

Beware of mashup indexes: how epidemic predictors got it all wrong

 glineq.blogspot.com/2021/01/beware-of-mashup-indexes-how-epidemic.html

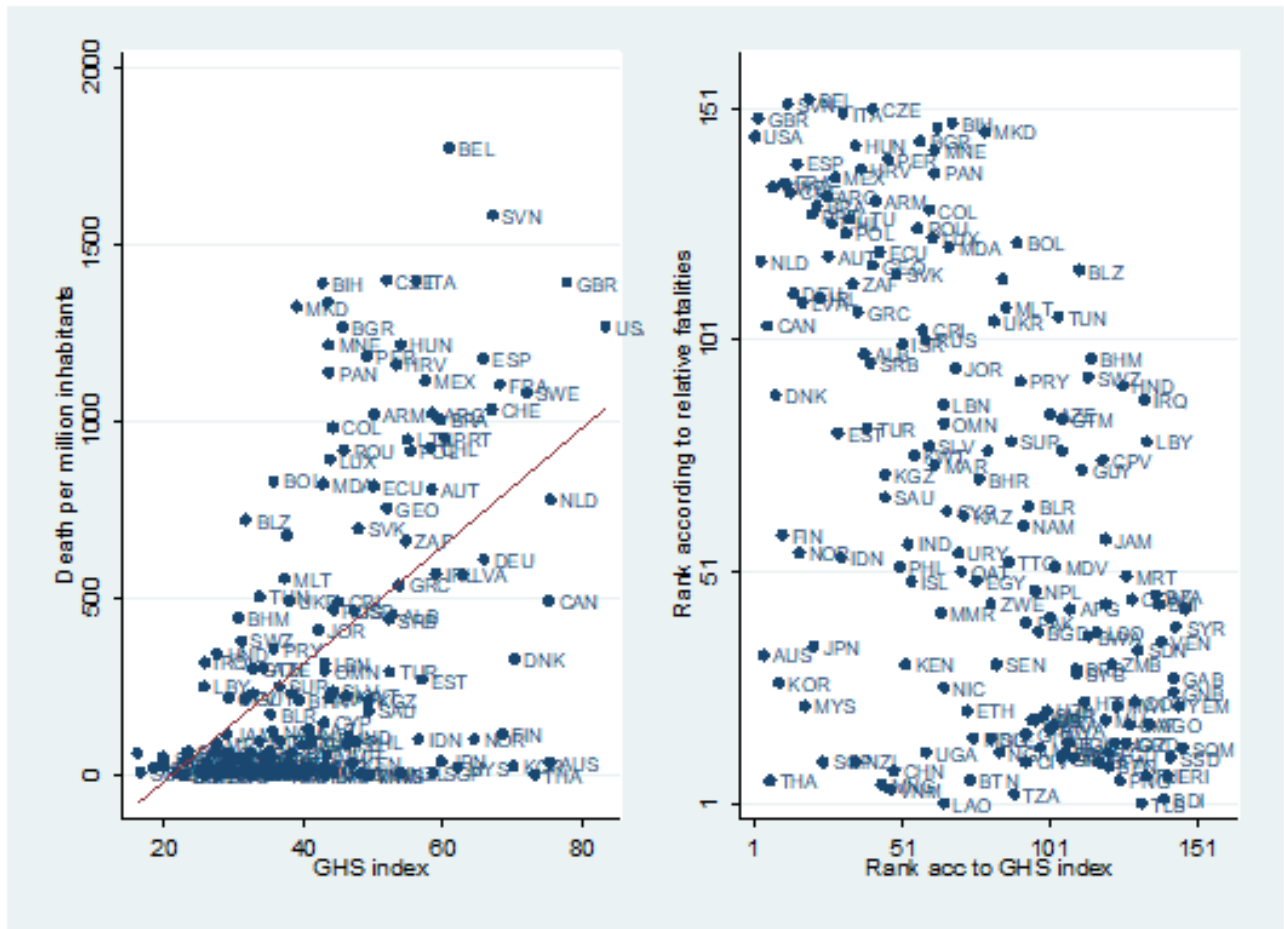
Branko Milanovic, January 2021

In October 2019, the Johns Hopkins University's Bloomberg School of Public Policy, Nuclear Threat Initiative, and the Economist Intelligence Unit published, with significant publicity, the first World Health Preparedness Report and Global Health Security Index. It claimed to study the degree of readiness to confront epidemics country-by-country. In a report of 324 pages (in addition to a Website that allows one to explore individual countries), the authors used six dimensions (or categories) to assess countries' overall preparedness: prevention of the emergence of pathogens, early detection, rapid response, robustness of the health system, commitment to improving national health capacity, and overall (country's) risk environment. The six categories themselves were built from 34 indicators, 85 subindicators and 140 questions. The authors then combined these six dimensions into an overall score, the index of Global Health Security (GHS). In this blog I shall be referring to that index.

The GHS index ranked 195 countries according to the number of points they have obtained from all the categories. The range was theoretically from 0 to 100, but the actual range went from 16.5 (the least prepared country; Equatorial Guinea) to 83.5 (the best prepared). The three top countries were the United States, UK, and the Netherlands.

As the "luck" would have it fewer than two months after the publication of the first global preparedness index, covid-19 struck the world with an unusual ferocity. So it is reasonable to ask how the experts' judgments about various countries' preparedness look compared to the actual outcomes in the fight against covid-19. For the latter, we use the number of covid-19 death per million inhabitants as of January 21, 2021. The data are collected from Worldometer. The death data are subject to many issues, from underestimation in many countries (as shown by the alternative statistic of excess deaths) to less frequent but possible overestimation. I will address these issues briefly below, and it will be indeed interesting to contrast GHS index with excess death data too.

If GHS were to predict the outcomes of covid well, we would expect that countries with a high score would have lower rates of fatalities. Or alternatively, we could disregard the cardinal measurement, and look at the ranks where we would expect that higher ranked countries according to GHS would be also higher ranked in terms of how successful they are in fighting the virus (i.e., they would have relatively fewer fatalities). The second comparison is in some sense better because its requirements are less: it requires that GHS has broadly gotten correct the ranking of countries, not necessarily that it has successfully captured the absolute differences in outcomes.



For many rich countries, the gaps between the predicted and observed performance are enormous: to give a few examples, for France 124 positions, Italy 119 positions, Canada 99 positions, Germany 97 positions. On the other hand, many countries' performances were much better than experts predicted: Vietnam was ranked No. 47, but in terms of performance it is No. 4; China 48 and 8; Cuba 95 and 19. There are thus many glaring discrepancies: Thailand and Sweden are ranked next to each other: the first recorded 1 death per million, the second 1078. Singapore and Argentina are also ranked together: Singapore had 5 deaths (per million), Argentina 1020. Several dozens similar comparisons can easily be made.

The exercise unmistakably shows that the predicted outcomes were far different (in some cases, the very opposite) of the actual outcomes. There are two possible defenses that the authors of the index can make.

First, it is very likely that the relative fatalities are mismeasured. But that argument is weakened by the fact that the differences in death rates between good and bad performers are enormous. They are often several orders of magnitude different: the deaths per one million were (as of January 21, 2021), 1266 in the USA, 1 in Thailand, 3 in China, 16 in Cuba. However, mismeasured deaths are in the latter three countries, they cannot be underestimated by 1200+ times in Thailand, 400 times in China, or 80 times in Cuba. Moreover, for the index to make sense, equalizing China, Thailand and Cuba with the US, UK and the Netherlands is not enough: one would need to show that China, Thailand and Cuba did (as the index predicted) much worse—so the death mismeasurement requirements become truly astronomical. Thus, the same exercise that might use excess death rates instead of reported deaths is almost certain to find the same lack of correlation between predicted and actual outcomes.

The second defense is that the predictions made here referred in general to epidemics, while covid-19 is a very specific epidemic that tends to be much more fatal for the elderly or obese population. According to that argument, had the authors known the characteristics of covid-19, they would have produced a better GHS index. This is quite possible. But that belies the very idea underlying the index. If each epidemic is very idiosyncratic, then what is the purpose of having a general GHS index? Suppose that the next epidemic kills people with blue eyes. Since we do not know that such an epidemic will happen, what useful information can be gleaned from GHS index? We could then as reasonably make an altogether random ranking of countries—if each epidemic is entirely specific and its effects cannot be forecast.

There is thus no escape from a sad conclusion that an index whose objective was to highlight strengths and weaknesses in the handling of potential epidemics has either entirely failed, or can be shown to have been useless. One can choose one or other of these two, equally damning, conclusions. But we should also make two additional points. First, study the (few) cases where the index successfully predicted the performance (they are in the SW corner of the second graph: Thailand, Australia, Singapore, Japan, Korea). Second, be wary of similar indexes that are produced for other variables like corruption, transparency in government, and the like. They too look “reasonable” until confronted by reality and may just reflect the experts’ echo chamber thinking.