

Analysis of Racial Bias in Healthcare Risk Prediction Algorithms

Introduction

In recent years, the integration of algorithms into healthcare systems has revolutionized the management of patient populations. These algorithms are designed to predict healthcare needs, enabling health systems to allocate resources efficiently and improve patient outcomes. However, the reliance on automated decision-making tools raises critical concerns about fairness and equity, particularly regarding potential biases that may inadvertently disadvantage certain demographic groups. The research article “Dissecting racial bias in an algorithm used to manage the health of populations” addresses this pressing issue by examining racial disparities embedded within a widely utilized health risk prediction algorithm. This midterm paper summarizes the methodology employed in Obermeyer et al.’s study and addresses the normative concerns arising from their findings. By analyzing the authors’ approach to identifying and quantifying racial bias present in healthcare-determining algorithms, this paper highlights the importance of ethical considerations in deploying predictive algorithms in healthcare.

Summary of the Paper’s Method

Obermeyer et al. investigate the presence of racial bias in a commercial risk prediction algorithm used by health systems to identify patients with complex health needs. The algorithm, which estimates future healthcare costs based on historical insurance claims data, is intended to allocate resources to high-risk patients through care management programs. In this context, a care management program is a coordinated system providing more in-depth and personalized healthcare. The study utilizes a comprehensive dataset from a large academic hospital comprising 6079 Black patients and 43,539 White patients, observed over 11,929 and 88,080 patient-years, where a patient-year represents data collected for an individual patient in a calendar year.

The Algorithm

The authors do not disclose the name of the algorithm used to obtain risk scores for the data associated with a patient. They claim that it is a widely used commercial tool that is

employed on a large scale.

Data Collection and Preparation

The researchers obtained algorithmic risk scores for each patient year based on claims data from the preceding year. The dataset included demographic information, insurance type, diagnosis and procedure codes, medications, and detailed cost data. Race was excluded from the algorithm’s input features to prevent direct racial discrimination.

Identifying Racial Disparities

The authors compared the predicted risk scores with actual health outcomes and healthcare costs to assess the algorithm’s calibration across races. Health outcomes were measured using a comprehensive score derived from electronic health records, encompassing diagnoses, laboratory results, and vital signs indicative of chronic illness severity. Healthcare costs were obtained from insurance claims data, including outpatient visits, emergency visits, hospitalizations, and total expenditures.

Analysis of Bias

The study focused on calibration bias, specifically examining whether the expected health outcome given a risk score was consistent across racial groups. The authors found that, at identical risk scores, Black patients exhibited significantly higher illness burdens than White patients. Put simply, Black people needed to be affected by many more health conditions to be enrolled in the health program compared to White patients. For instance, at the 97th percentile of risk scores, Black patients had an average of 4.8 chronic conditions compared to 3.8 in White patients. This disparity suggests that the algorithm underestimates the health needs of Black patients relative to their risk scores.

Mechanisms of Bias

The authors explored potential mechanisms underlying the observed bias, identifying that the algorithm’s reliance on predicted healthcare costs inadvertently introduced racial bias. Despite similar risk scores, Black patients generated lower healthcare costs than White patients due to systemic barriers to accessing care. This discrepancy indicates that the algorithm’s proxy for health needs—future costs—fails to account for unequal access and utilization of

health services among different racial groups. To quantify the impact of this bias, Obermeyer et al. conducted counterfactual simulations where the algorithm was adjusted to eliminate the predictive gap between Black and White patients. These simulations demonstrated that rectifying the bias could substantially increase the proportion of Black patients receiving additional care management support, from 17.7% to 46.5% at the highest risk threshold.

Summary of Findings

The study reveals that the algorithm, while ostensibly unbiased in predicting healthcare costs across races, inadvertently perpetuates significant racial disparities in health outcomes. Black patients, at any given risk score, exhibit a higher burden of chronic illnesses compared to their White counterparts. This finding underscores a critical flaw in using health care costs as a proxy for health needs, as it fails to capture the actual health status of marginalized populations who may face systemic barriers to accessing care. Consequently, the algorithm prioritizes White patients for care management programs more than Black patients with comparable or worse health needs, thereby exacerbating existing health inequities.

Description of Normative Consideration

The primary normative concern highlighted by Obermeyer et al. is the ethical implication of racial bias in predictive algorithms within healthcare. This concern centers on justice and equity in health care delivery. In this context, justice refers to the fair distribution of health resources, ensuring that all individuals receive appropriate care based on their health needs rather than socio-demographic factors.

Importance of Addressing Racial Bias

The presence of racial bias in health risk prediction algorithms has impactful consequences. It perpetuates existing disparities by systematically underestimating the health needs of Black patients, leading to inadequate resource allocation and support. Moreover, the ethical principle of non-maleficence, which mandates that actions should not harm individuals, is violated when algorithms contribute to unequal care. By failing to accurately identify high-risk Black patients, the algorithm indirectly causes harm by denying them the necessary support to manage their health conditions effectively.

Societal Implications

Beyond individual patient outcomes, racial bias in healthcare algorithms contributes to broader societal inequities. It reinforces structural racism within the healthcare system, eroding trust among marginalized communities and exacerbating health disparities. This mistrust can lead to reduced engagement with health services, further deteriorating health outcomes for these populations.

Conclusion

Obermeyer et al.'s study provides a compelling examination of racial bias in a widely used health risk prediction algorithm, revealing significant disparities that have substantial ethical and societal implications. By analyzing the algorithm's methodology and its impact on different racial groups, the authors underscore the necessity of integrating ethical considerations into the development and deployment of predictive tools in healthcare.

References

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>