

Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations

Obermeyer, Powers, Vogeli, Mullainathan
(2019)

You've trained your Precision Medicine model – now what?

I'm Bryce Rowland!

Background

Algorithms can reproduce racial and gender disparities

Study evaluating a commercial risk prediction algorithm

Algorithm identifies patients for “high-risk care management programs”

Assumption: Greatest care needs → Greatest benefit

Researchers accessed algorithm inputs, outputs, and eventual outcomes

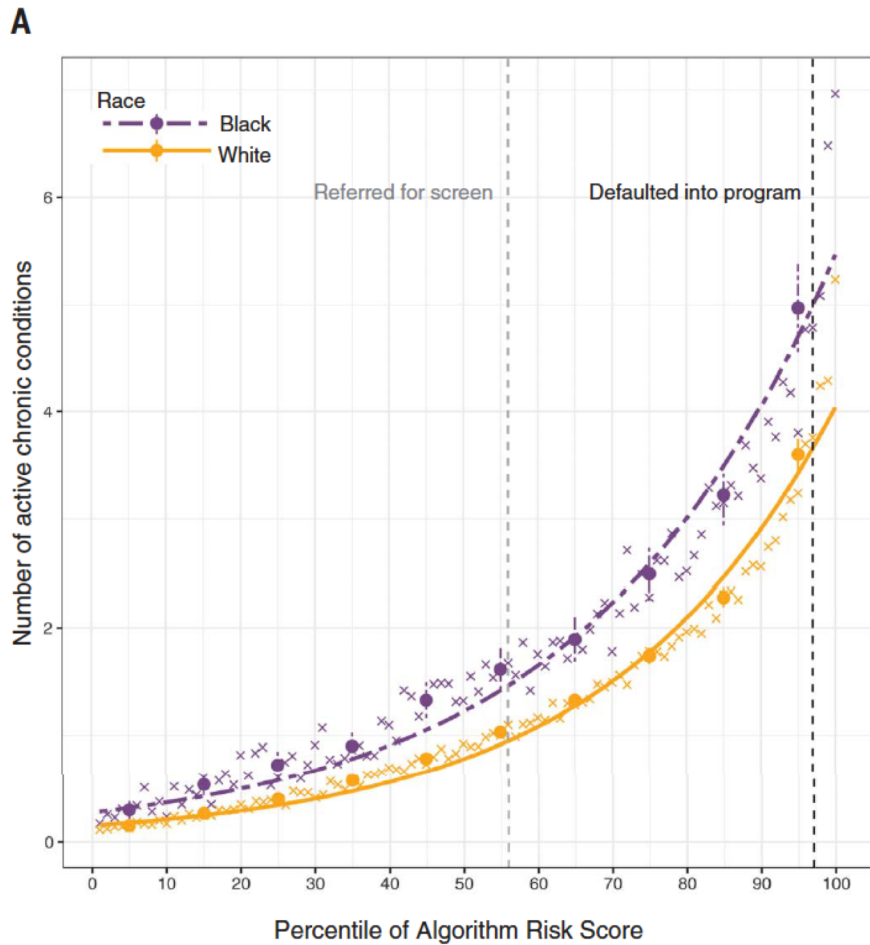
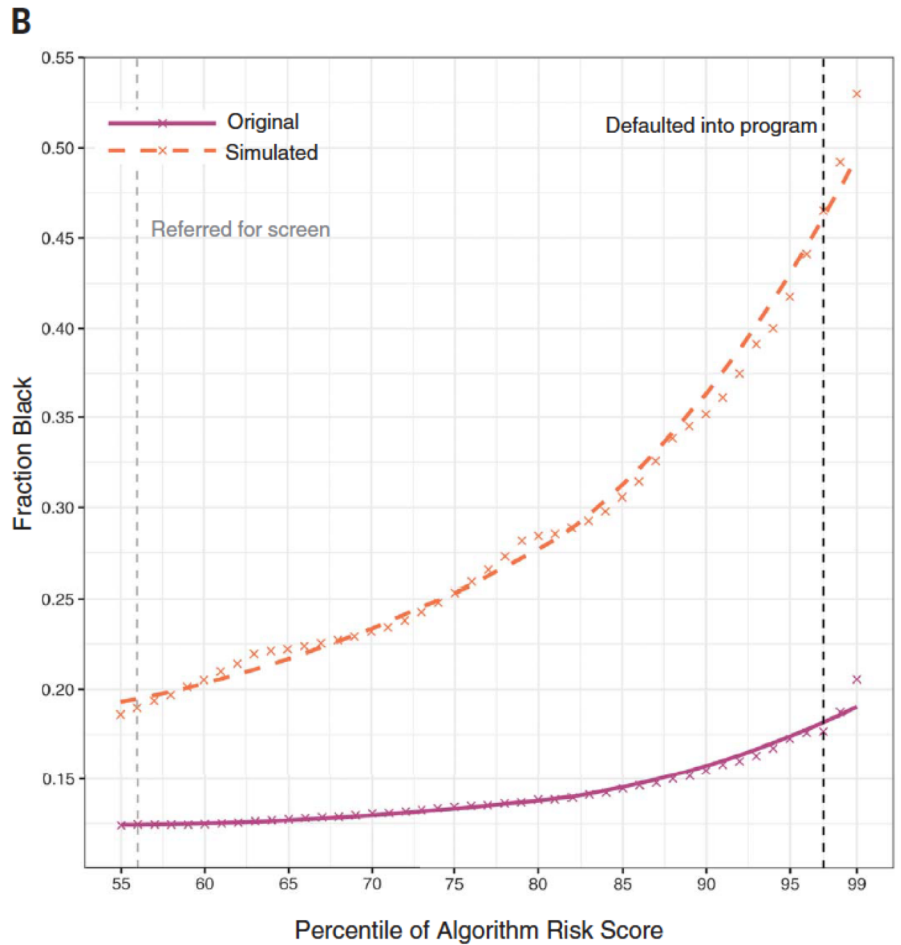


Fig. 1. Number of chronic illnesses versus algorithm-predicted risk, by race. (A) Mean number of chronic conditions by race, plotted against algorithm risk score. (B) Fraction of Black patients at or above a given risk score for the original algorithm (“original”) and for a simulated scenario that removes algorithmic bias (“simulated”: at each threshold of risk, defined at a given percentile on the x axis, healthier Whites above the threshold are



replaced with less healthy Blacks below the threshold, until the marginal patient is equally healthy). The × symbols show risk percentiles by race; circles show risk deciles with 95% confidence intervals clustered by patient. The dashed vertical lines show the auto-identification threshold (the black line, which denotes the 97th percentile) and the screening threshold (the gray line, which denotes the 55th percentile).

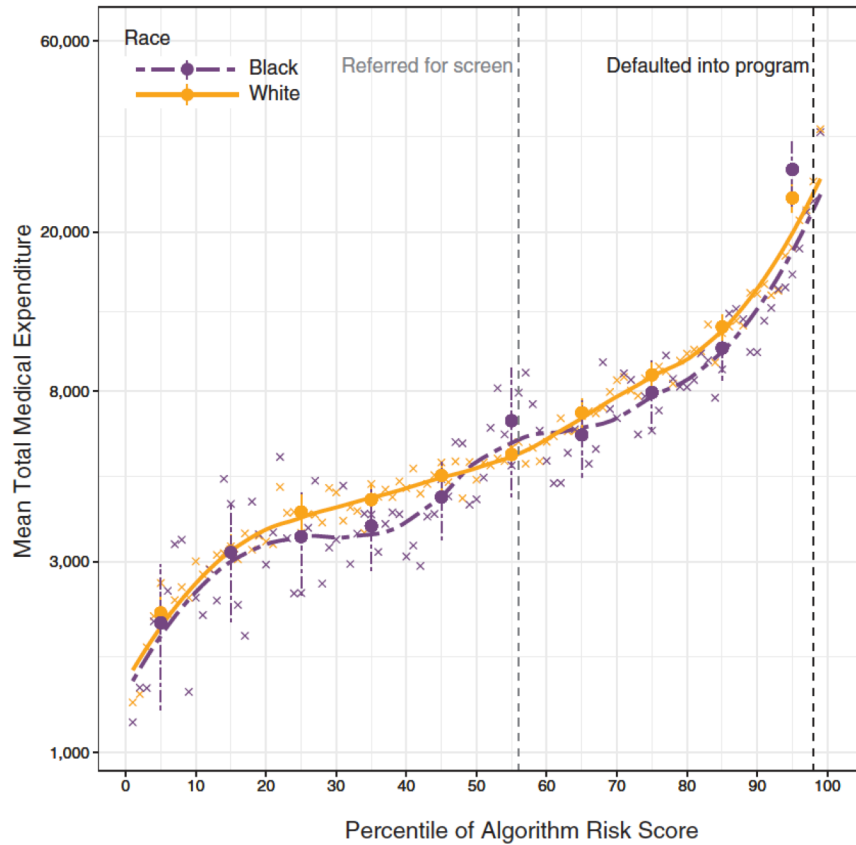
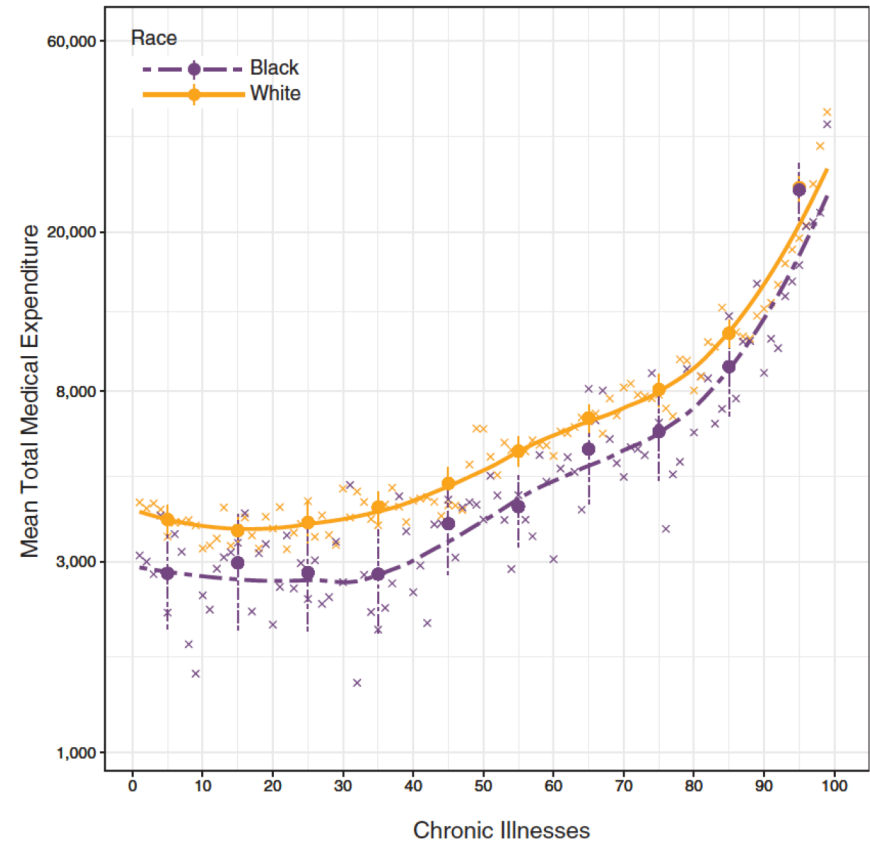
A**B**

Fig. 3. Costs versus algorithm-predicted risk, and costs versus health, by race. (A) Total medical expenditures by race, conditional on algorithm risk score. The dashed vertical lines show the auto-identification threshold (black line: 97th percentile) and the screening threshold (gray line: 55th percentile). (B) Total medical expenditures by race, conditional on number of chronic conditions. The × symbols show risk percentiles; circles show risk deciles with 95% confidence intervals clustered by patient. The y axis uses a log scale.

Take Home Message

Label Bias is a real danger in health algorithms.

Statistical assumptions, at scale, have societal implications.

Domain knowledge is essential to prevent the perpetuation of health inequity.