 21 October 2022

### References

1. GHS Index. The 2019 Global Health Security Index. *GHS Index* <https://www.ghsindex.org/> (2019).
2. Nuzzo, J. B. *et al.* What makes health systems resilient against infectious disease outbreaks and natural hazards? Results from a scoping review. *BMC Public Health* **19**, 1310 (2019).
3. Meyer, D. *et al.* A checklist to improve health system resilience to infectious disease outbreaks and natural hazards. *BMJ Glob. Health* **5**, e002429 (2020).
4. Ravi, S. J. *et al.* Establishing a theoretical foundation for measuring global health security: A scoping review. *BMC Public Health* **19**, 954 (2019).
5. Aitken, T., Chin, K. L., Liew, D. & Ofori-Asenso, R. Rethinking pandemic preparation: Global Health Security Index (GHSI) is predictive of COVID-19 burden, but in the opposite direction. *J. Infect.* **81**, 318–356 (2020).
6. Amadu, I. *et al.* Assessing sub-regional-specific strengths of healthcare systems associated with COVID-19 prevalence, deaths and recoveries in Africa. *PLoS ONE* **16**, e0247274 (2021).
7. Bouba, Y. *et al.* The determinants of the low COVID-19 transmission and mortality rates in Africa: A cross-country analysis. *Front. Public Health* **9**, 751197 (2021).
8. Chakravarty, S., Grover, G. & Aggarwal, S. Association of socioeconomic and demographic factors With COVID-19 related health outcomes in SAARC nations. *Stat. Appl.* **19**, 367–386 (2021).
9. Dewi, A. *et al.* Global policy responses to the COVID-19 pandemic: Proportionate adaptation and policy experimentation: A study of country policy response variation to the COVID-19 pandemic. *Health Promot. Perspect.* **10**, 359–365 (2020).
10. Duhon, J., Bragazzi, N. & Kong, J. D. The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: A cross-country study. *Sci. Total Environ.* **760**, 144325 (2021).
11. Haider, N. *et al.* The global health security index and joint external evaluation score for health preparedness are not correlated with countries' COVID-19 detection response time and mortality outcome. *Epidemiol. Infect.* **148**, e210 (2020).
12. Ji, Y. *et al.* Are we ready to deal with a global COVID-19 pandemic? Rethinking countries' capacity based on the Global Health Security Index. *Int. J. Infect. Dis.* **106**, 289–294 (2021).
13. Kim, J. *et al.* Factors associated with the difference between the incidence and case-fatality ratio of coronavirus disease 2019 by country. *Sci. Rep.* **11**, 18938 (2021).
14. Kumru, S., Yiğit, P. & Hayran, O. Demography, inequalities and global health security index as correlates of COVID-19 morbidity and mortality. *Int. J. Health Plan. Manag.* **37**(2), 944–962. <https://doi.org/10.1002/hpm.3384> (2022).
15. Leichtweis, B. G., de Faria Silva, L., da Silva, F. L. & Peterelli, L. A. How the global health security index and environment factor influence the spread of COVID-19: A country level analysis. *One Health* **12**, 100235 (2021).
16. Maraghi, E., Saki Malehi, A. & Rahim, F. Global health security capacity against COVID-19 outbreak: An analysis of annual data from 210 countries and territories. *Jundishapur J. Health Sci.* **12**, e110814 (2020).
17. Stribling, J., Clifton, A., McGill, G. & de Vries, K. Examining the UK Covid-19 mortality paradox: Pandemic preparedness, health-care expenditure, and the nursing workforce. *J. Adv. Nurs.* **76**, 3218–3227 (2020).
18. Wong, M. C., Huang, J., Wong, S. H. & Yuen-Chun Teoh, J. The potential effectiveness of the WHO International Health Regulations capacity requirements on control of the COVID-19 pandemic: a cross-sectional study of 114 countries. *J. R. Soc. Med.* **114**, 121–131 (2021).
19. Baum, F. *et al.* Explaining covid-19 performance: What factors might predict national responses?. *BMJ* **372**, n91 (2021).
20. Dalglish, S. L. COVID-19 gives the lie to global health expertise. *Lancet* **395**, 1189 (2020).
21. Baker, H. A., Safavynia, S. A. & Evered, L. A. The 'third wave': Impending cognitive and functional decline in COVID-19 survivors. *Br. J. Anaesth.* **126**, 44–47 (2021).
22. James, N., Menzies, M. & Radchenko, P. COVID-19 second wave mortality in Europe and the United States. *Chaos* **31**, 031105 (2021).
23. Stephenson, J. COVID-19 deaths helped drive largest drop in US life expectancy in more than 75 years. *JAMA Health Forum* **3**, e215286 (2022).
24. Ravi, S. J. *et al.* The value proposition of the global health security index. *BMJ Glob. Health* **5**, e003648 (2020).
25. Nuzzo, J. B., Bell, J. A. & Cameron, E. E. Suboptimal US response to COVID-19 despite robust capabilities and resources. *JAMA* **324**, 1391–1392 (2020).
26. Gangemi, S., Billeci, L. & Tonacci, A. Rich at risk: Socio-economic drivers of COVID-19 pandemic spread. *Clinic. Mol. Allergy* **18**, 12 (2020).
27. Djordjevic, M. *et al.* Inferring the main drivers of SARS-CoV-2 Global transmissibility by feature selection methods. *GeoHealth* **5**, e2021GH000432 (2021).
28. Böttcher, L., Xia, M. & Chou, T. Why case fatality ratios can be misleading: Individual- and population-based mortality estimates and factors influencing them. *Phys. Biol.* **17**, 065003 (2020).

29. Markovic, S. *et al.* COVID-19 severity determinants inferred through ecological and epidemiological modeling. *One Health* **13**, 100355 (2021).
30. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, New York, 2009).
31. Rose, S. M. *et al.* Analysing COVID-19 outcomes in the context of the 2019 Global Health Security (GHS) Index. *BMJ Glob. Health* **6**, e007581 (2021).
32. Beaney, T. *et al.* Excess mortality: The gold standard in measuring the impact of COVID-19 worldwide?. *J R Soc. Med.* **113**, 329–334 (2020).
33. Milani, F. COVID-19 outbreak, social response, and early economic effects: A global VAR analysis of cross-country interdependencies. *J. Popul. Econ.* **34**, 223–252 (2021).
34. Chaudhry, R., Dranitsaris, G., Mubashir, T., Bartoszko, J. & Riazi, S. A country level analysis measuring the impact of government actions, country preparedness and socioeconomic factors on COVID-19 mortality and related health outcomes. *EClinicalMedicine* **25**, 100464 (2020).
35. Kong, J. D., Tekwa, E. W. & Gignoux-Wolfsohn, S. A. Social, economic, and environmental factors influencing the basic reproduction number of COVID-19 across countries. *PLoS ONE* **16**, e0252373 (2021).
36. Calderón-Larrañaga, A. *et al.* High excess mortality in areas with young and socially vulnerable populations during the COVID-19 outbreak in Stockholm Region, Sweden. *BMJ Glob. Health* **5**, e003595 (2020).
37. Stokes, A. C. *et al.* Association of health care factors with excess deaths not assigned to COVID-19 in the US. *JAMA Netw. Open* **4**, e2125287 (2021).
38. Stokes, A. C. *et al.* COVID-19 and excess mortality in the United States: A county-level analysis. *PLoS Med.* **18**, e1003571 (2021).
39. Modi, C., Böhm, V., Ferraro, S., Stein, G. & Seljak, U. Estimating COVID-19 mortality in Italy early in the COVID-19 pandemic. *Nat. Commun.* **12**, 2729 (2021).
40. Sanmarchi, F. *et al.* Exploring the gap between excess mortality and COVID-19 deaths in 67 countries. *JAMA Netw. Open* **4**, e2117359 (2021).
41. Cohen, J. & Rodgers, Y. van der M. Contributing factors to personal protective equipment shortages during the COVID-19 pandemic. *Prev Med* **141**, 106263 (2020).
42. Ranney, M. L., Griffith, V. & Jha, A. K. Critical supply shortages — the need for ventilators and personal protective equipment during the Covid-19 pandemic. *N. Engl. J. Med.* **382**, e41 (2020).
43. GHS Index. The 2021 Global Health Security Index. *GHS Index* <https://www.ghsindex.org/> (2021).
44. Oshinubi, K., Rachdi, M. & Demongeot, J. Analysis of reproduction number R0 of COVID-19 using current health expenditure as gross domestic product percentage (CHE/GDP) across countries. *Healthcare* **9**, 1247 (2021).
45. Oshinubi, K., Rachdi, M. & Demongeot, J. Modeling of COVID-19 pandemic vis-à-vis some socio-economic factors. *Front. Appl. Math. Stat.* **7**, 786983. <https://doi.org/10.3389/fams.2021.786983> (2022).
46. Waku, J., Oshinubi, K. & Demongeot, J. Maximal reproduction number estimation and identification of transmission rate from the first inflection point of new infectious cases waves: COVID-19 outbreak example. *Math. Comput. Simul.* **198**, 47–64 (2022).
47. GHS Index. COVID-19: Identifying the most vulnerable countries using the GHS index and global flight data. *GHS Index* <https://www.ghsindex.org/news/covid-19-identifying-the-most-vulnerable-countries-using-the-ghs-index-and-global-flight-data/> (2020).
48. Worldometer. Coronavirus Pandemic. <https://www.worldometers.info/coronavirus> (2020).
49. Karlinsky, A. & Kobak, D. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *Elife* **10**, e69336 (2021).
50. Meinshausen, N. Relaxed lasso. *Comput. Stat. Data Anal.* **52**, 374–393 (2007).
51. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
52. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
53. MathWorks. Estimates of predictor importance for regression tree - MATLAB. <https://www.mathworks.com/help/stats/compactregressiontree.predictorimportance.html> (2022).
54. MathWorks. Estimates of predictor importance for regression ensemble - MATLAB. <https://www.mathworks.com/help/stats/compactregressionensemble.predictorimportance.html> (2022).
55. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
56. Wei, T. & Simko, V. R package 'corrplot': Visualization of a Correlation Matrix. (2021).
57. RStudio Team. RStudio: Integrated Development Environment for R. (2020).

## Author contributions

Conceptualization: M.D., I.S., and S.M. Formal analysis: S.M. and M.D. Methodology: S.M., M.D., and A.R. Software: S.M. Supervision: M.D. and I.S. Validation: all authors. Visualization: S.M. and M.D. Manuscript writing, review, and editing: all authors.

## Funding

This work was partially supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22578-2>.

**Correspondence** and requests for materials should be addressed to M.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022