

Perspective

Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19

Eliane Rösli,^{1,2} Brian Rice,³ and Tina Hernandez-Boussard^{2,4}

¹School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, ²Department of Medicine (Biomedical Informatics), Stanford University, Stanford, California, USA, ³Department of Emergency Medicine, Stanford University, Stanford, California, USA and ⁴Department of Biomedical Data Sciences, Stanford University, Stanford, California, USA

Corresponding Author: Tina Hernandez-Boussard, PhD, 1265 Welch Road, #245, Stanford, CA 94305-5246 (boussard@stanford.edu)

Received 11 August 2020; Editorial Decision 12 August 2020; Accepted 13 August 2020

ABSTRACT

The COVID-19 pandemic is presenting a disproportionate impact on minorities in terms of infection rate, hospitalizations, and mortality. Many believe artificial intelligence (AI) is a solution to guide clinical decision-making for this novel disease, resulting in the rapid dissemination of underdeveloped and potentially biased models, which may exacerbate the disparities gap. We believe there is an urgent need to enforce the systematic use of reporting standards and develop regulatory frameworks for a shared COVID-19 data source to address the challenges of bias in AI during this pandemic. There is hope that AI can help guide treatment decisions within this crisis; yet given the pervasiveness of biases, a failure to proactively develop comprehensive mitigation strategies during the COVID-19 pandemic risks exacerbating existing health disparities.

Key words: COVID, bias, artificial intelligence, disparities, reporting standards

INTRODUCTION

Although the COVID-19 pandemic was first thought to be the great equalizer, a mounting body of evidence suggests it further exacerbates health disparities among racial and ethnic minority groups. The disproportionate impact of this pandemic on minorities in terms of infection rate, hospitalizations, and mortality is striking.¹ Simultaneously, the global research community, placing high hopes in artificial intelligence (AI), is rushing to push out new findings as quickly as possible, creating a veritable research deluge. In this frenzy, the risk of producing biased prediction models due to unrepresentative datasets and other limitations during model development is higher than ever. If not properly addressed, propagating these biases under the mantle of AI has the potential to exaggerate the health disparities faced by minority populations already bearing the highest disease burden. We believe there is an urgent need to enforce the systematic use of reporting standards and broad sharing of code and data to address the challenges of bias in AI during the COVID-

19 pandemic. Stakeholders should come together to provide regulatory frameworks and incentives for creating a COVID-19 data resource which will help the medical community study this novel disease as quickly as possible in an environment that minimizes bias.

COVID-19 EXACERBATES RACIAL DISPARITIES

The profile of patients infected and hospitalized with COVID-19 highlights stark health disparities as racial and ethnic minority groups suffer a disproportionate burden of illness and death.¹ Two main factors are thought to underlie this burden. First, the economic and social circumstances of many minorities limit their ability to socially distance and increase their odds of infection. Second, the existing health disparities experienced by many minorities negatively affect the disease progression. Furthermore, implicit racial biases existing in clinical care are exaggerated by acute stressors such as the risk of personal infection among healthcare providers.

COVID-19 RESEARCH DELUGE

Given the unprecedented epidemiological challenges, high hopes are placed on AI to provide us with prognoses and guidance to fill the void in clinical evidence opened by the novel COVID-19 pandemic. Aggressive research efforts in the name of open science have generated a flood of non-peer-reviewed, preprint articles in online archives such as medRxiv and bioRxiv. However, this publication frenzy has recently culminated in the hasty retraction of reports in 2 of the most respected medical journals. Furthermore, a systematic review study of COVID-19 prediction models uncovered that serious shortcomings, particularly in regard to potential biases, are alarmingly widespread.² The PROBAST (Prediction model Risk Of Bias Assessment Tool) study tool rated all 66 screened models at “high” or “unclear” risk of bias. The most frequent problems encountered were unrepresentative data samples, high likelihood of model overfitting, and imprecise reporting of study populations and intended model use. There is an urgent need to balance the rapid dissemination of evidence to guide clinical decision-making with unbiased, high-quality models that truly benefit all populations in the COVID-19 era.

AI TO GUIDE RESOURCE ALLOCATION IS FRAUGHT WITH BIAS

In healthcare, bias manifests in domains involving particularly high stakes. A recent study illustrated how a widely used commercial algorithm to identify high-risk patients was significantly biased against Blacks.³ Using healthcare spending as a seemingly unbiased proxy to capture disease burden, this algorithm did not account for or ignored how systemic inequalities created from poorer access to care for Black patients resulted in less healthcare spending on Black patients relative to equally sick White patients.

Severe outbreaks of COVID-19 have sparked discussions about employing AI for optimal allocation of limited resources such as ventilators and ICU beds. However, heavy reliance on AI in such delicate decisions may provide a false sense of objectivity and fairness. These tools are built from biased data reflecting biased healthcare systems and are thus themselves also at high risk of bias—even if explicitly excluding sensitive attributes, such as race or gender. By example, models including comorbidities associated with worse outcomes in COVID-19 may perpetuate structural biases that have led to historically disadvantaged groups suffering those comorbidities disproportionately. Resource allocation models must therefore go beyond their basic utilitarian foundation to avoid further harming minority groups already suffering the most from COVID-19, based on health inequalities rooted in prior systemic discrimination. Properly developed models for care and resource allocation may be able to identify needs amongst minority patients who would otherwise remain undervalued in the current healthcare environment.

MANAGING THE CHALLENGES OF BIAS

We strongly believe transparency in reporting of AI algorithms is necessary to understand intended predictions, target populations, hidden biases, class imbalance problems, and their ability to generalize emerging technologies across hospital settings and populations. Reporting standards, such as MINIMAR, represent a first step towards accurate and responsible reporting of AI models in the COVID-19 pandemic.⁴ Intrinsic to transparency, the source code of any AI model should be shared publicly to ensure models can be

broadly applied, generalized, and compared. Strict reporting standards will facilitate the deployment of emerging COVID-19 AI models as clinical decision support tools by allowing healthcare systems to understand and mitigate potential biases associated with a particular AI model.

Simultaneously, we advocate the development of regulatory frameworks facilitating broad knowledge and data sharing, potentially through a national COVID-19 data resource. Never before has there been a greater need for timely, accurate and comprehensive healthcare data. Rich, granular information is required to better understand risk factors associated with COVID-19 infection and clinical escalation, including data capturing individual behavior (eg, social distancing practices), genetics, social and environmental factors, as well as clinical data.⁵ Without such a dataset, our ability to produce and validate generalizable models applicable across systems is limited. Instead, what we will find—and what we are seeing—are models that are developed at elite, affluent academic healthcare systems that are not representative of the general US population and fail external validity. COVID-related data are being produced at incredible speed but these data remain siloed within each country or academic institute, in part due to a lack of interoperability and in part due to a lack of appropriate incentives.⁶ During the COVID-19 pandemic, there is an opportunity to break down current barriers of data sharing and build a dataset that genuinely represents a diverse population that is accessible to the global research community. As an example, the Medical Information Mart for Intensive Care (MIMIC) is a publicly available, deidentified, and broadly studied dataset for critical care patients.⁷ A MIMIC-equivalent for COVID-19 from diverse data sources could incentivize urgently needed data sharing and interoperability to enable diverse, population-based tailored therapy—a step that could decisively reduce bias and disparities in healthcare while bolstering clinical judgment and decision-making.

CONCLUSION

The COVID-19 pandemic has left healthcare systems struggling to evolve care in real time while operating in a void of evidence. There is hope that AI can help guide treatment decisions, including the allocation of scarce resources within this crisis. However, the hasty adoption of AI tools bears great risk due to biased, unrepresentative training data and a lack of a regulated COVID-19 data resource for validation purposes. Given the pervasiveness of biases, a failure to proactively develop comprehensive mitigation strategies during the COVID-19 pandemic risks exacerbating existing health disparities and hindering the adoption of AI tools capable of improving patient outcomes.

AUTHOR CONTRIBUTIONS

Dr. Hernandez-Boussard takes responsibility for the integrity of the data and the accuracy of the data analysis.

- Concept and design: All authors
- Interpretation of data: All authors
- Drafting of the manuscript: ER
- Critical revision of the manuscript: All authors.
- Administrative, technical, or material support: THB.
- Study supervision: THB.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES:

1. Pan D, Sze S, Minhas JS, *et al.* The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. *EClinicalMedicine* 2020; 23: 100404.
2. Wynants L, Van Calster B, Bonten MMJ, *et al.* Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020; 369: m1328.
3. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019; 366 (6464): 447–53.
4. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (Minimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc* 2020. doi: 10.1093/jamia/ocaa088.
5. Artiga S, Hinton E. Beyond health care: the role of social determinants in promoting health and health equity. *Health* 2019; 20: 10.
6. Cosgriff CV, Ebner DK, Celi LA. Data sharing in the era of COVID-19. *Lancet Digit Health* 2020; 2 (5): e224.
7. Johnson AE, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3 (1): 160035.