

COMMENTARY

Open Access



# Does it matter that standard preparedness indices did not predict COVID-19 outcomes?

Michael A. Stoto<sup>1\*</sup> , Christopher D. Nelson<sup>2</sup> and John D. Kraemer<sup>1</sup>

## Abstract

A number of scientific publications and commentaries have suggested that standard preparedness indices such as the Global Health Security Index (GHSI) and Joint External Evaluation (JEE) scores did not predict COVID-19 outcomes. To some, the failure of these metrics to be predictive demonstrates the need for a fundamental reassessment which better aligns preparedness measurement with operational capacities in real-world stress situations, including the points at which coordination structures and decision-making may fail. There are, however, several reasons why these instruments should not be so easily rejected as preparedness measures.

From a methodological point of view, these studies use relatively simple outcome measures, mostly based on cumulative numbers of cases and deaths at a fixed point of time. A country's "success" in dealing with the pandemic is highly multidimensional – both in the health outcomes and type and timing of interventions and policies – is too complex to represent with a single number. In addition, the comparability of mortality data over time and among jurisdictions is questionable due to highly variable completeness and representativeness. Furthermore, the analyses use a cross-sectional design, which is poorly suited for evaluating the impact of interventions, especially for COVID-19.

Conceptually, a major reason that current preparedness measures fail to predict pandemic outcomes is that they do not adequately capture variations in the presence of effective political leadership needed to activate and implement existing system, instill confidence in the government's response; or background levels of interpersonal trust and trust in government institutions and country ability needed to mount fast and adaptable responses. These factors are crucial; capacity alone is insufficient if that capacity is not effectively leveraged. However, preparedness metrics are intended to identify gaps that countries must fill. As important as effective political leadership and trust in institutions, countries cannot be held accountable to one another for having good political leadership or trust in institutions. Therefore, JEE scores, the GHSI, and similar metrics can be useful tools for identifying critical gaps in capacities and capabilities that are necessary but not sufficient for an effective pandemic response.

**Keywords** Public health preparedness, Preparedness measures, Joint External Evaluation, Global Health Security Index, COVID-19 outcomes

\*Correspondence:

Michael A. Stoto  
stotom@georgetown.edu

<sup>1</sup> Department of Health Management and Policy, School of Health,  
Georgetown University, Washington DC, USA

<sup>2</sup> RAND Corporation and, RAND Pardee Graduate School, Santa Monica,  
CA, USA



© The Author(s) 2023. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Since the start of the pandemic, a number of scientific publications and commentaries have suggested that scores based on the WHO's States Party Self-Assessment Annual Reporting tool (SPAR) and the Joint External Evaluation (JEE) tool and the Global Health Security Index (GHSI) did not predict COVID-19 outcomes [1–6]. Citing such results, the Global Preparedness Monitoring Board, in its 2020 report, notes that “The ultimate test of preparedness is mounting an effective response,” suggesting that “our understanding of pandemic preparedness has been inadequate” [7]. To the Independent Panel for Pandemic Preparedness and Response, “the failure of these metrics to be predictive demonstrates the need for a fundamental reassessment which better aligns preparedness measurement with operational capacities in real-world stress situations, including the points at which coordination structures and decision-making may fail” [8].

## Main text

But do the analyses comparing mortality to SPAR and JEE scores and the GHSI really prove that these measures are not valid measures of a country's preparedness [9, 10]? Predictive validity, the degree to which a measure statistically predicts desirable outcomes, is a common way to assess performance measures. For example, measures of a hospital's adherence to infection control protocols should be associated with lower surgical mortality rates. Nevertheless, there are several reasons that these instruments should not be so easily rejected as preparedness measures.

First, as Stoto and colleagues [11] note, the comparability of such data (which come from the Johns Hopkins University COVID-19 Dashboard, Worldometer, and similar sources) over time and among jurisdictions is questionable due to highly variable completeness and representativeness [12, 13]. Indeed, countries that have stronger public health systems—and thus higher scores for surveillance in particular and preparedness overall—may be more likely to count COVID-19 cases and deaths completely. This would create a correlation in the wrong direction.

Second, these studies use relatively simple outcome measures, mostly based on cumulative numbers of cases and deaths at the country level and at a fixed point of time. However, a country's “success” in dealing with the pandemic is highly multidimensional – both in the health outcomes and type and timing of interventions and policies. Performance as measure by cumulative numbers of cases or deaths early in the pandemic might be contradicted by performance in later stages (e.g. when the vaccine became available). Total cases and deaths also

do not reflect differences within countries by socio-economic groups, geography, or other factors. In addition, limiting social and economic disruption is an important policy aim that is not addressed by case/death counts. In other words, the impact of capabilities that were highly effective for some groups at a specific time might not be observed in cumulative mortality figures.

Third, the analyses to assess predictive validity use cross-sectional designs with outcome data aggregated over multiple epidemic phases, which is poorly suited for evaluating the impact of interventions, especially for COVID-19 policies, where part of the challenge was to adjust policies to the emergence of new variables and new socio-economic impacts over the long course of the pandemic [11, 14]. Cross-sectional studies are even more problematical where, as in the case of COVID-19, both outcomes and interventions are highly multidimensional, as discussed in the previous paragraph.

Beyond questions of data quality and study design, there is a serious conceptual issue: a major reason that current preparedness measures fail to predict pandemic outcomes is that they do not adequately capture variations in the presence of effective political leadership needed to activate and implement existing system, instill confidence in the government's response; or background levels of interpersonal trust and trust in government institutions and country ability needed to mount fast and adaptable responses. As Bell, Fukayama, and others have noted, these factors are crucial; capacity alone is insufficient if that capacity isn't effectively leveraged [15, 16]. These factors might be labeled “social capital, and represent the difference between preparedness and resilience [17].

Ledesma and colleagues [18] recently analyzed the relationship between GHSI scores and COVID-19 outcomes and found a quite different result than the others cited in this section: higher GHSI scores were associated with lower COVID-19 deaths<sup>18</sup>. Part of the reason for the difference is that these authors avoided problems of undercounting by focusing on excess mortality estimates. They also adjusted for population age and looked at mortality over two years (2020 and 2021) rather than a limited window. Ledesma and colleagues further analyzed the results in a multivariate regression exploring the impact of the six components of the GHSI. Controlling for the other components, they found that the “risk environment score” had a stronger relationship to COVID-19 mortality than any of the others. This score includes government effectiveness, public confidence in governance, trust in medical and health advice, and related factors. Thus, it differs conceptually from the other five components, which are more traditional preparedness measures. This suggests that the GHSI is a good measure of resilience, measuring both preparedness and social capital.

## Conclusions

As always in assessing measurement systems, the purpose is critical. If the goal is, as suggested by the Independent Panel, to identify “the points at which coordination structures and decision-making may fail” [8], analyses of summary metrics can be useful.

Preparedness metrics, however, were not intended to predict outcomes. Rather, the WHO’s SPAR tool was developed to hold countries accountable for fulfilling their obligations under the International Health Regulations [19]. Other measurement systems, such as the JEE tool, are intended to identify gaps in preparedness systems and to allow countries to engage with donors and partners such as UN agencies, local and international nongovernmental organizations to target resources effectively [10]. In other words, it is about what countries do to enhance preparedness capacities, not the outcomes they achieve. For these purposes, the question is not whether the SPAR, JEE, the GHSI and similar metrics predict overall COVID-19 outcomes, but rather whether they identify gaps in preparedness capacities and capabilities that are necessary, but not sufficient, to guarantee good outcomes. As important as effective political leadership and trust in institutions are, countries cannot hold one another accountable for having good political leadership or trust in institutions.

It is also important to consider the nature of the systems that we are seeking to measure and improve. Predictive validity is a perfectly reasonable approach in systems where cause-effect relationships are relatively stable and knowable. However, such reliable “if-then” knowledge is harder to come by in highly complex and changing systems, and where the impact of a factor X may be highly conditional on the any number of contextual factors. Prediction would, of course, be very desirable in such situations. Here we are reminded of Berwick’s widely-read essay “The Science of Improvement,” where he argues that for complex social interventions – “whose effectiveness ... is sensitive to an array of influences: leadership, changing environments, details of implementation, organizational history, and much more” [20] – it is necessary to understand the complexities through detailed examples of processes and dynamics. We fear that too much focus on predictive validity – while perhaps understandable – may distract us from this task.

## Abbreviations

COVID-19	Coronavirus disease 2019 (the illness caused by the SARS-CoV-2 virus)
GHSI	Global Health Security Index
JEE	Joint External Evaluation tool
SPAR	States Party Self-Assessment Annual Reporting tool

## Authors’ contributions

All of the authors contributed to conceptualizing this research and drafting the manuscript. All authors read and approved the final manuscript.

## Funding

No external funding was received for this research.

## Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

### Ethics approval and consent to participate

This Comment does not involve human subjects research.

### Consent for publication

NA.

### Competing interests

The authors declare that they have no competing interests.

Received: 18 June 2023 Accepted: 18 September 2023

Published online: 23 September 2023

## References

- Abbey EJ, Khalifa BAA, Oduwale MO, et al. The Global Health Security Index is not predictive of coronavirus pandemic responses among Organization for Economic Cooperation and Development countries. Ashkenazi I, ed. *PLoS One*. 2020;15(10):e0239398. <https://doi.org/10.1371/journal.pone.0239398>
- Bollyky T, Crosby S, Kiernan S. Fighting a pandemic requires trust. *Foreign Affairs*. <https://www.foreignaffairs.com/articles/united-states/2020-10-23/coronavirus-fighting-requires-trust>. Published October 23, 2020.
- Haider N, Yavlinsky A, Chang YM, 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*. 2020;148:e210. <https://doi.org/10.1017/S0950268820002046>
- Milanovic B. Beware of mashup indexes: How Epidemic Predictors got it All Wrong. *Global Policy*. <https://www.globalpolicyjournal.com/blog/28/01/2021/beware-mashup-indexes-how-epidemic-predictors-got-it-all-wrong>. Published January 28, 2021. Accessed 21 Sept 2023.
- Duong DB, King AJ, Grépin KA, et al. Strengthening national capacities for pandemic preparedness: a cross-country analysis of COVID-19 cases and deaths. *Health Policy Plan*. 2022;37(1):55–64. <https://doi.org/10.1093/heapol/czab122>.
- Alhassan RK, Nketiah-Amponsah E, Afaya A, Salia SM, Abuosi AA, Nutor JJ. Global Health Security Index not a proven surrogate for health systems capacity to respond to pandemics: The case of COVID-19. *J Infect Public Health*. 2023;16(2):196–205. <https://doi.org/10.1016/j.jiph.2022.12.011>.
- Global Preparedness Monitoring Board. A world in disorder: global preparedness monitoring board annual report 2020. *Global Preparedness Monitoring Board*; 2020. <https://apps.who.int/iris/handle/10665/351720>. Accessed 21 Sept 2023.
- The Independent Panel. COVID-19: Make It the Last Pandemic; 2021. [https://theindependentpanel.org/wp-content/uploads/2021/05/COVID-19-Make-it-the-Last-Pandemic\\_final.pdf](https://theindependentpanel.org/wp-content/uploads/2021/05/COVID-19-Make-it-the-Last-Pandemic_final.pdf). Accessed 21 Sept 2023.
- World Health Organization. International Health Regulations (2005): State party self-assessment annual reporting tool. *World Health Organization*; 2005. <https://www.who.int/publications/i/item/9789240040120>. Accessed 21 Sept 2023.
- World Health Organization. Joint external evaluation tool: International Health Regulations (2005). 3rd ed. *World Health Organization*; 2016. Accessed 2 May 2023. <https://apps.who.int/iris/handle/10665/246107>. Accessed 21 Sept 2023.

11. Stoto MA, Woolverton A, Kraemer J, Barlow P, Clarke M. COVID-19 data are messy: analytic methods for rigorous impact analyses with imperfect data. *Glob Health*. 2022;18(1):2. <https://doi.org/10.1186/s12992-021-00795-0>.
12. Johns Hopkins University of Medicine. Johns Hopkins Coronavirus Resource Center. Published online March 10, 2023. <https://coronavirus.jhu.edu/>. Accessed 21 Sept 2023.
13. Worldometer. Covid-19 Coronavirus Pandemic Worldometer. Published online 2023. <https://www.worldometers.info/coronavirus/>. Accessed 21 Sept 2023.
14. Haber NA, Clarke-Deelder E, Salomon JA, Feller A, Stuart EA. Impact evaluation of coronavirus disease 2019 policy: a guide to common design issues. *Am J Epidemiol*. 2021;190(11):2474–86. <https://doi.org/10.1093/aje/kwab185>.
15. Bell J. The U.S. and COVID-19: Leading the world by ghs index score, not by response. Nuclear Threat Initiative. <https://www.nti.org/atomic-pulse/us-and-covid-19-leading-world-ghs-index-score-not-response/>. Published April 21, 2020. Accessed 21 Sept 2023.
16. Fukayama F. The pandemic and political order. *Foreign Affairs*. <https://www.foreignaffairs.com/articles/world/2020-06-09/pandemic-and-political-order>. Published June 9, 2020. Accessed 21 Sept 2023.
17. Stoto MA, Nelson CD. Measuring and Assessing Public Health Emergency Preparedness: A Methodological Primer. Published online August 18, 2023. <https://ssrn.com/abstract=4538548> or <http://dx.doi.org/10.2139/ssrn.4538548>. Accessed 21 Sept 2023.
18. Ledesma JR, Isaac CR, Dowell SF, et al. Evaluation of the Global Health Security Index as a predictor of COVID-19 excess mortality standardised for under-reporting and age structure. *BMJ Glob Health*. 2023;8(7):e012203. <https://doi.org/10.1136/bmjgh-2023-012203>.
19. World Health Organization. International Health Regulations. IHR 2005. Published online. 2005;2008:74.
20. Berwick DM. The Science of Improvement. *JAMA*. 2008;299(10):1182. <https://doi.org/10.1001/jama.299.10.1182>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

