



CEGA

Center for Effective Global Action

Diagnosing bias with machine learning

Ziad Obermeyer
UC Berkeley

Joint work with:

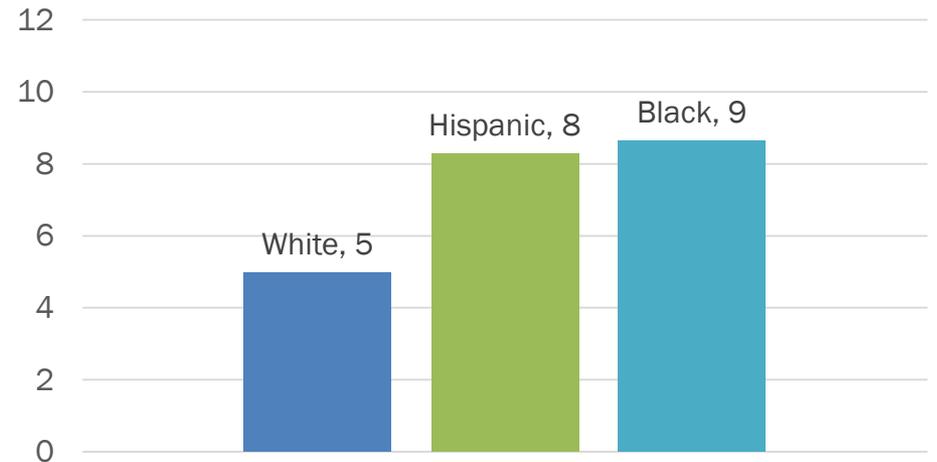
Emma Pierson, Jure Leskovec,
David Cutler, Sendhil Mullainathan



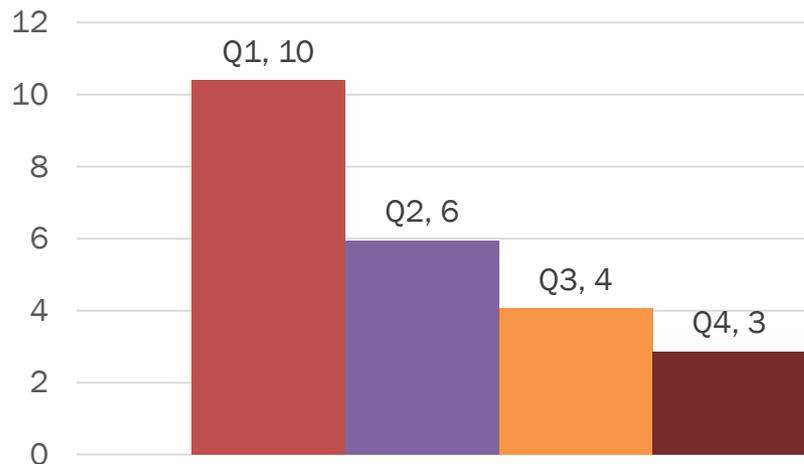
Pain is concentrated in society's most disadvantaged

- Large pain gradients
 - Race
 - Income
 - Education

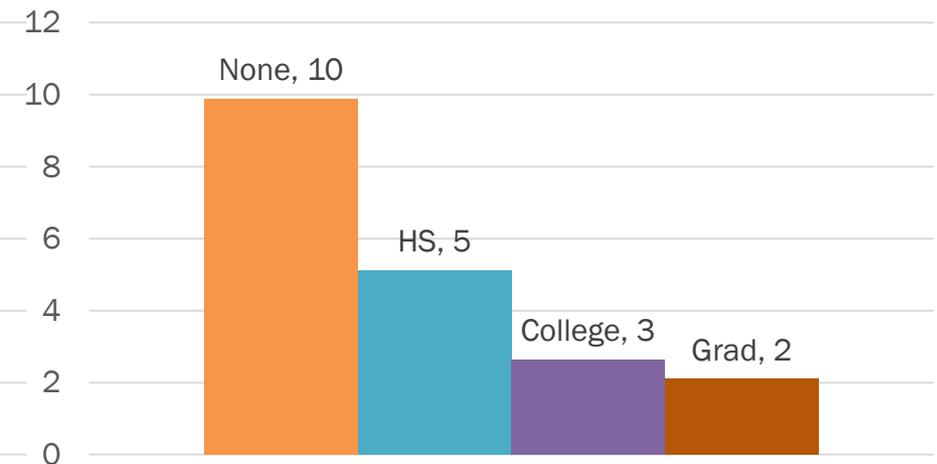
Severe pain (%): By race



Severe pain (%): By income

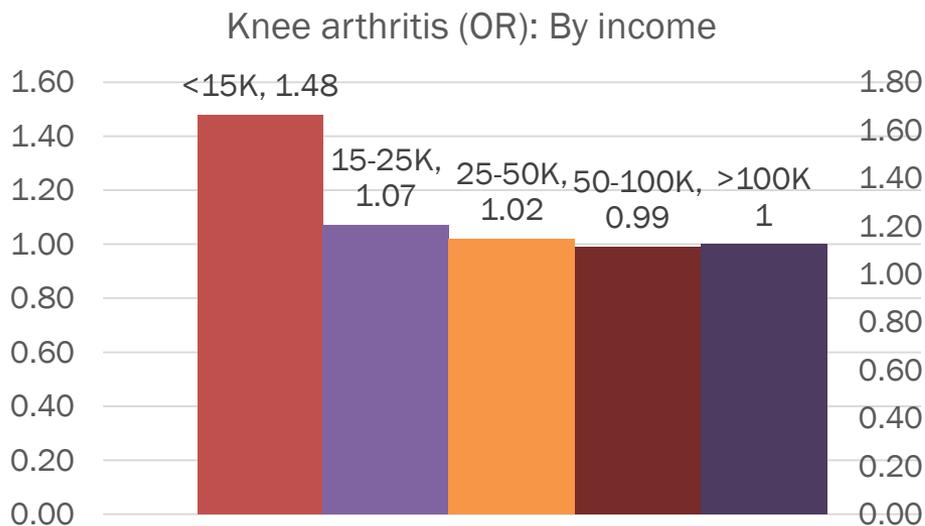


Severe pain (%): By education

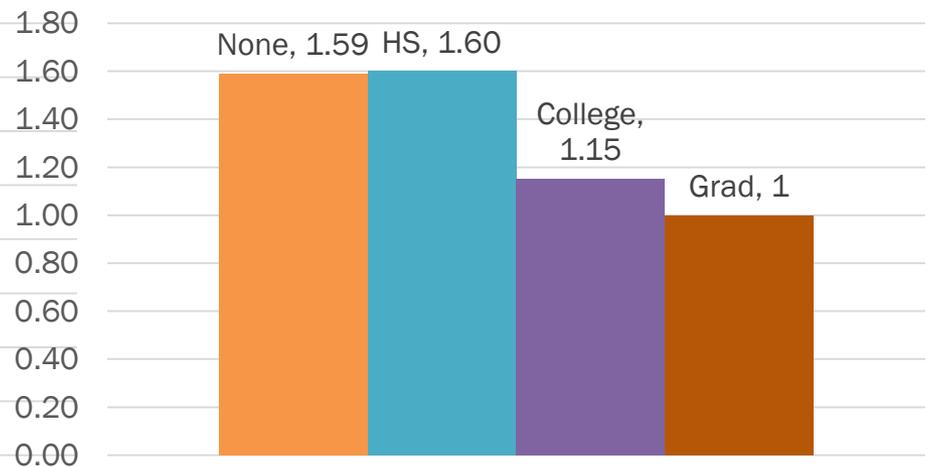


An obvious explanation

- Higher prevalence of painful conditions, e.g., in the knee
 - By income
 - By education



Knee arthritis (OR): By education



But this doesn't seem to be the whole story

- We can adjust for this
 - Compare pain, holding disease severity constant
- e.g., for knee osteoarthritis:
 - Two patients, similar x-rays
 - Who has more pain?
- Key finding: **Black, low-income, low-education have more pain**
 - Even after adjusting for their disease severity



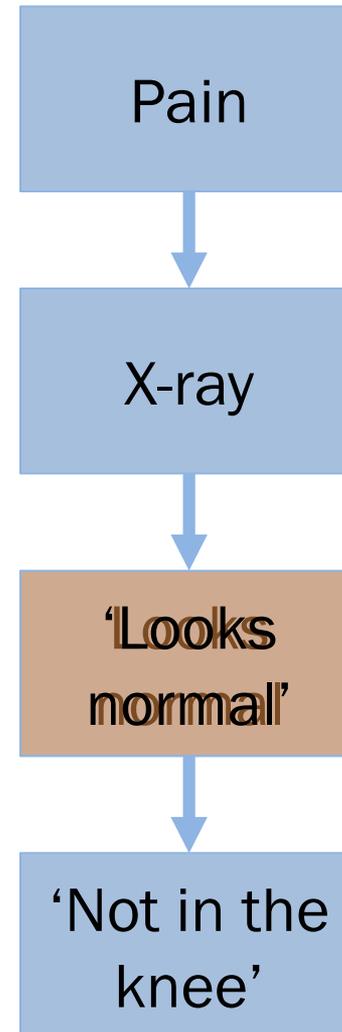
Some explanations from the literature



- If it's not in their knees...
- Maybe it's in their heads?
 - Stress
 - Psychosomatic complaints
 - Coping strategies
- Or lack of access to medical care

An observation about the “pain gap”

- Concretely, what does this look like in practice?
- Underserved patients x-rays look less severe
 - Given their level of pain
- This depends heavily on how we measure severity
 - So how do we?



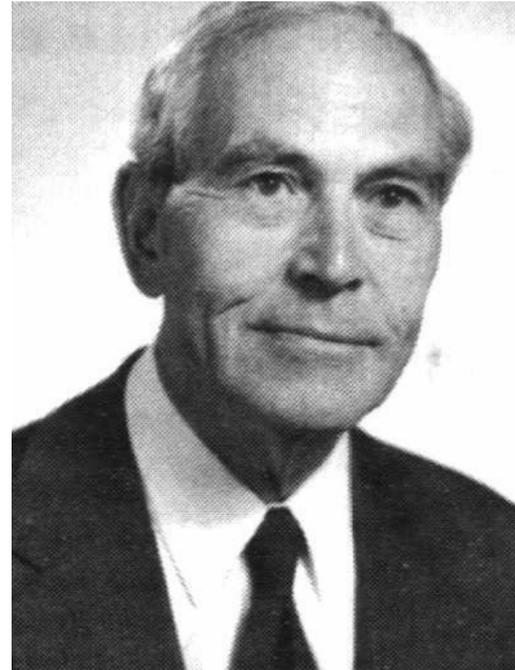
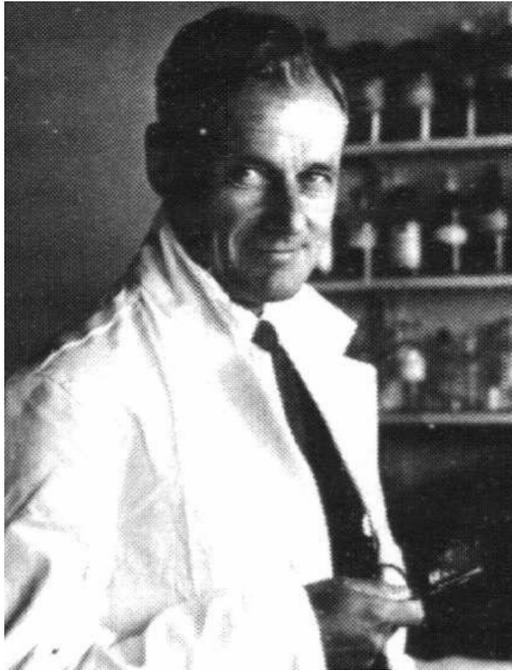
Are we missing something?



- What if there is an objective problem in the x-ray
 - But doctors just don't recognize it
- A patient with severe pain, 'normal' x-ray
 - Could it be in the knee after all?
- This isn't that implausible
 - Remember how medical knowledge gets produced

Measuring osteoarthritis severity

- We use objective grading systems to read x-rays
 - Most common: Kellgren-Lawrence, 1957

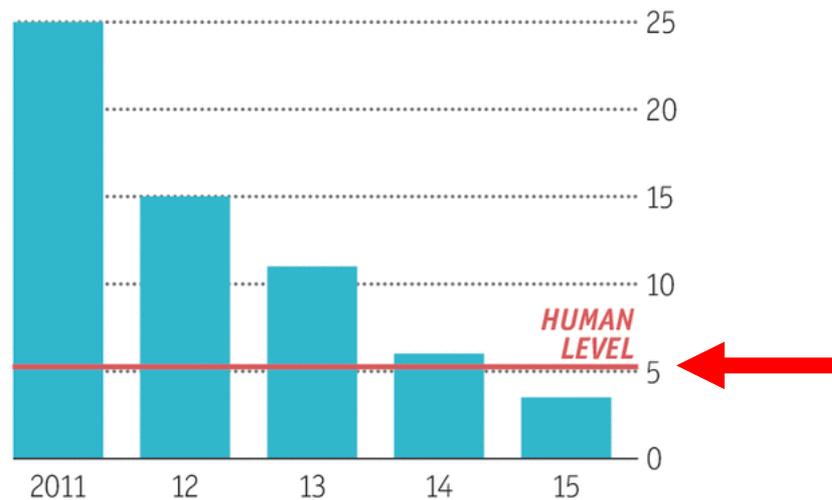


- Based on studies of coal miners in Lancashire, England
 - Original papers: no mention of subjects' race

A good job for an algorithm?

Ever cleverer

Error rates on ImageNet Visual Recognition Challenge, %



Sources: ImageNet; Stanford Vision Lab

Economist.com

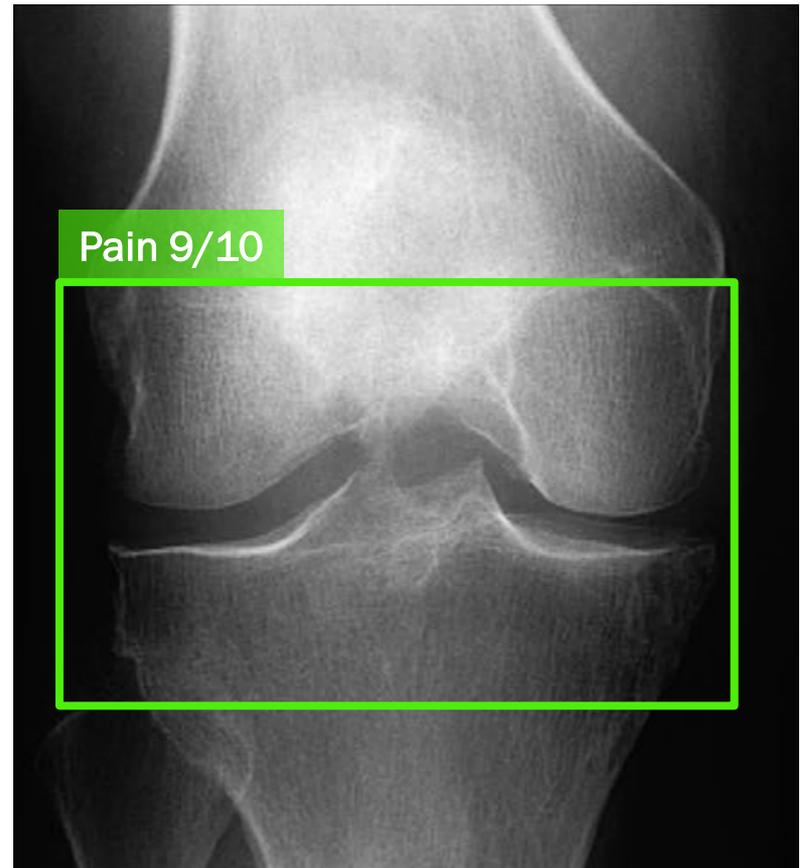
- Human radiologists may overlook causes of pain in disadvantaged groups
- We'd like to train an algorithm to help—but how?
 - Typical ML approach: train on radiologists' interpretation
- Exactly what we don't want to do!

What if there were another way to train the algorithm?

What if instead of learning from the radiologist...

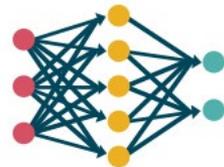


We trained the algorithm to listen to the patient?



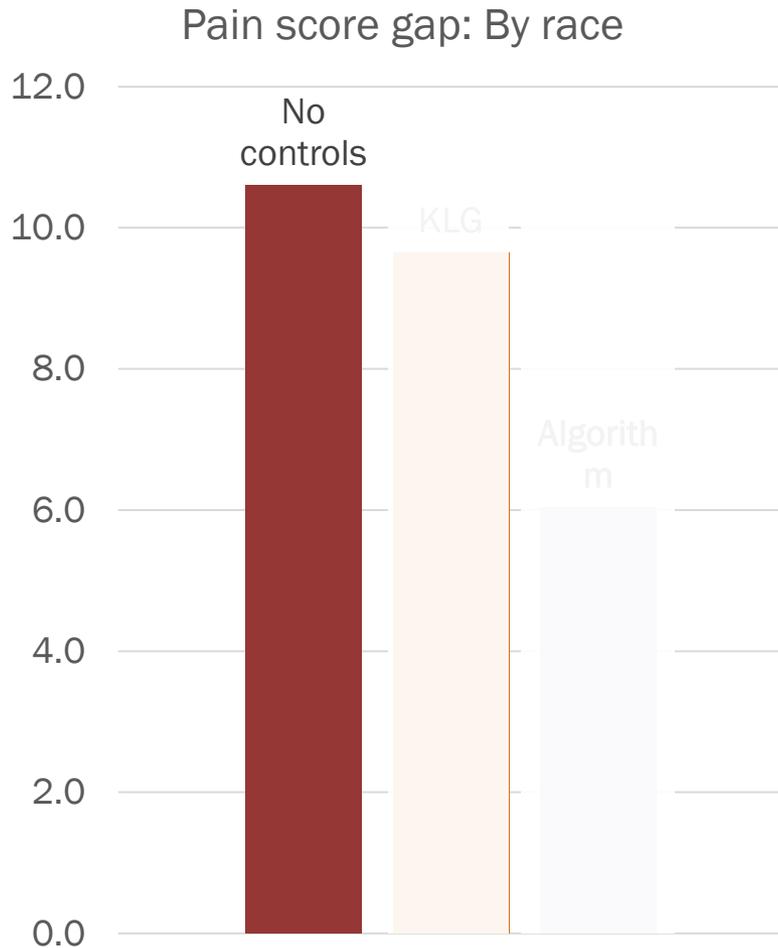
Finding the data: Not straightforward

- Easy to find dataset of x-rays + radiologist interpretation
 - Sitting on every hospital's PACS system
- Much harder to find x-rays + patient pain experience
 - Funny story: How we got these data
- But once we have data: a very straightforward ML problem



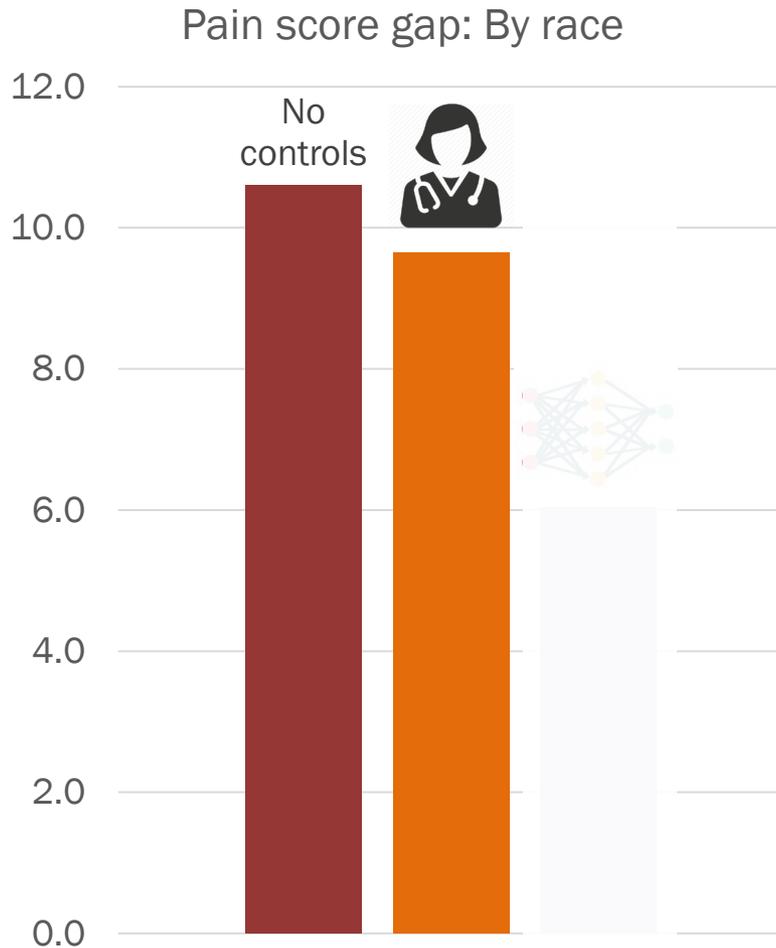
- Does algorithm or doctor better explain extra pain among disadvantaged?

Starting point: the Black–White pain gap



- Self-reported pain scores
 - Best: 100
 - Severe pain: 86
- The pain gap
 - Blacks have 10.6 points higher pain
- Similar for other groups
 - Low-income: 4.2
 - Low-education: 5.3

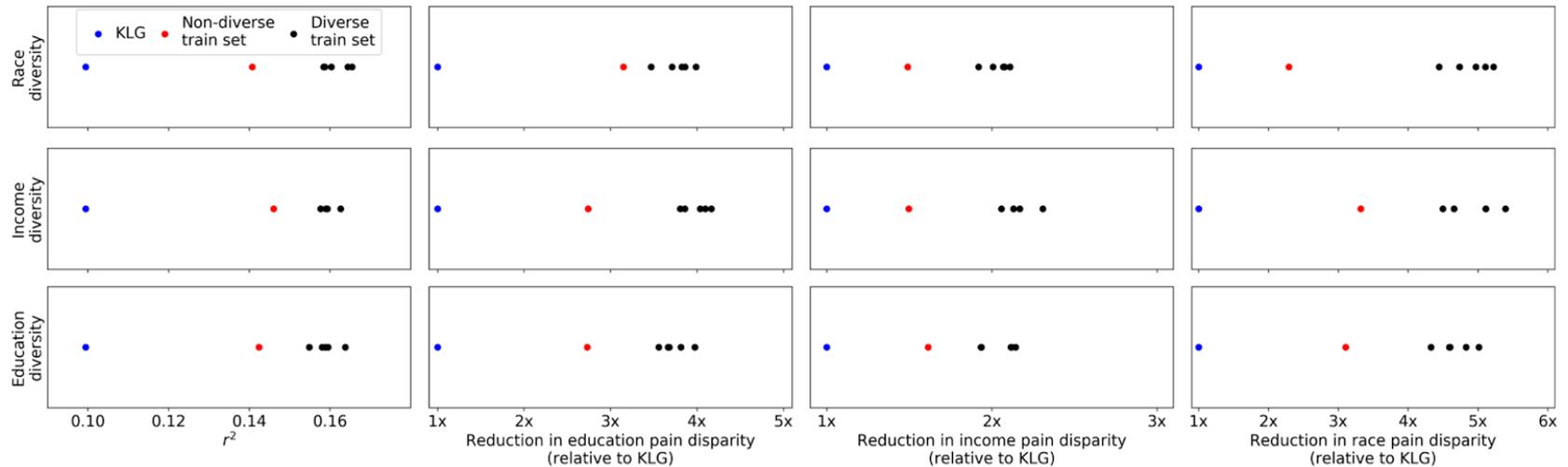
Some results: Closing the Black–White pain gap



- Adjusting for standard severity measure: **-9%**
- Adjusting for algorithmic severity measure: **-43%**
 - 4.7x more than standard measure
 - 95% CI: 3.2-11.8
- Similar results for
 - Income: 2.0x
 - Education: 3.6x

What can we say about the model?

- A key strength: **diversity** of the training set
 - Removing non-white patients hurts model performance



- It appears to be finding **real signal**
 - Not specific to KLG, or a more granular version; MRIs too
 - Not just re-learning/re-weighting radiologist features
 - Works across data study sites

The stakes here are high

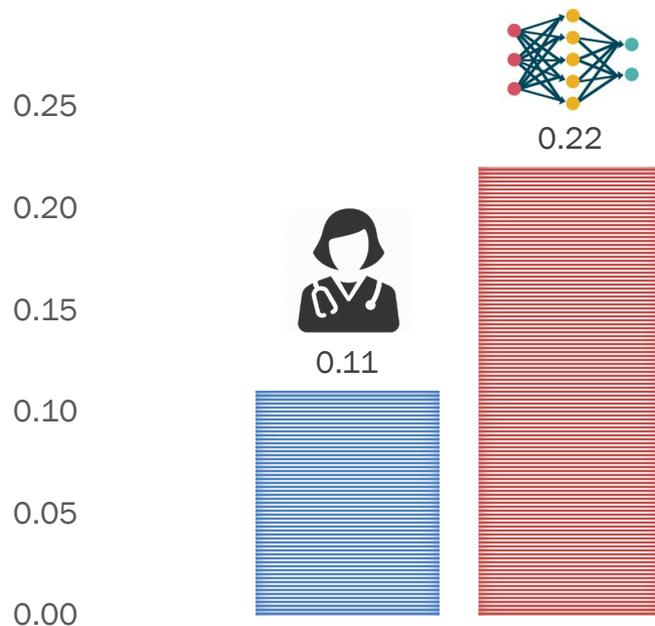
- Knee arthritis is very treatable
 - With knee replacement surgery
 - This is “life changing”
- We offer surgery to those with severe disease
 - Not just those with severe pain
 - Surgery only works if the problem is in your knee
- But what about disadvantaged patients with
 - Severe pain, an unimpressed radiologist
 - ... and high algorithmic pain scores?



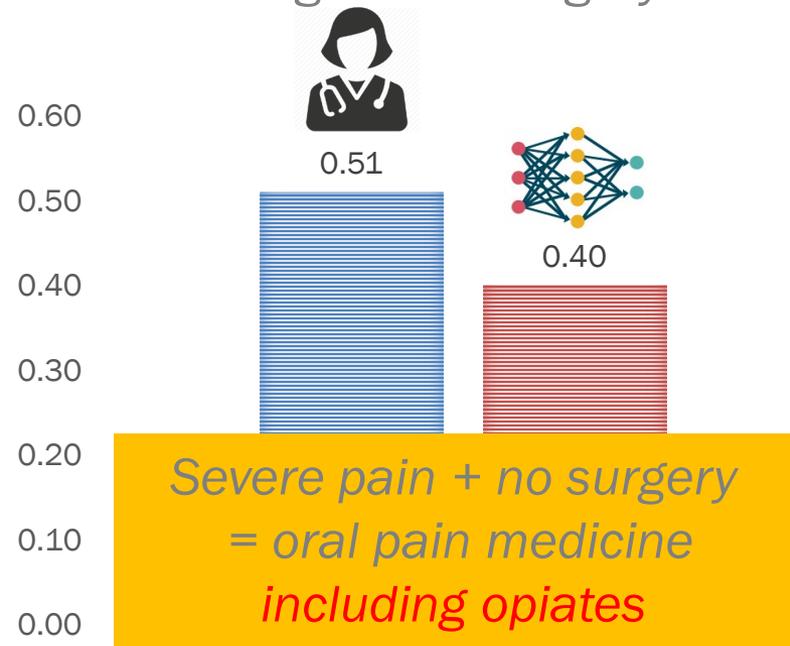
Simulation: Who would get surgery... if the algorithm were in charge, not the doctor?

- Identify patients with severe pain and
 - High disease severity according to human
 - High disease severity according to algorithm (same n)

Black knees eligible for surgery



Black knees, severe pain but ineligible for surgery



Some closing thoughts

- We worry a lot about **bias in algorithms**
 - With good reason
 - But there's also reason for optimism
- Algorithms can learn **biased labels**
 - Or they can **learn to listen to patients**
 - Which one is up to us, and how we train them
- It also depends on access to **high-quality data**
 - Which is why it's frustrating that so many medical datasets are **inaccessible**

The Nightingale Project



- An open data platform for research, bringing together
 - Leading health organizations
 - World class researchers communities
- Around interesting datasets
 - High-dimensional medical data
 - ECGs from Bay Area to Tanzania
 - Chest x-rays from Peru to Wales
 - Linked to real outcomes
- Greatest social good these algorithms can do
 - ‘Where there is no doctor’