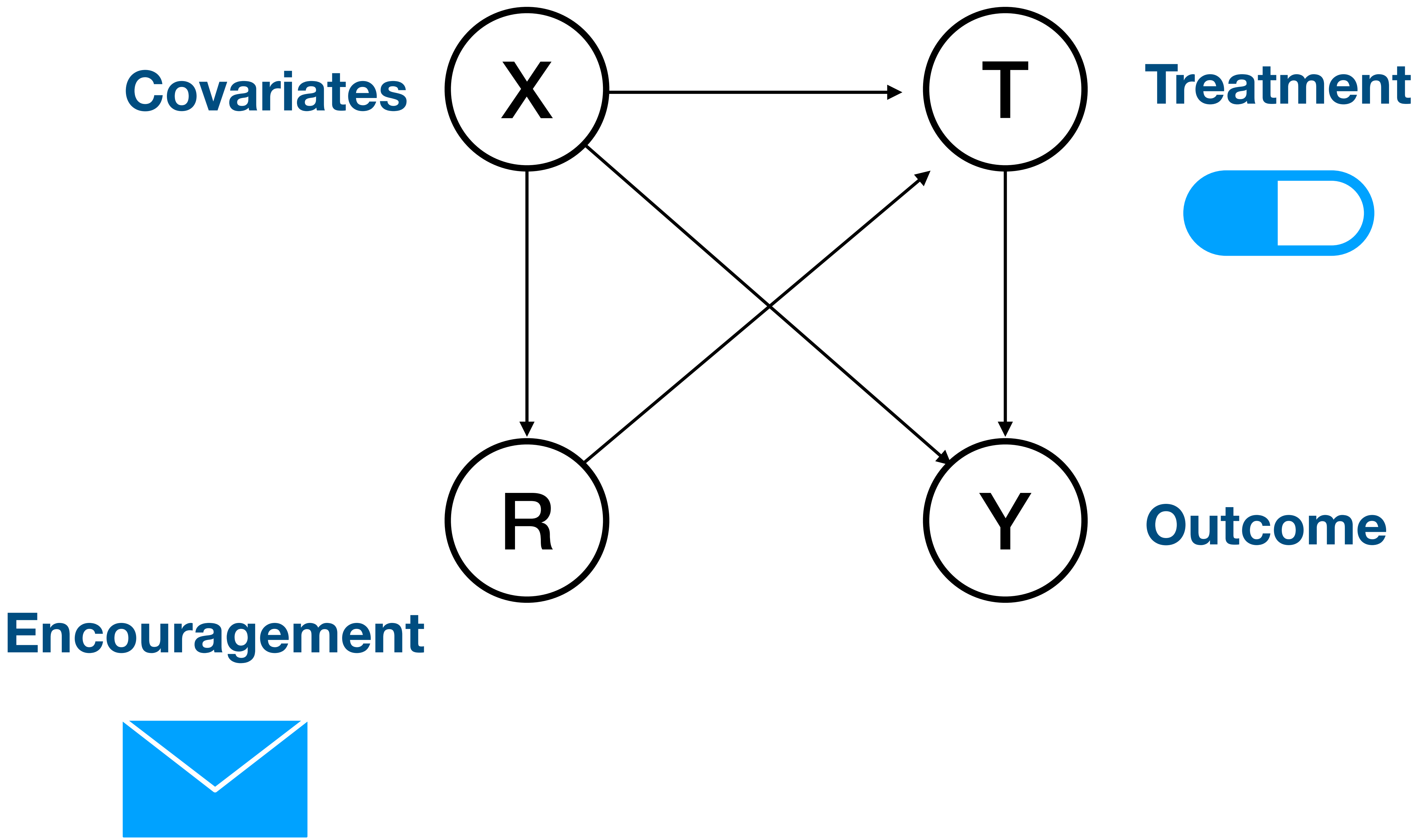
Optimal and Fair Encouragement Policy Evaluation and Learning

Angela Zhou

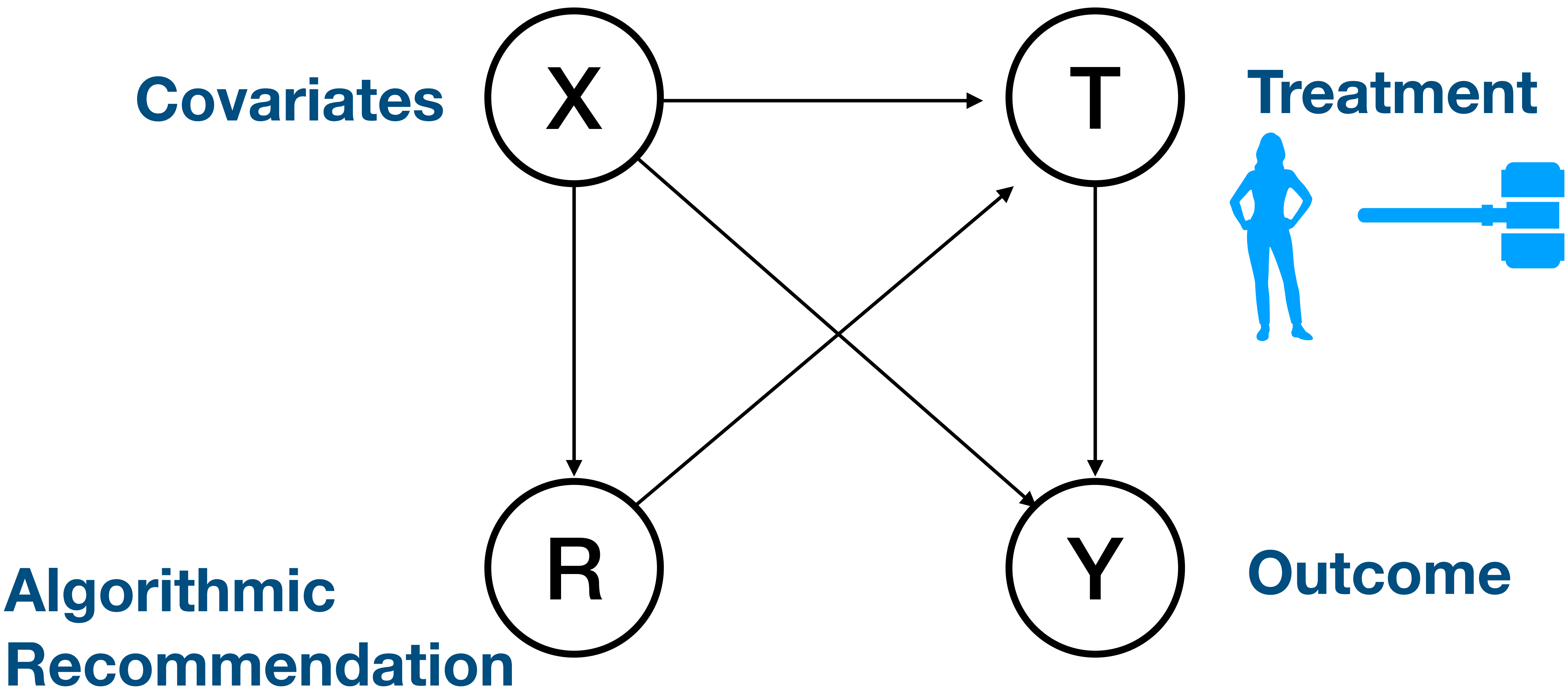
University of Southern California,
Marshall School of Business
Data Sciences and Operations
zhoua@usc.edu

- “Program evaluation view of algorithmic accountability”
- Two settings
  - Encouragement designs
    i.e. targeted interventions to improve service delivery
  - Algorithmic recommendations
- Contributions:
  Modeling using “off-policy learning”, methodology
  Applied context:
  Targeted service delivery interventions to reduce disparities

Setting 1: Encouragement Designs



Setting 2: Algorithmic Recommendations



"High-risk", "low-risk", etc.

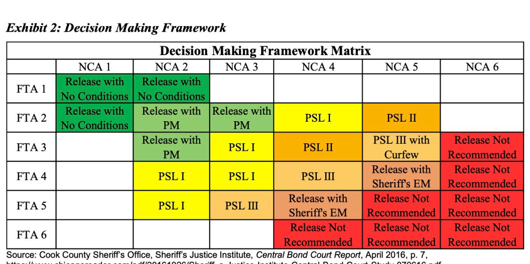


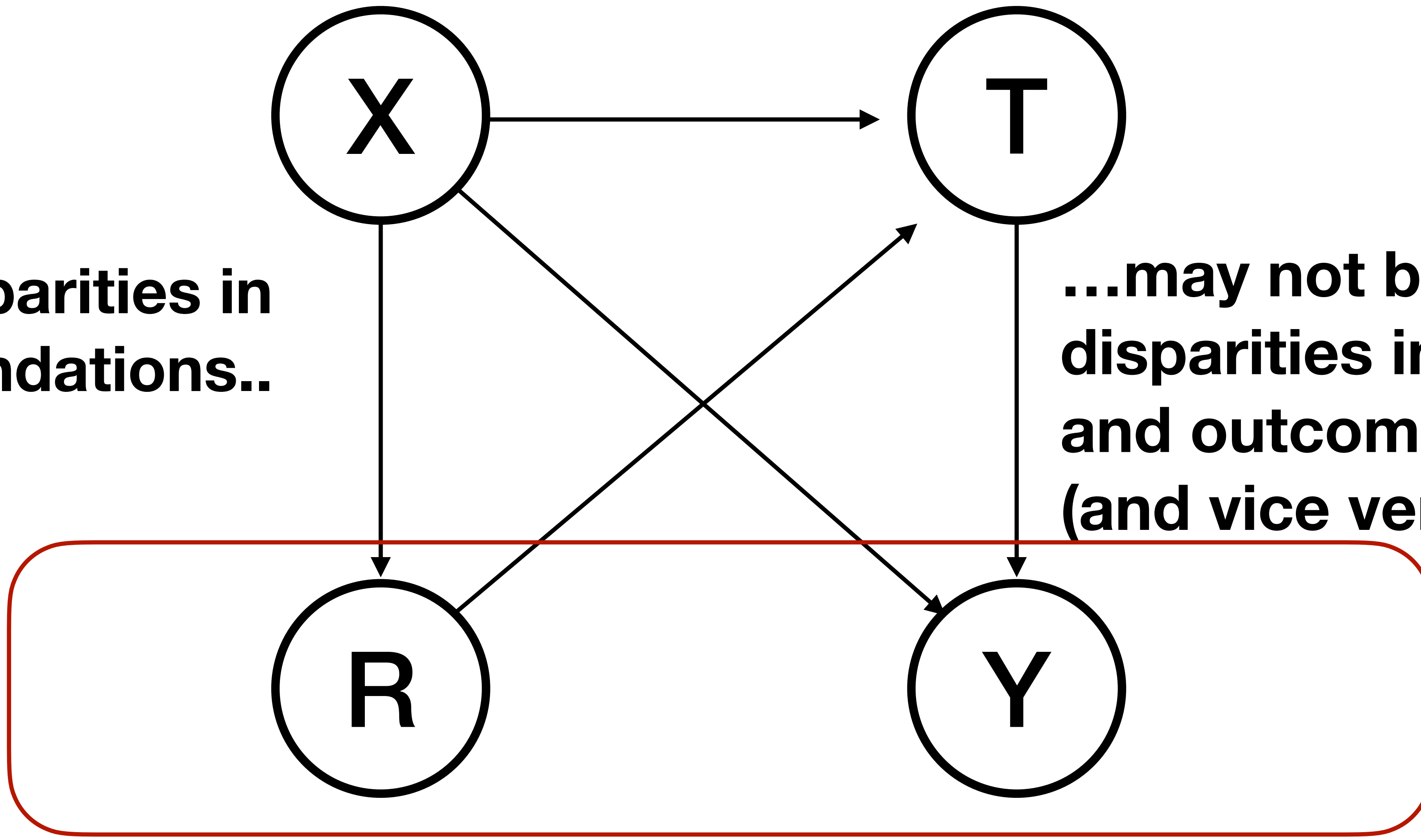
Exhibit 2: Decision Making Framework

Decision Making Framework Matrix						
	NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
FTA 1	Release with No Conditions	Release with No Conditions				
FTA 2	Release with No Conditions	Release with PM	Release with PM	PSL I	PSL II	
FTA 3		Release with PM	PSL I	PSL II	PSL III with Curfew	Release Not Recommended
FTA 4		PSL I	PSL I	PSL III	Release with Sheriff's EM	Release Not Recommended
FTA 5		PSL I	PSL III	Release with Sheriff's EM	Release Not Recommended	Release Not Recommended
FTA 6				Release Not Recommended	Release Not Recommended	Release Not Recommended

Source: Cook County Sheriff's Office, Sheriff's Justice Institute, *Central Bond Court Report*, April 2016, p. 7, https://www.chicagoreader.com/pdf/20161026/Sheriff_s-Justice-Institute-Central-Bond-Court-Study-070616.pdf

Algorithmic recommendations need a human in the loop in consequential domains: social services, medicine, law …

Disparities in recommendations..



…may not be disparities in treatment and outcome (and vice versa)

Usual algorithmic auditing focuses on recommendations only, or assumes recs = treatment!

Optimal encouragement designs

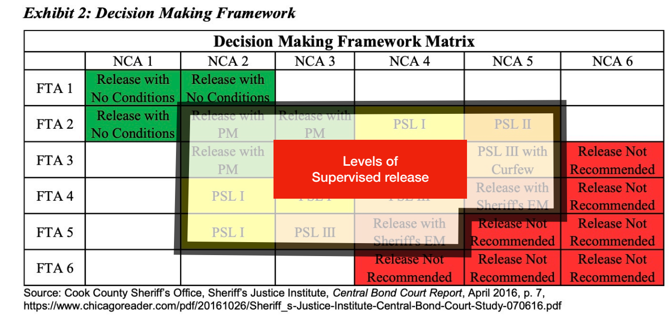


Exhibit 2: Decision Making Framework

Decision Making Framework Matrix						
	NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
FTA 1	Release with No Conditions	Release with No Conditions				
FTA 2	Release with No Conditions	Release with PM	Release with PM	PSL I	PSL II	
FTA 3		Release with PM	[illegible]	[illegible]	PSL III with Curfew	Release Not Recommended
FTA 4		PSL I	[illegible]	[illegible]	Release with Sheriff's EM	Release Not Recommended
FTA 5		PSL I	PSL III	Release with Sheriff's EM	Release Not Recommended	Release Not Recommended
FTA 6				Release Not Recommended	Release Not Recommended	Release Not Recommended

Source: Cook County Sheriff's Office, Sheriff's Justice Institute, *Central Bond Court Report*, April 2016, p. 7, https://www.chicagoreader.com/pdf/20161026/Sheriff_s-Justice-Institute-Central-Bond-Court-Study-070616.pdf

- Goal: A data-driven optimal decision rule $\pi(X)$
  … recommend treatment
  … to optimize <u>*population expected utility*</u>

  … subject to policy-relevant concerns
  (<u>*parity in beneficial resources*</u>)

-

$$\max\ \mathbb{E}[u(\pi, T(\pi), Y(\pi))]$$
$$\mathbb{E}[T(\pi) \mid A = a] - \mathbb{E}[T(\pi) \mid A = b] \leq \epsilon$$
$$\mathbb{E}[T(\pi)] \leq \kappa$$

Case study: Oregon insurance study

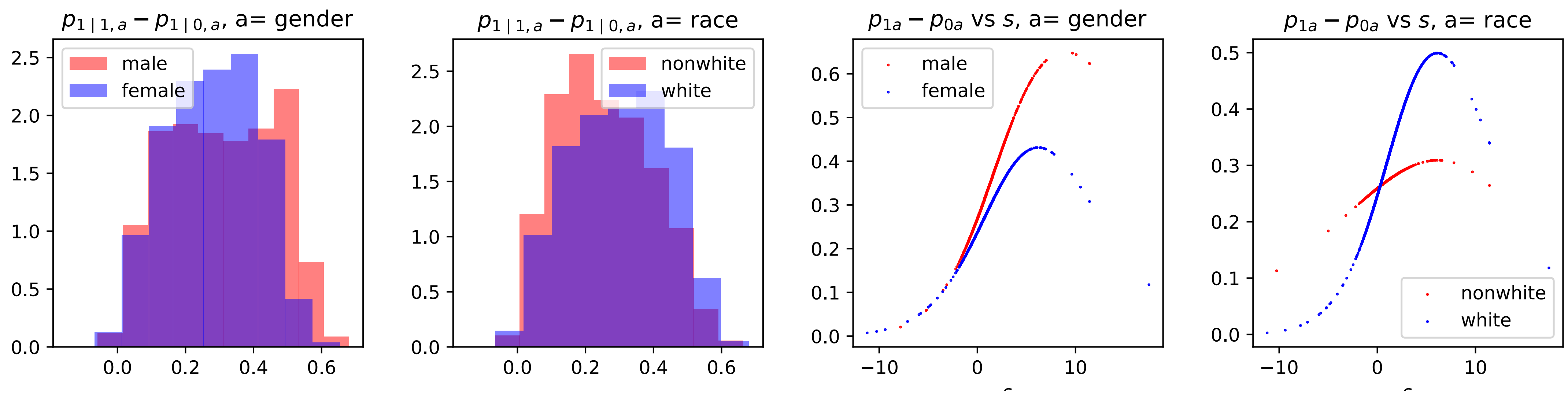


Figure 1: Distribution of lift in treatment probabilities $p_{1|1,a} - p_{1|0,a} = P(T = 1 \mid R = 1, A = a, X) - P(T = 1 \mid R = 0, A = a, X)$, and plot of $p_{1|1,a} - p_{1|0,a}$ vs. τ.

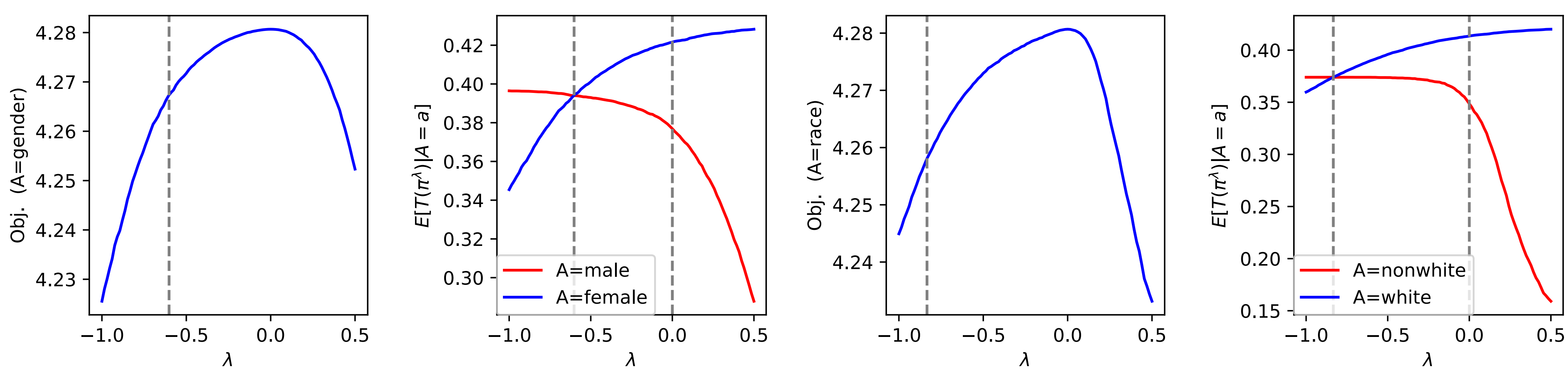


Figure 2: Policy value $V(\pi^{\lambda})$, treatment value $\mathbb{E}[T(\pi^{\lambda}) \mid A = a]$, for A = race, gender.

Case study: Supervised release

- Pretrial risk assessment in criminal justice
  R: Recommendation for supervised release
  T: Electronic monitoring (EM)
  Y: Failure to appear (FTA)
  Y(t(r)): does someone FTA when T=t,R=r?

- Causal impact of ***electronic monitoring*** (EM) on reducing failure to appear for court date
  Heterogeneity: EM could mean
  beneficial services or
  burdensome surveillance (paying for EM, losing job)

PSA-DMF
*(Coarsened) data from Chicago

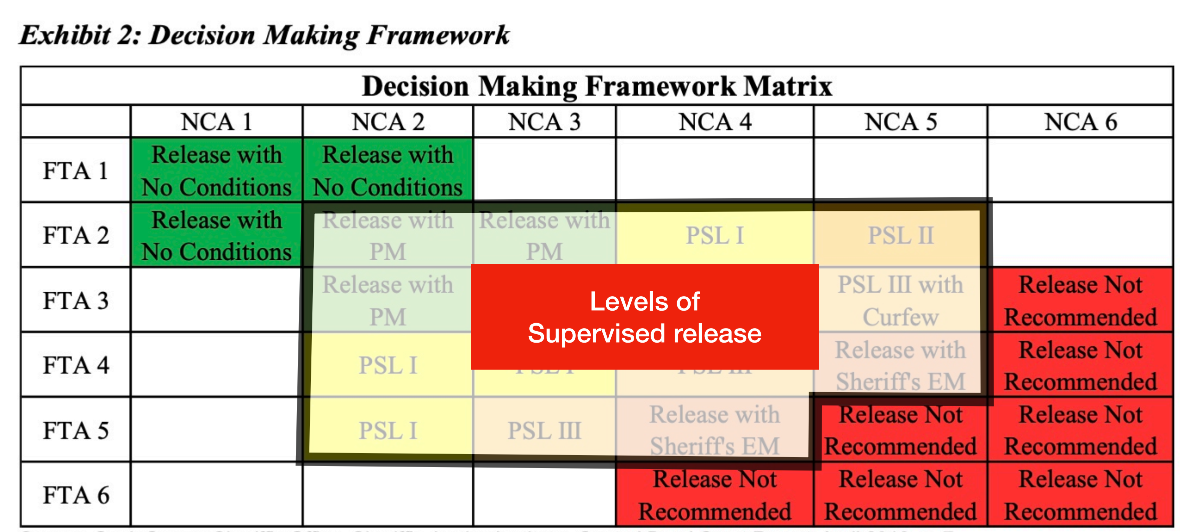


Exhibit 2: Decision Making Framework

Decision Making Framework Matrix						
	NCA 1	NCA 2	NCA 3	NCA 4	NCA 5	NCA 6
FTA 1	Release with No Conditions	Release with No Conditions				
FTA 2	Release with No Conditions	Release with PM	Release with PM	PSL I	PSL II	
FTA 3		Release with PM	[illegible]	[illegible]	PSL III with Curfew	Release Not Recommended
FTA 4		PSL I	[illegible]	[illegible]	Release with Sheriff's EM	Release Not Recommended
FTA 5		PSL I	PSL III	Release with Sheriff's EM	Release Not Recommended	Release Not Recommended
FTA 6				Release Not Recommended	Release Not Recommended	Release Not Recommended

Source: Cook County Sheriff's Office, Sheriff's Justice Institute, *Central Bond Court Report*, April 2016, p. 7,
https://www.chicagoreader.com/pdf/20161026/Sheriff_s-Justice-Institute-Central-Bond-Court-Study-070616.pdf

- “Program evaluation view of algorithmic accountability”
- Contributions: modeling, methodology

Applied context:
Targeted service delivery interventions to reduce disparities