

Dissecting Algorithmic Bias

Ziad Obermeyer
UC Berkeley

Joint work with:
David Cutler,
Jure Leskovec,
Sendhil Mullainathan,
Emma Pierson,
Brian Powers,
Christine Vogeli

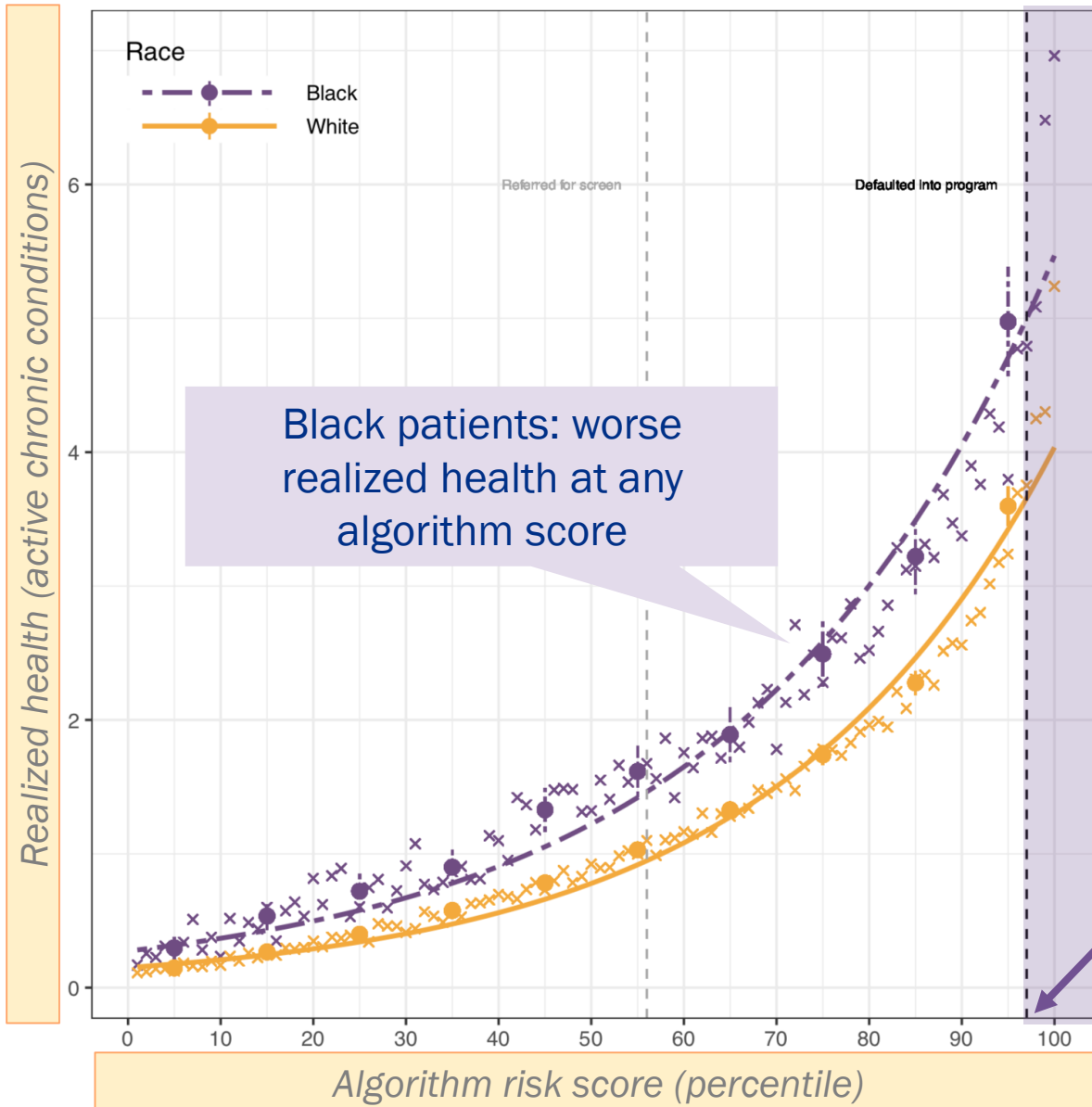
Algorithms in health care

- ▶ Many great uses of algorithms in health
 - Risk prediction: What will happen
 - Diagnosis: Likelihood that patient has a disease
 - ...
- ▶ Many ways this can go wrong
 - Algorithms can reproduce, scale up bias

Example: Targeting extra help for complex patients

- ▶ Complex, chronically ill patients have high costs, poor care
 - ‘High-risk care management’ can help
 - But expensive – so targeting critical
- ▶ Algorithms are used everywhere for this
 - Specific software we study: 70 million patients/year (US)
 - Market estimates: 150-200 million patients/year (US)
- ▶ Common goal: Find patients who are going to get sick
 - As measured by future health care costs
 - Target help to highest-risk now (and screen out low risk)

We studied 'racial bias'



- ▶ Principle: Same score
→ Treated the same
 - Should have same needs
- ▶ Color of their skin should not matter
- ▶ How much bias?
 - Fast track today: 18% Black
 - **Without bias: 47% Black**

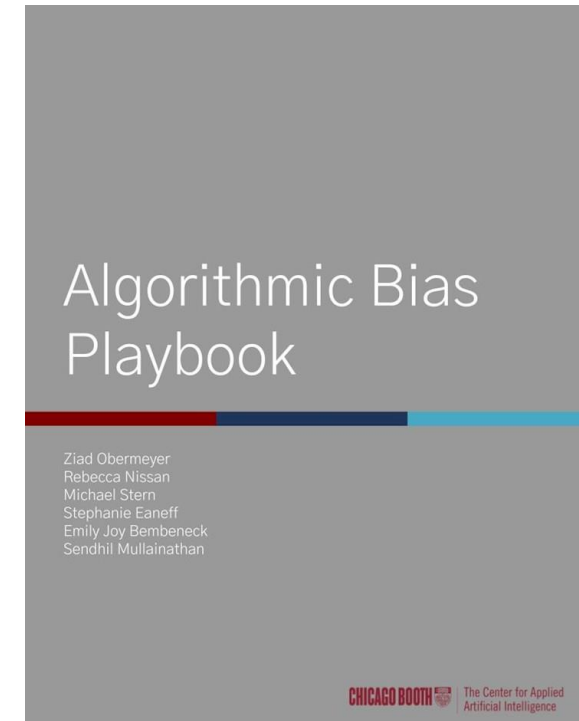
Dissecting the bias

- ▶ Algorithm predicts health costs accurately
 - For Black and White patients alike
 - ...Exactly as we designed it to do
- ▶ But cost is a biased proxy for health
 1. White patients have better access to health care
 2. The health system treats Black patients differently
- ▶ Result: Black patients with the same health cost less
 - So accurate cost prediction = biased health prediction

Our 'playbook'—inspired by work over past 2 years

- ▶ Bad news: We found bias almost everywhere we looked
 - Population health resource allocation
 - Clinical disease prediction
 - Operational decisions
- ▶ Good news: Almost all fixable
 - By retraining on less biased target

[DOWNLOAD THE PLAYBOOK](#)



Lessons for preventing bias in the real world

1. Inventory all algorithms in use
 - Common finding: no one knows what's being used



State of California
Office of the Attorney General

ROB BONTA
ATTORNEY GENERAL

August 31, 2022

Dear Hospital CEO:

I write today regarding our shared interest in ensuring that California healthcare consumers are able to access medical services that meet their needs, and are not disproportionately limited by race or other protected characteristics. To that end, the Office of the Attorney General seeks to ascertain how healthcare facilities and other providers are addressing racial and ethnic disparities in commercial decision-making tools and algorithms.

Lessons for preventing bias in the real world

1. Inventory all algorithms in use
 - Common finding: no one knows what's being used
2. Document performance
 - Ideal target performance overall and by group
 - Biased or inaccurate algorithms must be fixed or deleted
3. Organizational structure for algorithm oversight
 - No one in C-suite has responsibility
 - Oversight is a strategic priority—even bigger than just bias

Lessons we took away from this

- ▶ Getting the exact target for algorithms right matters a lot
 - Cost is a bad proxy for need
- ▶ This might sound familiar to you
 - Because the same issues come up in quality measures
- ▶ Health equity strategy—how to pick the right metrics?
 - HbA1c... what about undiagnosed diabetics?
 - Statin adherence... what about those never prescribed?

Pain is concentrated in society's most disadvantaged

- ▶ Black patients: more pain
- ▶ Is this just because they have more problems like arthritis?
 - Two patients, similar x-rays
 - Compare pain scores
- ▶ Black, lower-income, lower-education: still have more pain
- ▶ Where does this 'pain gap' come from?



Two Views

Society



The knee



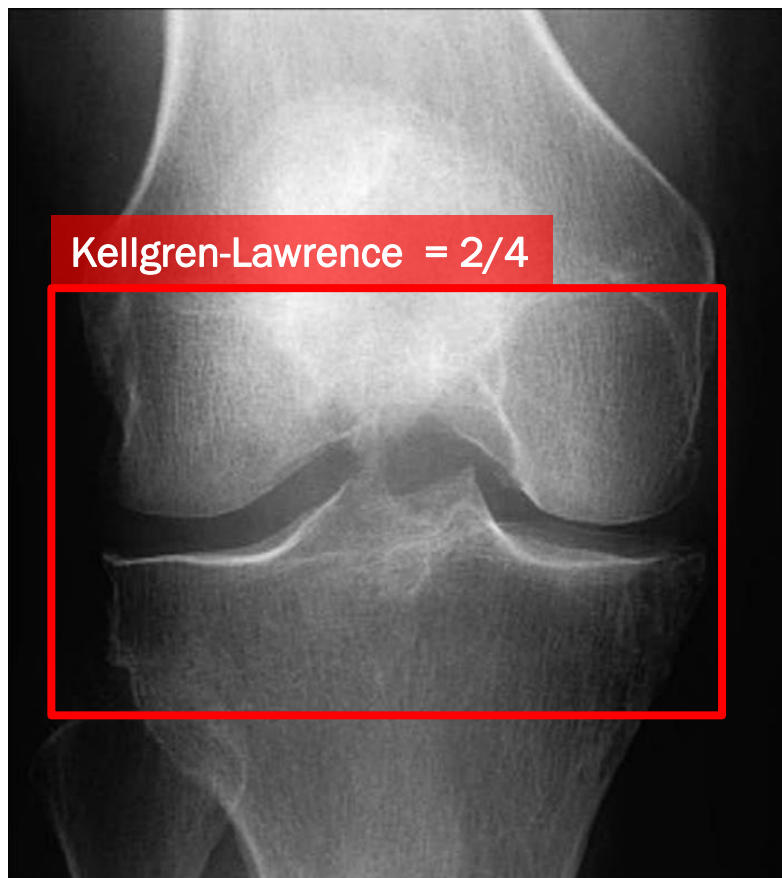
Could the current state of the art be missing something?

- ▶ Objective grading scales, based on x-ray appearance
 - Most common: Kellgren-Lawrence (KLG)
- ▶ Original studies on coal miners in Lancashire, England
 - No mention of subjects' race, sex



Finding a better target for prediction

Learn from the radiologist

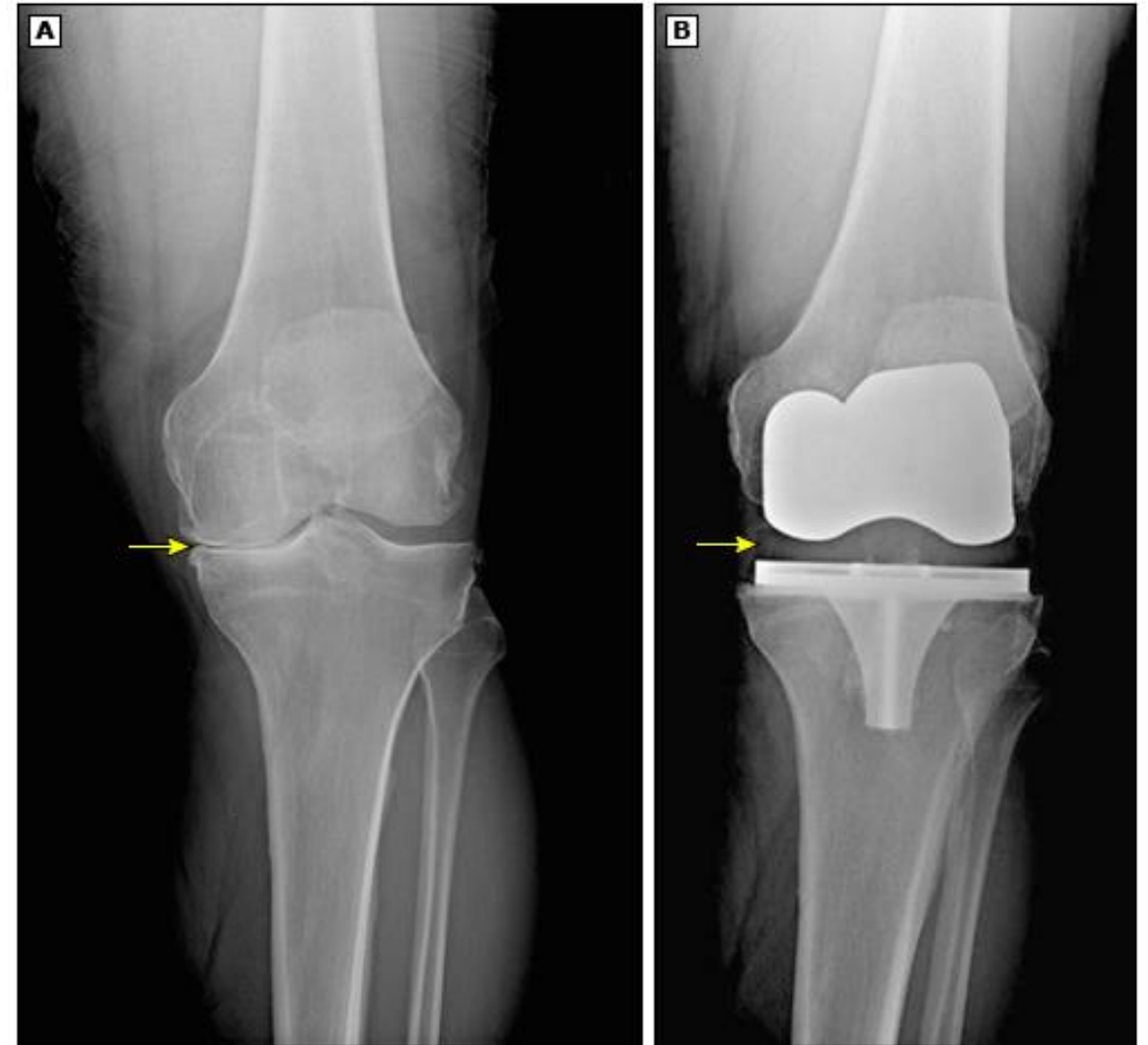


Listen to the patient



The stakes are high: Who gets a new knee?

- ▶ Patients with **severe pain**:
 - Only eligible if they have **severe arthritis**
- ▶ But who decides?
- ▶ Algorithm would **double fraction of Black knees eligible** vs. radiologist



Closing thoughts

- ▶ The promise of algorithms: **Doing better than humans**
 - Not just reproducing our errors and biases
- ▶ Algorithms should **learn from patients**, not humans
- ▶ **Small-seeming choices have large consequences**
 - Make sure your organization is **doing it right**