Inequity: Algorithm Case Study

Simone Collier

University of Toronto

Bias in Algorithms

Many widely used algorithms reproduce gender and racial disparities.

- Algorithms used for job searches are less likely to present highly paid positions to women.
- Facial recognition systems used in law enforcement have a harder time recognizing the faces of women and people of colour.
- Language is encoded with bias towards gender and race in natural language processing algorithms.

These biases are generally introduced into the algorithms by the methods used to build them or though the data used to train them.

Algorithms Health Systems

Algorithms are used in the healthcare system in the US to identify patients that could benefit from high-risk care management programs.

- The programs are meant to improve the care of patients with high health risk.
- The algorithm predicts health care costs, not health condition.
- The prediction tool is applied to $\sim$200 million people in the US yearly.

Algorithms Health Systems

The are several assumptions and methods that under pin the algorithm itself.

- Assumes that those in need of the most care will benefit most from the program.
- Assumes cost of care is a proxy for health.
- Past data is used to predict future health care needs.

Recreating the Algorithm

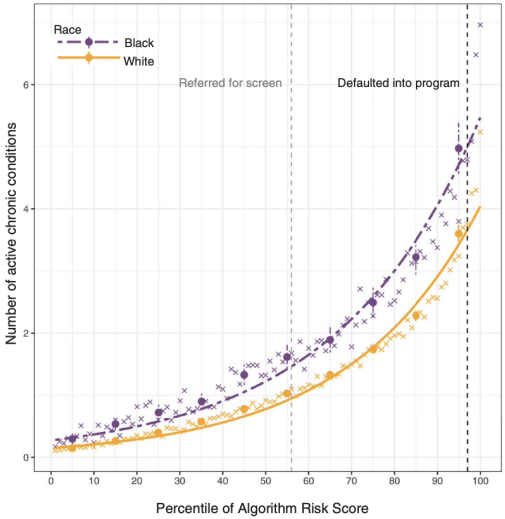
The algorithms such as these are proprietary, meaning they cannot be examined by outsiders for potential issues.

However, Obsermeyer et al. (2019) created an algorithm similar to the ones used in US health care.

- The algorithm gives a risk score to each primary care patient enrolled in a risk-based contract.
- The risk score is calculated based on insurance claims data from the previous year and the patients health.
- Patient health is predicted based on health record data including diagnoses, lab studies, and vital signs that capture the severity of chronic conditions.

Algorithm Predictions

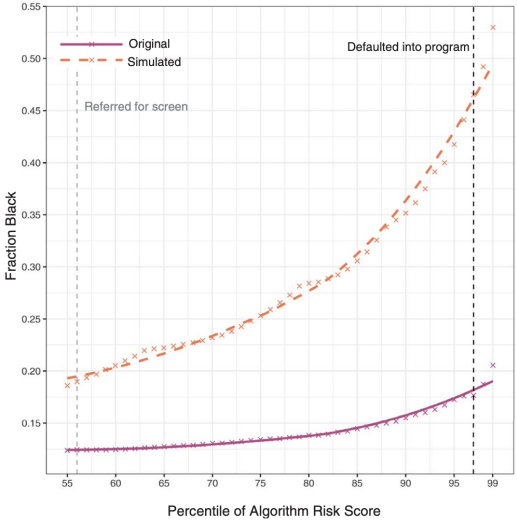
- The algorithm predicts the same health risk (cost) for White people in greater health than Black people in poorer health.
- Black people have 23% more chronic illnesses than white people at the point where both groups are auto-identified for program enrolment (97th percentile of risk score).



Algorithm Predictions

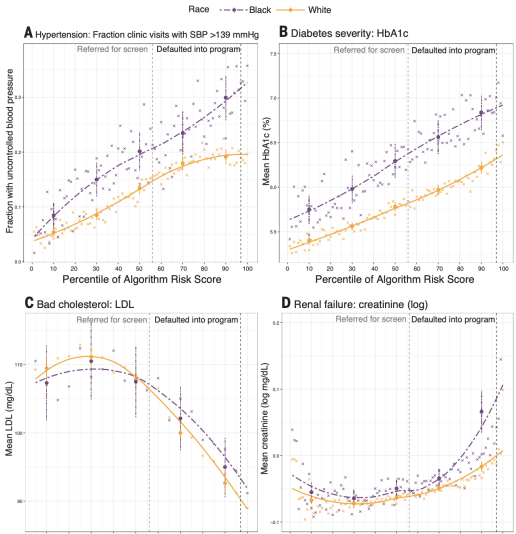
The fraction patients at or above a given risk score that are black.

- The "original" curve is from the original algorithm.
- The "simulated" curve is simulated from a scenario that removes algorithmic bias.
  - At each threshold of risk, healthier white patients above the threshold are replaced with less healthy black patients until the two groups are equal.



Algorithm Predictions

- Biomarkers indicate the severity of common chronic illnesses.
- At any level of algorithm predictions, black patients are substantially less healthy than white patients according to the biomarkers.
- The differences in severity of diabetes imply differences in mortality of 30%!



Reasons for Bias

The algorithm's prediction on health is a prediction on health *costs*.

- Cost seems a good measure of health since sicker patients require and receive more care *on average*.
- At every level of algorithm predicted risk, the predicted health cost is the same for black and white patients and it is an accurate prediction.
- So, there are substantial disparities in health conditional on risk but little disparity on costs.

Black patients generate lower health care costs than white patients with the same heath conditions.

- Socioeconomic barriers to accessing health care.
- Mistrust in health care system leading to lower uptake of recommendations from providers.

Other methods

This highlights the importance of the choice of label on which the algorithm is trained. Possible label alternatives for this algorithm:

- Avoidable future costs such as emergency visits and hospitalizations.
- Health indicators such as the number of chronic health conditions.

The authors of this study contacted the algorithm manufacturer and they worked together to create a new model with 84% less bias.

Reference

Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. Science 366: 447-453.