Those designing healthcare algorithms must become actively anti-racist

Kellie Owens[1], **Alexis Walker**[2,✉]

[1]Data & Society Research Institute, New York, NY, USA.

[2]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, USA.

Abstract

Many widely used health algorithms have been shown to encode and reinforce racial health inequities, prioritizing the needs of white patients over those of patients of color. Because automated systems are becoming so crucial to access to health, researchers in the field of artificial intelligence must become actively anti-racist. Here we list some concrete steps to enable anti-racist practices in medical research and practice.

Calls for attention to the effects of systemic racism in and on medicine have populated health literature for years. But as the murders of Breonna Taylor, George Floyd and countless other Black Americans have sparked protests for Black liberation across the country, the world faces a moment of reckoning that few can ignore—including the US healthcare system.

Medical and public-health scholars have consistently made clear that police violence is a health issue[1,2]. But social scientists have demonstrated that discriminatory policing can be done without police officers, as it has been built deeply into widespread algorithmic systems[3–5]. Unfortunately, there are many examples of racially discriminatory algorithms in medicine as well; multiple automated health technologies have been shown to encode and reinforce health inequities—from a heart-failure risk score that inappropriately categorizes Black patients as being in need of less care[6] to algorithms that are poor at detecting cancers in people of color[7].

This trend has been exemplified well by the research from Ziad Obermeyer and colleagues that found that a widely used commercial algorithm identifying patients for extra support with complex health needs prioritizes white patients above Black patients[8]. This algorithm is meant to predict which patients will have the greatest health needs and thus would derive the most benefit from enrolling in 'high-risk care-management' programs. In theory, the algorithm would help health systems allocate resources to achieve optimal health outcomes by using existing patient data to predict future healthcare needs. Importantly, the algorithm does not use race as a predictor in its model.

How, then, does the algorithm produce stark racial discrimination? That is because it uses health costs as a proxy for health needs. Patients with higher predicted health costs are presumed to be sicker, and thus receive a higher risk score, and may be recommended for enrollment in special programming. But due to histories of racism, the US health system tends to spend less money on Black patients than on white patients. Black patients have less access to health services and are generally valued less than white patients by health providers and systems[9]. By assuming that health costs are a suitable and race-neutral proxy for health needs, the algorithm indicated that Black patients at a given risk score were significantly sicker than white patients were at the same score. The authors estimated that this miscalculation reduced the number of Black patients who should have been enrolled in high-risk care programs by more than half.

It can be presumed that the algorithm's developers did not explicitly intend to perpetuate discrimination against Black patients. But lack of social consciousness can still produce disastrous results. Obermeyer et al. suggest a fairly simple way to reduce racial bias in these types of health algorithms: changing the data fed into algorithms and the labels given to that data[8]. But producing proper labels requires a concerted and in-depth understanding of how structural discrimination operates in society. Health researchers and providers often do not have the training and expertise to identify or address these structural factors.

It is not enough for researchers to make their analyses 'race neutral' by eliminating race as a variable in their prediction models; instead, researchers need to take a proactive, explicitly anti-racist approach to data collection, analysis and prediction. Just as 'implicit bias' training for police does little to change racist behavior[10]—in large part because departmental cultures do not fully support the lessons of anti-racism—healthcare must also go beyond 'window dressing' training. Health systems need to be held accountable for equitable outcomes. Because ignorance can so easily lead to the perpetuation of systemic racism, health researchers and providers need to receive long-term, in-depth training—not just a short 'bias' training—to ensure deep, critical thinking about systemic racism.

Fortunately, there are a number of existing social-science fields with decades of experience with this type of scholarship and education, including critical race studies, critical data studies, and science and technology studies[3,4,9,11–13]. Experts from these fields should be included in the education of new health researchers and practitioners, and they should become active members of research teams testing new models for predicting health outcomes. This is a crucial element in the development of an anti-racist culture.

The proliferation of racist healthcare algorithms also identifies weaknesses in the current applications of medical ethics. The institutionalized medical ethics frameworks tend to focus on ensuring the good intentions and behaviors of researchers and providers. Research and clinical ethics infrastructures focus on identifying and eliminating conflicts of interest, misconduct or questionable scientific practices as a means of ensuring ethical outputs. Perhaps this would seem sufficient to prevent racial discrimination in a previous era during which racist intentions were often quite explicit. But failing to anticipate the structural bias in a dataset or the social implications of a product is not likely to be qualified as scientific misconduct—although perhaps it should be.

Restructuring medical ethics to better address structural discrimination in healthcare would require that justice be truly centered as a key principle in both clinical ethics and research ethics. Racial justice means achieving equity in the face of centuries of discrimination and violence against communities of color; equity cannot be achieved through equal treatment but requires over-compensation in prioritizing benefits to those who have so long been marginalized. This involves structural changes, such as universal, affordable healthcare, to help address the differences in spending that underlie the injustices highlighted in the study by Obermeyer et al.[8] And although social scientists and ethicists have called for such transformations[13,14], many health practitioners and researchers continue to lack proficiency in the basic terminologies and concepts of racial justice. These competencies need to be required for medical licensing and accreditation.

A radically socially conscious approach is needed to eliminate subtle but widespread discrimination. We urge each reader to take action in their institution today and, with continued vigor in the upcoming months and years, to induce a true culture shift that would stop algorithmic design from perpetuating inequities. For this, systemic education of health practitioners and investigators on issues of racial justice is needed, as well as standards for anti-racist development and analysis of research.

References

1. Hardeman R, Medina E & Boyd RN Engl. J. Med 383, 197–199 (2020).

2. Cooper H & Fullilove MJ Urban Health 93, 1–7 (2016).

3. Benjamin R Race After Technology: Abolitionist Tools for the New Jim Code (John Wiley & Sons, 2019).

4. Eubanks V Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (St. Martin's Press, 2018).

5. O'neil C Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Broadway Books, 2016).

6. Vyas DA, Eisenstein LG & Jones DSN Engl. J. Med 10.1056/NEJMms2004740 (2020).

7. Noor P Br. Med. J 368, m363 (2020). [PubMed: 32051165]

8. Obermeyer Z, Powers B, Vogeli C & Mullainathan S Science 366, 447–453 (2019). [PubMed: 31649194]

9. Benjamin R Science 366, 421–422 (2019). [PubMed: 31649182]

10. Forscher PS, Lai CK & Axt JR et al. J. Pers. Soc. Psychol 117, 522–559 (2019). [PubMed: 31192631]

11. Noble S Algorithms of Oppression: How Search Engines Reinforce Racism (NYU Press, 2018).

12. Boyd D & Crawford K Inf. Commun. Soc 15, 662–679 (2012).

13. Benjamin R Sci. Technol. Human Values 41, 967–990 (2016).

14. Powers M & Faden R Social Justice: The Moral Foundations of Public Health and Health Policy (Oxford University Press, 2006).