The hidden effects of algorithmic recommendations

Alex Albright
Opportunity and Inclusive Growth Institute
Federal Reserve Bank of Minneapolis

Algorithms in high-stakes decision-making

- **Hiring:** resume scores
- **Consumer finance:** credit scores
- **Housing**: housing readiness
- **Health**: risk scores for mental health
- **Justice**: risk scores for pretrial misconduct

Algorithms in high-stakes decision-making

- **Hiring:** resume scores
- **Consumer finance:** credit scores
- **Housing**: housing readiness
- **Health**: risk scores for mental health
- **Justice**: risk scores for pretrial misconduct

But human decision-makers often make the final decision

(e.g., hiring managers, loan officers, therapists, judges)

How are algorithms given to decision-makers?

Consider a hiring manager reviewing a job applicant…

Resume score: high

Algorithm output

(prediction from algorithm)

How are algorithms given to decision-makers?

Consider a hiring manager reviewing a job applicant...

Resume score: high	+	*Hire this person*
Algorithm output *(prediction from algorithm)*		Algorithmic recommendation *(mapping from prediction to action)*

How are algorithms given to decision-makers?

Consider a hiring manager reviewing a job applicant...

Resume score: high	+	*Hire this person*
Algorithm output *(prediction from algorithm)*		Algorithmic recommendation *(mapping from prediction to action)*

Studying "the effect of algorithms" on decisions often confounds these two components

How are algorithms given to decision-makers?

Consider a hiring manager reviewing a job applicant...

Resume score: high + *Hire this person*

Algorithm output
(prediction from algorithm)

Algorithmic recommendation
(mapping from prediction to action)

Studying "the effect of algorithms" on decisions often confounds these two components

Empirical challenge: usually introduced at the same time

The hidden effects of algorithmic recommendations

- Predictions generated by algorithms are distinct from algorithmic recommendations, which are often overlooked
  - **This paper: isolate the effects of recommendations**

The hidden effects of algorithmic recommendations

- Predictions generated by algorithms are distinct from algorithmic recommendations, which are often overlooked
    - **This paper: isolate the effects of recommendations**

How? => Leverage a setting *(judges making bail decisions in CJS)* where

1. algorithmic predictions *(risk scores)* available to decision-makers stay the same
2. BUT algorithmic recommendations vary

Preview of Results

1. Basic fact: **Recommendations matter**
    - Algorithmic recommendations impact decisions
      (a lenient recommendation increases lenient bail for marginal cases by 50+%)
    - Recommendations have independent effects from algorithms themselves

Preview of Results

1. Basic fact: **Recommendations matter**
    - Algorithmic recommendations impact decisions
      (a lenient recommendation increases lenient bail for marginal cases by 50+%)
    - Recommendations have independent effects from algorithms themselves
2. Testing mechanisms: **Recommendations can change private costs of errors**
    - Making mistakes is less costly when decision consistent with recommendation
      (lenient recommendations from a social planner provide “cover” for judges)
    - Impacting payoffs rather than just risk predictions

Preview of Results

1. Basic fact: **Recommendations matter**
    - Algorithmic recommendations impact decisions
      (a lenient recommendation increases lenient bail for marginal cases by 50+%)
    - Recommendations have independent effects from algorithms themselves
2. Testing mechanisms: **Recommendations can change private costs of errors**
    - Making mistakes is less costly when decision consistent with recommendation
      (lenient recommendations from a social planner provide "cover" for judges)
    - Impacting payoffs rather than just risk predictions
3. Heterogeneity: **Recommendations may not impact all groups equally**
    - Judges deviate from lenient recommendation more for Black defendants than for white defendants with identical algorithmic scores

Roadmap

1. Algorithmic Systems Background
2. Empirical setting: Kentucky bail decisions
3. What are the effects of algorithmic recommendations?
4. What is the mechanism behind the effect?
5. Heterogeneous effects of recommendations and implications for racial inequality

Algorithmic Systems Background

Algorithmic systems

Discretion;
No algorithm

No discretion;
Dictated by
algorithm-based rule

Algorithmic systems

Berk (2017), Jung et al. (2017), Mullainathan and Obermeyer (2022), Kleinberg et al. (2018)

Discretion;
No algorithm

No discretion;
Dictated by
algorithm-based rule

These papers: algorithms can outperform human decisions

...but what about when humans are involved?

Algorithmic systems

Berk (2017), Jung et al. (2017), Mullainathan and Obermeyer (2022), Kleinberg et al. (2018)

Discretion;
No algorithm

Discretion;
Informed by algorithm

No discretion;
Dictated by algorithm-based rule

Algorithmic systems

Berk (2017), Jung et al. (2017), Mullainathan and Obermeyer (2022), Kleinberg et al. (2018)

Discretion;
No algorithm

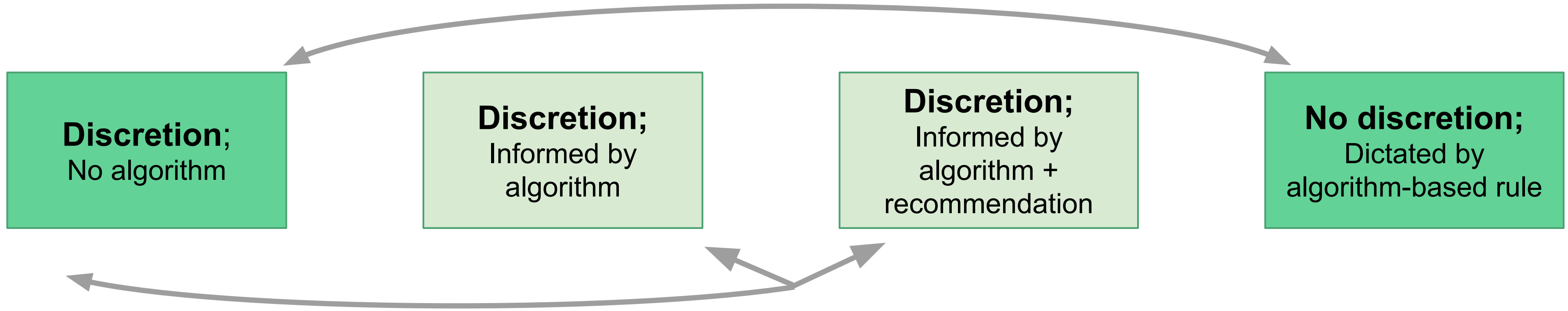
Discretion;
Informed by algorithm

Discretion;
Informed by algorithm + recommendation

No discretion;
Dictated by algorithm-based rule

Algorithmic systems

Berk (2017), Jung et al. (2017), Mullainathan and Obermeyer (2022), Kleinberg et al. (2018)

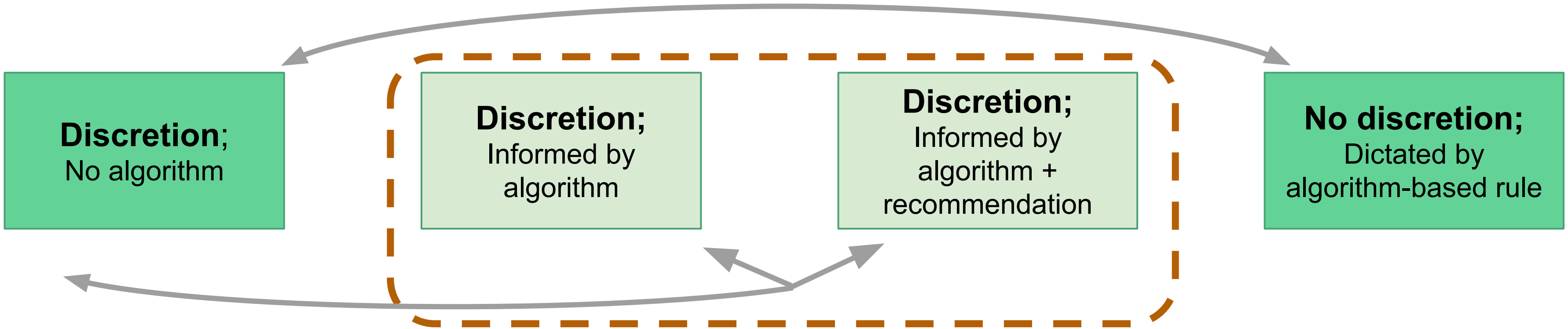


Sloan, Naufal, and Caspers (2021), Stevenson (2018), Doleac and Stevenson (2021), Garrett and Monahan (2018), DeMichele et al. (2018), Cowgill and Tucker (2019)

These papers:
how does human use of algorithms change outcomes?

Algorithmic systems

Berk (2017), Jung et al. (2017), Mullainathan and Obermeyer (2022), Kleinberg et al. (2018)



Sloan, Naufal, and Caspers (2018), Stevenson (2018), Doleac and Stevenson (2021), Garrett and Monahan (2018), DeMichele et al. (2018), Cowgill and Tucker (2019)

Today: focus on distinction between these

Criminal justice + bail context

Algorithms in criminal justice

Predictive policing, pretrial risk assessment, sentencing, prison management, parole

Algorithms in criminal justice

Predictive policing, **pretrial risk assessment**, sentencing, prison management, parole

Algorithms in criminal justice

Predictive policing, **pretrial risk assessment**, sentencing, prison management, parole

STATE	TYPE/SCOPE OF USE
Alabama	VPRAI / Jefferson County
Alaska	State Created / Statewide
Arizona	PSA / Statewide | VPRAI / 2 County Superior Courts
Arkansas	State Created / Statewide
California (Sample risk assessment documents from San Francisco, and Napa County)	PSA / 3 counties | PRRS II / 2 Counties
Colorado (sample risk assessment documents)	CPAT / Statewide | ODARA for DV / Statewide
Connecticut	State created / Statewide
Delaware	State created (DELPAT) / Statewide
District of Columbia	Developed with Urban Institute and Maxarth
Florida	PSA / Volusia County | COMPAS - Sentencing / Statewide | State Created FPRAI Being piloted / 6 Counties
Georgia	State created / Some counties
Hawaii	PSA / Statewide | ORAS-PAT / Statewide
Idaho (see FOI documents below)	State created / Statewide | Ada County / Revised IPRAI
Illinois	PSA / 4 counties | VPRAI/RVRA / Most Courts
Indiana (sample risk assessment documents)	Mandatory use of IRAS and IYAS / Statewide
Iowa	PSA / 4 Counties via Pilot Program | IRR
Kansas	State created / Johnson County
Kentucky	PSA / Statewide
Louisiana	PSA / New Orleans
Maine	ODARA (sex offenders) / Statewide | 2019 Task Force for expansion
Maryland	State created / Most counties
Massachusetts	COMPAS / Statewide LS/CMI / Statewide
Michigan	COMPAS for Sentencing / Statewide
Minnesota (see Pretrial Release Evaluation Form and Bench Card)	MNPAT / Statewide
Mississippi	CRJ (Crime Justice Institute) / Statewide
Missouri	PSA / 1 County | Statewide / State created | Separate statewide system for Juvenile and Sex Offenders | Use Oregon Public Safety Checklist for Sentencing
Montana	PSA / 2 Counties and 5 Pilot Counties
Nebraska	STRONG-R
Nevada	State created / Statewide Mar. 2019 by NV Supreme Court
New Hampshire	Yes
New Jersey	PSA / Statewide
New Mexico	PSA / 4 Counties | ODARA for DV
New York	(NYC) City Created / Citywide |State Created / State-wide for Parole
North Carolina	PSA / 1 County | Developing another statewide one
Ohio	PSA / 3 Counties | ORAS-PAT / Statewide
Oklahoma	ORAS for Pretrial Services Program + LSI/R / Statewide
Oregon (sample assessments)	Public Safety Checklist
Pennsylvania	PSA / Allegheny County | State created / 1 County
Rhode Island	PSA / Statewide
South Carolina	State Created - Cash Bail Use
South Dakota	PSA / 2 Counties
Tennessee	PSA / 2 Counties | State Created / One Judicial District Test
Texas (sample assessments)	PSA / Harris + Dallas County | PRAISTX (derivative of ORAS) / Statewide Parole Board
Utah	PSA / Statewide
Vermont	ORAS
Virginia	VPRAI revised by Luminosity / Statewide | Use Oregon Public Safety Checklist for Sentencing
Washington	PSA / 3 Counties
West Virginia	LS/CMI
Wisconsin (See sample assessment documents)	PSA / 4 Counties | COMPAS / Statewide
Wyoming	COMPAS for Prisoners / Statewide
Federal	PTRA

Source: Epic (2020)

Algorithms in criminal justice

Predictive policing, **pretrial risk assessment**, sentencing, prison management, parole

STATE	TYPE/SCOPE OF USE
Alabama	VPRAI / Jefferson County
Alaska	State Created / Statewide
Arizona	PSA / Statewide | VPRAI / 2 County Superior Courts
Arkansas	State Created / Statewide
California (Sample risk assessment documents from San Francisco, and Napa County)	
Colorado (sample risk assessment documents)	CPAT / Statewide | ODARA for DV / Statewide
Connecticut	State created / Statewide
Delaware	State created (DELPAT) / Statewide
District of Columbia	Developed with Urban Institute and Maxarth
Florida	PSA / Volusia County | COMPAS - Sentencing / Statewide | State Created FPRAI Being piloted / 6 Counties
Georgia	State created / Some counties
Hawaii	PSA / Statewide | ORAS-PAT / Statewide
Idaho (see FOI documents below)	State created / Statewide | Ada County / Revised IPRAI
Illinois	PSA / 4 counties | VPRAI/RVRA / Most Courts
Indiana (sample risk assessment documents)	Mandatory use of IRAS and IYAS / Statewide
	PSA / 4 Counties via Pilot
Maine	Statewide | 2019 Task Force for expansion
Maryland	State created / Most counties
Massachusetts	COMPAS / Statewide LS/CMI / Statewide
Michigan	COMPAS for Sentencing / Statewide
Minnesota (see Pretrial Release Evaluation Form and Bench Card)	MNPAT / Statewide
Mississippi	CRJ (Crime Justice Institute) / Statewide
Missouri	PSA / 1 County | Statewide / State created | Separate statewide system for Juvenile and Sex Offenders | Use Oregon Public Safety Checklist for Sentencing
Montana	PSA / 2 Counties and 5 Pilot Counties
Nebraska	STRONG-R
Nevada	State created / Statewide Mar. 2019 by NV Supreme Court
New Hampshire	Yes
New Jersey	PSA / Statewide
New Mexico	PSA / 4 Counties | ODARA for DV
	(NYC) City Created / Citywide
Ohio	Statewide
Oklahoma	ORAS for Pretrial Services Program + LSI/R / Statewide
Oregon (sample assessments)	Public Safety Checklist
Pennsylvania	PSA / Allegheny County | State created / 1 County
Rhode Island	PSA / Statewide
South Carolina	State Created - Cash Bail Use
South Dakota	PSA / 2 Counties
Tennessee	PSA / 2 Counties | State Created / One Judicial District Test
Texas (sample assessments)	PSA / Harris + Dallas County | PRAISTX (derivative of ORAS) / Statewide Parole Board
Utah	PSA / Statewide
Vermont	ORAS
Virginia	VPRAI revised by Luminosity / Statewide | Use Oregon Public Safety Checklist for Sentencing
Washington	PSA / 3 Counties
West Virginia	LS/CMI
Wisconsin (See sample assessment documents)	PSA / 4 Counties | COMPAS / Statewide
Wyoming	COMPAS for Prisoners / Statewide
Federal	PTRA

Goal: "data-driven way to advance pretrial release"

Source: Epic (2020)

Algorithms in criminal justice

Predictive policing, **pretrial risk assessment**, sentencing, prison management, parole

STATE	TYPE/SCOPE OF USE
Alabama	VPRAI / Jefferson County
Alaska	State Created / Statewide
Arizona	PSA / Statewide | VPRAI / 2 County Superior Courts
Arkansas	State Created / Statewide
California (Sample risk assessment documents from San Francisco, and Napa County)	
Colorado (sample risk assessment documents)	CPAT / Statewide | ODARA for DV / Statewide
Connecticut	
Delaware	
District of Columbia	Developed with Urban Institute and Maxarth
Florida	PSA / Volusia County | COMPAS - Sentencing / Statewide | State Created FPRAI Being piloted / 6 Counties

STATE	TYPE/SCOPE OF USE
Georgia	State created / Some counties
Hawaii	PSA / Statewide | ORAS-PAT / Statewide
Idaho (see FOI documents below)	State created / Statewide | Ada County / Revised IPRAI
Illinois	PSA / 4 counties | VPRAI/RVRA / Most Courts
Indiana (sample risk assessment documents)	Mandatory use of IRAS and IYAS / Statewide
	PSA / 4 Counties via Pilot
Maine	Statewide | 2019 Task Force for expansion
Maryland	State created / Most counties
Massachusetts	COMPAS / Statewide
Mississippi	CRJ (Crime Justice Institute) / Statewide
Missouri	PSA / 1 County | Statewide / State created | Separate statewide system for Juvenile and Sex Offenders | Use Oregon Public Safety Checklist for Sentencing

Montana	PSA / 2 Counties and 5 Pilot Counties
Nebraska	STRONG-R
Nevada	State created / Statewide Mar. 2019 by NV Supreme Court
New Hampshire	Yes
New Jersey	PSA / Statewide
New Mexico	PSA / 4 Counties | ODARA for DV
Ohio	Statewide
Oklahoma	ORAS for Pretrial Services Program + LSI/R / Statewide
South Carolina	State Created - Cash Bail Use
South Dakota	PSA / 2 Counties
Tennessee	PSA / 2 Counties | State Created / One Judicial District Test
Texas (sample assessments)	PSA / Harris + Dallas County | PRAISTX (derivative of ORAS) / Statewide Parole Board
Utah	PSA / Statewide
Vermont	ORAS

Virginia	VPRAI revised by Luminosity / Statewide | Use Oregon Public Safety Checklist for Sentencing
shington	PSA / 3 Counties
st Virginia	LS/CMI
consin (See sample essment uments)	PSA / 4 Counties | COMPAS / Statewide
Wyoming	COMPAS for Prisoners / Statewide
Federal	PTRA

Goal: "data-driven way to advance pretrial release"

Predict misconduct based on observable data

Source: Epic (2020)

“A data-driven way to advance pretrial release”

Econ-speak: more efficiently allocate detention

> “one policy simulation *[where an algorithm makes decisions instead of judges]* shows [pretrial rearrest] reductions up to 24.7% with no change in jailing rates”
>
> (Kleinberg et al. 2018)

“A data-driven way to advance pretrial release”

Econ-speak: more efficiently allocate detention

> “one policy simulation *[where an algorithm makes decisions instead of judges]* shows [pretrial rearrest] reductions up to 24.7% with no change in jailing rates”
>
> (Kleinberg et al. 2018)

But judges pick more than the allocation of defendants to bail decisions…

(They pick the rate of bail decisions)

Judge’s decision *(before algorithmic recommendations)*

Judge j has a choice between release or detain – she wants to pick the less costly action

Judge's decision *(before algorithmic recommendations)*

Judge j has a choice between release or detain – she wants to pick the less costly action

- If detain, that costs c_{jd}
- If release, there is some probability, predicted by judge j, $P_{ji}(m)$ of misconduct in case i
  - misconduct always has cost c_{jm}

integrates information about score for case i $s_i(m)$

Judge releases iff $c_{jd} > P_{ji}(m)c_{jm}$

Judge's decision *(before algorithmic recommendations)*

Judge j has a choice between release or detain – she wants to pick the less costly action

- If detain, that costs c_{jd}
- If release, there is some probability, predicted by judge j, $P_{ji}(m)$ of misconduct in case i
    - misconduct always has cost c_{jm}

integrates information about score for case i $s_i(m)$

Judge releases iff $c_{jd} > P_{ji}(m)c_{jm}$ or $c_{jd} / c_{jm} > P_{ji}(m)$

$P_{j1}(m)$ $P_{j2}(m)$ … … … $P_{jN-1}(m)$ $P_{jN}(m)$

Lowest predicted risk — Highest predicted risk

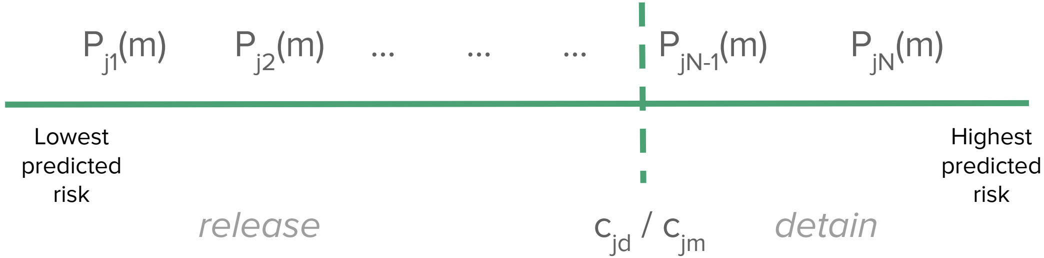
Judge's decision *(before algorithmic recommendations)*

Judge j has a choice between release or detain – she wants to pick the less costly action

- If detain, that costs c_{jd}
- If release, there is some probability, predicted by judge j, $P_{ji}(m)$ of misconduct in case i
    - misconduct always has cost c_{jm}

integrates information about score for case i $s_i(m)$

Judge releases iff $c_{jd} > P_{ji}(m)c_{jm}$ or $c_{jd} / c_{jm} > P_{ji}(m)$



Why might recommendations matter?

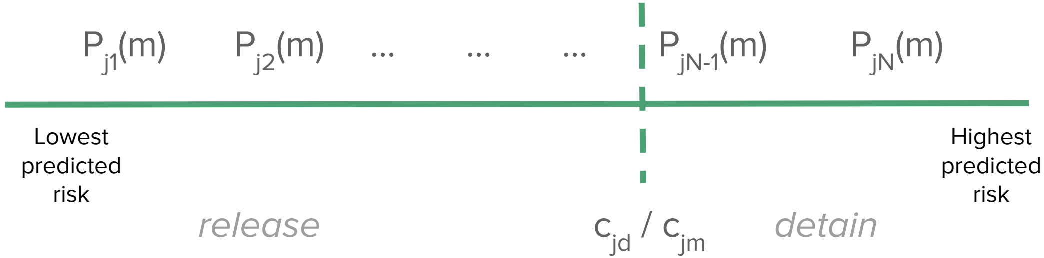
Algorithms predict outcomes… but ***many*** choice sets can be consistent with algorithmic output

Algorithmic recommendations are normative – they are informative about preferred trade-offs of the algorithm designer

Why might recommendations matter?

Algorithms predict outcomes… but ***many*** choice sets can be consistent with algorithmic output

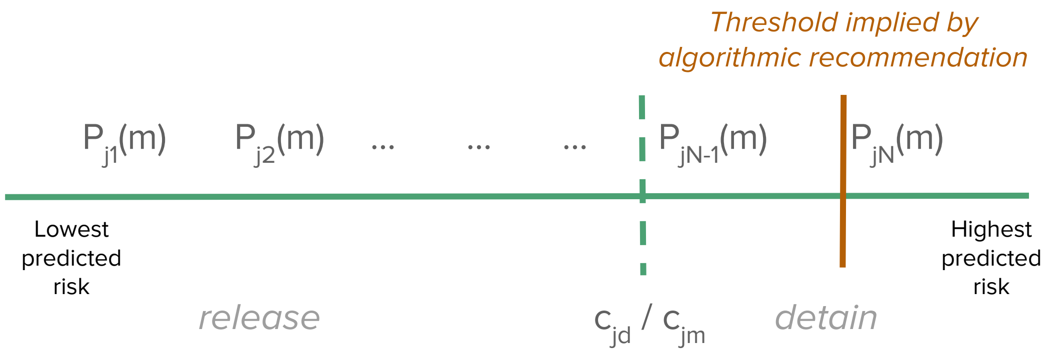
Algorithmic recommendations are normative – they are informative about preferred trade-offs of the algorithm designer



Why might recommendations matter?

Algorithms predict outcomes… but ***many*** choice sets can be consistent with algorithmic output

Algorithmic recommendations are normative – they are informative about preferred trade-offs of the algorithm designer



This algorithmic recommendation is more lenient than judge j's natural threshold

Empirical Setting: Kentucky Bail Decisions

The algorithm: Kentucky pretrial risk score

After person booked, pretrial services officer calculates a risk score

Kentucky Pretrial Risk Assessment tool
(March 2011-May 2013)

The algorithm: Kentucky pretrial risk score

After person booked, pretrial services officer calculates a risk score

Kentucky Pretrial Risk Assessment tool
(March 2011-May 2013)

- Not complex black-box ML tool – it is a “checklist tool” (or “rule-based formula”)
- Total points and convert to levels:
  - *0-5: low*
  - *6-13: moderate*
  - *14-24: high*
- Scores have relative, not absolute meaning *(e.g., high is riskier than low)*
- **Only levels shared with judges**

Risk Component	Points
No verified address	2
No verified means of support	1
ABC Felony charge	1
Pending case	7
Prior/active mis/felony FTA	2
Prior FTA traffic violation	1
Prior misdemeanors	2
Prior felonies	1
Prior violent convictions	1
History of drug/alcohol abuse	2
Prior felony escape conviction	3
On probation/parole	1

House Bill introduces recommendation for some cases

Before June 2011:

- Judge receives info about defendant, incident, risk level and makes a bail decision in a few minutes
  - *Risk level: Kentucky Pretrial Risk Assessment tool*
  - Judge decides **whether to set money bail** *(requires defendant to post money for release)*

House Bill introduces recommendation for some cases

Before June 2011:

- Judge receives info about defendant, incident, risk level and makes a bail decision in a few minutes
  - *Risk level: Kentucky Pretrial Risk Assessment tool*
  - Judge decides **whether to set money bail** *(requires defendant to post money for release)*

Starting June 2011:

- House Bill *(legislature action)* **recommends no money bail ("lenient bail") for low and moderate risk level cases**
  - Judges could deviate by saying a few words *(no large admin cost)*
  - No recommendation for high risk cases

Variation in recommendation over time and scores

No recommendation

June 2011

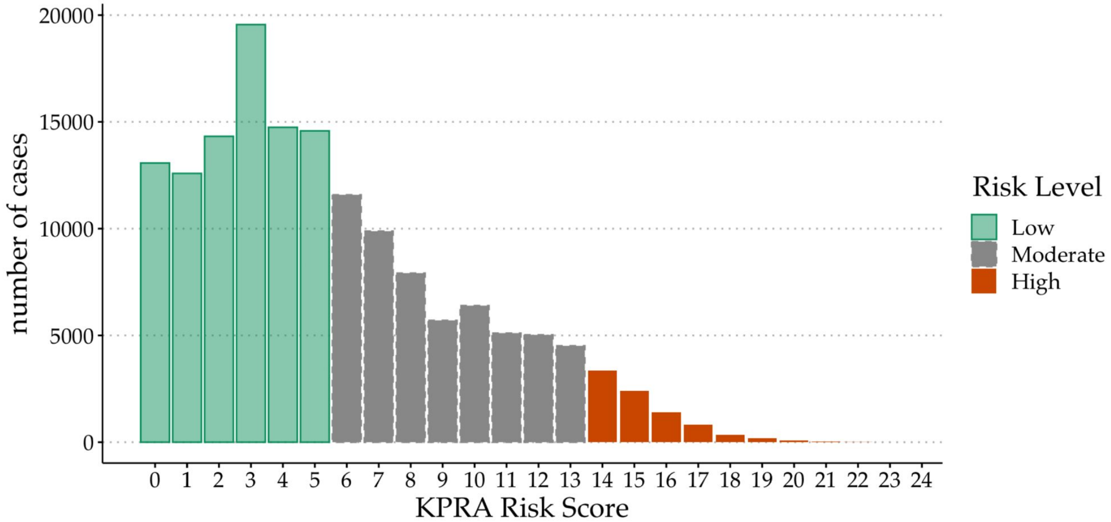
Recommendation
(lenient bail for low/moderate cases)

Variation in recommendation over time and scores

No recommendation

June 2011

Recommendation
(lenient bail for low/moderate cases)

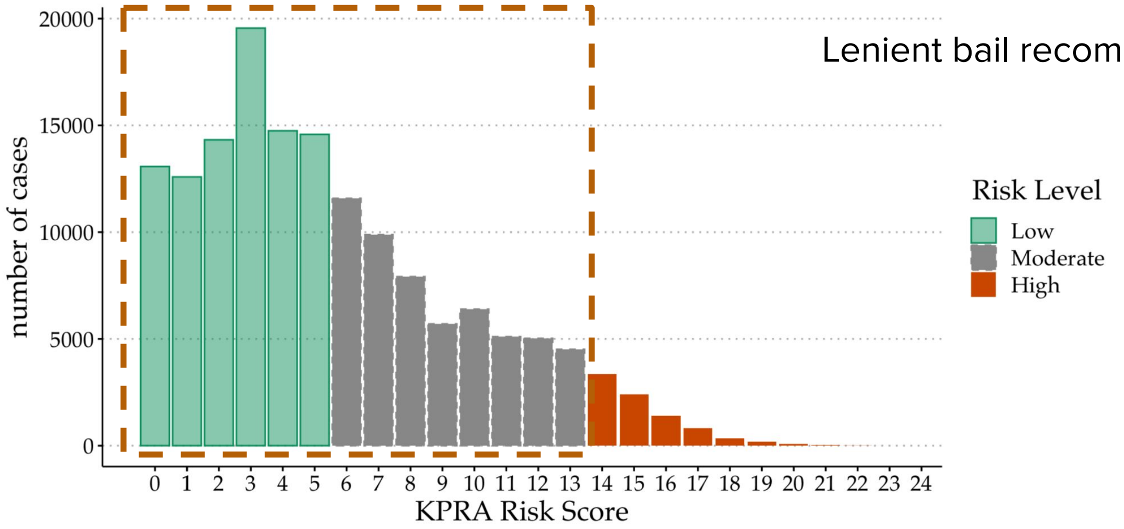


Variation in recommendation over time and scores

No recommendation

June 2011

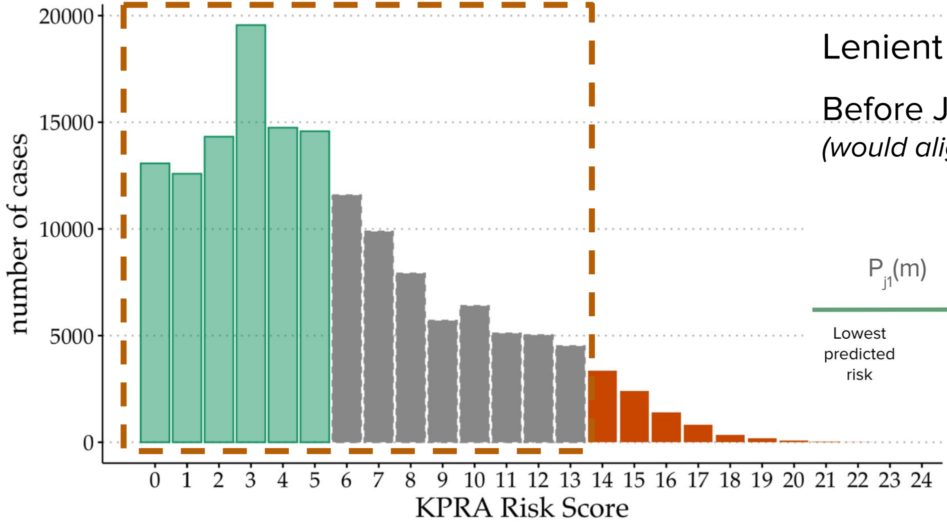
Recommendation
(lenient bail for low/moderate cases)



Lenient bail recommended for **<u>90% of cases</u>**

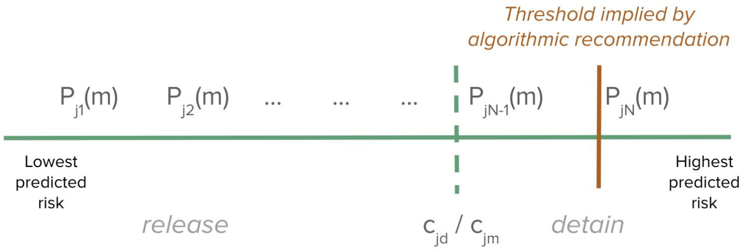
Variation in recommendation over time and scores

No recommendation | June 2011 | Recommendation *(lenient bail for low/moderate cases)*



Lenient bail recommended for **<u>90% of cases</u>**

Before June 2011, <u>only **32%**</u> got lenient bail
(would align with a threshold of score<4)



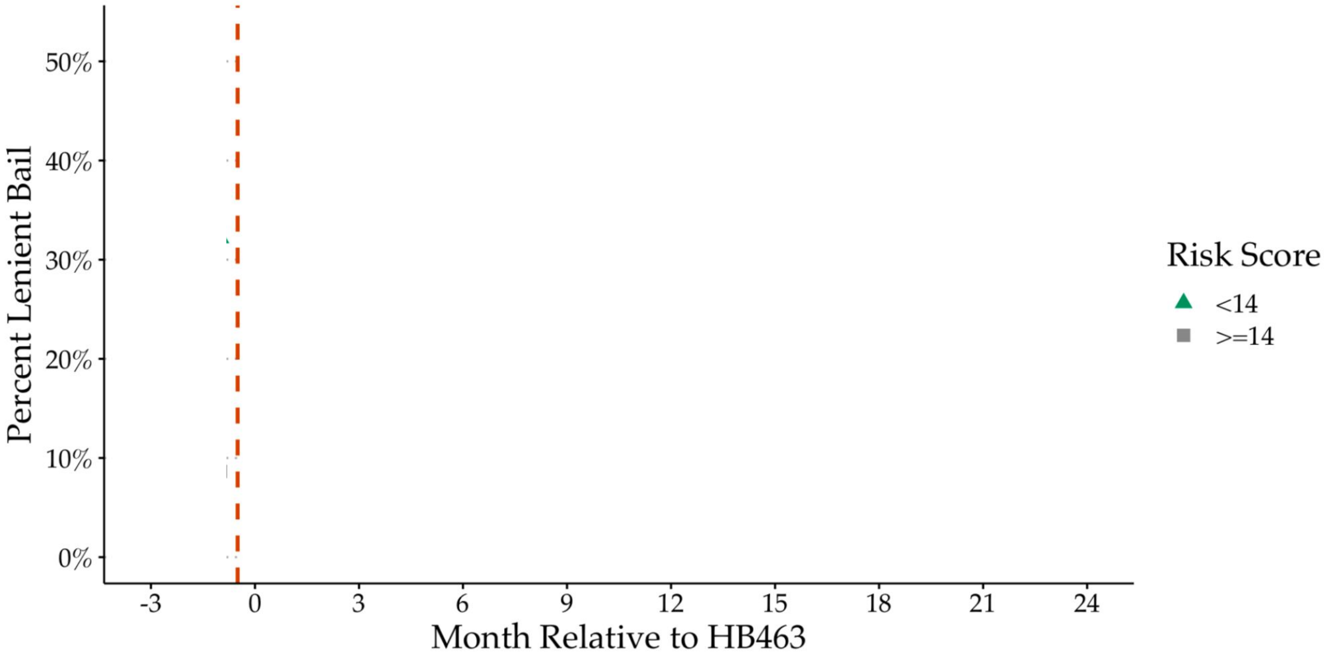
What are the effects of recommendations?

Difference-in-differences approach

- Cases with scores < 14 get a lenient recommendation
- Cases with scores >= 14 do not

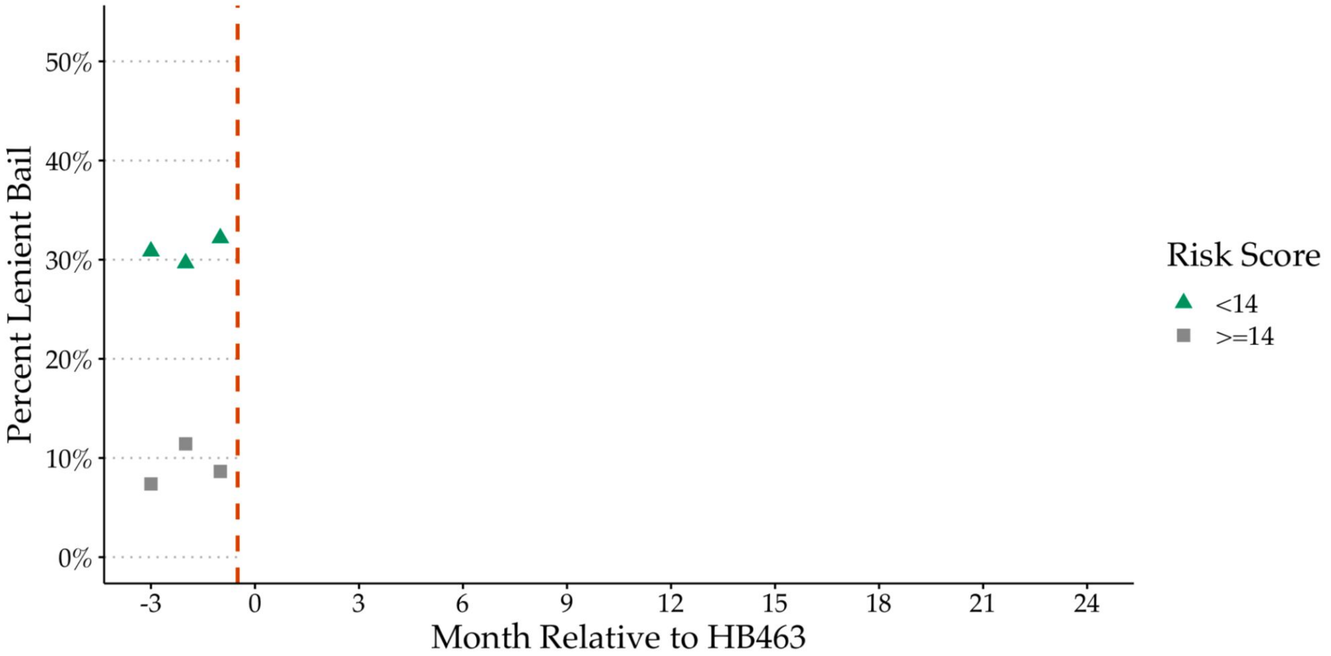
Difference-in-differences approach

- Cases with scores < 14 get a lenient recommendation
- Cases with scores >= 14 do not



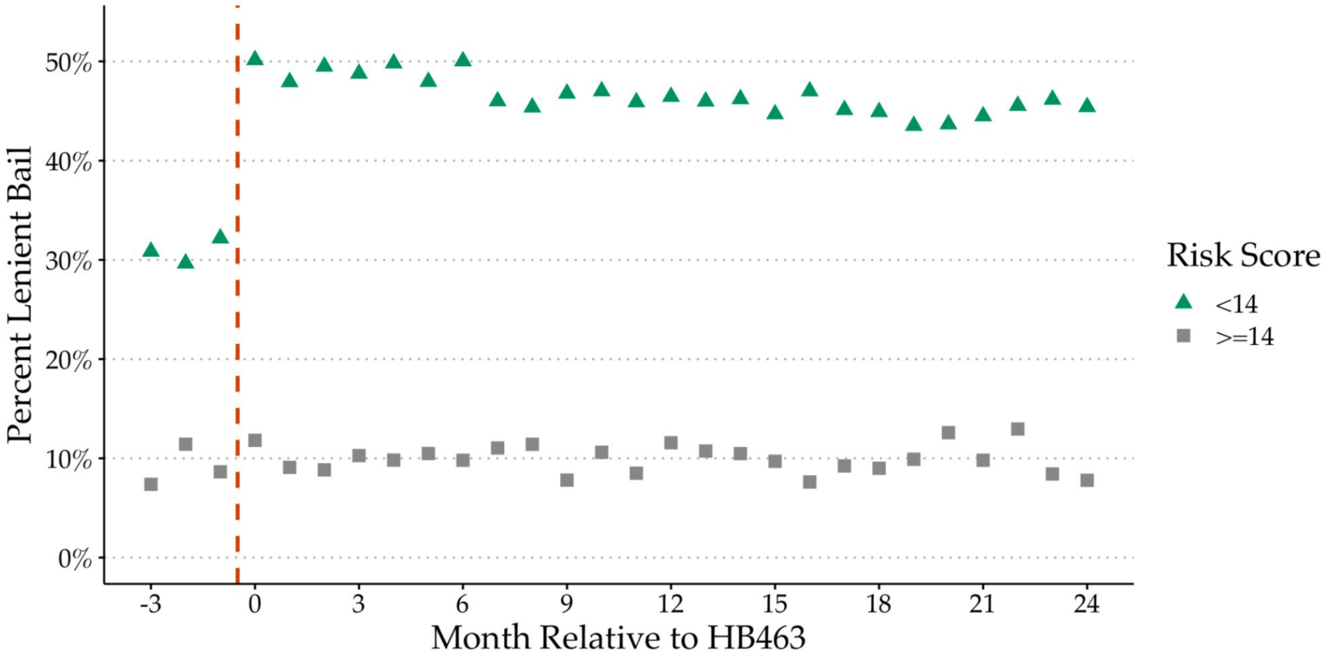
Difference-in-differences approach

- Cases with scores < 14 get a lenient recommendation
- Cases with scores >= 14 do not



Difference-in-differences approach

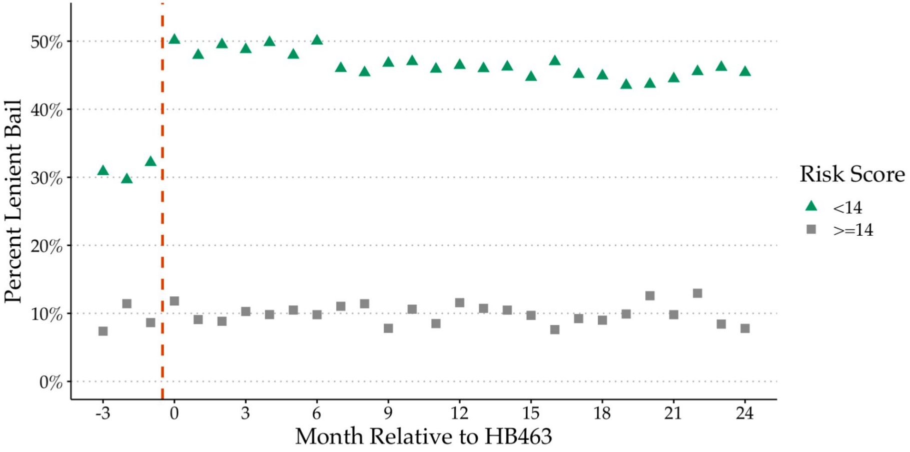
- Cases with scores < 14 get a lenient recommendation
- Cases with scores >= 14 do not

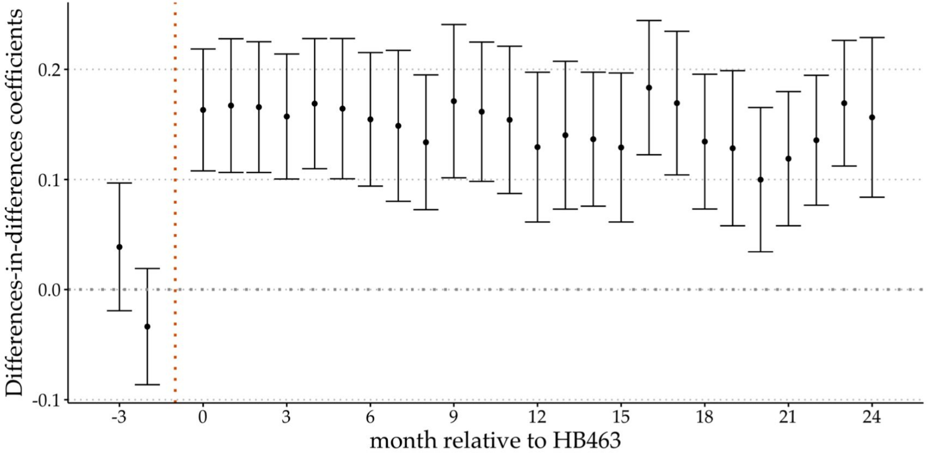


Difference-in-differences approach

- Cases with scores < 14 get a lenient recommendation
- Cases with scores >= 14 do not

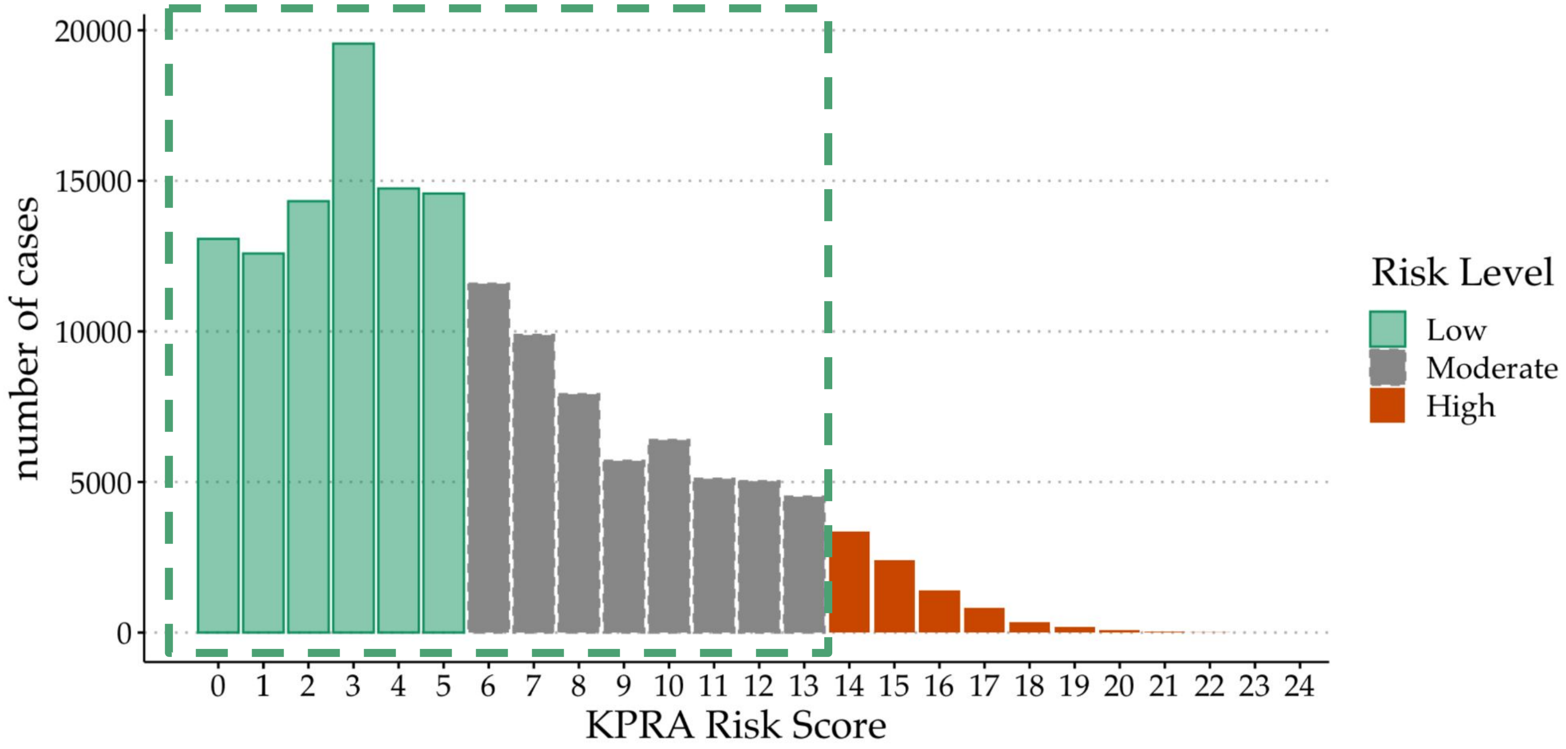
$$lenient_{itj} = \sum_{m \neq -1} [\beta_m \times I(score_i < 14)] + X_{itj} + \epsilon_{itj}$$



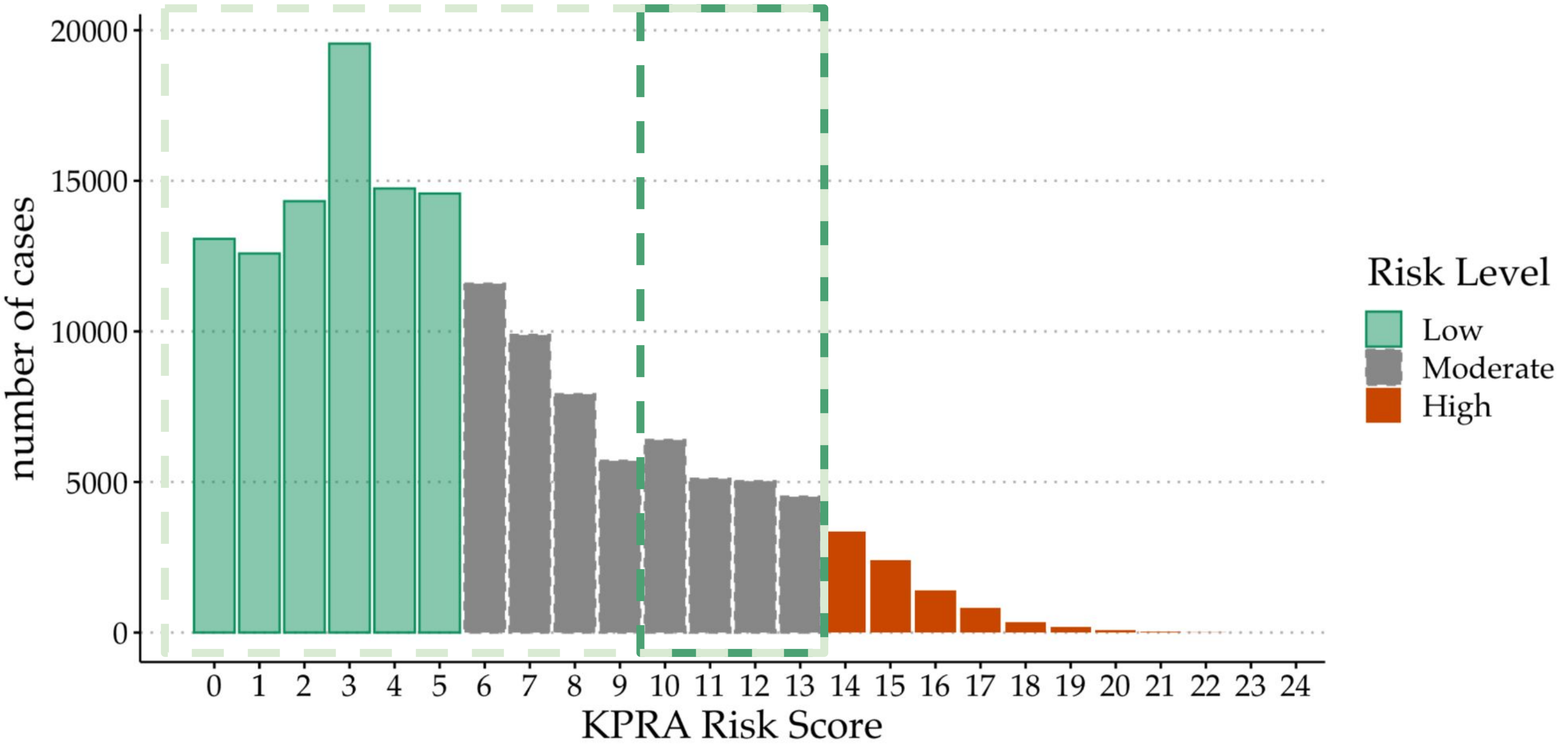


<u>Pooled DD:</u> **15 pp increase / 50% increase** (off the 30% baseline)

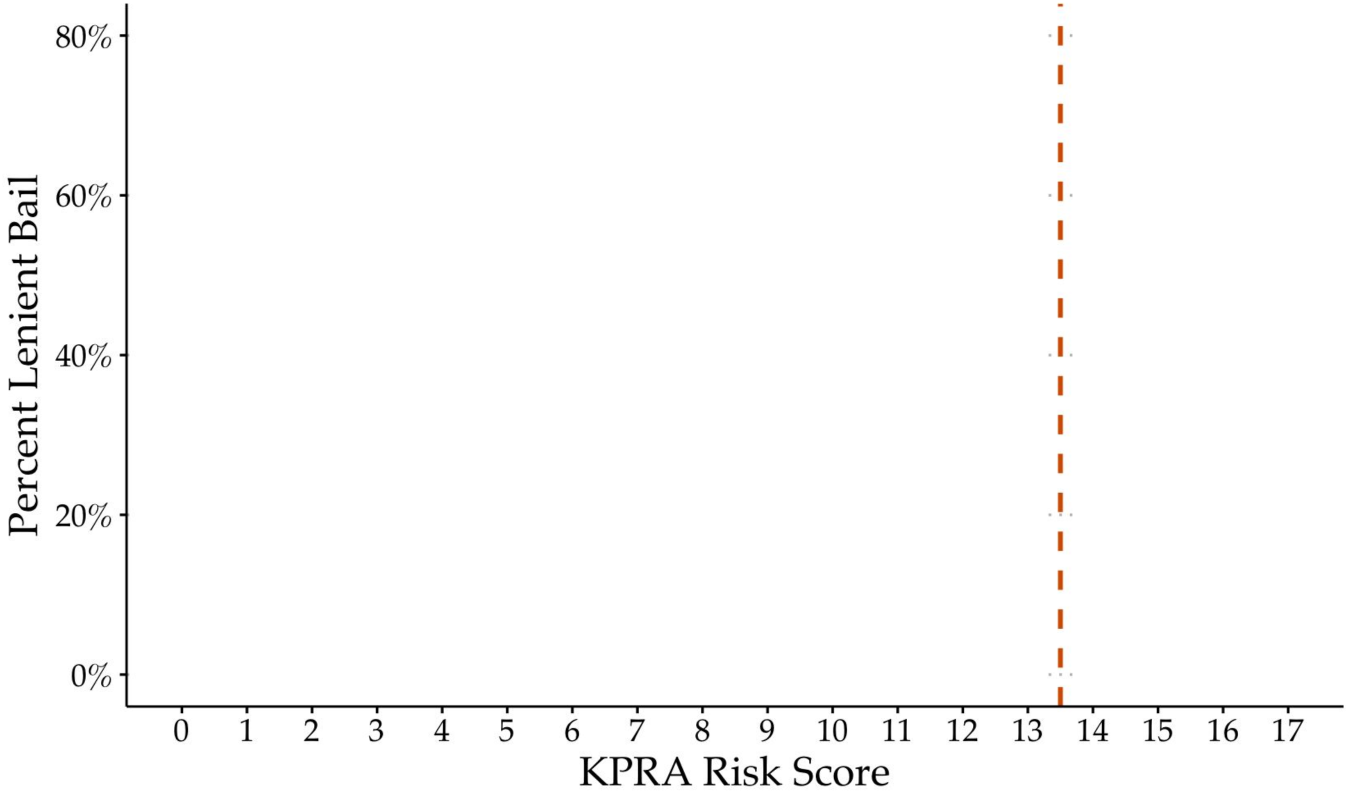
Heterogeneity in effects across the risk score distribution



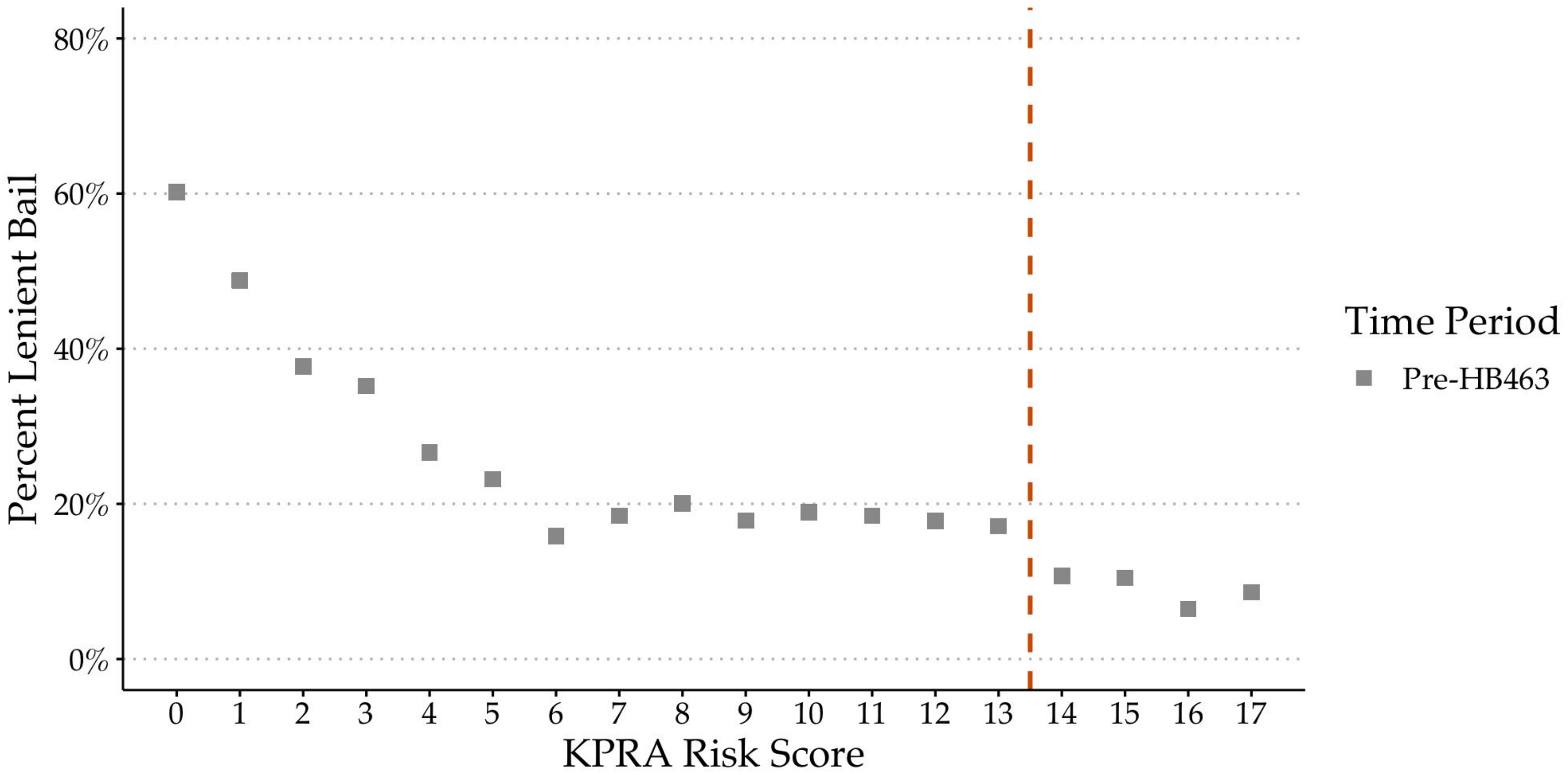
Heterogeneity in effects across the risk score distribution



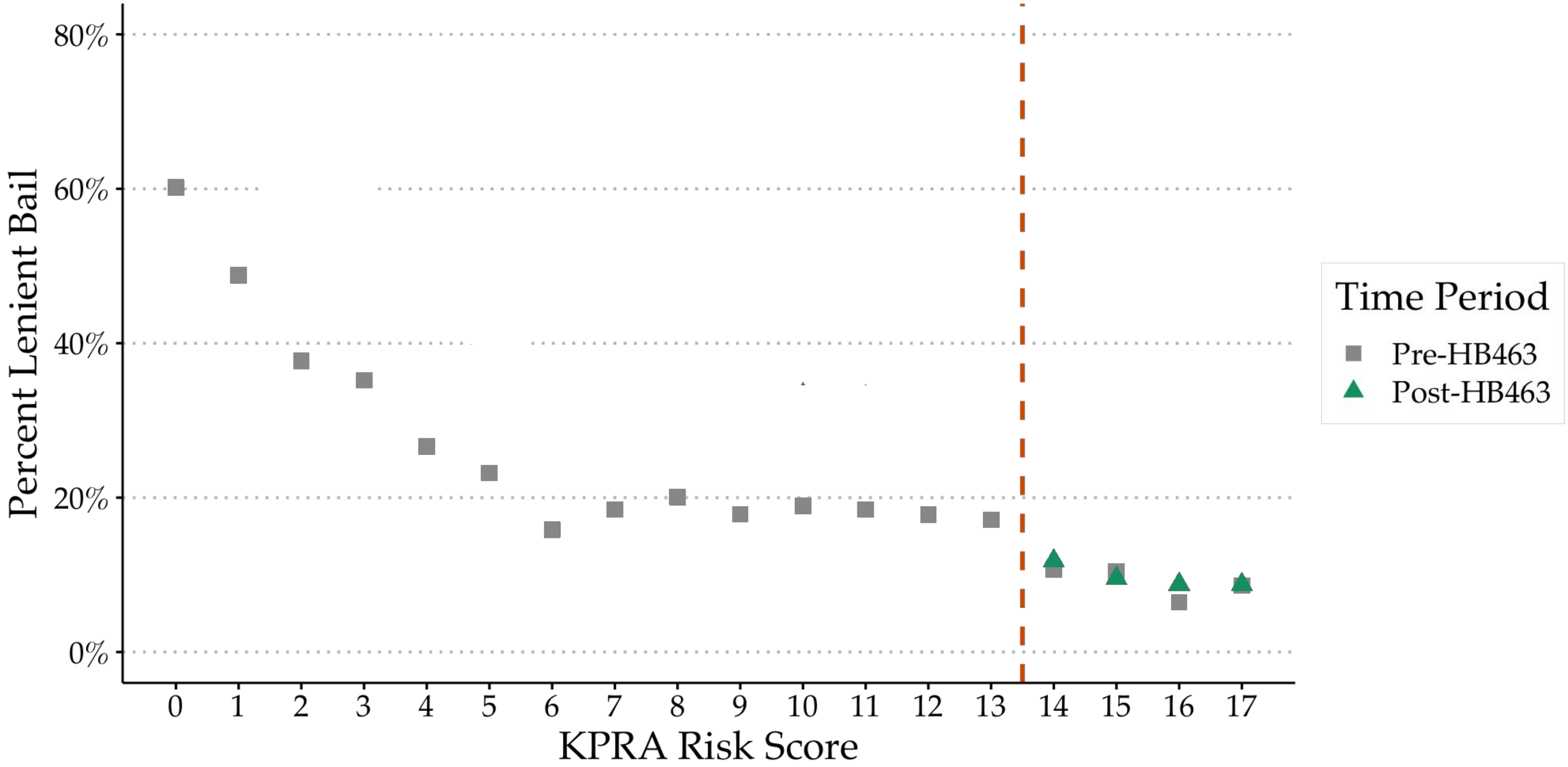
Heterogeneity in effects across the risk score distribution



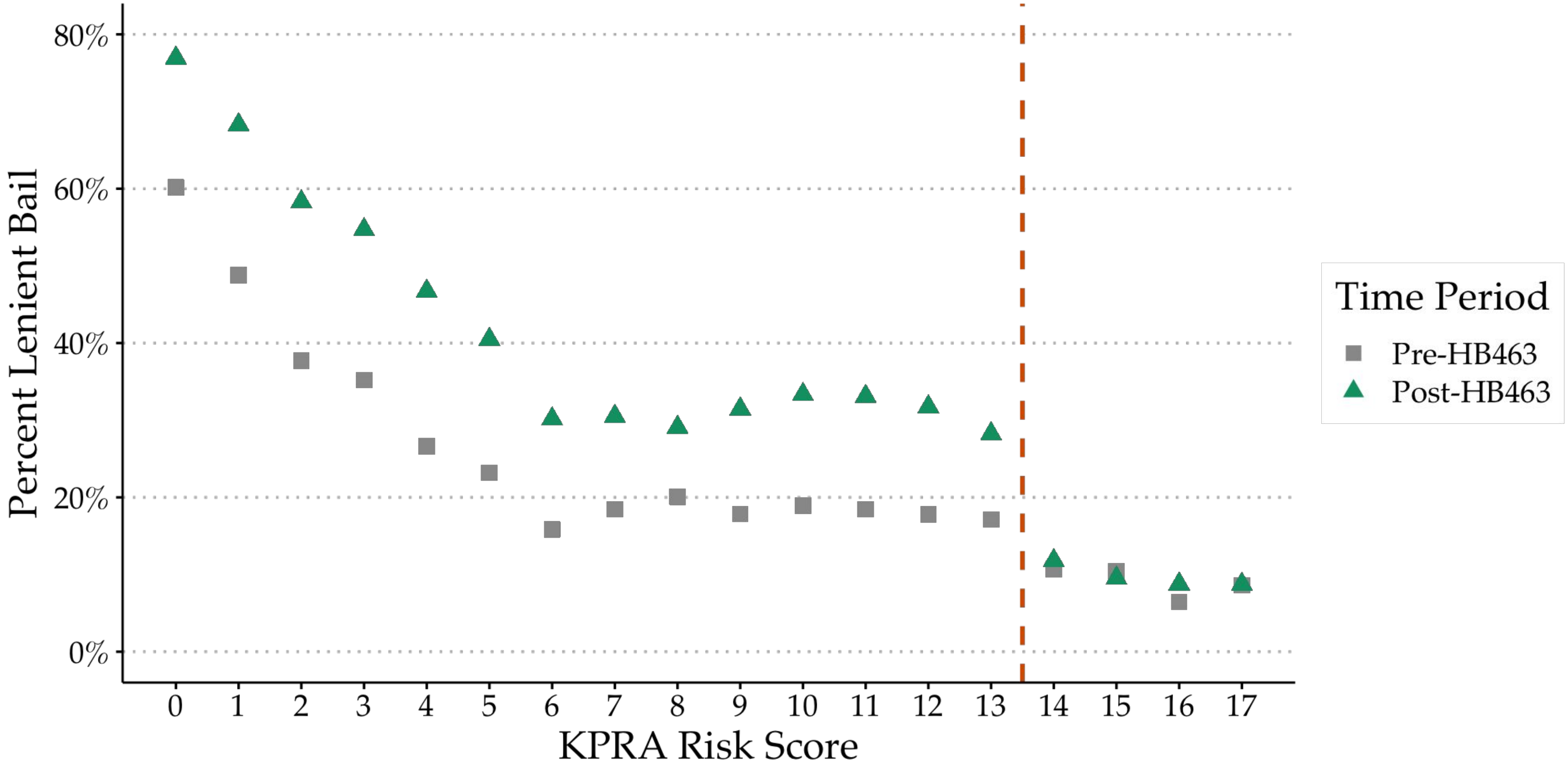
Heterogeneity in effects across the risk score distribution



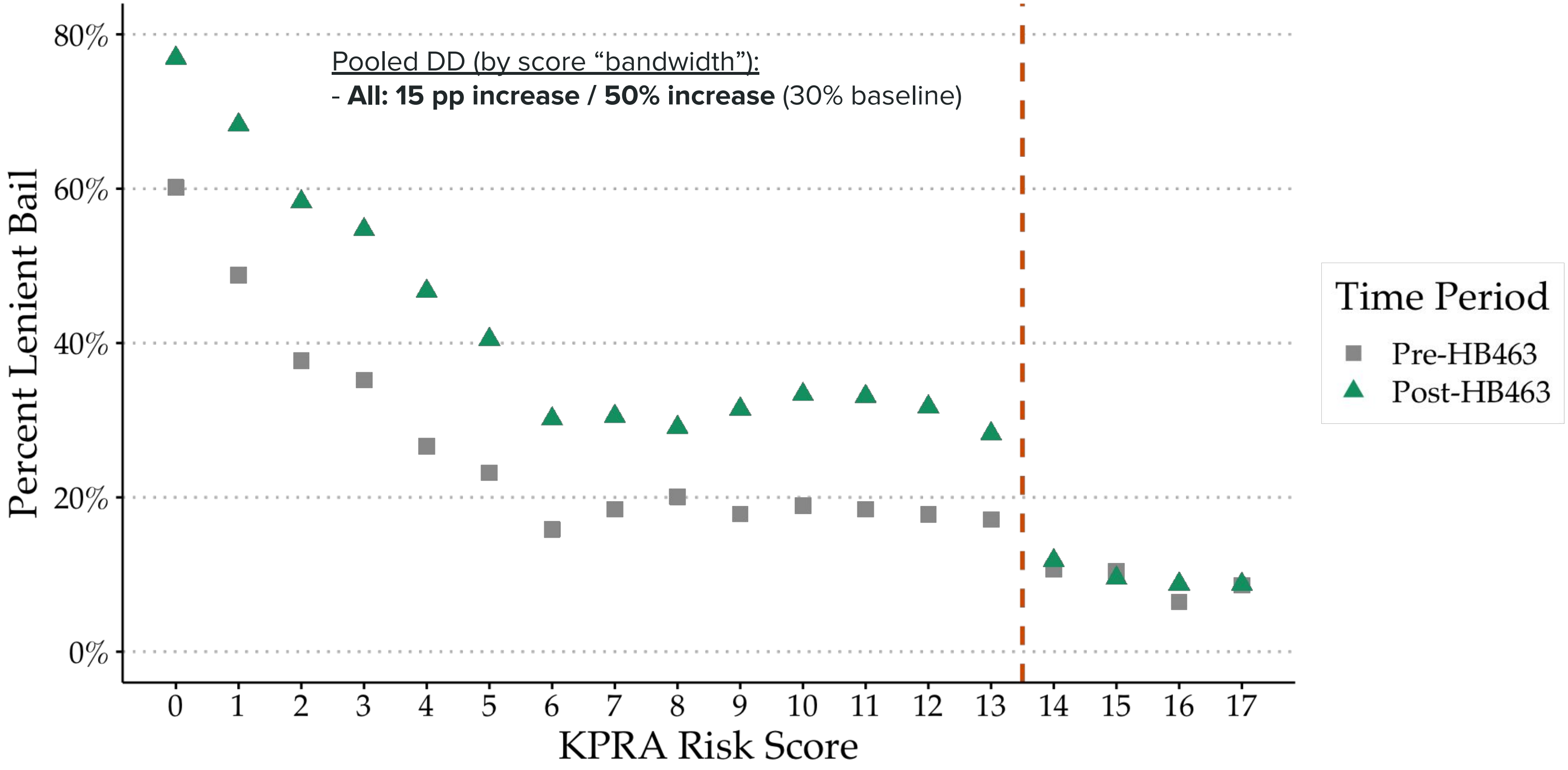
Heterogeneity in effects across the risk score distribution



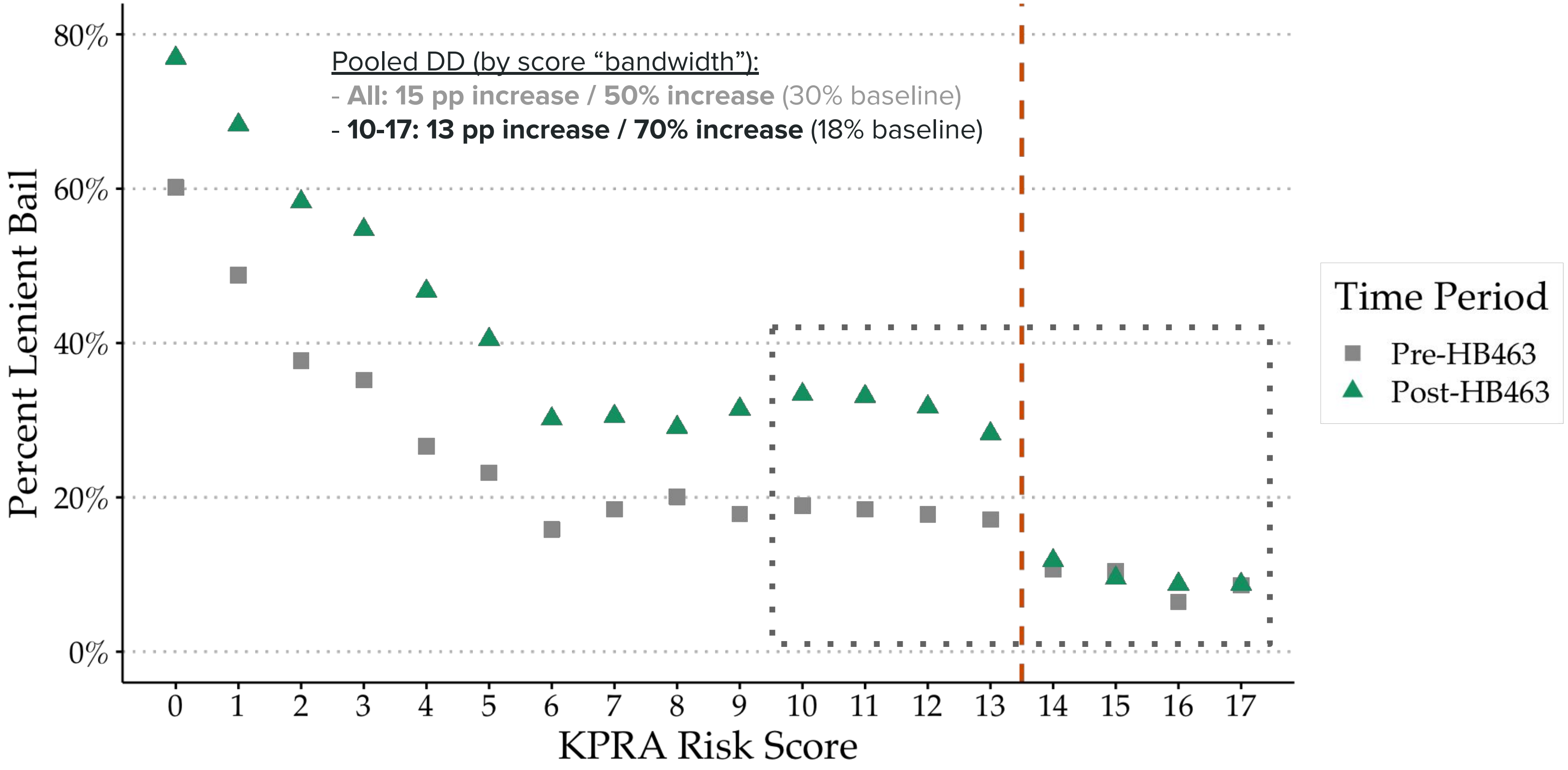
Heterogeneity in effects across the risk score distribution



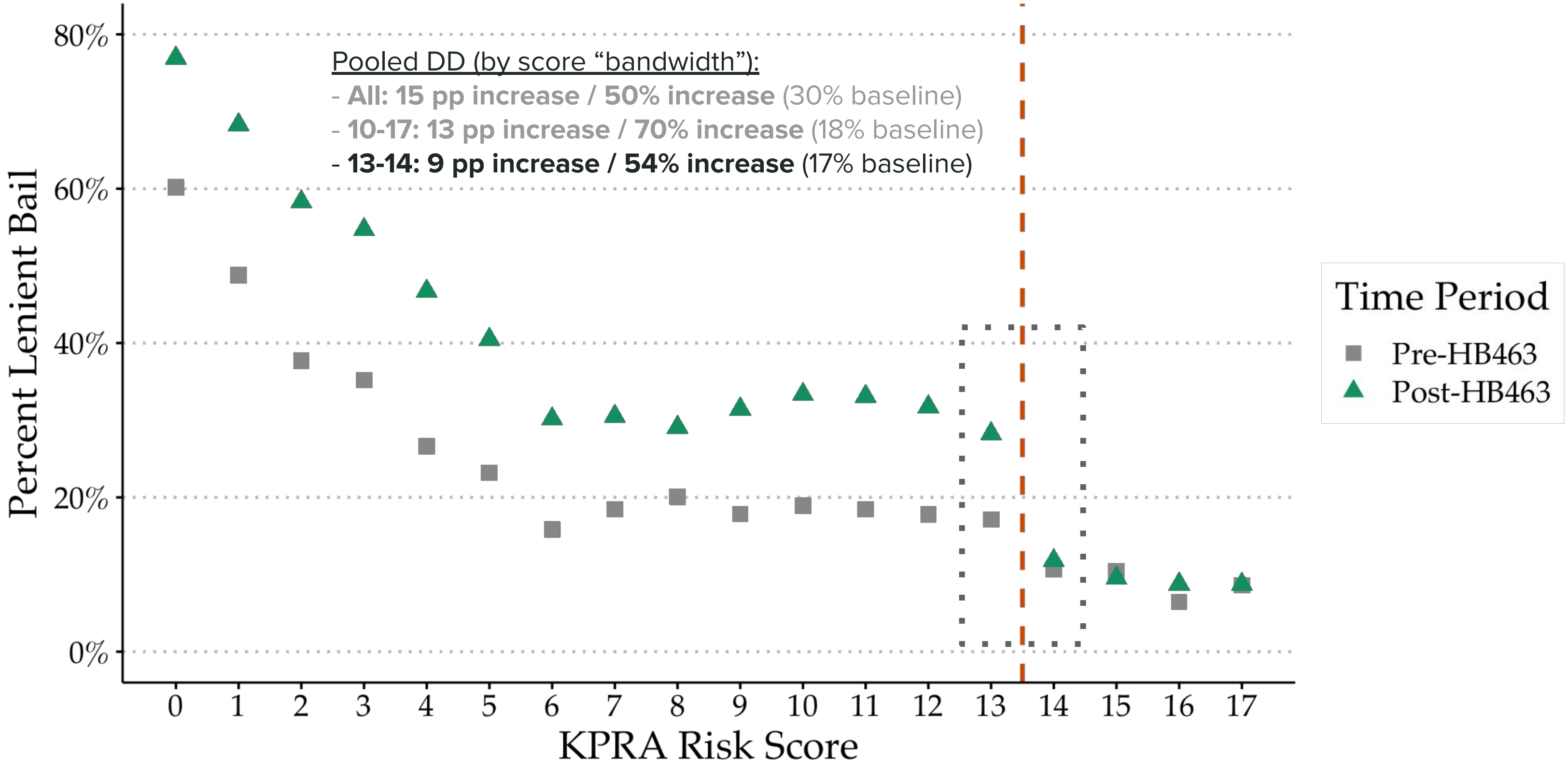
Heterogeneity in effects across the risk score distribution



Heterogeneity in effects across the risk score distribution



Heterogeneity in effects across the risk score distribution



What is the mechanism behind the effect?

Why might recommendations *change judge decisions*?

1. Administrative cost to deviate

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

Why might recommendations *change judge decisions*?

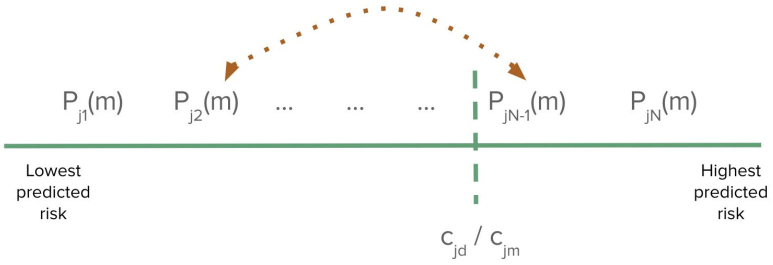
1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)



=> change in *allocation* but not *composition* of decisions, **<u>decrease in lenient bail for high risk</u>**

Why might recommendations *change judge decisions?*

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

   => change in *allocation* but not *composition* of decisions, **<u>decrease in lenient bail for high risk</u>**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

   => change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)

NEWS

Darrell Brooks Should Not Have Been Released on Low Bail, Milwaukee DA Admits

BY KATHERINE FUNG ON 11/22/21 AT 2:02 PM EST

"[Bail] in this case is not consistent with … the risk assessment of the defendant prior to the setting of bail."

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

   => change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)
   - Not as risky if recommended release (algorithm designer gives reputational cover)

In New York City court observations,

"judges routinely stated that they only ordered people to be released [...] because the law forced them to."
(Corvert 2022)

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

   => change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

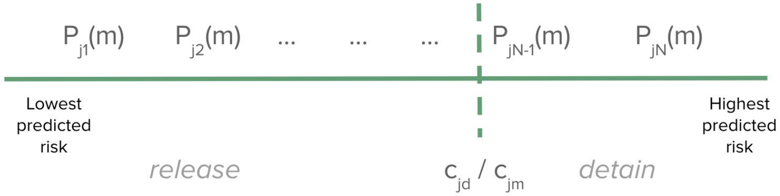
3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)
   - **Not as risky if recommended release (algorithm designer gives reputational cover)**

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

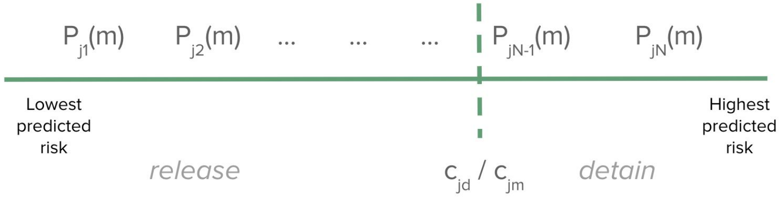
low/moderate risk



=> change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
    - Sticking neck out more if recommended detention (additional penalty)
    - **Not as risky if recommended release (algorithm designer gives reputational cover)**

high risk

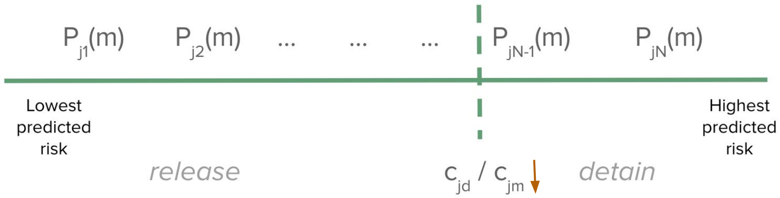


Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

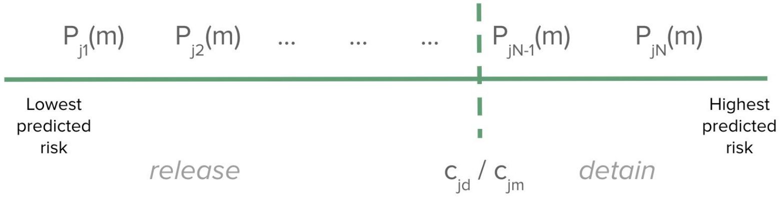
low/moderate risk



=> change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)
   - **Not as risky if recommended release (algorithm designer gives reputational cover)**

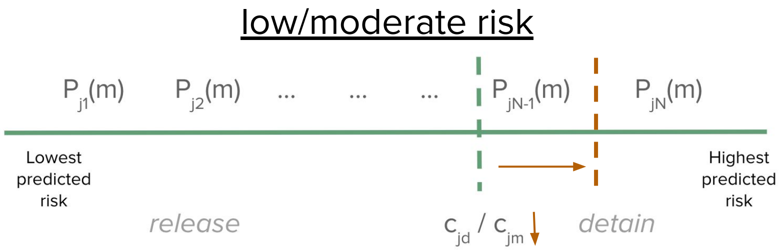
high risk



Why might recommendations *change judge decisions*?

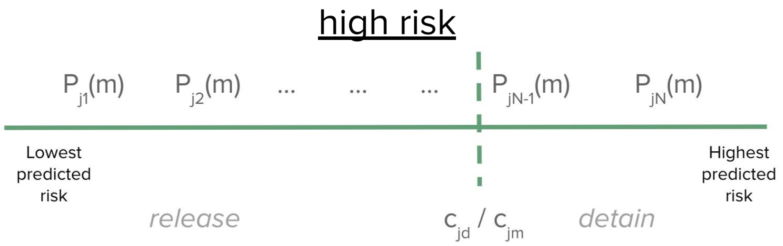
1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)



=> change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)
   - **Not as risky if recommended release (algorithm designer gives reputational cover)**



=> change in *composition* of decisions, **no effect for high risk**

Testing the dueling predictions

Want to estimate effect of recommendation introduction on high risk group *(not covered by recommendation)*

Testing the dueling predictions

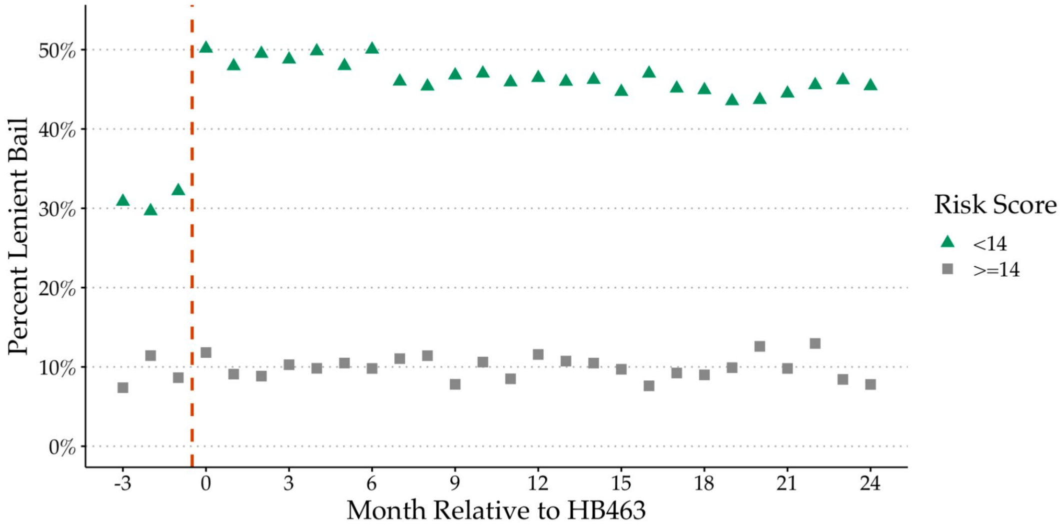
Want to estimate effect of recommendation introduction on high risk group *(not covered by recommendation)*

Methodology = **leverage variation over time (RD in time)**

Testing the dueling predictions

Want to estimate effect of recommendation introduction on high risk group *(not covered by recommendation)*

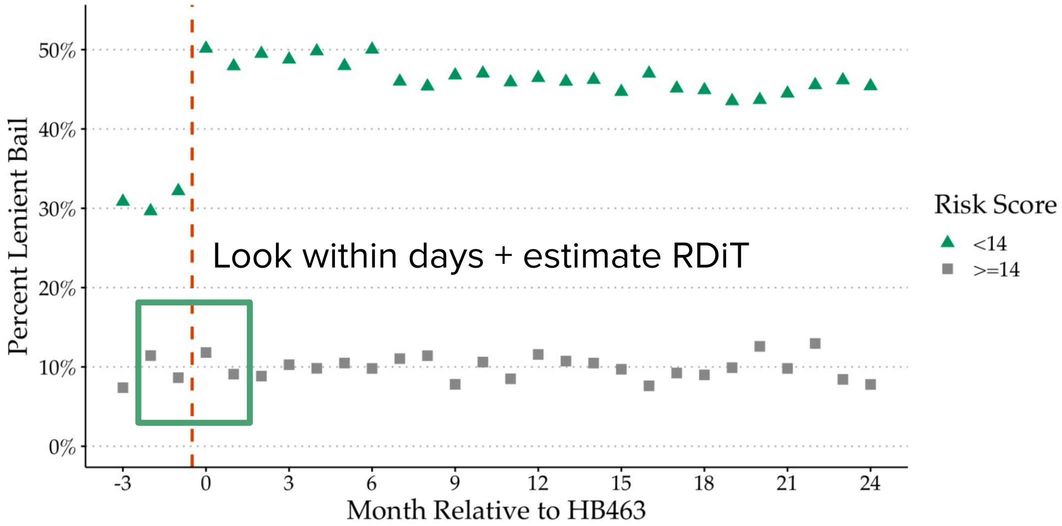
Methodology = **leverage variation over time (RD in time)**



Testing the dueling predictions

Want to estimate effect of recommendation introduction on high risk group *(not covered by recommendation)*

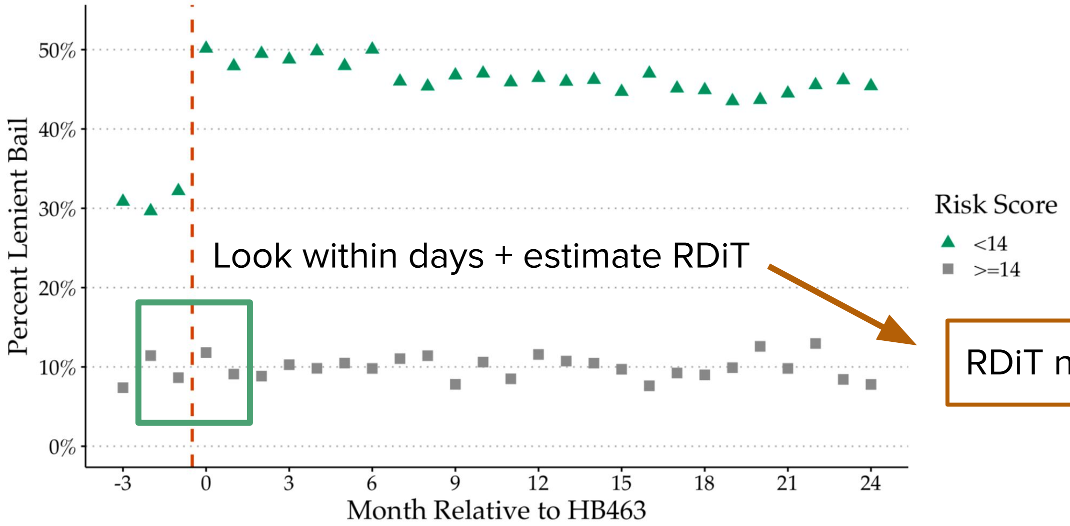
Methodology = **leverage variation over time (RD in time)**



Testing the dueling predictions

Want to estimate effect of recommendation introduction on high risk group *(not covered by recommendation)*

Methodology = **leverage variation over time (RD in time)**



RDiT not statistically significant

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

   => change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)
   - **Not as risky if recommended release (algorithm designer gives reputational cover)**

   => change in *composition* of decisions, **no effect for high risk**

Why might recommendations *change judge decisions*?

1. ~~Administrative cost to deviate~~

2. Recommendations increase algorithms use in setting perceived probabilities (upweight algorithm output)

   => change in *allocation* but not *composition* of decisions, **decrease in lenient bail for high risk**

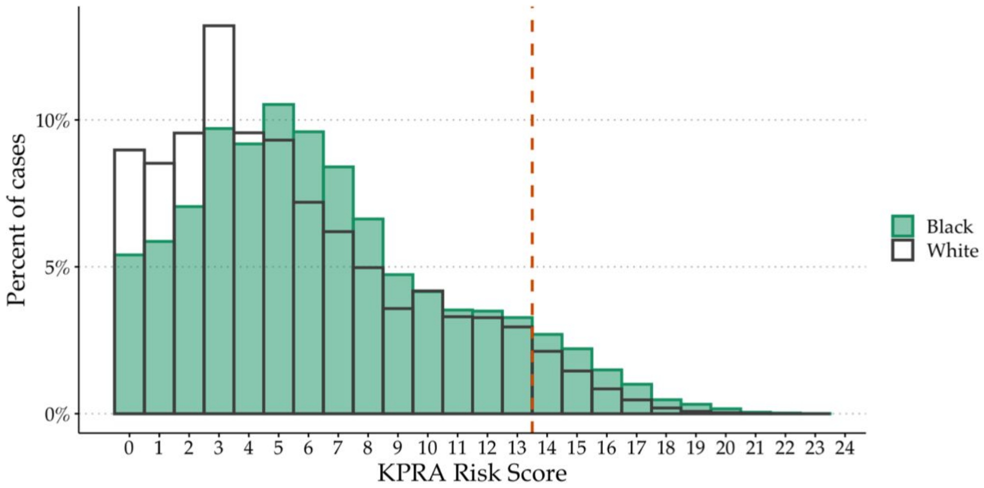
3. Recommendations change misconduct costs in the event of a bad outcome *(someone was released and commits misconduct; "type II errors")*
   - Sticking neck out more if recommended detention (additional penalty)
   - **Not as risky if recommended release (algorithm designer gives reputational cover)**

   => change in *composition* of decisions, **no effect for high risk**

Heterogeneous recommendation effects and implications for racial inequality

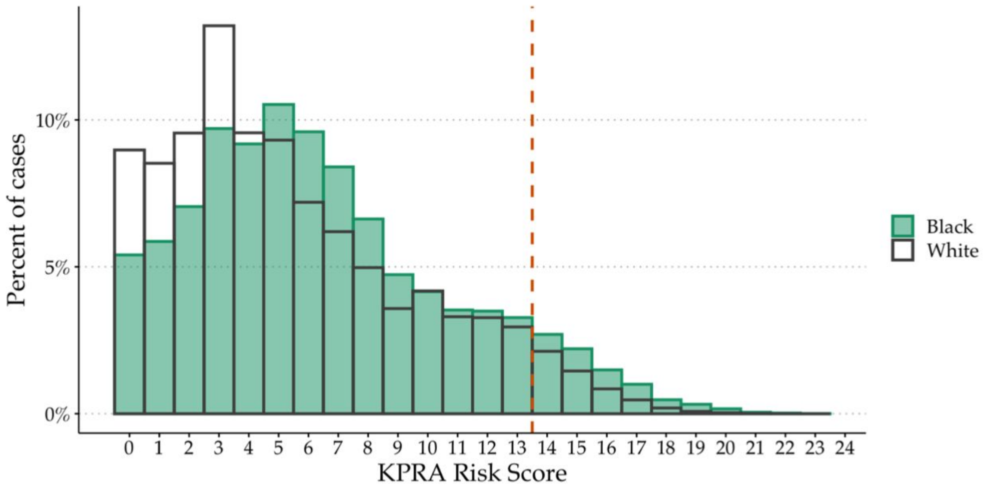
Racial disparities in risk scores, recommendations, and outcomes

Concern about the algorithm distribution:
usage might widen racial disparities



Racial disparities in risk scores, recommendations, and outcomes

Concern about the algorithm distribution:
usage might widen racial disparities

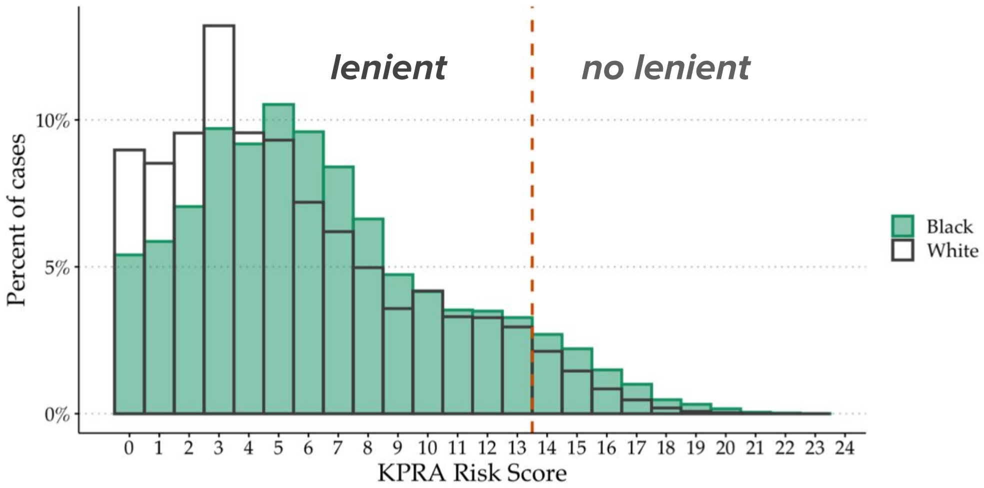


After the recommendations implemented,

Black people were **9.3 pp less likely** to get lenient bail (36.7% vs. 46%) than white people

Racial disparities in risk scores, recommendations, and outcomes

Concern about the algorithm distribution:
usage might widen racial disparities



After the recommendations implemented,

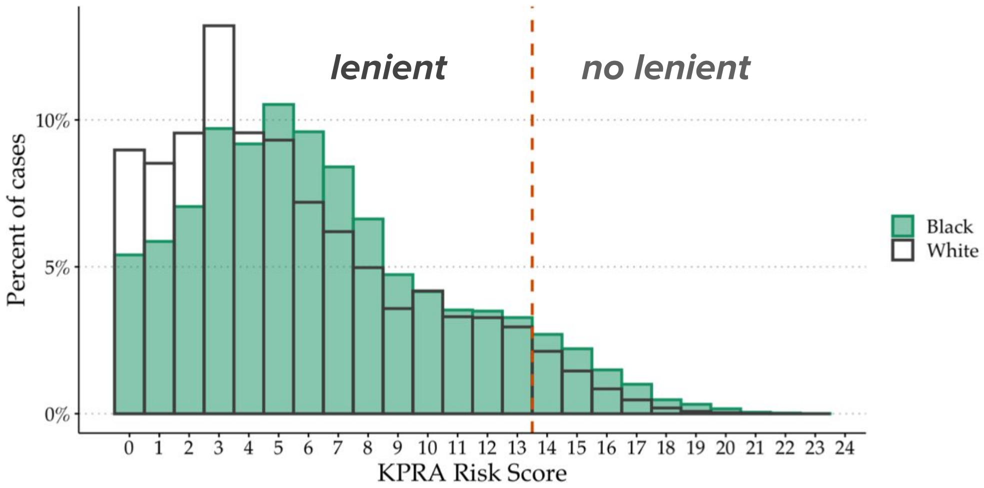
Black people were **9.3 pp less likely** to get lenient bail (36.7% vs. 46%) than white people

If bail automatically set by recommendations (low/mod => lenient; high => no lenient),

Black people would've been **3.3 pp less likely** to get lenient bail (91.5% vs. 94.8%) than white people

Racial disparities in risk scores, recommendations, and outcomes

Concern about the algorithm distribution:
usage might widen racial disparities



After the recommendations implemented,

Black people were **9.3 pp less likely** to get lenient bail (36.7% vs. 46%) than white people

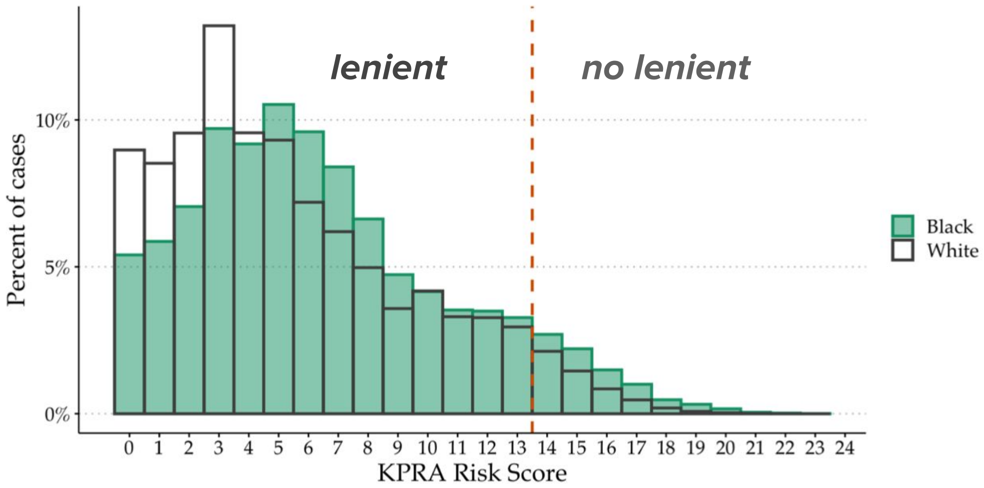
If bail automatically set by recommendations (low/mod => lenient; high => no lenient),

Black people would've been **3.3 pp less likely** to get lenient bail (91.5% vs. 94.8%) than white people

(automating to recommendation generates a 65% smaller racial gap than observed)

Racial disparities in risk scores, recommendations, and outcomes

Concern about the algorithm distribution:
usage might widen racial disparities



After the recommendations implemented,

Black people were **9.3 pp less likely** to get lenient bail (36.7% vs. 46%) than white people

If bail automatically set by recommendations (low/mod => lenient; high => no lenient),

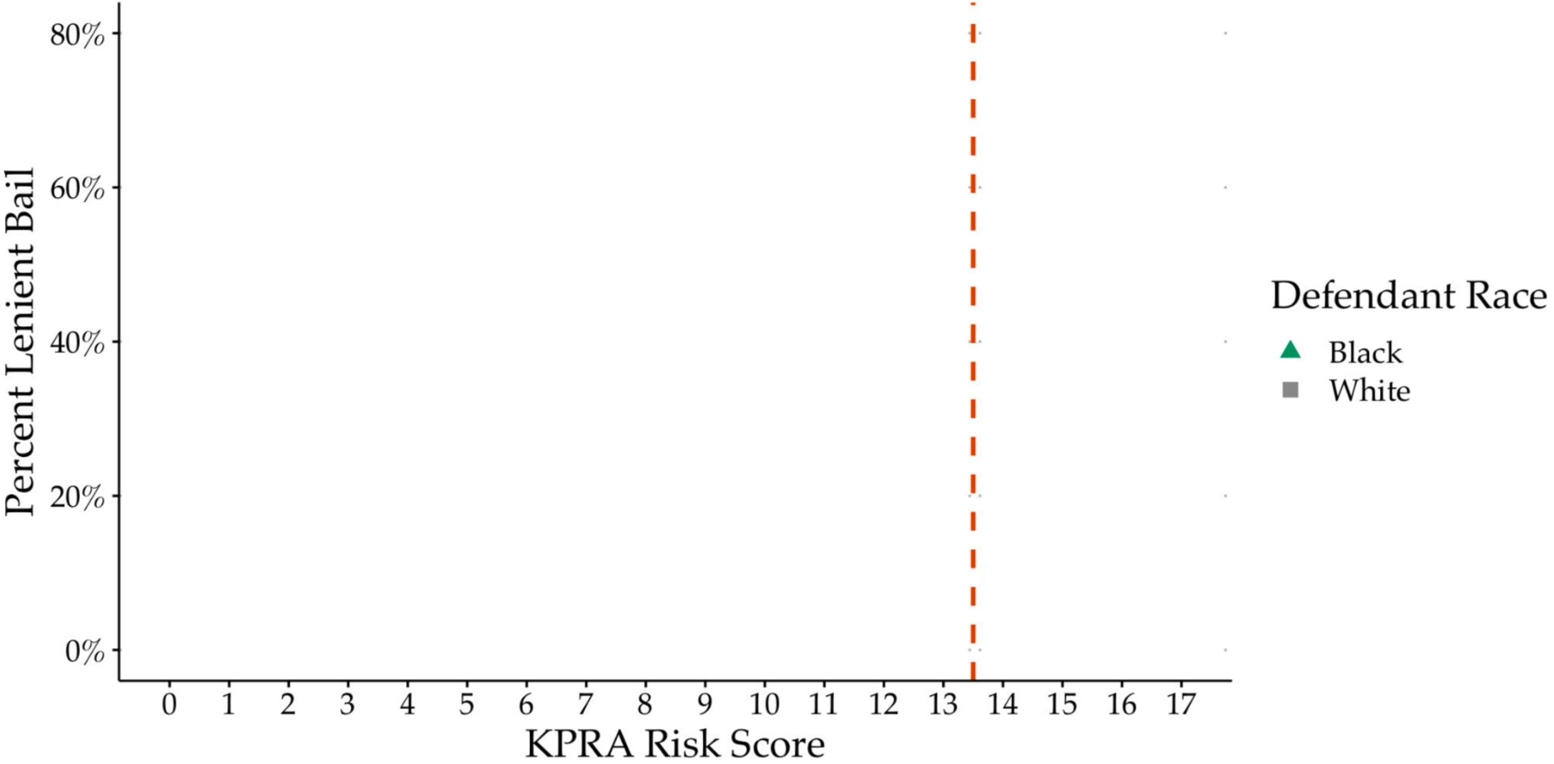
Black people would've been **3.3 pp less likely** to get lenient bail (91.5% vs. 94.8%) than white people

(automating to recommendation generates a 65% smaller racial gap than observed)

corollary: deviations from lenient recommendation vary by defendant race

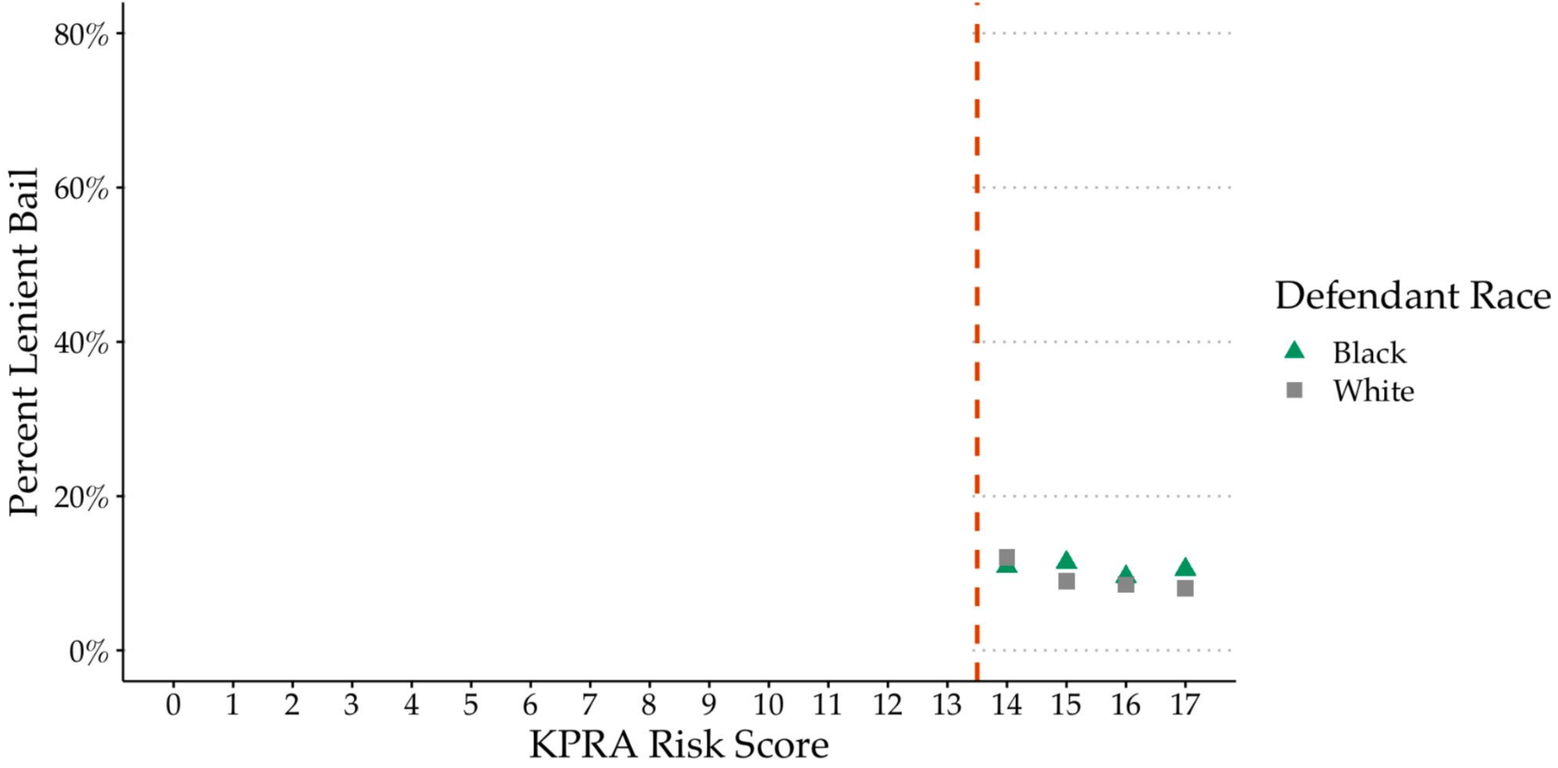
Are differences in deviations due to different risk score distributions?

Using data after algorithmic recommendations in place



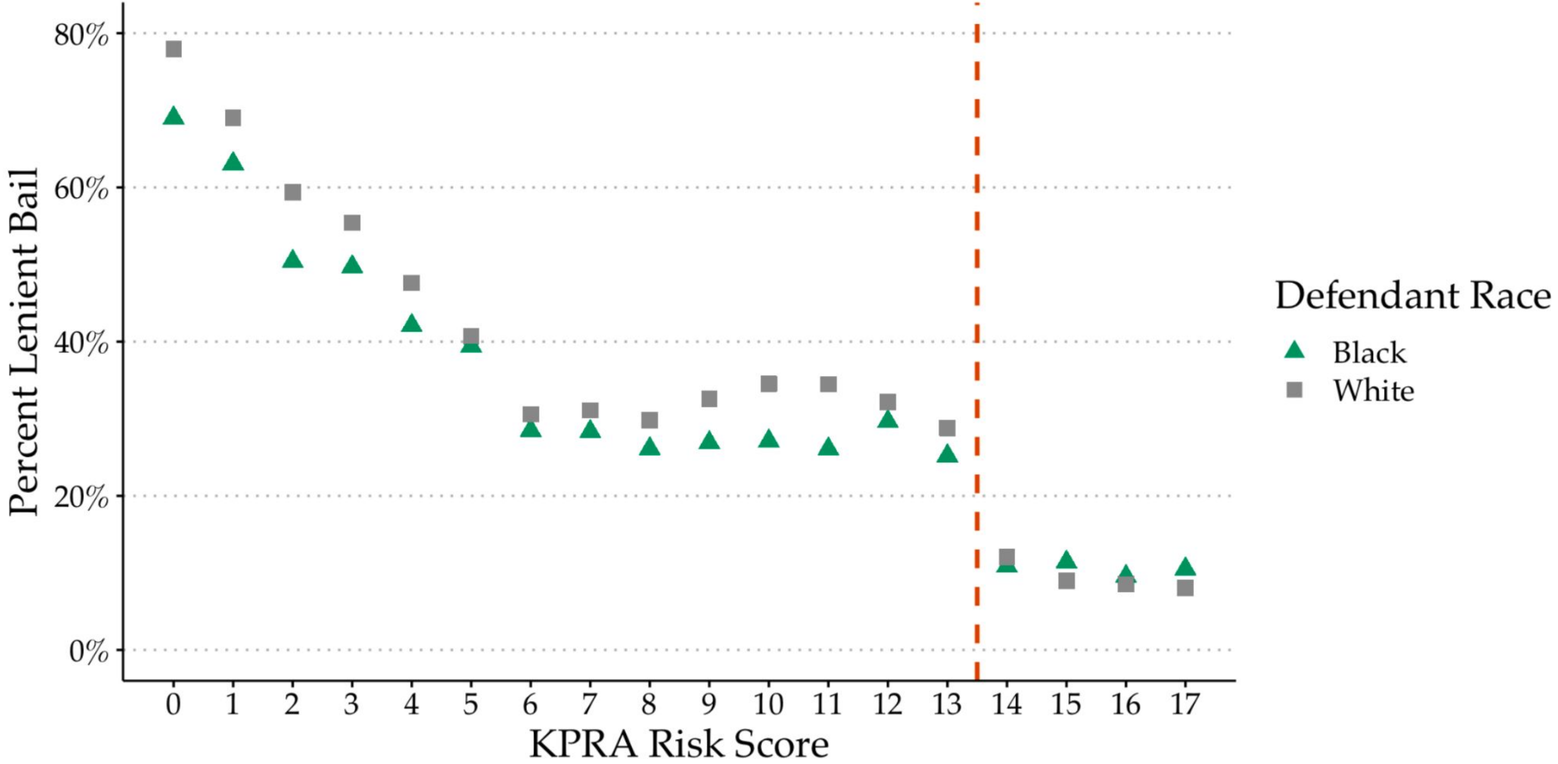
Are differences in deviations due to different risk score distributions?

Using data after algorithmic recommendations in place



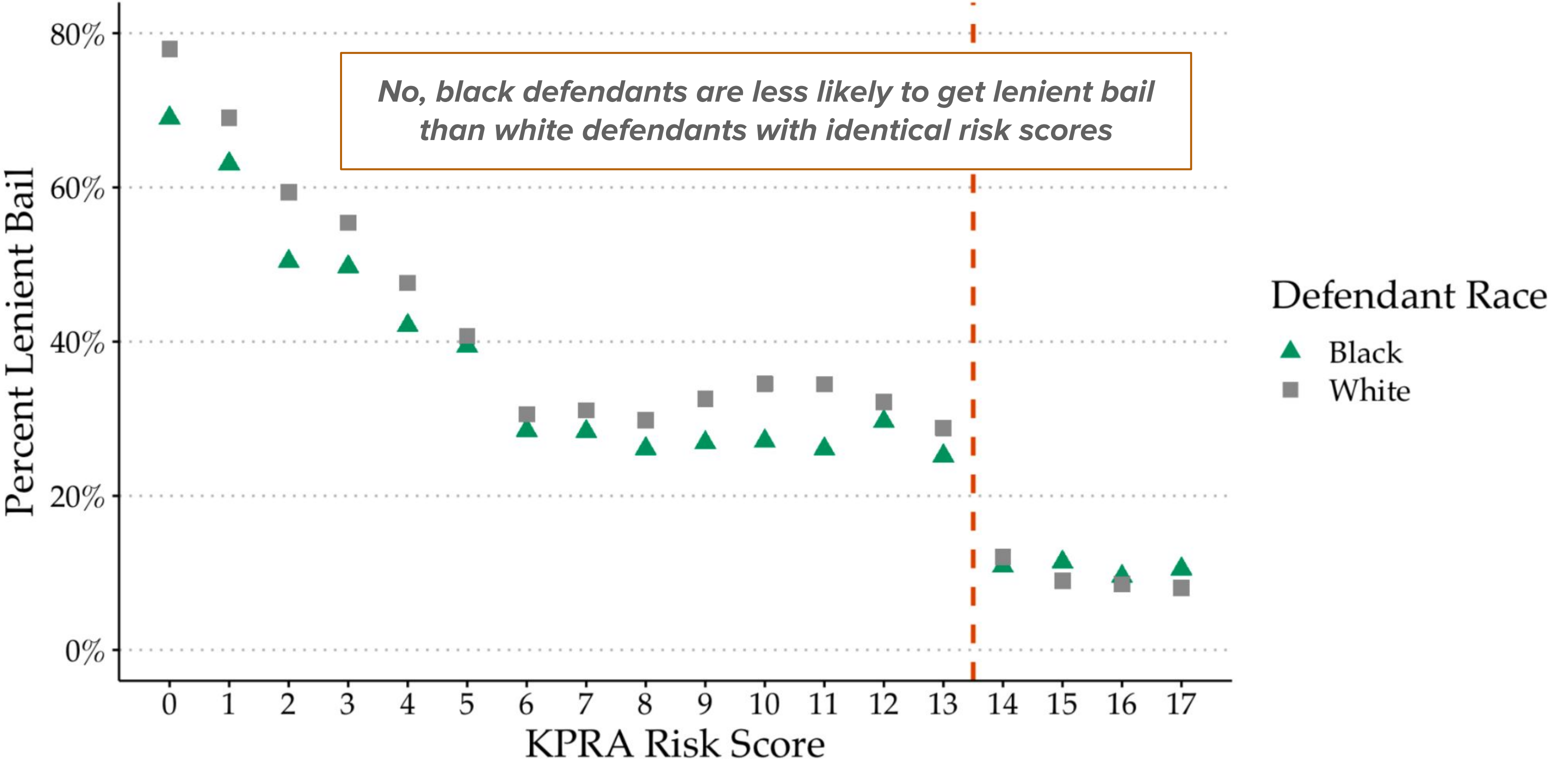
Are differences in deviations due to different risk score distributions?

Using data after algorithmic recommendations in place



Are differences in deviations due to different risk score distributions?

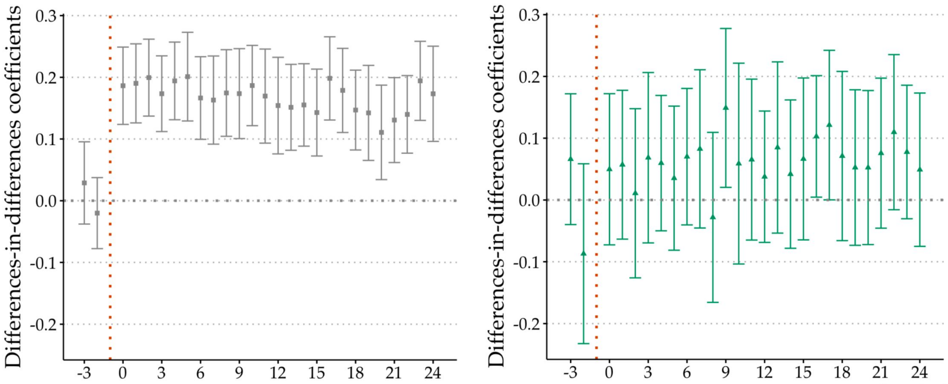
Using data after algorithmic recommendations in place



Effects of algorithmic recommendation differ by defendant race

	Dependent variable: I(lenient bail)	
	DD *(White Defendants)*	DD *(Black Defendants)*
	(1)	(2)

Effects of algorithmic recommendation differ by defendant race

	Dependent variable: I(lenient bail)	
	DD *(White Defendants)*	DD *(Black Defendants)*
	(1)	(2)
I(score<14) x Post	0.168*** (0.020)	0.088*** (0.034)
		
Avg Dep Var (Pre-HB463)	0.312	0.297

Effects of algorithmic recommendation differ by defendant race

	Dependent variable: I(lenient bail)		
	DD *(White Defendants)*	DD *(Black Defendants)*	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.168***	0.088***	
	(0.020)	(0.034)	
Avg Dep Var (Pre-HB463)	0.312	0.297	0.309

$$lenient_{itj} = \beta_1[I(score_i < 14) \times Post_t] + \beta_2[I(score_i < 14) \times Black_i] + \beta_3[Post_t \times Black_i] + \beta_4[I(score_i < 14) \times Post_t \times Black_i] + X_{itj} + \epsilon_{itj}$$

Effects of algorithmic recommendation differ by defendant race

	Dependent variable: I(lenient bail)		
	DD *(White Defendants)*	DD *(Black Defendants)*	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.168***	0.088***	0.167***
	(0.020)	(0.034)	(0.020)
I(score<14) x Black			0.032
			(0.029)
Post x Black			0.009
			(0.031)
I(score<14) x Post x Black			−0.083**
			(0.033)
Avg Dep Var (Pre-HB463)	0.312	0.297	0.309

What explains this heterogeneity?

	Dependent variable: I(lenient bail)
	DDD
	(1)
I(score<14) x Post	0.167^{***}
	(0.020)
I(score<14) x Black	0.032
	(0.030)
Post x Black	0.009
	(0.031)
I(score<14) x Post x Black	-0.083^{**}
	(0.034)
Extra FEs	NA

What explains this heterogeneity?

	Dependent variable: I(lenient bail)
	DDD
	(1)
I(score<14) x Post	0.167^{***}
	(0.020)
I(score<14) x Black	0.032
	(0.030)
Post x Black	0.009
	(0.031)
I(score<14) x Post x Black	-0.083^{**}
	(0.034)
Extra FEs	NA

Are these differences
within judges or ***between*** judges?

What explains this heterogeneity?

	Dependent variable: I(lenient bail)	
	DDD	DDD
	(1)	(2)
I(score<14) x Post	0.167***	
	(0.020)	
I(score<14) x Black	0.032	
	(0.030)	
Post x Black	0.009	
	(0.031)	
I(score<14) x Post x Black	−0.083**	
	(0.034)	
Extra FEs	NA	judge x under14 x post

Allow for time-score-varying judge FEs

What explains this heterogeneity?

	Dependent variable: I(lenient bail)	
	DDD	DDD
	(1)	(2)
I(score<14) x Post	0.167***	
	(0.020)	
I(score<14) x Black	0.032	−0.006
	(0.030)	(0.035)
Post x Black	0.009	0.006
	(0.031)	(0.029)
I(score<14) x Post x Black	−0.083**	$\boxed{-0.020}$
	(0.034)	$\boxed{(0.034)}$
Extra FEs	NA	judge x under14 x post

What explains this heterogeneity?

	Dependent variable: I(lenient bail)		
	DDD	DDD	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.167^{***}		
	(0.020)		
I(score<14) x Black	0.032	−0.006	−0.010
	(0.030)	(0.035)	(0.028)
Post x Black	0.009	0.006	0.008
	(0.031)	(0.029)	(0.024)
I(score<14) x Post x Black	-0.083^{**}	−0.020	−0.026
	(0.034)	(0.034)	(0.029)
Extra FEs	NA	judge x under14 x post	county x under14 x post

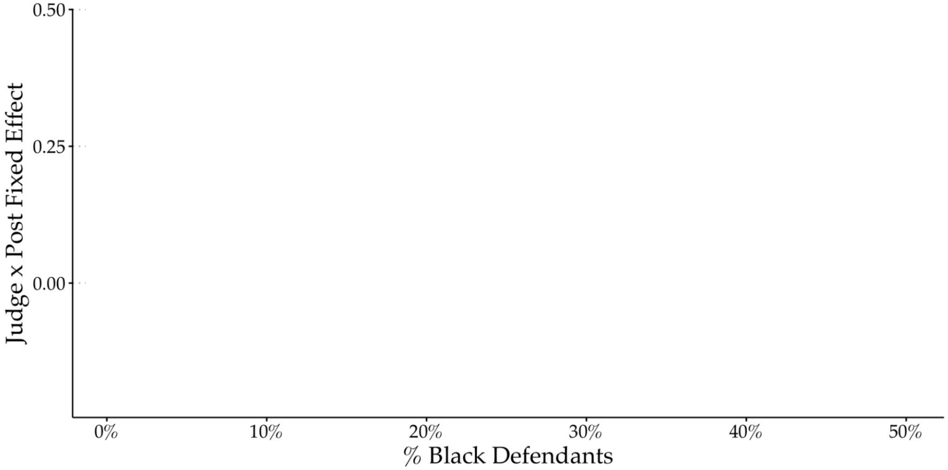
Judges with more Black defendants respond less to lenient recommendations

	Dependent variable: I(lenient bail)		
	DDD	DDD	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.167^{***}		
	(0.020)		
I(score<14) x Black	0.032	−0.006	−0.010
	(0.030)	(0.035)	(0.028)
Post x Black	0.009	0.006	0.008
	(0.031)	(0.029)	(0.024)
I(score<14) x Post x Black	-0.083^{**}	−0.020	−0.026
	(0.034)	(0.034)	(0.029)
Extra FEs	NA	judge x under14 x post	county x under14 x post

Judges with more Black defendants respond less to lenient recommendations

	Dependent variable: I(lenient bail)		
	DDD	DDD	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.167***		
	(0.020)		
I(score<14) x Black	0.032	−0.006	−0.010
	(0.030)	(0.035)	(0.028)
Post x Black	0.009	0.006	0.008
	(0.031)	(0.029)	(0.024)
I(score<14) x Post x Black	−0.083**	−0.020	−0.026
	(0.034)	(0.034)	(0.029)
Extra FEs	NA	judge x under14 x post	county x under14 x post

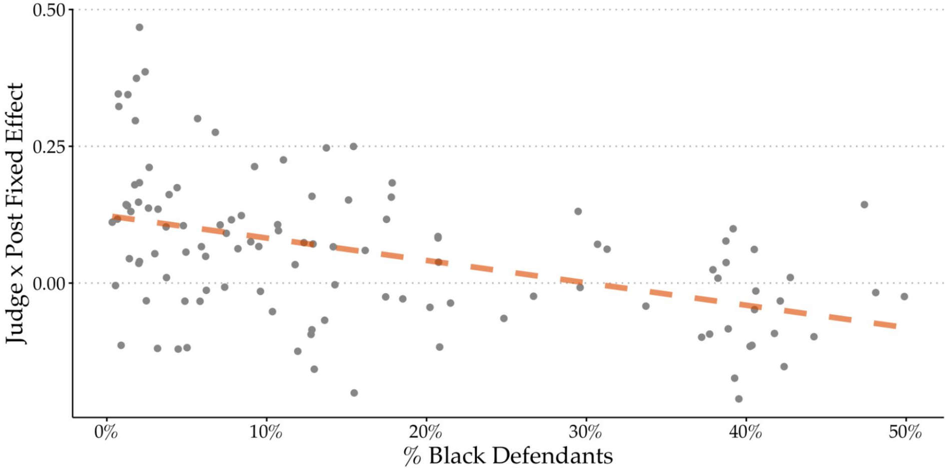
Subset to cases with score <14,
Estimate judge x post FEs
(judge FE for post period)



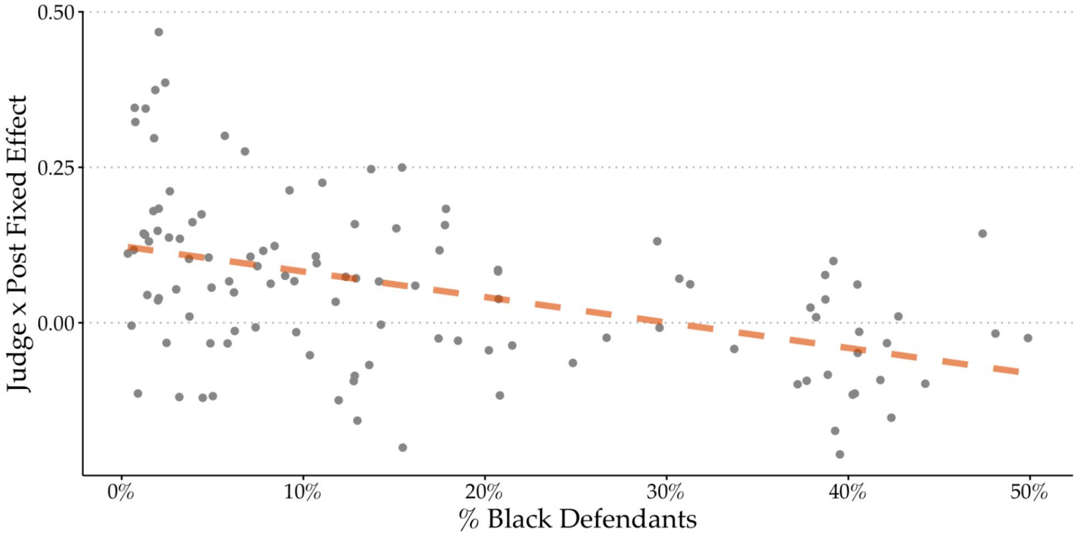
Judges with more Black defendants respond less to lenient recommendations

	Dependent variable: I(lenient bail)		
	DDD	DDD	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.167***		
	(0.020)		
I(score<14) x Black	0.032	−0.006	−0.010
	(0.030)	(0.035)	(0.028)
Post x Black	0.009	0.006	0.008
	(0.031)	(0.029)	(0.024)
I(score<14) x Post x Black	−0.083**	−0.020	−0.026
	(0.034)	(0.034)	(0.029)
Extra FEs	NA	judge x under14 x post	county x under14 x post

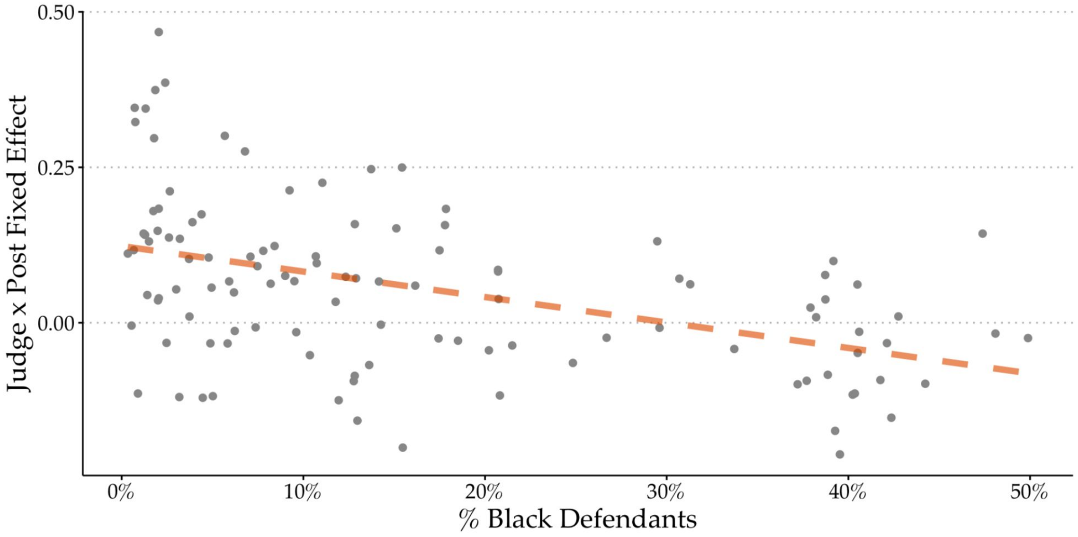
Subset to cases with score <14,
Estimate judge x post FEs
(judge FE for post period)



Why do they respond less?



Why do they respond less?

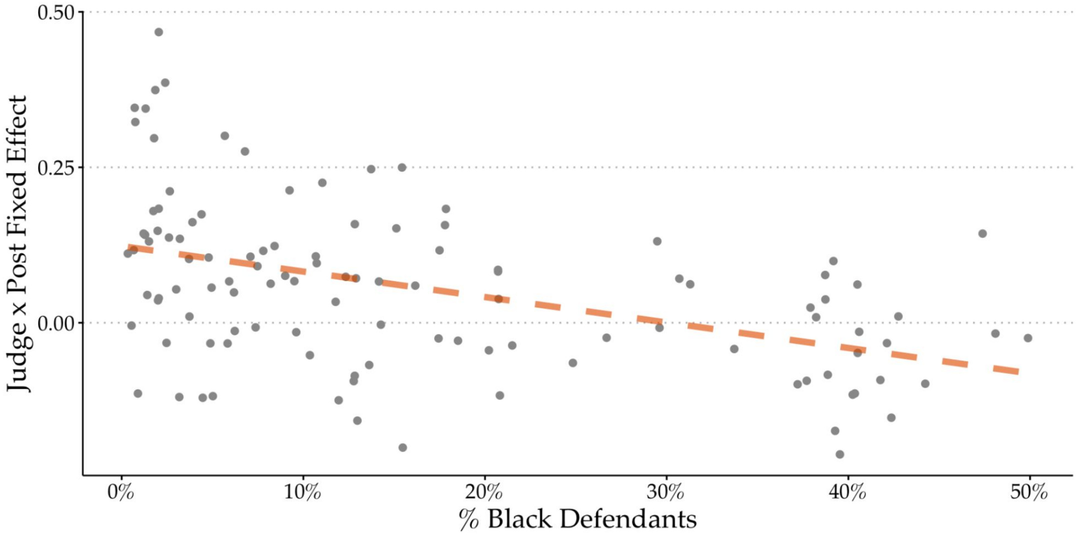


Could this relationship be explained by...

Judge characteristics?
- *Demographics (race, gender)*
- *Experience (years as judge)*
- *Election competitiveness*
- *Misconduct rates*

County characteristics?
- *Population*
- *Crime rates*

Why do they respond less?



Could this relationship be explained by...

Judge characteristics?

- *Demographics (race, gender)*
- *Experience (years as judge)*
- *Election competitiveness*
- *Misconduct rates*

County characteristics?

- *Population*
- *Crime rates*

Data sources:

- **Judge demographics/experience:** hand-collect data from public profiles online, interviews with staff
- **Election competitiveness:** hand-collect data on 2010 local election PDFs
- **Misconduct rates:** calculate FTA/re-arrest rates by judge in pre-period
- **Population and crime rates:** county-level data from 2010 UCR data

Why do they respond less?

	Dependent Variable = Judge x Post FE					
	(1)	(2)	(3)	(4)	(5)	(6)
Share Black Defendants	-0.374^{***}					
	(0.081)					

Judges who see 10 pp more Black defendants respond to the recommendation 3.7 pp less

(25% drop from the 15 pp baseline effect)

Why do they respond less?

	Dependent Variable = Judge x Post FE					
	(1)	(2)	(3)	(4)	(5)	(6)
Share Black Defendants	−0.374*** (0.081)					
		+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge + Election contest + Contest in district + log(voters)	+ Judge race + Judge gender + Years as judge + Election contest + Contest in district + log(voters) + FTA rate pre- + Rearrest rate pre-	+ Judge race + Judge gender + Years as judge + Election contest + Contest in district + log(voters) + FTA rate pre- + Rearrest rate pre- + County pop + Rural indicator	+ Judge race + Judge gender + Years as judge + Election contest + Contest in district + log(voters) + FTA rate pre- + Rearrest rate pre- + County pop + Rural indicator + Crime rate + Index crime rate + Prop crime rate + Violent crime rate

Judges who see 10 pp more Black defendants respond to the recommendation 3.7 pp less

(25% drop from the 15 pp baseline effect)

Other judge- and county-level covariates do not explain this

	Dependent Variable = Judge x Post FE					
	(1)	(2)	(3)	(4)	(5)	(6)
Share Black Defendants	−0.374***	−0.384***	−0.377**	−0.323**	−0.307*	−0.374**
	(0.081)	(0.085)	(0.144)	(0.149)	(0.169)	(0.178)
		+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge
			+ Election contest + Contest in district + log(voters)	+ Election contest + Contest in district + log(voters)	+ Election contest + Contest in district + log(voters)	+ Election contest + Contest in district + log(voters)
				+ FTA rate pre- + Rearrest rate pre-	+ FTA rate pre- + Rearrest rate pre-	+ FTA rate pre- + Rearrest rate pre-
					+ County pop + Rural indicator	+ County pop + Rural indicator
						+ Crime rate + Index crime rate + Prop crime rate + Violent crime rate

Judges who see 10 pp more Black defendants respond to the recommendation 3.7 pp less

(25% drop from the 15 pp baseline effect)

Other judge- and county-level covariates do not explain this

	Dependent Variable = Judge x Post FE					
	(1)	(2)	(3)	(4)	(5)	(6)
Share Black Defendants	−0.374***	−0.384***	−0.377**	−0.323**	−0.307*	−0.374**
	(0.081)	(0.085)	(0.144)	(0.149)	(0.169)	(0.178)
		+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge	+ Judge race + Judge gender + Years as judge
			+ Election contest + Contest in district + log(voters)	+ Election contest + Contest in district + log(voters)	+ Election contest + Contest in district + log(voters)	+ Election contest + Contest in district + log(voters)
				+ FTA rate pre- + Rearrest rate pre-	+ FTA rate pre- + Rearrest rate pre-	+ FTA rate pre- + Rearrest rate pre-
					+ County pop + Rural indicator	+ County pop + Rural indicator
						+ Crime rate + Index crime rate + Prop crime rate + Violent crime rate

Judges who see 10 pp more Black defendants respond to the recommendation 3.7 pp less

(25% drop from the 15 pp baseline effect)

Suggestive evidence:
Reputational cover recommendations provide depends on county demographics

similar to Feigenberg + Miller (2021) finding of higher CJS severity in more racially heterogeneous places

Wrap-up

Wrap-up

- Algorithmic recommendations are common practice in decision-making settings
- Designing algorithmic recommendations =/= solving a prediction problem; it's making normative decisions (i.e., about trade-offs between detention and misconduct)

> *"predictive analytics are best understood as political decision-making machines [... which] encourages us to perceive deeply political decisions as natural and inevitable"* - Eubanks (2019)

Wrap-up

- Algorithmic recommendations are common practice in decision-making settings
- Designing algorithmic recommendations =/= solving a prediction problem; it's making normative decisions (i.e., about trade-offs between detention and misconduct)

  "predictive analytics are best understood as political decision-making machines [... which] encourages us to perceive deeply political decisions as natural and inevitable" - Eubanks (2019)

- The effects of algorithmic recommendations are often hidden when we study effects of algorithms
  - But ***algorithmic recommendations have independent effects***

Wrap-up

Results consistent with model of recommendations changing private costs of type II errors (i.e., pretrial misconduct upon release)

- Recommendations can shift more than allocation of decisions, can shift composition of decisions
- Broad idea: ***Recommendations can better align decision-maker objective functions with algorithm designer/social planner objective functions*** (building on McLaughlin et al 2022)

Wrap-up

Results consistent with model of recommendations changing private costs of type II errors (i.e., pretrial misconduct upon release)

- Recommendations can shift more than allocation of decisions, can shift composition of decisions
- Broad idea: ***Recommendations can better align decision-maker objective functions with algorithm designer/social planner objective functions*** (building on McLaughlin et al 2022)

Decision-makers may deviate from recommendations in ways that complicate effects on racial inequality

- ***Discretion matters even though algorithmic systems aim to limit its importance***

`\end{talk}`

Thanks for coming!

You can email me here:

alex@albrightalex.com

More info:

albrightalex.com