Reading group - Obermeyer et al (2019)

Dissecting racial bias in an algorithm used to manage the health of populations

Laura Morris
9/30/2021

Starting question

- For encoding racial bias in a decision-making process, is the ‘left’ or ‘right’ hand side more important?
- (ie: which **predictors** are used, or which **outcome**)

Rest of this presentation:

- Findings of this paper
- How the analysis worked
- Context of current debates in applied ML/AI

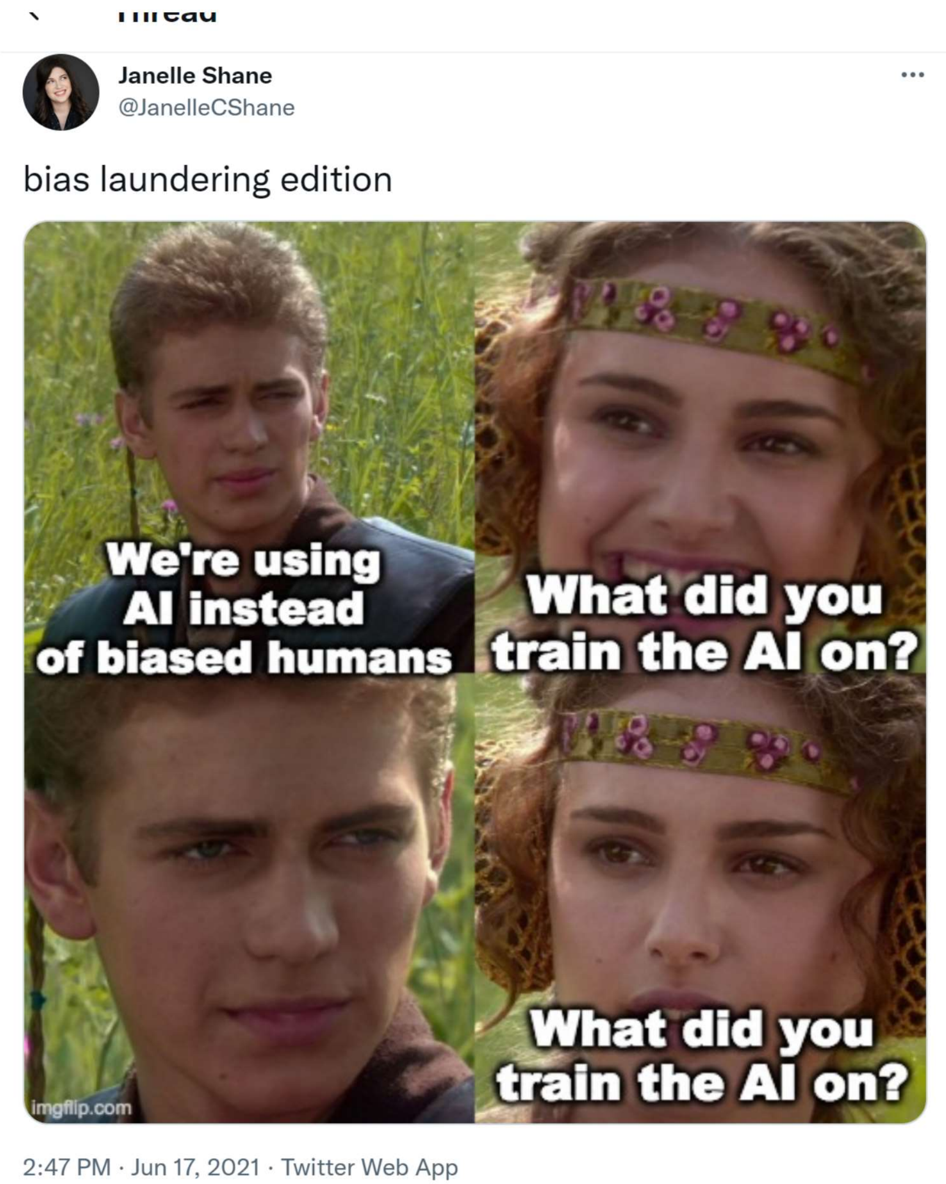
Paper summary

- An algorithm designed to identify and help patients with complex health needs was shown to exhibit racial bias by under-treating Black patients.
- Bias arose because the algorithm predicted health care costs rather than direct severity of illness
- Unequal access to care means that we spend less money caring for Black patients than for White patients.

Context: algorithms in health care

- Increasingly used to target care
- Particularly common for high-cost care and directing patients to “high-risk care management” programs
- Great example of public use of algorithms for research: high-volume data is typically more available than other sectors
- Scope of what outcomes are ‘predictable’ is expanding over time
- Generates theoretical insights: e.g. origins of low-value care (moral hazard or just misprediction?)

Context: algorithmic bias



- Are algorithms biased? Common defence: no, it's the data
- 'bias laundering' = using flawed historical data to train algorithms, later deployed as 'neutral' or 'objective' policy instruments
- Examples in other sectors: facial recognition algorithms, predictive policing

Bias in health care algorithms

“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…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” - Ruha Benjamin (Science, 2019)

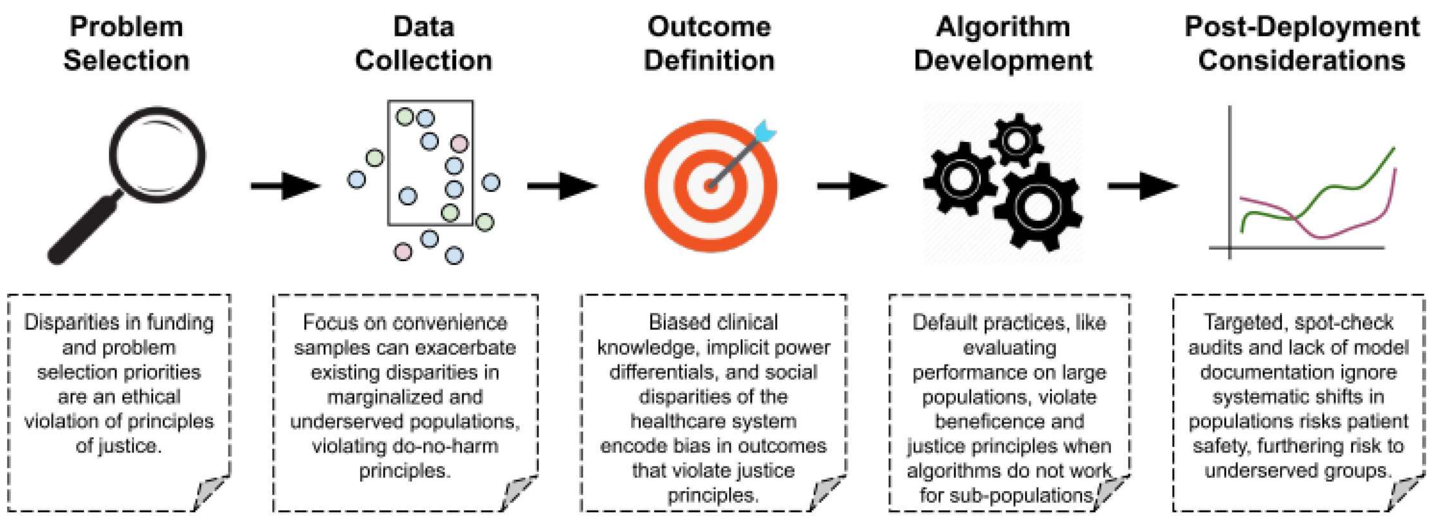


Figure 1

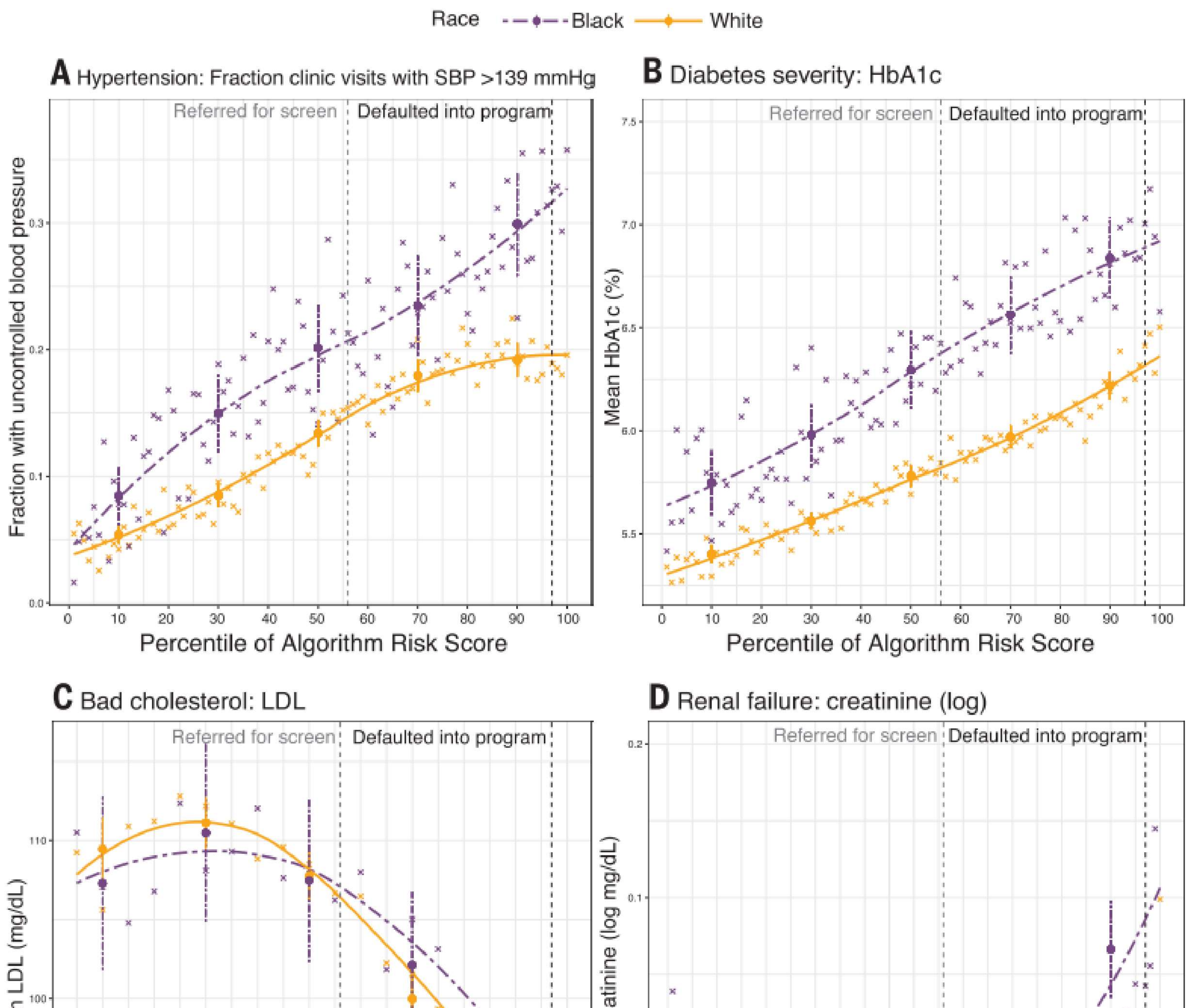
We motivate the five steps in the ethical pipeline for health care model development. Each stage

Source: Chen, I. et al. (2020). Ethical Machine Learning in Healthcare

What this paper shows

- Black patients with same risk score as White patients tend to be much sicker, because providers spend less on their care
- If the predictive tool were recalibrated to need (number and severity of active chronic illnesses), **twice as many** Black patients would be identified for intervention

Findings hold for other biomarkers of health



Algorithm is successfully predicting health care costs

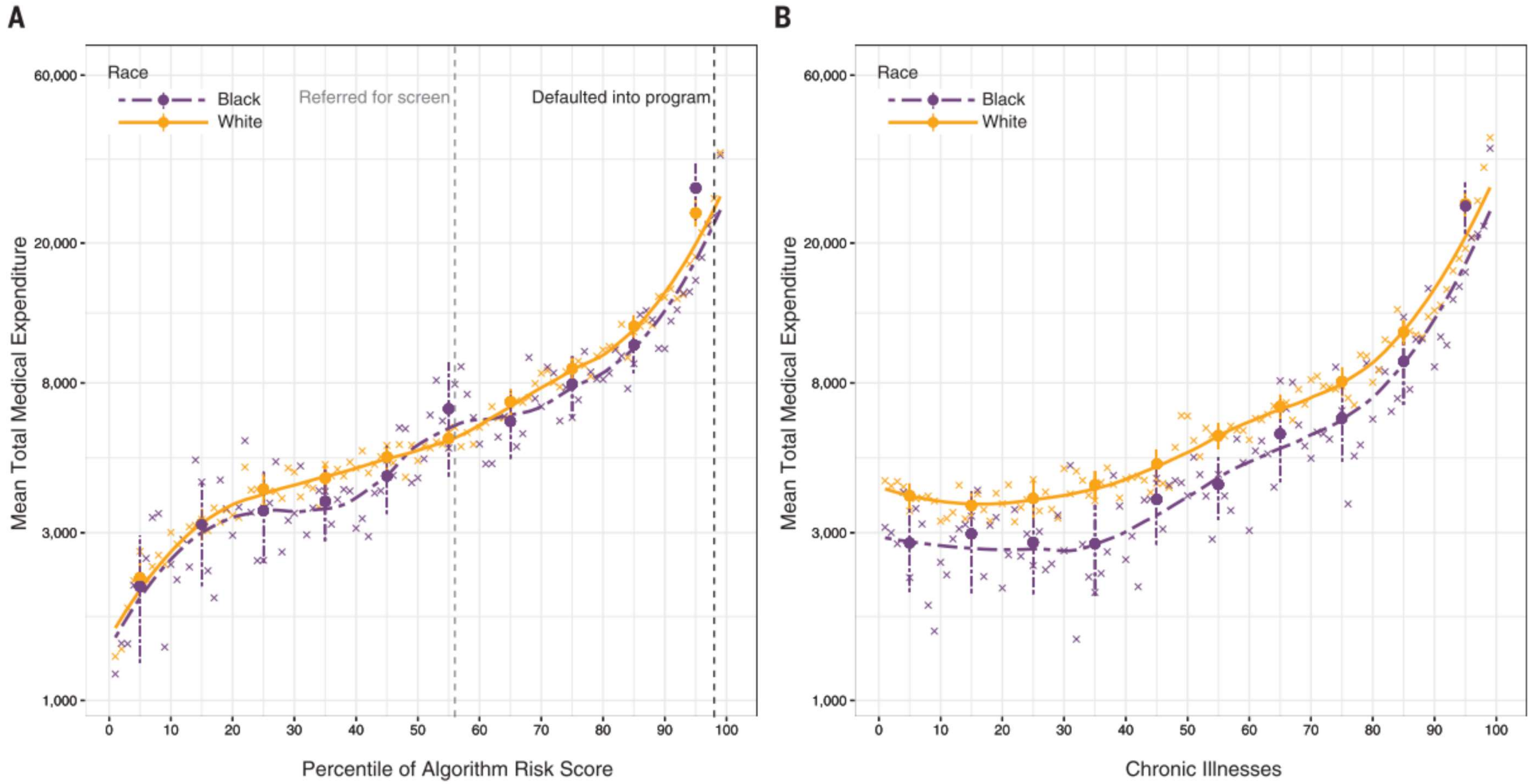


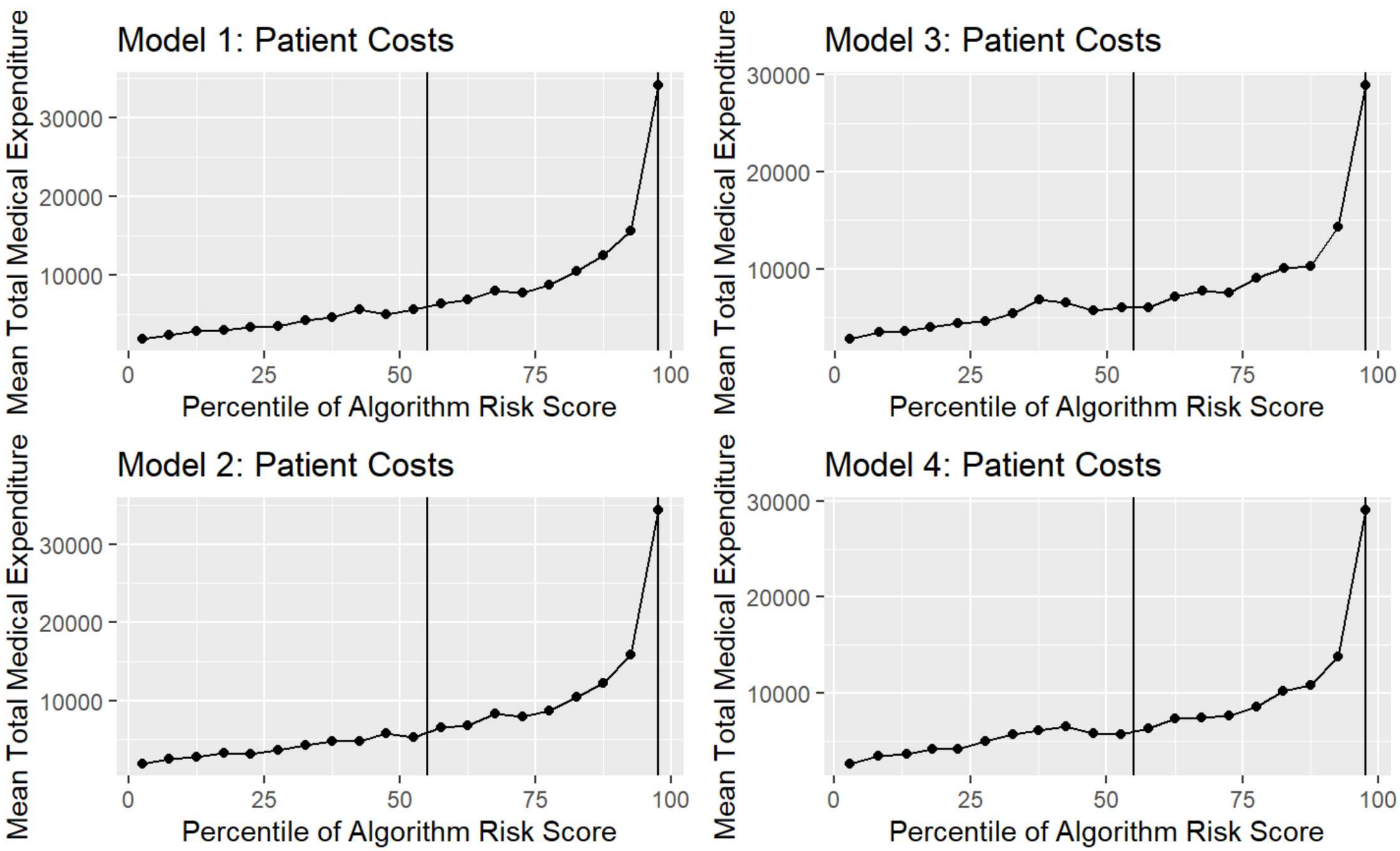
Fig. 3. Costs versus algorithm-predicted risk, and costs versus health, by race. (**A**) Total medical expenditures by race, conditional on algorithm risk score. The dashed vertical lines show the auto-identification threshold (black line: 97th percentile) and the screening threshold (gray line: 55th 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.

How the results were generated

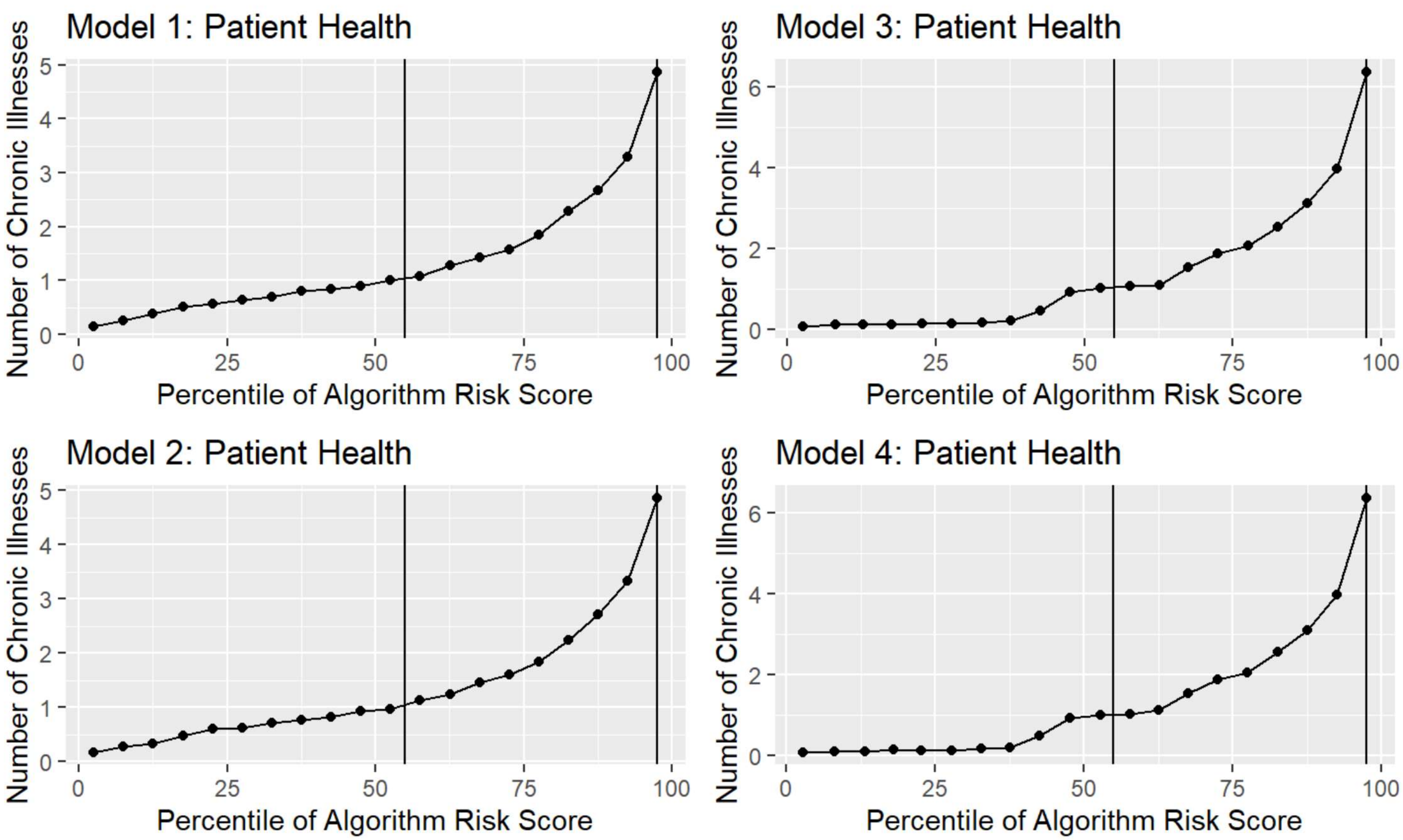
- Using insurance provider data, train several prediction algorithms: some including the patient’s race and others explicitly leaving out the patient’s race.
- Statistical models estimated using the training data set, both with and without race as predictor:
- Random forests to predict the “label” of patient costs
- Random forest to predict the “label” of patient health
- Predictions from each algorithm converted into percentile ranks (risk scores)

Graphs: patient costs as outcome

(Code for demo: EC50a teaching [Bruich], Sp 2021)



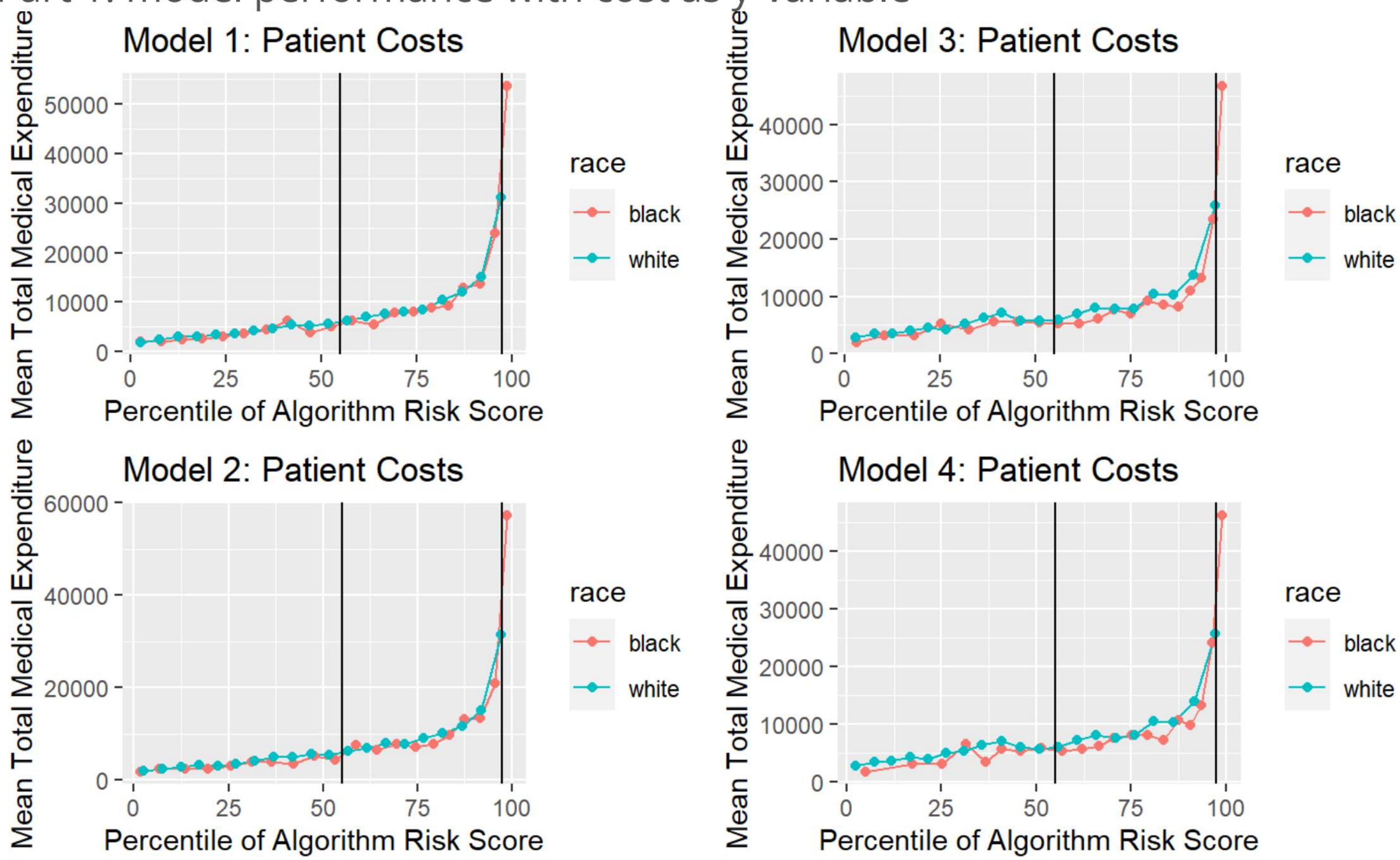
Graphs 2: patient health as outcome



Code for demo: EC50a teaching [Bruich], Sp 2021

Graphs stratified by race (1)

Part 1: model performance with cost as y-variable



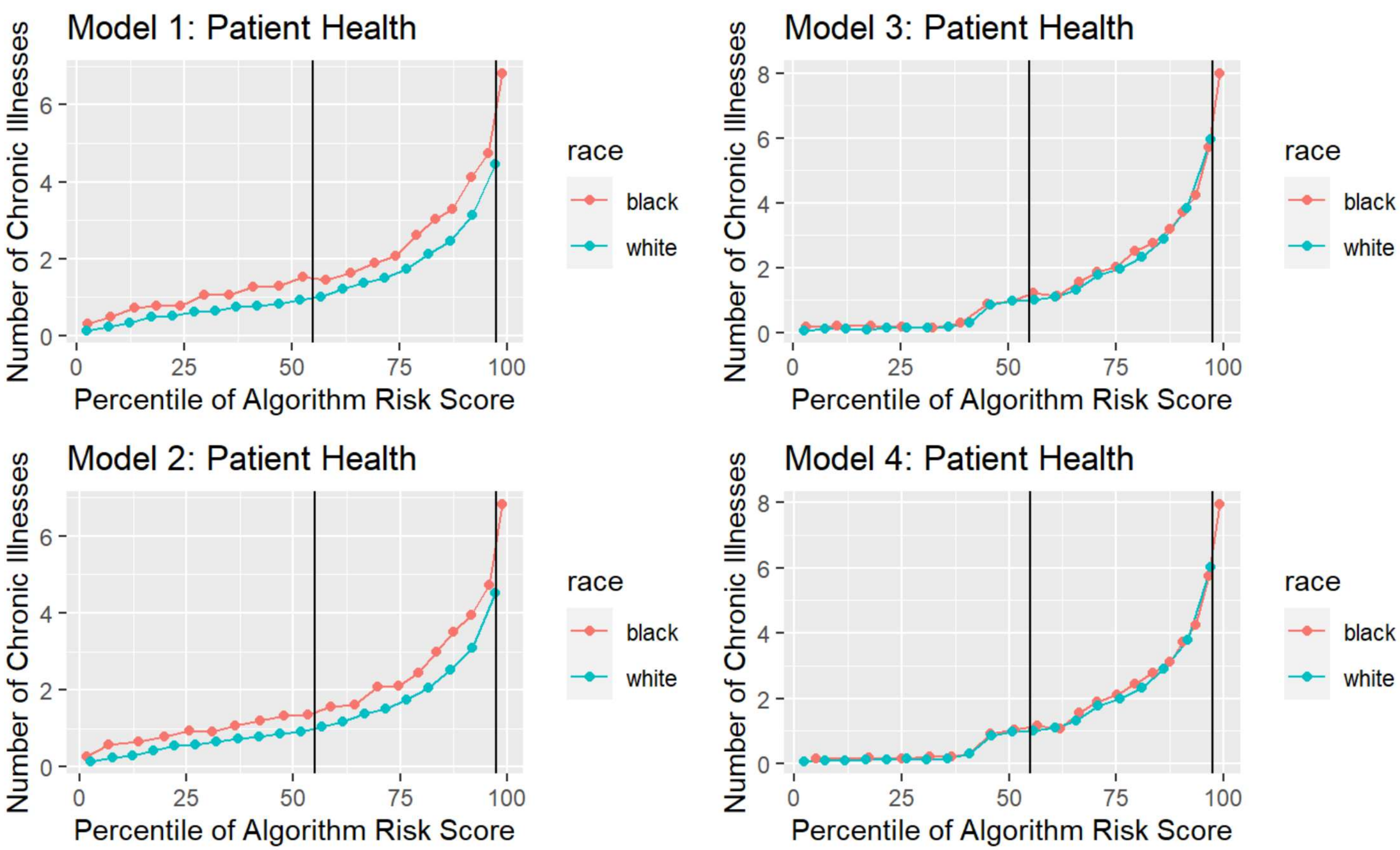
(Left column: models trained to predict costs)

(Top pair: models excluding race as predictor)

Code for demo: EC50a teaching [Bruich], Sp 2021

Graphs stratified by race (2)

Part 2: model performance with health as y-variable



(Left column: models trained to predict costs)

(Top pair: models excluding race as predictor)

Code for demo: EC50a teaching [Bruich], Sp 2021

Contributions of paper

- Changed health group practice! Researchers actually constructed new + more fine grained algorithm to nearly eliminate racial inequity
- Paper engaged lit on ‘problem formulation’ in data science, which stresses importance of carefully defining the problem to be solved (e.g. lowering costs vs increasing access)
- Frames algorithmic bias clearly, evidence for not worrying too much about predictors.

Caveat/question: is this actually something very new? Label choice bias is well established in health policy.

WHAT’S MEASURED IS WHAT MATTERS: TARGETS AND GAMING IN THE ENGLISH PUBLIC HEALTH CARE SYSTEM

GWYN BEVAN AND CHRISTOPHER HOOD

What's next?

- Can algorithms de-bias biased data? (Rambachan et al, 2020)
- Are algorithm-driven processes *actually* less transparent than human-driven? What about accountable?
- How to regulate development/use of algorithms? May still be optimal to let them use any predictors in training data.

Links / related reading

- Algorithmic Justice League https://www.ajl.org/ (Joy Buolamwini, MIT)
- Algorithmic Bias Playbook @ Chicago Booth (Obermeyer et al): written for C-suite leaders and technical teams in health orgs, offers guide to audit processes for algorithmic bias
- Chen, I. et al (2020). Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science*, 4
- Kleinberg, J. et al (2019). Algorithms as discrimination detectors. *PNAS* 117(48)?
- Rajkomar, A. et al (2018). Ensuring Fairness in Machine Learning to Advance Health Equity. *Annals of internal medicine* 169(12), 866–872
- Rambachan, A. et al (2020). An Economic Perspective on Algorithmic Fairness. *AEA Papers and Proceedings 2020*, 110: 91–95