Algorithms Can Fight Bias

Ziad Obermeyer

UC Berkeley

Today

- Algorithms can reproduce, **scale up bias**
  - In health care
  - In criminal justice, finance, education, …

- Less attention to possibility that they can **fight bias**

- I'll tell you about some work in this area
  - And **how we (and you!) can contribute**

Pain is concentrated in society's most disadvantaged

- Story isn't as simple as it looks

- Typical exercise in literature, e.g., for knee osteoarthritis:
  - Two patients, similar x-rays
  - Compare pain scores

- Black, lower-income, lower-education: <u>still</u> have more pain
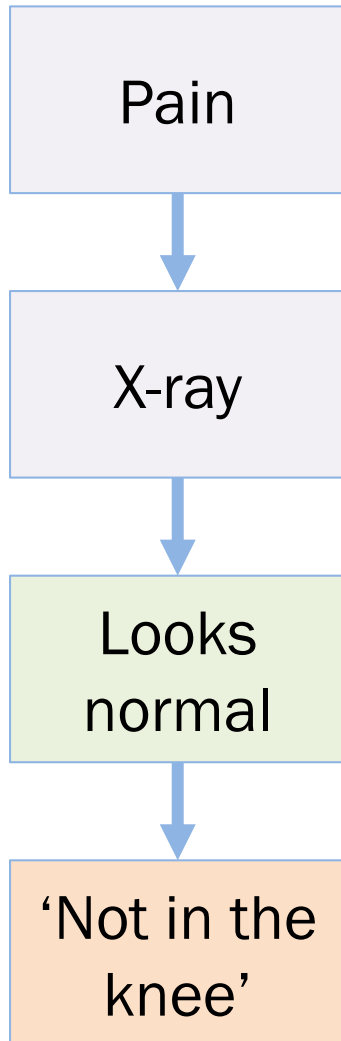 - At every level of x-ray graded disease severity

Some explanations from the literature

- If it's not **in their knees**…

- Maybe it's **in their heads?**
  - Stress makes similar stimuli more painful
  - Psychosomatic factors
  - Coping skills

- Or in the **medical system**
  - Access to therapies

Concrete clinical scenario

Pain

↓

X-ray

↓

Looks normal

↓

'Not in the knee'

- Implication of literature
  - Black patients' pain out or proportion to severity

- Leads to questions about stress, depression, …
  - And allocation of non-knee-based treatments

- But what do we mean by 'disease severity'?
  - How do we measure it?

Current SOTA

Measuring osteoarthritis severity

- Objective grading scales, based on x-ray appearance
  – Most common: Kellgren-Lawrence, 1957 (KLG)
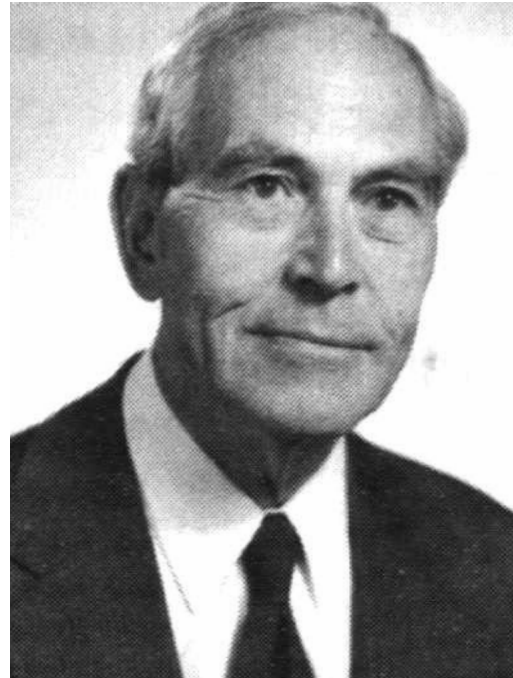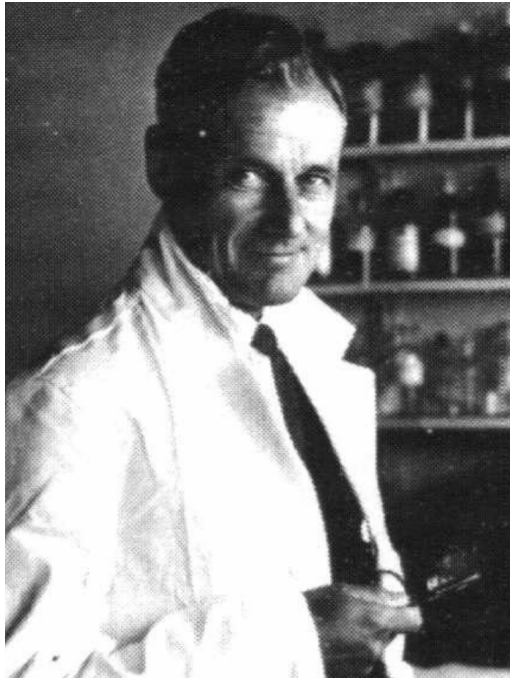
- Original studies on coal miners in Lancashire, England
  – No mention of subjects' race, sex

A good job for an algorithm?

Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images

Aniket A. Tolpadi[1,2], Jinhee J. Lee[2], Valentina Pedoia[2] & Sharmila Majumdar[2]*
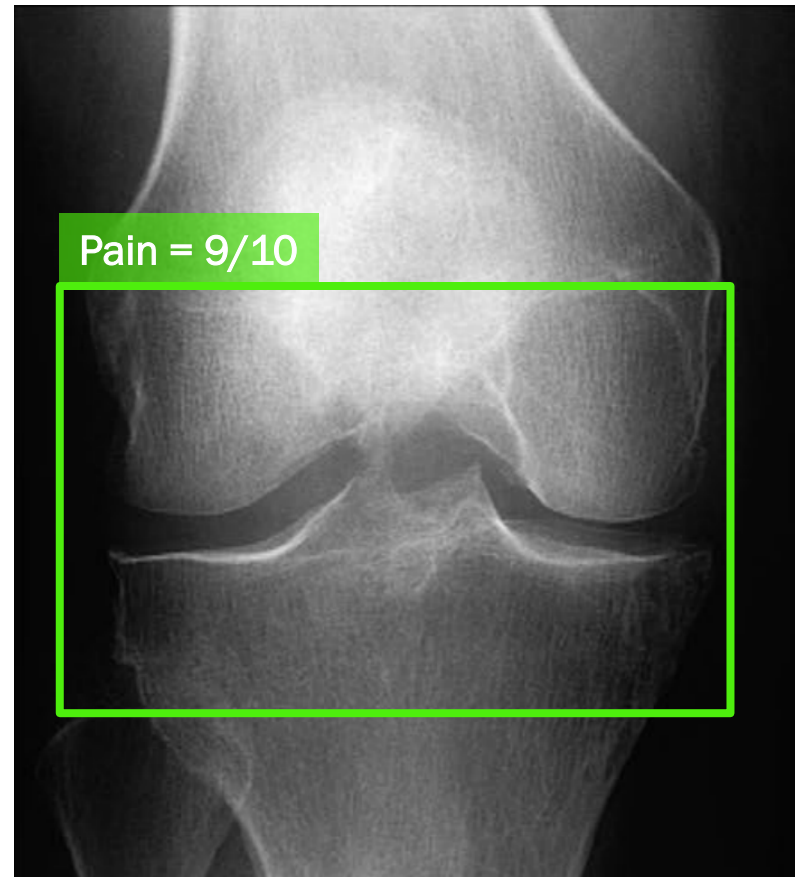
- Human radiologists may **overlook causes of pain** in disadvantaged groups

- We'd like an algorithm to help—but...

  Typical approach: train to **match human performance**

- Exactly what we **don't want to do!**

Finding a better target for prediction
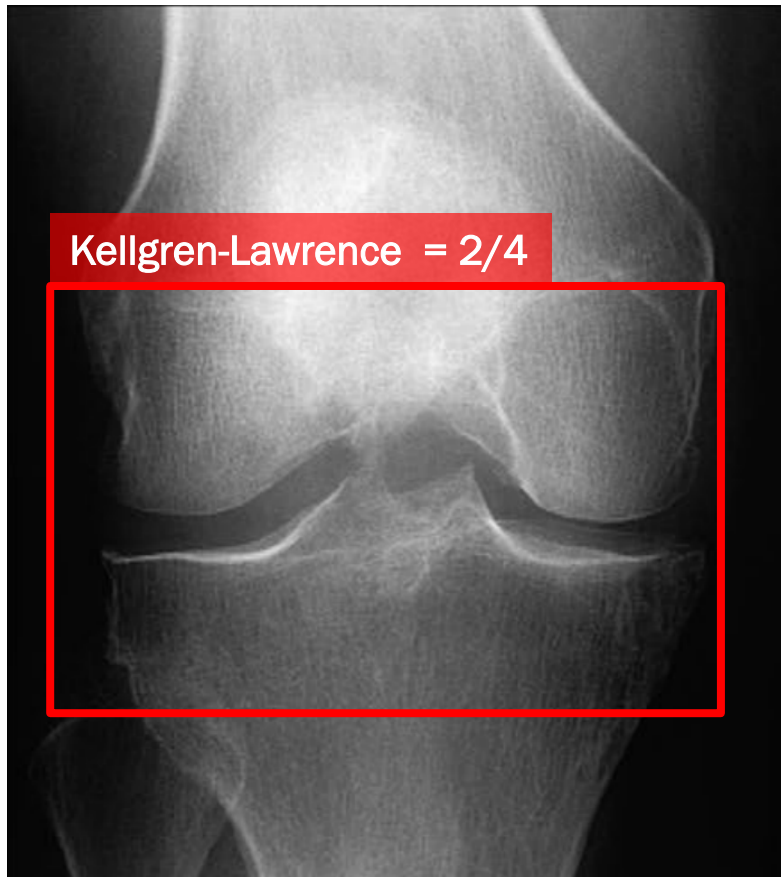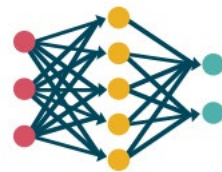
Finding the data: Not straightforward

- Easy to find: x-rays + **radiologist interpretation**
  - Sitting on every hospital's PACS system

- Much harder to find: x-rays + **patient pain experience**

- But once we have data: a very straightforward ML problem

  - If pain is predictable from knee image
    - ...Pain is **in the knee** (not in the head, coping, ...)

Algorithm explains nearly half the pain gap

Pain score gap: By race
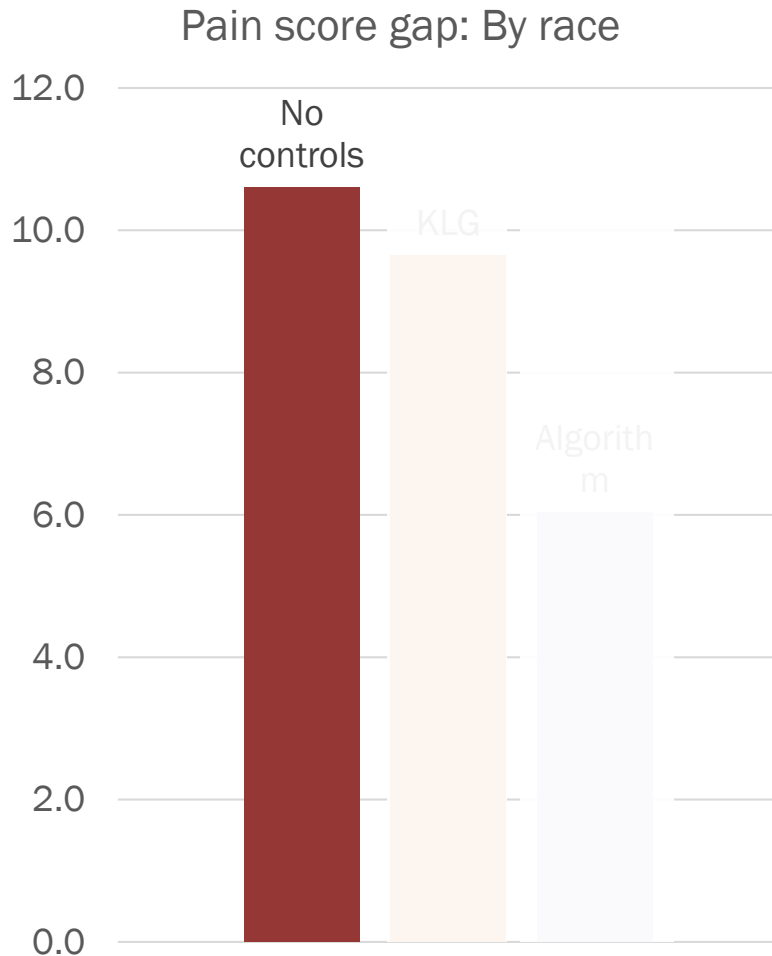
- Adjusting for standard severity measure: −9%

- 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

Koos score: Max 100, severe ≤86

The stakes are high

- Take patients with severe pain

- Eligibility: swap in **algorithm** severity, not **radiologist**

- Double fraction of Black knees eligible for **surgery**
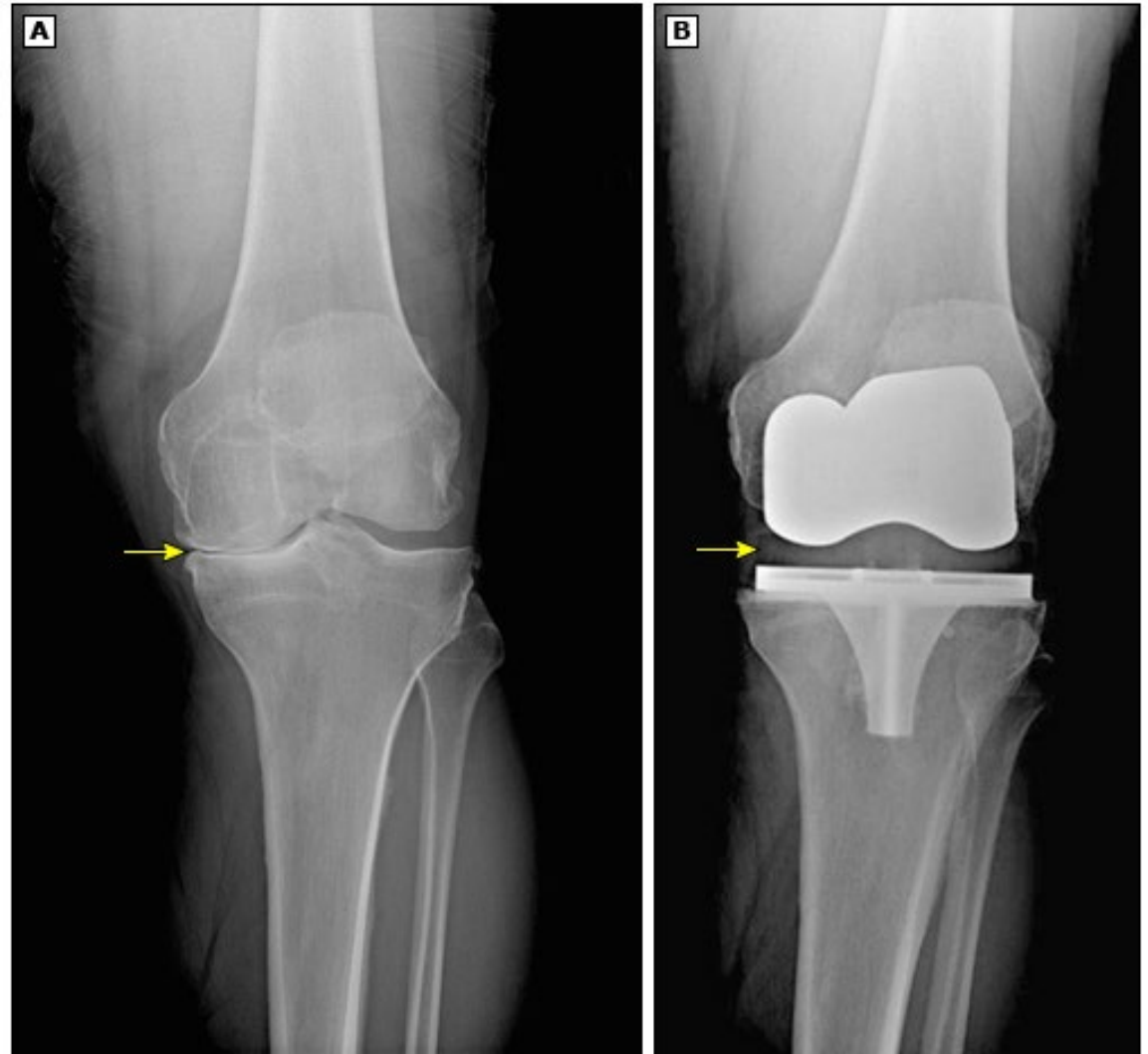
Finding better proxies for medicine

- The promise of algorithms: **Doing better than humans**
  - Not just reproducing their errors and biases

- Algorithms should **learn from nature**, not humans
  - But data on patient outcomes, experiences are siloed
  - The 'market' can't match good researchers to good data

- This lack of data is holding back a scientific field
  - We are working towards a solution

Building the data backbone for equitable algorithms

NIGHTINGALE
OPEN SCIENCE

- We work with health systems, companies, governments, ...
  - To **build** up data infrastructure
  - And **curate** datasets around unsolved medical problems
  - Focusing on **high-priority** problems and populations

- We then **open up the data** to non-profit researchers
  - Deidentified images + clinical outcomes
  - On a cloud platform, free