

Fairness: A few notions

Su Lin Blodgett

Abram Handler

Katie

Keith

Individual and group fairness

Individual fairness:

Individuals who are similar should be treated similarly.

Group fairness:

Different groups should be treated similarly, on average.

Recidivism

In the context of these articles, a person who has been arrested ***recidivates*** if they are arrested again for another crime.

Detour: confusion matrix

Did not recidivate

True negative

False positive

Recidivated

False negative

True positive

Predicted low risk

Predicted high risk

Detour: confusion matrix

<i>Did not recidivate</i>	True negative	False positive
<i>Recidivated</i>	False negative	True positive
	<i>Predicted low risk</i>	<i>Predicted high risk</i>

- What do different stakeholders want?
 - Defendant?
 - Law enforcement?
 - Society?

Cancer screening?

Search engine?

Self-driving car red light detector?

UMass admissions decisions?

Detour: confusion matrix

<i>Did not recidivate</i>	True negative	False positive
<i>Recidivated</i>	False negative	True positive
	<i>Predicted low risk</i>	<i>Predicted high risk</i>

Precision

$$\frac{TP}{TP + FP}$$

Detour: confusion matrix

<i>Did not recidivate</i>	True negative	False positive
<i>Recidivated</i>	False negative	True positive
	<i>Predicted low risk</i>	<i>Predicted high risk</i>

Precision

$$\frac{TP}{TP + FP}$$

False positive rate

$$\frac{FP}{FP + TN}$$

Detour: confusion matrix

<i>Did not recidivate</i>	True negative	False positive
<i>Recidivated</i>	False negative	True positive
	<i>Predicted low risk</i>	<i>Predicted high risk</i>

Precision

$$\frac{TP}{TP + FP}$$

False positive rate

$$\frac{FP}{FP + TN}$$

False negative rate

$$\frac{FN}{FN + TP}$$

Worksheet!

Group fairness: many metrics!

Predictive parity

Precision across
groups is the same

Group fairness: many metrics!

Predictive parity

Precision across
groups is the same

Equal false positive rate

Defendants of different
groups are equally likely
to be incorrectly
detained

Group fairness: many metrics!

Predictive parity

Precision across groups is the same

Equal false positive rate

Defendants of different groups are equally likely to be incorrectly detained

Equal false negative rate

Defendants of different groups are equally likely to be incorrectly released

Group fairness: many metrics!

Predictive parity

Precision across groups is the same

Equal false positive rate

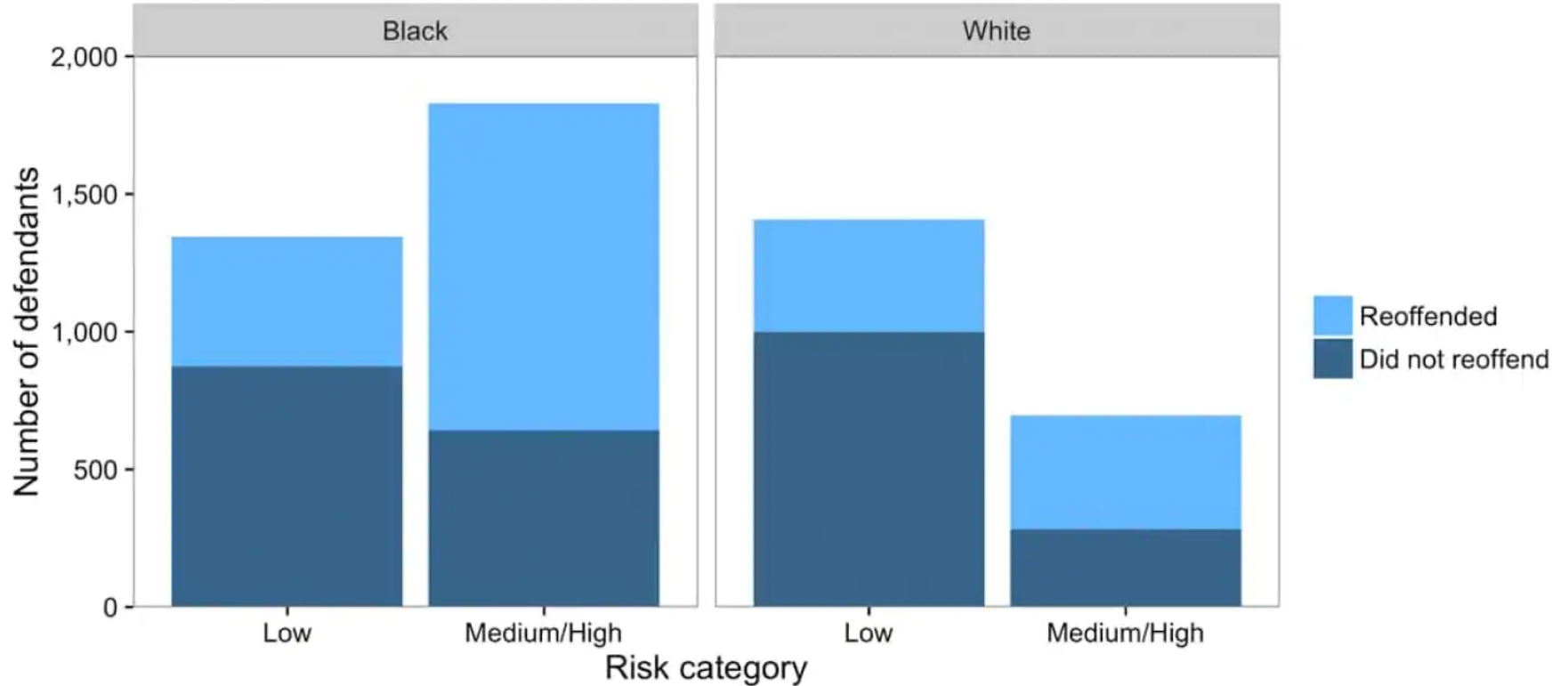
Defendants of different groups are equally likely to be incorrectly detained

Equal false negative rate

Defendants of different groups are equally likely to be incorrectly released

Impossibility Theorem: *If prevalence across groups is different, then we can't satisfy all three of these!*

Detour: confusion matrix



Other thoughts

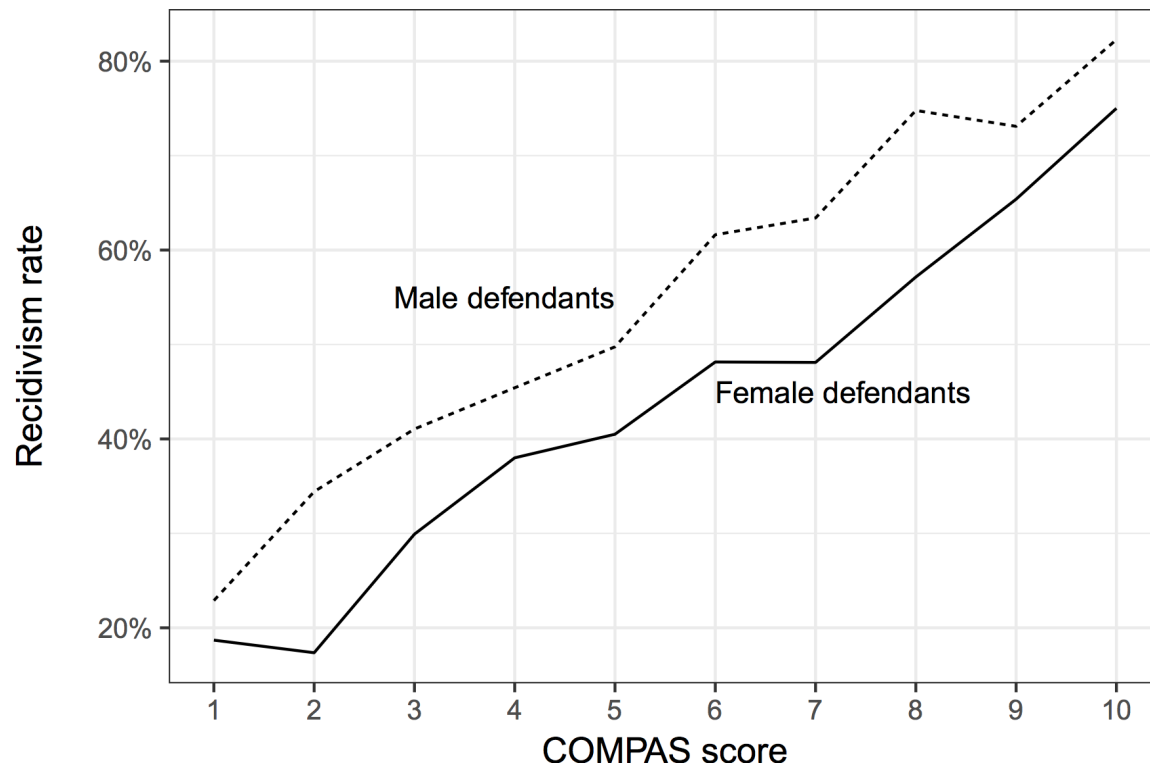
What about ignoring protected attributes altogether?

Why is prevalence across groups different, anyway?

- Measurement bias
- Historical prejudice

Who should be doing this work?

Other thoughts



How would you define fairness? Try to use the definitions of true positive and false positive?

What definitions are reasonable for criminal justice?

Does removing knowledge of a protected attribute make a system fair?

Think of a system where algorithms might be deployed (school admissions, financial aid decisions, audit decisions from the IRS). How would different stakeholders define fairness?