Racial bias in cost data leads an algorithm to underestimate health care needs of Black patients.

PHOTO: FATCAMERA/GETTY IMAGES

SOCIAL SCIENCE

Assessing risk, automating racism

A health care algorithm reflects underlying racial bias in society

By **Ruha Benjamin**

As more organizations and industries adopt digital tools to identify risk and allocate resources, the automation of racial discrimination is a growing concern. Social scientists have been at the forefront of studying the historical, political, economic, and ethical dimensions of such tools (*1*–*3*). But most analysts do not have access to widely used proprietary algorithms and so cannot typically identify the precise mechanisms that produce disparate outcomes. On page 447 of this issue, Obermeyer *et al.* (*4*) report one of the first studies to examine the outputs and inputs of an algorithm that predicts health risk, and influences treatment, of millions of people. They found that because the tool was designed to predict the cost of care as a proxy for health needs, Black patients with the same risk score as White patients tend to be much sicker, because providers spend much less on their care overall. This study contributes greatly to a more socially conscious approach to technology development, demonstrating how a seemingly benign choice of label (that is, health cost) initiates a process with potentially life-threatening results. Whereas in a previous era, the intention to deepen racial inequities was more explicit, today coded inequity is perpetuated precisely because those who design and adopt such tools are not thinking carefully about systemic racism.

Obermeyer *et al.* gained access to the training data, algorithm, and contextual data for one of the largest commercial tools used by health insurers to assess the health profiles for millions of patients. The purpose of the tool is to identify a subset of patients who require additional attention for complex health needs before the situation becomes too dire and costly. Given increased pressure by the Affordable Care Act to minimize spending, most hospital systems now utilize predictive tools to decide how to invest resources. In addition to identifying the precise mechanism that produces biased predictions, Obermeyer *et al.* were able to quantify the racial disparity and create alternative algorithmic predictors.

Practically speaking, their finding means that if two people have the same risk score that indicates they do not need to be enrolled in a "high-risk management program," the health of the Black patient is likely much worse than that of their White counterpart. According to Obermeyer *et al.*, if the predictive tool were recalibrated to actual needs on the basis of the number and severity of active chronic illnesses, then twice as many Black patients would be identified for intervention. Notably, the researchers went well beyond the algorithm developers by constructing a more fine-grained measure of health outcomes, by extracting and cleaning data from electronic health records to determine the severity, not just the number, of conditions. Crucially, they found that so long as the tool remains effective at predicting costs, the outputs will continue to be racially biased by design, even as they may not explicitly attempt to take race into account. For this reason, Obermeyer *et al.* engage the literature on "problem formulation," which illustrates that depending on how one defines the problem to be solved—whether to lower health care costs or to increase access to care—the outcomes will vary considerably.

To grasp the broader implications of the study, consider this hypothetical: The year is 1951 and an African American mother of five, Henrietta Lacks, goes to Johns Hopkins Hospital with pain, bleeding, and a knot in her stomach. After Lacks is tested and treated with radium tubes, she is "digitally triaged" (*2*) using a new state-of-the-art risk assessment tool that suggests to hospital staff the next course of action. Because the tool assesses risk using the predicted cost of care, and because far less has commonly been spent on Black patients despite their actual needs, the automated system underestimates the level of attention Lacks needs. On the basis of the results, she is discharged, her health rapidly deteriorates,

Department of African American Studies, Princeton University, Princeton, NJ, USA. Email: ruha@princeton.edu

Downloaded from http://science.sciencemag.org/ on February 4, 2020

Published by AAAS

and, by the time she returns, the cancer has advanced considerably, and she dies.

This fictional scenario ends in much the same way as it did in reality, as those familiar with Lacks's story know well (*5*–*7*). But rather than getting assessed by a seemingly race-neutral algorithm applied to all patients in a colorblind manner, she was admitted into the Negro wing of Johns Hopkins Hospital during a time when explicit forms of racial discrimination were sanctioned by law and custom—a system commonly known as Jim Crow. However, these are not two distinct processes, but rather Jim Crow practices feed the "New Jim Code"—automated systems that hide, speed, and deepen racial discrimination behind a veneer of technical neutrality (*1*).

Data used to train automated systems are typically historic and, in the context of health care, this history entails segregated hospital facilities, racist medical curricula, and unequal insurance structures, among other factors. Yet many industries and organizations well beyond health care are incorporating automated tools, from education and banking to policing and housing, with the promise that algorithmic decisions are less biased than their human counterpart. But human decisions comprise the data and shape the design of algorithms, now hidden by the promise of neutrality and with the power to unjustly discriminate at a much larger scale than biased individuals.

For example, although the Fair Housing Act of 1968 sought to protect people from discrimination when they rent or buy a home, today social media platforms allow marketers to explicitly target advertisements by race, excluding racialized groups from the housing market without penalty (*8*). Although the federal government brought a suit against Facebook for facilitating digital discrimination in this manner, more recently the U.S. Department of Housing and Urban Development introduced a rule that would make it harder to fight algorithmic discrimination by lenders, landlords, and others in the housing industry. And unlike the algorithm studied by Obermeyer *et al*., which used a proxy for race that produced a racial disparity, targeted ads allow for explicit racial exclusion, which violates Facebook's own policies. Yet investigators found that the company continued approving ads excluding "African Americans, mothers of high school kids, people interested in wheelchair ramps, Jews, expats from Argentina and Spanish speakers," all within minutes of an ad submission (*8*). So, whether it is a federal law or a company policy, top-down reform does not by itself dampen discrimination.

Labels matter greatly, not only in algorithm design but also in algorithm analysis. Black patients do not "cost less," so much as they are valued less (*9*). It is not "something about the interactions that Black patients have with the healthcare system" that leads to poor care, but the persistence of structural and interpersonal racism. Even health care providers hold racist ideas, which are passed down to medical students despite an oath to "do no harm" (*10*). The trope of the "noncompliant (Black) patient" is yet another way that hospital staff stigmatize those who have reason to question medical authority (*11*, *12*). But a "lack of trust" on the part of Black patients is not the issue; instead, it is a lack of trustworthiness on the part of the medical industry (*13*). The very designation "Tuskegee study" rather than the official name, U.S. Public Health Service Syphilis Study at Tuskegee, continues to hide the agents of harm. Obermeyer *et al*. mention some of this context, but passive and sanitized descriptions continue to hide the very social processes that make their study consequential. Labels matter.

As researchers build on this analysis, it is important that the "bias" of algorithms does not overshadow the discriminatory context that makes automated tools so important in the first place. If individuals and institutions valued Black people more, they would not "cost less," and thus this tool might work similarly for all. Beyond this case, it is vital to develop tools that move from assessing individual risk to evaluating the production of risk by institutions so that, ultimately, the public can hold them accountable for harmful outcomes. ■

REFERENCES AND NOTES

1. R. Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Polity Press, 2019).
2. V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018).
3. S. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (NYU Press, 2018).
4. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, *Science* **366**, 447 (2019).
5. K. Holloway, *Private Bodies, Public Texts: Race, Gender, and a Cultural Bioethics* (Duke Univ. Press, 2011).
6. H. Landecker, *Sci. Context* **12**, 203 (1999).
7. R. Skloot, *The Immortal Life of Henrietta Lacks* (Broadway Books, 2011).
8. J. Angwin, A. Tobin, M. Varner, "Facebook (still) letting housing advertisers exclude users by race," *ProPublica*, 21 November 2017; www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin.
9. E. Glaude Jr., *Democracy in Black: How Race Still Enslaves the American Soul* (Crown Publishers, 2016).
10. K. M. Bridges, *Reproducing Race: An Ethnography of Pregnancy as a Site of Racialization* (Univ. California Press, 2011).
11. A. Nelson, *Body and Soul: The Black Panther Party and the Fight Against Medical Discrimination* (Univ. Minnesota Press, 2011).
12. H. Washington, *Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present* (Harlem Moon, Broadway Books, 2006).
13. R. Benjamin, *People's Science: Bodies and Rights on the Stem Cell Frontier* (Stanford Univ. Press, 2013).

10.1126/science.aaz3873

Published by AAAS