Discussion: Bias, Fairness, and Inequality in an Algorithmic Age

Kosuke Imai
Harvard University

CSSS 25th Anniversary Celebration
University of Washington

May 17, 2024

Three Diverse Presentations

1. Chu: How algorithmic ranking affects students' perception of colleges
   - important study given the recent controversy of college rankings
   - ranking keeps low-SES students from preferring highly ranked colleges

2. Nabi: How to define fairness and achieve it
   - definition of fairness should be based on context and causality
   - statistical methods to quantify fairness and develop fair algorithms

3. Hoff: How to combine Bayes and Frequentist for powerful inference
   - develop a Bayesian credible interval with frequentist coverage property
   - the proposed tests and credible intervals are much more powerful

Conjoint Experiment

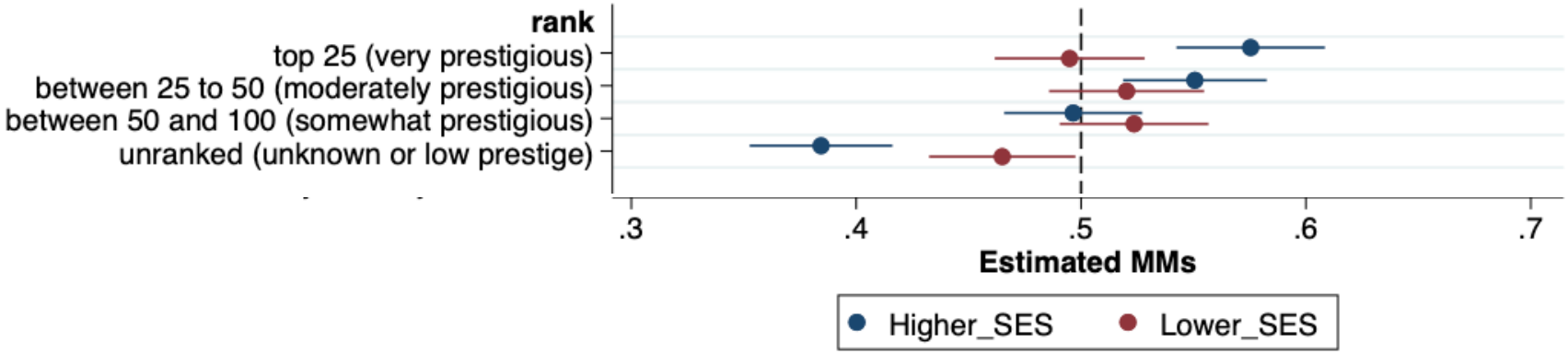
Please compare the following two colleges. You might feel like you do not have enough information to answer all the questions, but please answer with your best guess.

College A		College B
top 25 (very prestigious)	**College Rank**	between 25 to 50 (moderately prestigious)
public	**Type**	private
small (<3K)	**Size of Student Pop.**	small (<3K)
30-50%	**Graduation Rate**	70-90%
urban	**Location**	suburban
about 1 faculty to every 10 students	**Student to Faculty Ratio**	about 1 faculty to every 10 students

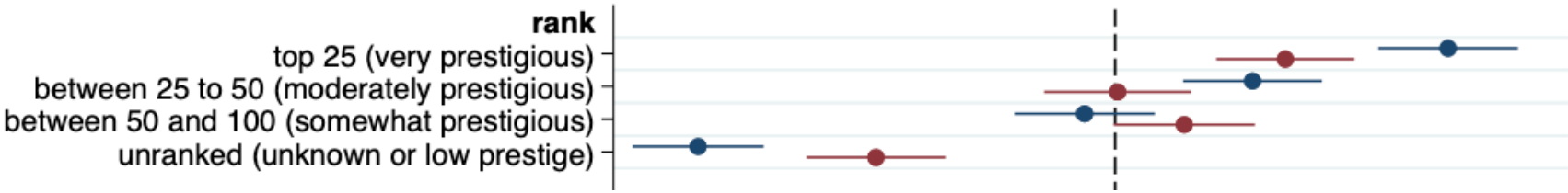
Do you think the following descriptions apply more to College A or College B?

Differential Impacts of Ranking on College Perception

- High ranking makes students with low SES feel less likely to apply:



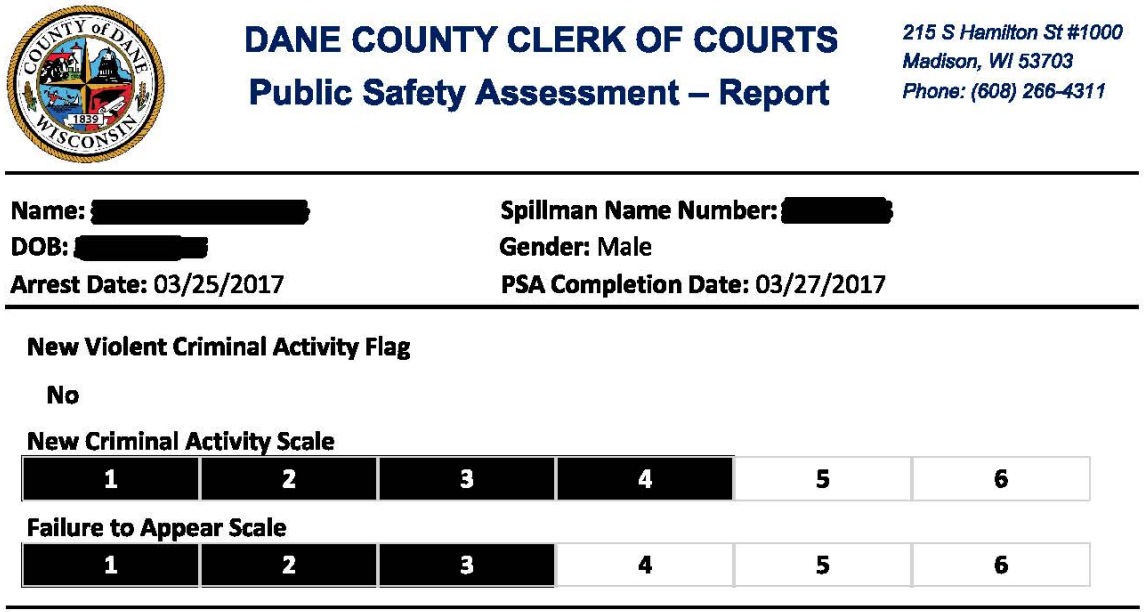
- Even if college is free, the gap remains:



- Mechanisms:
  - Lower-SES students (incorrectly) interpret high rank as high cost
  - Higher-SES students interpret high rank as more opportunities
- Questions: Make conjoint analysis more realistic? Beyond conjoint?

Implication: Algorithms can Have Differential Impacts

- Different people may interpret the same algorithmic outputs differently
- Experimental evaluation of pretrial risk assessment score: (Imai et al. 2023)



DANE COUNTY CLERK OF COURTS
Public Safety Assessment – Report

215 S Hamilton St #1000
Madison, WI 53703
Phone: (608) 266-4311

Name: [redacted] **Spillman Name Number:** [redacted]
DOB: [redacted] **Gender:** Male
Arrest Date: 03/25/2017 **PSA Completion Date:** 03/27/2017

New Violent Criminal Activity Flag

No

New Criminal Activity Scale

1	2	3	4	5	6

Failure to Appear Scale

1	2	3	4	5	6

- Too many researchers focus on the accuracy and fairness of algorithms
  - but humans, not machines, make final decisions when stakes are high
  - we should study heterogeneous impacts of algorithms on humans

Fairness Based on Path-Specific Effects

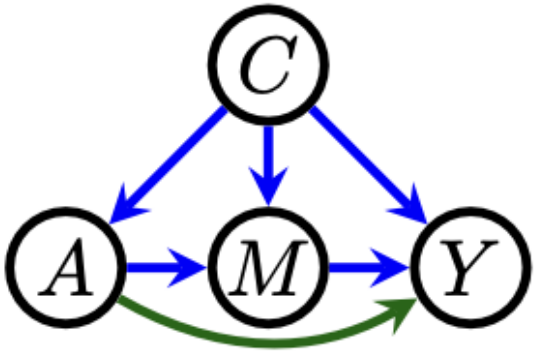
- Motivating idea taken from a legal literature:

  The central question in any employment-discrimination case is whether the employer would have taken the same action had the employee been of a different race (age, sex, religion, national origin etc.) and everything else had been the same.

- This leads to the use of natural direct/indirect (path-specific) effects:
  - formalize fairness using a causal model
  - identification under a certain set of assumptions
  - statistical learning for fair algorithmic decisions using the observed data

COMPAS Score Application (Nabi and Shpitser 2018)

- A: race
- M: prior convictions
- C: demographics (age, gender, etc.)
- Y: recidivism
- Natural direct effect (NDE): $A \to Y$



- Build a prediction model for Y subject to the constraint NDE is small

- Practical considerations:
  - Does C capture all confounders?
  - Does M capture all "fair" path-specific effects of race?
  - Selective labels problem: recidivism is affected by judge's decision
  - Identification of NDE requires the absence of post-treatment confounders that affect M and are affected by A

Alternative Causal Approach to Fairness

- Potential outcomes framework:
  - Judge's decision: detain ($D = 1$) or release ($D = 0$)
  - Pretrial risk scores such as COMPAS are supposed to predict $Y(0)$
- Racial disparity in judge's decision:

$$\Pr(D = 1 \mid Y(0) = y, A = \text{black}) \stackrel{?}{=} \Pr(D = 1 \mid Y(0) = y, A = \text{white})$$

- Classification framework

		Decision	
		Negative ($D = 0$)	Positive ($D = 1$)
Outcome	Negative ($Y(0) = 0$)	True Negative (TN)	False Positive (FP)
	Positive ($Y(0) = 1$)	False Negative (FN)	True Positive (TP)

- We can obtain informative bounds on racial disparity using the single-blinded experiment without additional assumptions (Ben-Michael et al. 2024)

FAB (Frequentist, Assisted by Bayes)

- We want a frequentist confidence interval:

$$\Pr(\theta_j \in C(\boldsymbol{y}) \mid \boldsymbol{\theta}) \;=\; 1-\alpha$$

  that is Bayes-optimal, minimizing

$$\mathbb{E}[|C_j(\boldsymbol{y})|] \;=\; \int |C_j(\boldsymbol{y})| \underbrace{p(\boldsymbol{y} \mid \boldsymbol{\theta})}_{\text{likelihood}} d\boldsymbol{y} \underbrace{\pi(\boldsymbol{\theta})}_{\text{prior}} d\boldsymbol{\theta}$$

- Idea: shorter intervals for the region of $\boldsymbol{y}$ with a high prior density
- A variety of applications:
  1. mean difference test
  2. regression
  3. multi-level models, small area estimation, etc.
  4. R package: FABInference

Fair Inference in Multilevel Data Analysis

- FAB is a powerful idea that is widely applicable
  - factorial experiments like conjoint analysis
  - variable selection methods

- How can FAB enhance fairness?
- Fair statistical inference: all groups have a proper type I control
- This does not guarantee fair statistical power, which depends on sample size and variance for each group
- Failure to detect racial disparity does not necessary imply its absence

- FAB power analysis? Designs of experiments and sample surveys

Past and Future of Statistics and the Social Sciences

- Quantitative social science (QSS) as an intersection between statistics and social sciences
- CSSS has played a leadership role over the past 25 years

- Current state of QSS:
  - technical progress in statistics and machine learning has been incredible
  - increasing data availability led to many opportunities in social sciences
  - yet, the gap between technical and substantive fields is widening fast

- We need more people who can bridge between the two fields