The Hidden Effects of Algorithmic Recommendations

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

Predictive algorithms and high-stakes decisions

- Algorithms predicting:

  default / self-harm / re-arrest

  ...are used in:

  loan / medical / criminal justice decisions

Thomas Fuchs

Predictive algorithms and high-stakes decisions

- Algorithms predicting:

  default / self-harm / re-arrest

  ...are used in:

  loan / medical / criminal justice decisions

- But, humans – not algorithms – usually make final decisions

  (loan officers / therapists / judges)

Thomas Fuchs

Predictive algorithms and high-stakes decisions

- Algorithms predicting:

  default / self-harm / re-arrest

  ...are used in:

  loan / medical / criminal justice decisions

- But, humans – not algorithms – usually make final decisions

  (loan officers / therapists / judges)

Thomas Fuchs

=> understanding how algorithms change these systems requires understanding *how algorithms change human decisions*

How do predictive algorithms change human decisions?

Conventional wisdom:

algorithms provide decision-makers
with data-driven predictions

"algorithmic predictions"

- *loan officer's algorithm prediction: "high risk"*
- *therapist's algorithm prediction: "high risk"*
- *judge's algorithm prediction: "high risk"*

How do predictive algorithms change human decisions?

Conventional wisdom:

algorithms provide decision-makers with data-driven predictions

"algorithmic predictions"

- *loan officer's algorithm prediction: "high risk"*
- *therapist's algorithm prediction: "high risk"*
- *judge's algorithm prediction: "high risk"*

But:

algorithms often provide more than predictions – they provide recommendations

"algorithmic recommendations"

- *loan officer's algorithm recommends rejection*
- *therapist's algorithm recommends hospitalization*
- *judge's algorithm recommends jail*

How do predictive algorithms change human decisions?

Conventional wisdom:

algorithms provide decision-makers with data-driven predictions

"algorithmic predictions"

- *loan officer's algorithm prediction: "high risk"*
- *therapist's algorithm prediction: "high risk"*
- *judge's algorithm prediction: "high risk"*

But:

algorithms often provide more than predictions – they provide recommendations

"algorithmic recommendations"

- *loan officer's algorithm recommends rejection*
- *therapist's algorithm recommends hospitalization*
- *judge's algorithm recommends jail*

Studying the effect of "algorithms" on decisions conflates these two components

This paper:

Conventional wisdom:

algorithms provide decision-makers with data-driven predictions

"algorithmic predictions"

But:

algorithms often provide more than predictions – they provide recommendations

"algorithmic recommendations"

This paper: demonstrates independent effects of recommendations

Conventional wisdom:

algorithms provide decision-makers with data-driven predictions

"algorithmic predictions"

But:

algorithms often provide more than predictions – they provide recommendations

"algorithmic recommendations"

This paper: demonstrates independent effects of recommendations

Conventional wisdom:

algorithms provide decision-makers with data-driven predictions

"algorithmic predictions"

But:

algorithms often provide more than predictions – they provide recommendations

"algorithmic recommendations"

Empirical challenges: opaque institutional details around algorithm construction and implementation + simultaneous introduction of the predictions and recommendations

This paper: demonstrates independent effects of recommendations

Conventional wisdom:

algorithms provide decision-makers with data-driven predictions

"algorithmic predictions"

But:

algorithms often provide more than predictions – they provide recommendations

"algorithmic recommendations"

Empirical challenges: opaque institutional details around algorithm construction and implementation + simultaneous introduction of the predictions and recommendations

Leverage a natural experiment *(judges making bail decisions in CJS)* where

1. algorithmic predictions given to decision-makers stayed the same
2. BUT use of algorithmic recommendations changed

Preview of results

1. **Recommendations change decisions**
    - Recommendations have independent effects from algorithm predictions themselves
    - **Lenient recommendations increase lenient bail by 50%**

Preview of results

1. **Recommendations change decisions**
    - Recommendations have independent effects from algorithm predictions themselves
    - **<u>Lenient recommendations increase lenient bail by 50%</u>**

2. Why? **Recommendations can change private costs of errors**
    - Making mistakes is less costly when decision consistent with recommendation *(lenient recommendations provide "cover" for judges)*
    - **<u>Algorithms can impact decision-maker incentives, rather than just predictions</u>**

Preview of results

1. **Recommendations change decisions**
    - Recommendations have independent effects from algorithm predictions themselves
    - **Lenient recommendations increase lenient bail by 50%**

2. Why? **Recommendations can change private costs of errors**
    - Making mistakes is less costly when decision consistent with recommendation *(lenient recommendations provide "cover" for judges)*
    - **Algorithms can impact decision-maker incentives, rather than just predictions**

3. Heterogeneity: **Recommendations may not impact all groups equally**
    - **Judges deviate from lenient recommendation more for Black defendants** than for white defendants with the same algorithmic risk

Roadmap

1. Background on algorithms and bail decisions
2. Empirical setting: Kentucky bail decisions
3. Toy model and theoretical predictions
4. Causal effects of algorithmic recommendations
5. Addressing identification concerns
6. Heterogeneous effects by defendant race

Background on algorithms and bail decisions

Algorithms in decision-making

No algorithmic information given to humans

Algorithm-based rules dictate outcomes

Algorithms in decision-making

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

No algorithmic information given to humans

These papers:
algorithms can outperform human decisions

...but what about when humans are involved?

Algorithm-based rules dictate outcomes

Algorithms in decision-making

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

No algorithmic information given to humans

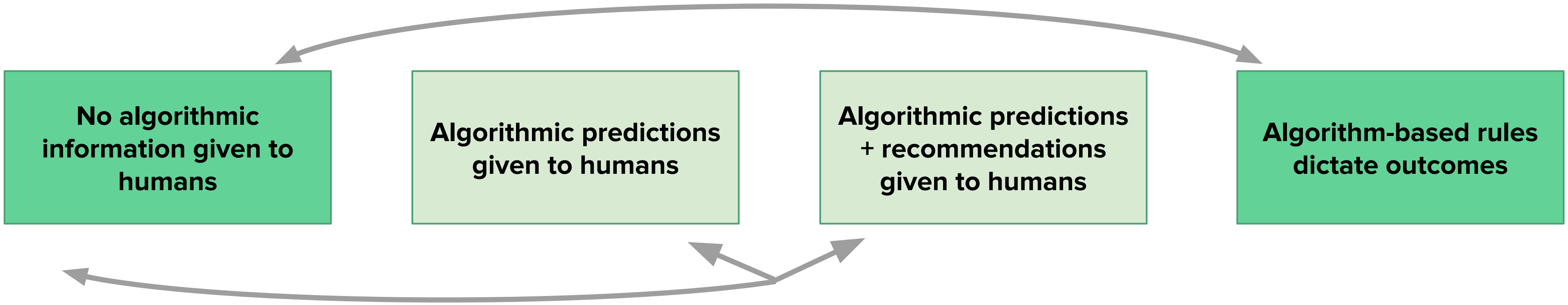
Algorithmic predictions given to humans

Algorithmic predictions + recommendations given to humans

Algorithm-based rules dictate outcomes

Algorithms in decision-making

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

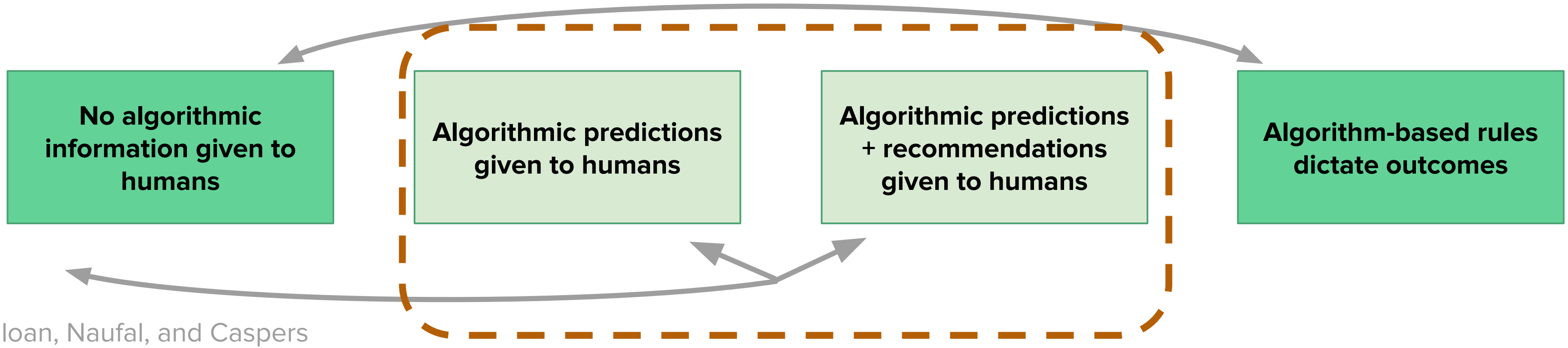


Sloan, Naufal, and Caspers (Forthcoming), Stevenson (2018), Doleac and Stevenson (Forthcoming), Garrett and Monahan (2018), DeMichele et al. (2018), Cowgill and Tucker (2019)

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

Algorithms in decision-making

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



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

Today: highlight the importance of the distinction between these intermediate options

Bail system in the US

- Incarceration before any conviction common in the US
- 65% of people in US jails in pretrial detention (~500,000 people)

Arrest => **Bail conditions set** => Conviction determination

- Bail's purpose: minimum conditions to ensure court appearance + public safety
- Most salient example of bail: money bail
  - Requires financial deposit for jail release
  - Goal: incentivize returning to court/no rearrest (i.e., good conduct)

Bail decisions and algorithms

Judge objective: minimize bail conditions, minimize pretrial misconduct

Lever: setting money bail (requires defendant to post money for release from jail)

Algorithms:

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)

Bail decisions and algorithms

Judge objective: minimize bail conditions, minimize pretrial misconduct

Lever: setting money bail (requires defendant to post money for release from jail)

Algorithms:

Common goal: "data-driven way to advance pretrial release"

STATE	TYPE/SCOPE OF USE
Alabama	VPRAI / Jefferson County
Alaska	
Arizona	
Arkansas	
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
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
	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)

Bail decisions and algorithms

Judge objective: minimize bail conditions, minimize pretrial misconduct

Lever: setting money bail (requires defendant to post money for release from jail)

Algorithms:

STATE	TYPE/SCOPE OF USE
Alabama	VPRAI / Jefferson County
Alaska	
Arizona	
Arkansas	
California (Sample risk assessment documents from San Francisco, and Napa County)	PSA / 3 counties | PRRS II / 2 Counties
Colorado (sample risk assessment documents)	CPAT / Statev DV / Statewid
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

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
Louisiana	PSA / New Orleans
Maine	ODARA (sex offenders) / Statewide | 2019 Task Force
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
	another statewide one
Ohio	PSA / 3 Counties | ORAS-PAT / Statewide
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
	PSA / 3 Counties
	LS/CMI
	PSA / 4 Counties | COMPAS / Statewide
Wyoming	COMPAS for Prisoners / Statewide
Federal	PTRA

Common goal: "data-driven way to advance pretrial release"

Predict misconduct based on observable data

Source: Epic (2020)

How algorithms matter depends on the setting

Example 1: allocating housing

- People are scored (e.g., according to need or housing readiness)
- Generates a ranked list
- Available housing allocated down the list

How algorithms matter depends on the setting

Example 1: allocating housing

- People are scored (e.g., according to need or housing readiness)
- Generates a ranked list
- Available housing allocated down the list

Supply of housing **fixed**

=> algorithms only change allocation

How algorithms matter depends on the setting

Example 1: allocating housing

- People are scored (e.g., according to need or housing readiness)
- Generates a ranked list
- Available housing allocated down the list

Supply of housing **fixed**

=> algorithms only change allocation

Example 2: setting bail after arrest

- People are scored (e.g., according to risk of failing to appear in court)
- Scores, recommendations given to judges
- Judges decide how to set bail

How algorithms matter depends on the setting

Example 1: allocating housing

- People are scored (e.g., according to need or housing readiness)
- Generates a ranked list
- Available housing allocated down the list

Supply of housing **fixed**

=> algorithms only change allocation

Example 2: setting bail after arrest

- People are scored (e.g., according to risk of failing to appear in court)
- Scores, recommendations given to judges
- Judges decide how to set bail

Supply of bail is **not fixed**

=> algorithms can change allocation AND composition

Empirical setting: Kentucky bail decisions

Pre-Period: judges set bail without recommendations

Judges make bail decisions via brief phone calls with pretrial officers (admin court employees)

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

The Kentucky Algorithm

After person booked, pretrial services officer calculates a risk score

- 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

June 2011: House Bill introduces recommendation for some cases

Judges make bail decisions via brief phone calls with pretrial officers (admin court employees)

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

After 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

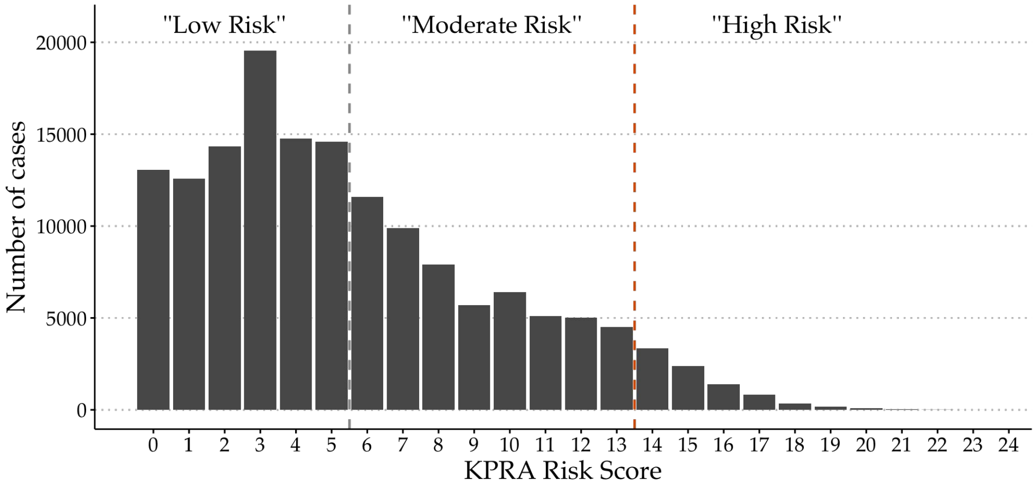
Recommendation
(lenient bail for low/moderate cases)

Variation in recommendation over time and scores

No recommendation

June 2011

Recommendation
(lenient bail for low/moderate cases)

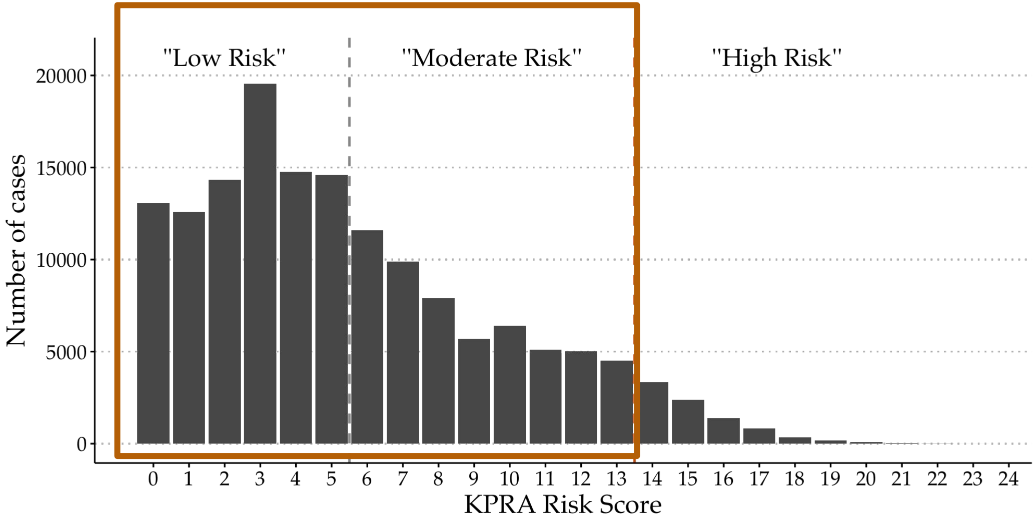


Variation in recommendation over time and scores

No recommendation

June 2011

Recommendation
(lenient bail for low/moderate cases)



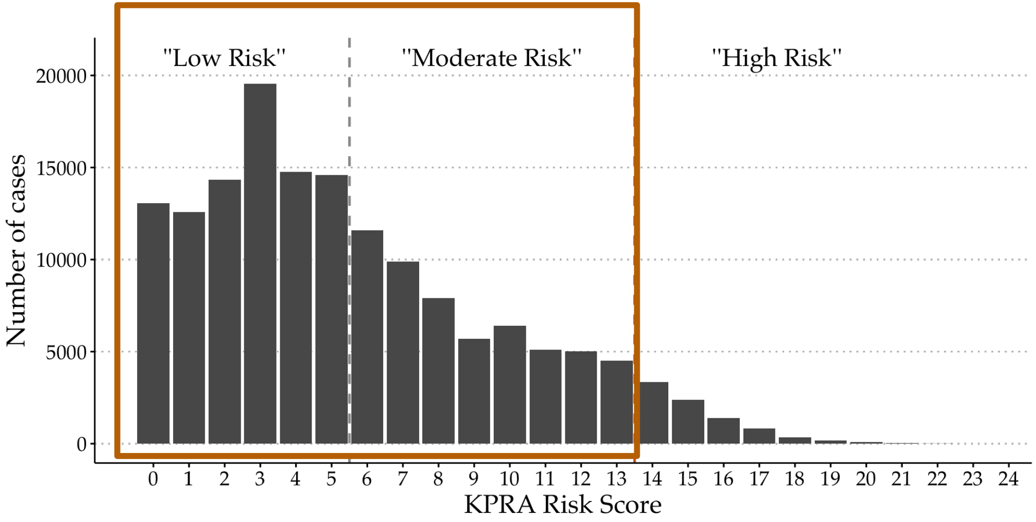
Lenient bail recommended for **<u>90%</u>** <u>of cases</u>

Variation in recommendation over time and scores

No recommendation

June 2011

Recommendation
(lenient bail for low/moderate cases)



Lenient bail recommended for **90% of cases**

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

Toy model and theoretical predictions

Status quo bail decisions

Legal bail objective: set lowest possible bail to ensure court appearance, public safety
=> want to set bail low but also want low misconduct

Judge has choice between lenient (no money bail; $\underline{b=l}$) and harsh bail (money bail; $\underline{b=h}$)

Status quo bail decisions

Legal bail objective: set lowest possible bail to ensure court appearance, public safety
=> want to set bail low but also want low misconduct

Judge has choice between lenient (no money bail; <u>b=l</u>) and harsh bail (money bail; <u>b=h</u>)

Judge costs:

$$\underbrace{(1-P(d|b=h))}_{\text{probability of release}} \times \underbrace{P(m|b=h)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=h)}_{\text{cost of misconduct}} \qquad \underbrace{P(d|b=h)}_{\text{probability of detention}} \times \underbrace{c(d|b=h)}_{\text{cost of detention}}$$

Status quo bail decisions

Legal bail objective: set lowest possible bail to ensure court appearance, public safety
=> want to set bail low but also want low misconduct

Judge has choice between lenient (no money bail; b=l) and harsh bail (money bail; b=h)

Judge costs:

$$\underbrace{(1-P(d|b=h))}_{\text{probability of release}} \times \underbrace{P(m|b=h)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=h)}_{\text{cost of misconduct}} + \underbrace{P(d|b=h)}_{\text{probability of detention}} \times \underbrace{c(d|b=h)}_{\text{cost of detention}}$$

Status quo bail decisions

Legal bail objective: set lowest possible bail to ensure court appearance, public safety
=> want to set bail low but also want low misconduct

Judge has choice between lenient (no money bail; b=l) and harsh bail (money bail; b=h)

Judge costs:

$$\underbrace{P(m|b=l)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=l)}_{\text{cost of misconduct}} \qquad \underbrace{(1-P(d|b=h))}_{\text{probability of release}} \times \underbrace{P(m|b=h)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=h)}_{\text{cost of misconduct}} + \underbrace{P(d|b=h)}_{\text{probability of detention}} \times \underbrace{c(d|b=h)}_{\text{cost of detention}}$$

Status quo bail decisions

Legal bail objective: set lowest possible bail to ensure court appearance, public safety
=> want to set bail low but also want low misconduct

Judge has choice between lenient (no money bail; <u>b=l</u>) and harsh bail (money bail; <u>b=h</u>)

Judge costs:

$$\underbrace{P(m|b=l)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=l)}_{\text{cost of misconduct}} \qquad \underbrace{(1-P(d|b=h))}_{\text{probability of release}} \times \underbrace{P(m|b=h)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=h)}_{\text{cost of misconduct}} + \underbrace{P(d|b=h)}_{\text{probability of detention}} \times \underbrace{c(d|b=h)}_{\text{cost of detention}}$$

Judges do not face costs when make *"correct decision"*

Status quo bail decisions

Legal bail objective: set lowest possible bail to ensure court appearance, public safety
=> want to set bail low but also want low misconduct

Judge has choice between lenient (no money bail; b=l) and harsh bail (money bail; b=h)

Judge costs:

$$\underbrace{P(m|b=l)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=l)}_{\text{cost of misconduct}} \qquad \cancel{\underbrace{(1-P(d|b=h))}_{\text{probability of release}} \times \underbrace{P(m|b=h)}_{\text{probability of misconduct}} \times \underbrace{c(m|b=h)}_{\text{cost of misconduct}}} + \underbrace{P(d|b=h)}_{\text{probability of detention}} \times \underbrace{c(d|b=h)}_{\text{cost of detention}}$$

Judges do not face costs when make *"correct decision"*

=> no misconduct costs when harsh and released (but no way to "verify" detention choice because misconduct unobserved)

Status quo bail decisions

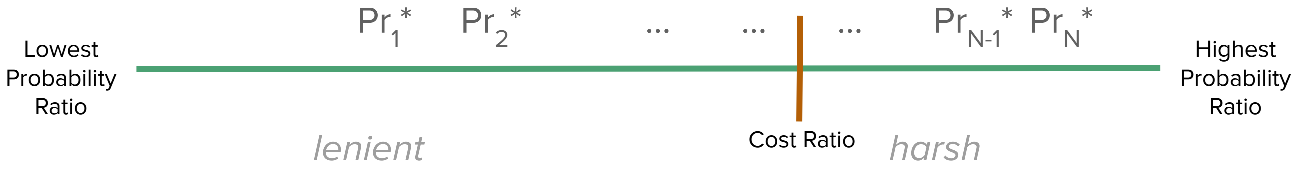
Judge sets bail based on threshold rule:

$$b = \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases}$$

Status quo bail decisions

Judge sets bail based on threshold rule:

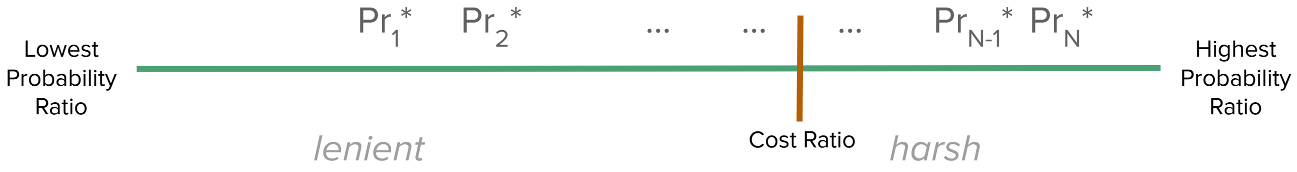
$$b = \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases}$$



Status quo bail decisions

Judge sets bail based on threshold rule:

$$b = \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases}$$



How does the judge predict P(m|b=l)?

- Vector of case information: X
- Risk level from algorithm r^A in {low, moderate, high}
  - Transformation of $P^A(m|b=l)$, algorithm’s prediction of misconduct under lenient bail
- $P(m|b=l)=f(X, r^A)$

Decisions with algorithm recommendations

Introduce algorithm recommendation R, which is based on r^A

$$R = \begin{cases} b = l, & \text{if } r^A \in \{low, moderate\} \\ -, & \text{otherwise} \end{cases}$$

Decisions with algorithm recommendations

Introduce algorithm recommendation R, which is based on r^A

$$R = \begin{cases} b = l, & \text{if } r^A \in \{low, moderate\} \\ -, & \text{otherwise} \end{cases}$$

Theory 1: Recommendation impacts judge predictions only

- R: b=l tells judge that r^A in {low, moderate}
    - Judge already knew this because P(m|b=l)=f(X, rA)
        - **<u>Prediction: no changes to behavior</u>**

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

Harsh recommendation makes lenience **more** costly

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

Harsh recommendation makes lenience **more** costly

School therapists re: mental health algorithmic recommendations

"I'd feel nervous about the liability... You have this thing telling you someone is high risk, and you're just going to let them go?"

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

Harsh recommendation makes lenience **more** costly

CAN A.I. TREAT MENTAL ILLNESS?

New computer systems aim to peer inside our heads—and to help us fix what they find there.

By Dhruv Khullar

February 27, 2023

School therapists re: mental health algorithmic recommendations

"I'd feel nervous about the liability... You have this thing telling you someone is high risk, and you're just going to let them go?"

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."

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

Lenient recommendation makes lenience **less** costly

WHY NEW YORK JAIL POPULATIONS ARE RETURNING TO PRE-PANDEMIC LEVELS

THE APPEAL

Bryce Covert
Jan 20, 2022

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)

Harsh recommendation makes lenience **more** costly

CAN A.I. TREAT MENTAL ILLNESS?

New computer systems aim to peer inside our heads—and to help us fix what they find there.

By Dhruv Khullar
February 27, 2023

School therapists re: mental health algorithmic recommendations

"I'd feel nervous about the liability... You have this thing telling you someone is high risk, and you're just going to let them go?"

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."

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

- $c(m|b=l)$ becomes $c(m|b=l, R)$; *in this case,* $c(m|b=l, R) = c(m|b=l, R^{b=l})$

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

- $c(m|b=l)$ becomes $c(m|b=l, R)$; *in this case,* $c(m|b=l, R) = c(m|b=l, R^{b=l})$

Judges set bail based on two threshold rules (depending on if recommendation applies or not):

$$b = \begin{cases} R = b = l, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l, R=b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \\ R = -, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \end{cases}$$

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

- c(m|b=l) becomes c(m|b=l, R); *in this case, c(m|b=l, R) = c(m|b=l, $R^{b=l}$)*

Judges set bail based on two threshold rules (depending on if recommendation applies or not):

$$b = \begin{cases} R = b = l, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l,R=b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \\ R = -, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \end{cases}$$

c(m|b=l, $R^{b=l}$)<c(m|b=l) because there is less liability when a mistake is in line with a recommendation

Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

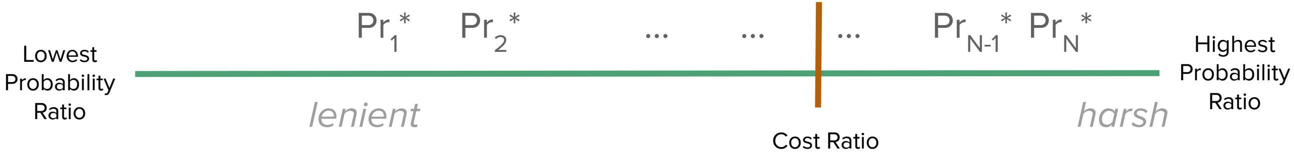
- c(m|b=l) becomes c(m|b=l, R); *in this case, c(m|b=l, R) = c(m|b=l, $R^{b=l}$)*

Judges set bail based on two threshold rules (depending on if recommendation applies or not):

$$b = \begin{cases} R = b = l, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l,R=b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \\ R = -, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \end{cases}$$

c(m|b=l, $R^{b=l}$)<c(m|b=l) because there is less liability when a mistake is in line with a recommendation

- Critical threshold shifts right for low/moderate cases **<u>(increase in lenient bail setting rate)</u>**



Decisions with algorithm recommendations

Theory 2: Recommendation changes judge error costs

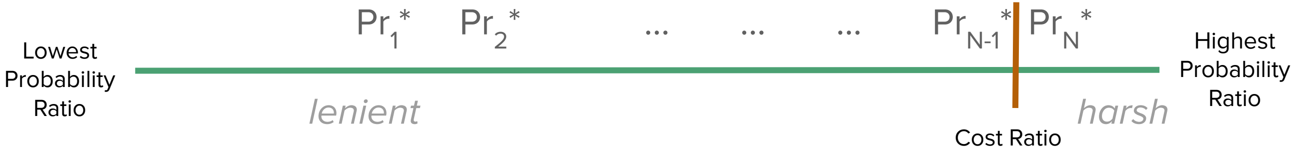
- c(m|b=l) becomes c(m|b=l, R); *in this case, c(m|b=l, R) = c(m|b=l,* $R^{b=l}$*)*

Judges set bail based on two threshold rules (depending on if recommendation applies or not):

$$b = \begin{cases} R = b = l, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l,R=b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \\ R = -, \begin{cases} h, & \text{if } \frac{c(d|b=h)}{c(m|b=l)} < \frac{Pr(m|b=l)}{Pr(d|b=h)} \\ l, & \text{otherwise} \end{cases} \end{cases}$$

c(m|b=l, $R^{b=l}$)<c(m|b=l) because there is less liability when a mistake is in line with a recommendation

- Critical threshold shifts right for low/moderate cases **(increase in lenient bail setting rate)**



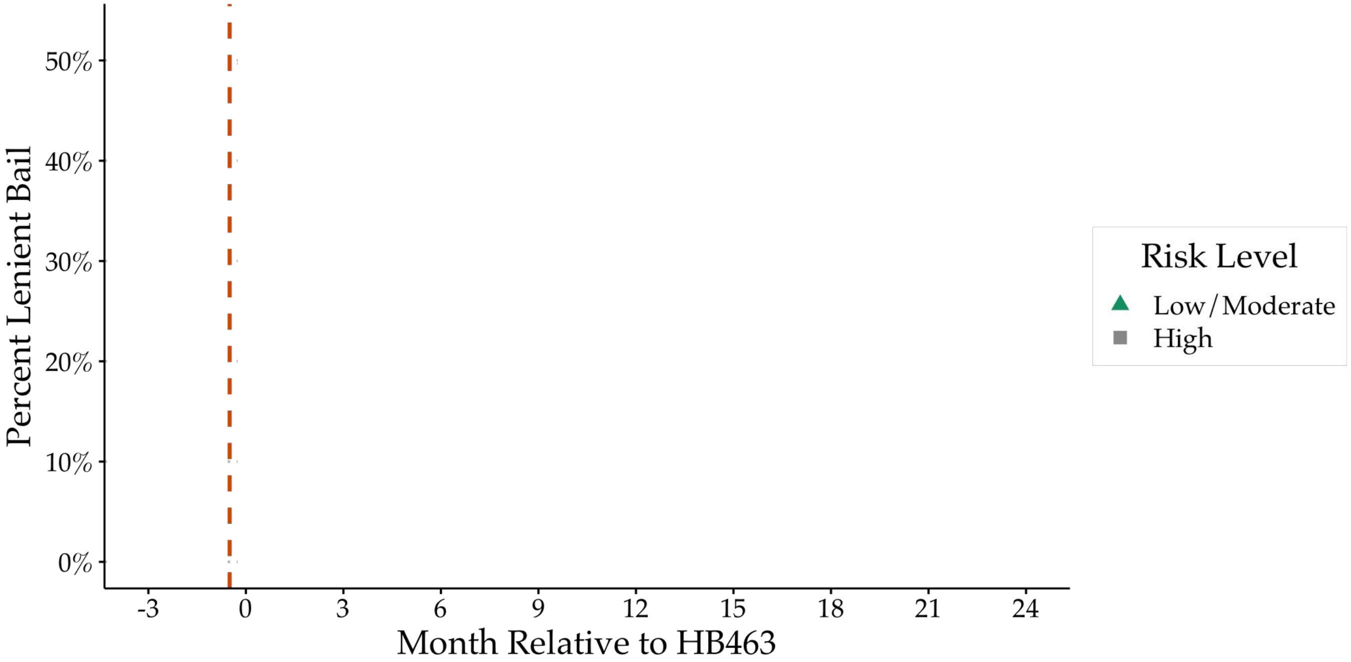
Causal effects of algorithmic recommendations

Difference-in-differences approach

- Low/moderate risk level cases get a lenient recommendation
- High risk level cases do not

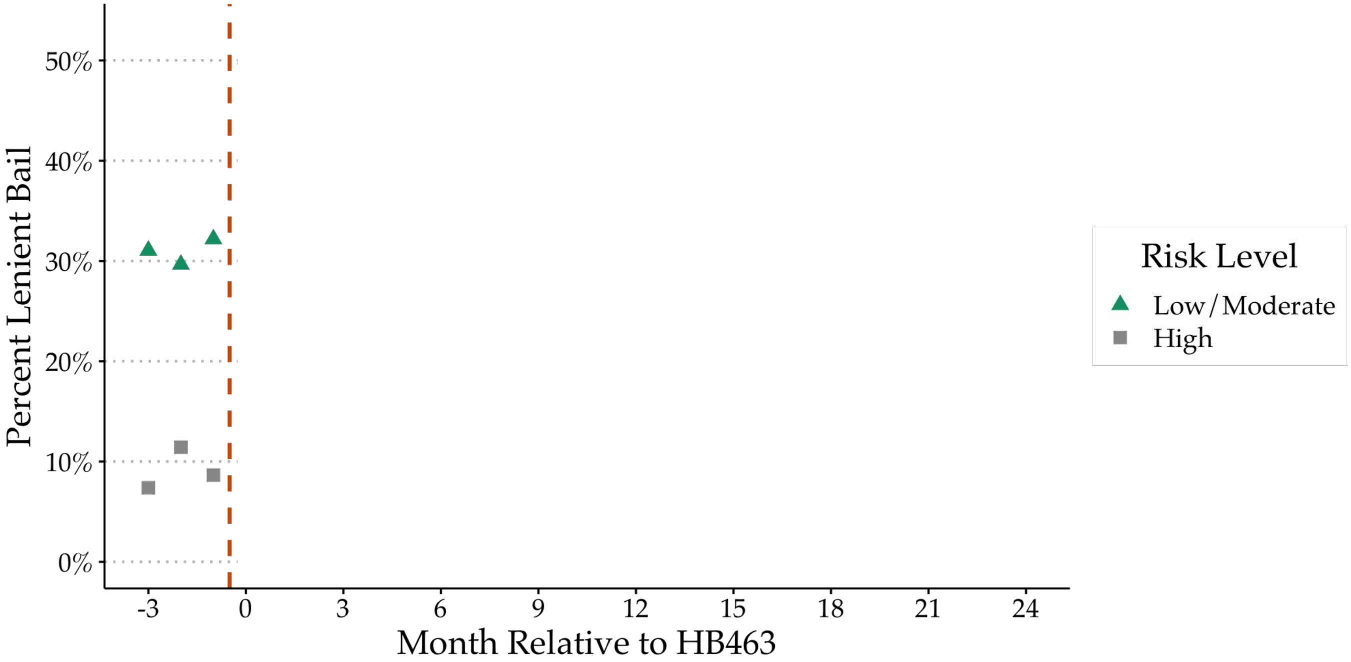
Difference-in-differences approach

- Low/moderate risk level cases get a lenient recommendation
- High risk level cases do not



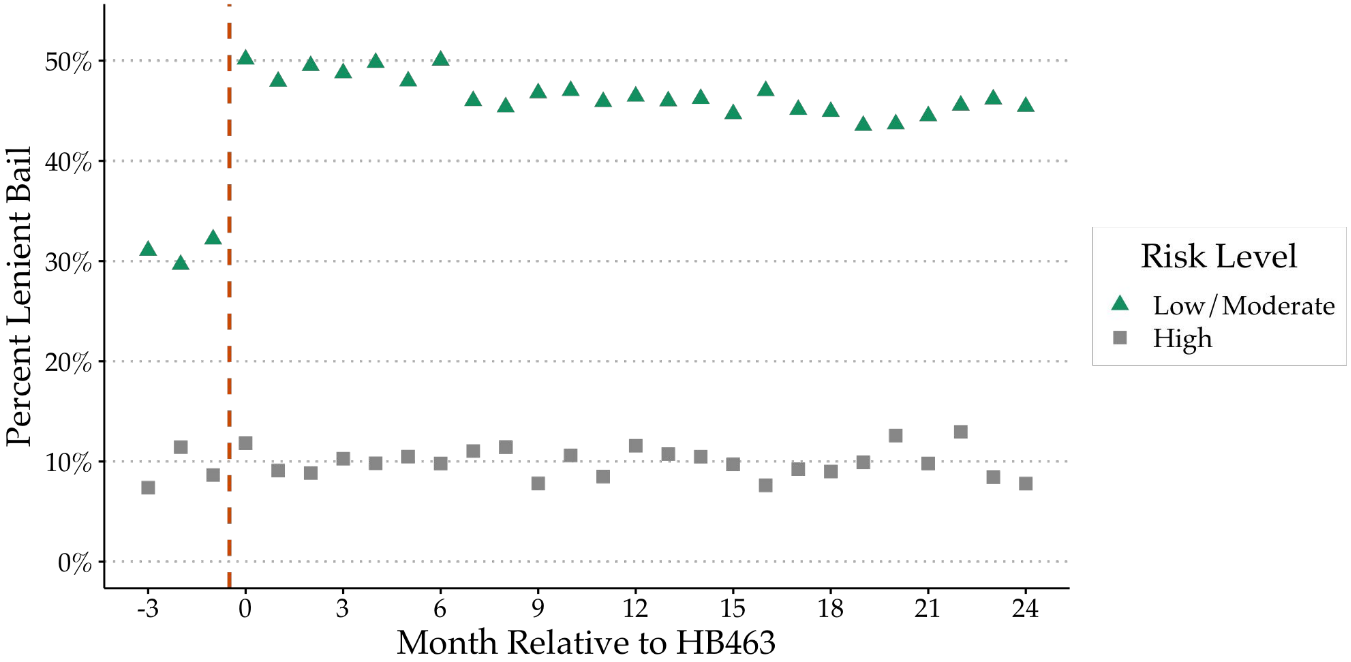
Difference-in-differences approach

- Low/moderate risk level cases get a lenient recommendation
- High risk level cases do not



Difference-in-differences approach

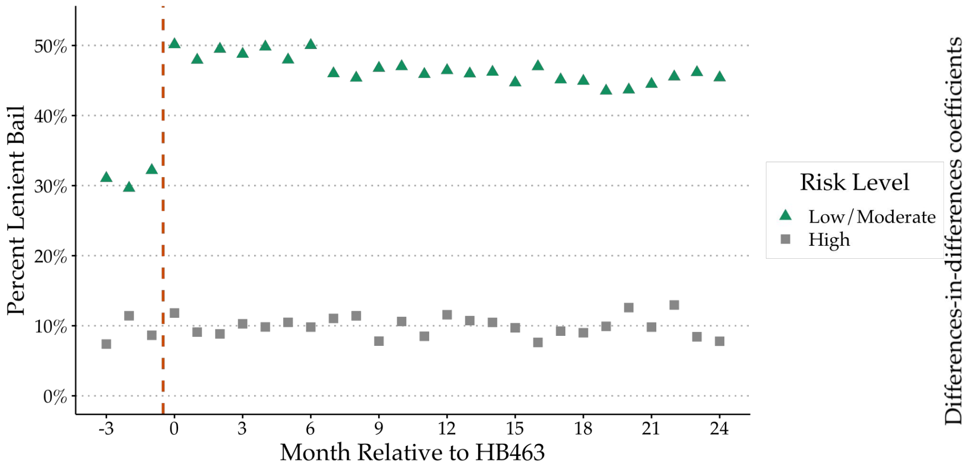
- Low/moderate risk level cases get a lenient recommendation
- High risk level cases do not

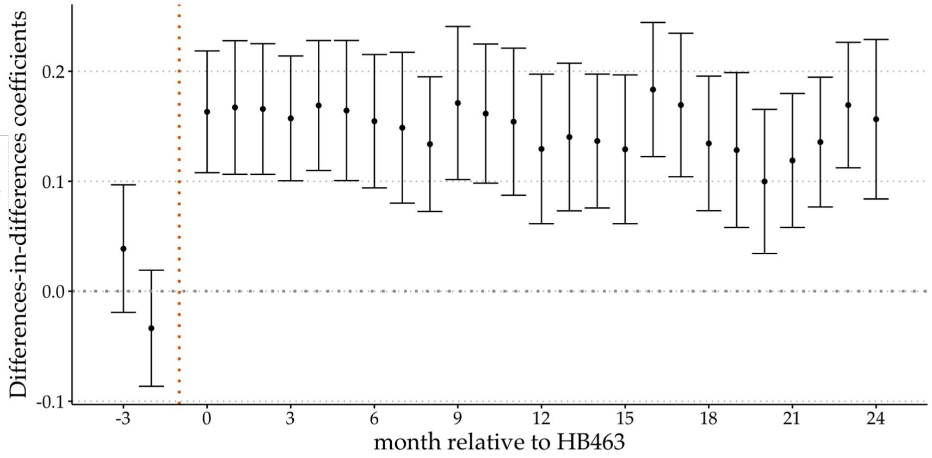


Difference-in-differences approach

- Low/moderate risk level cases get a lenient recommendation
- High risk level cases do not

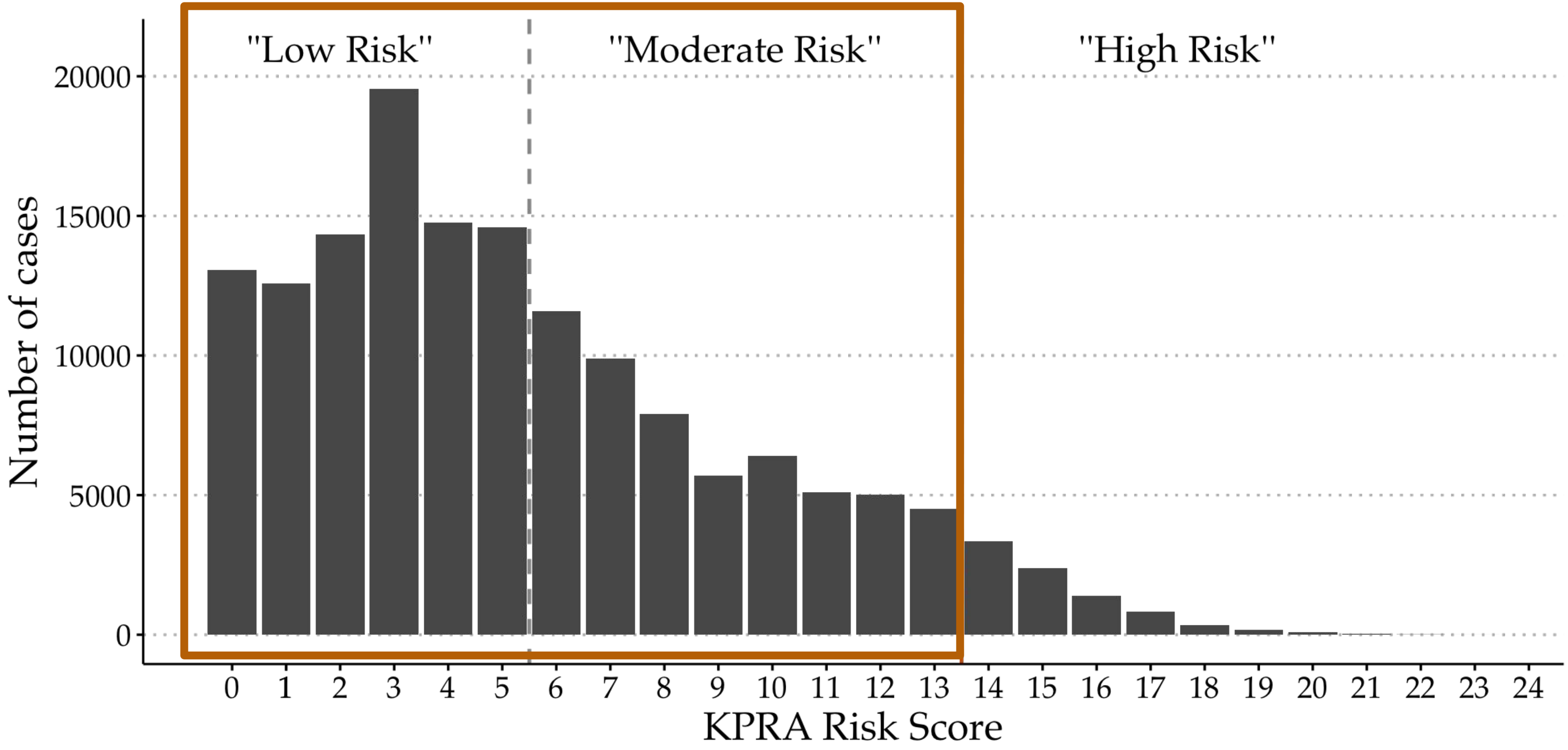
$$lenient_{itj} = \sum_{m \neq -1} [\beta_m \times I(score_i < 14)] + X_{itj} + \epsilon_{itj}$$



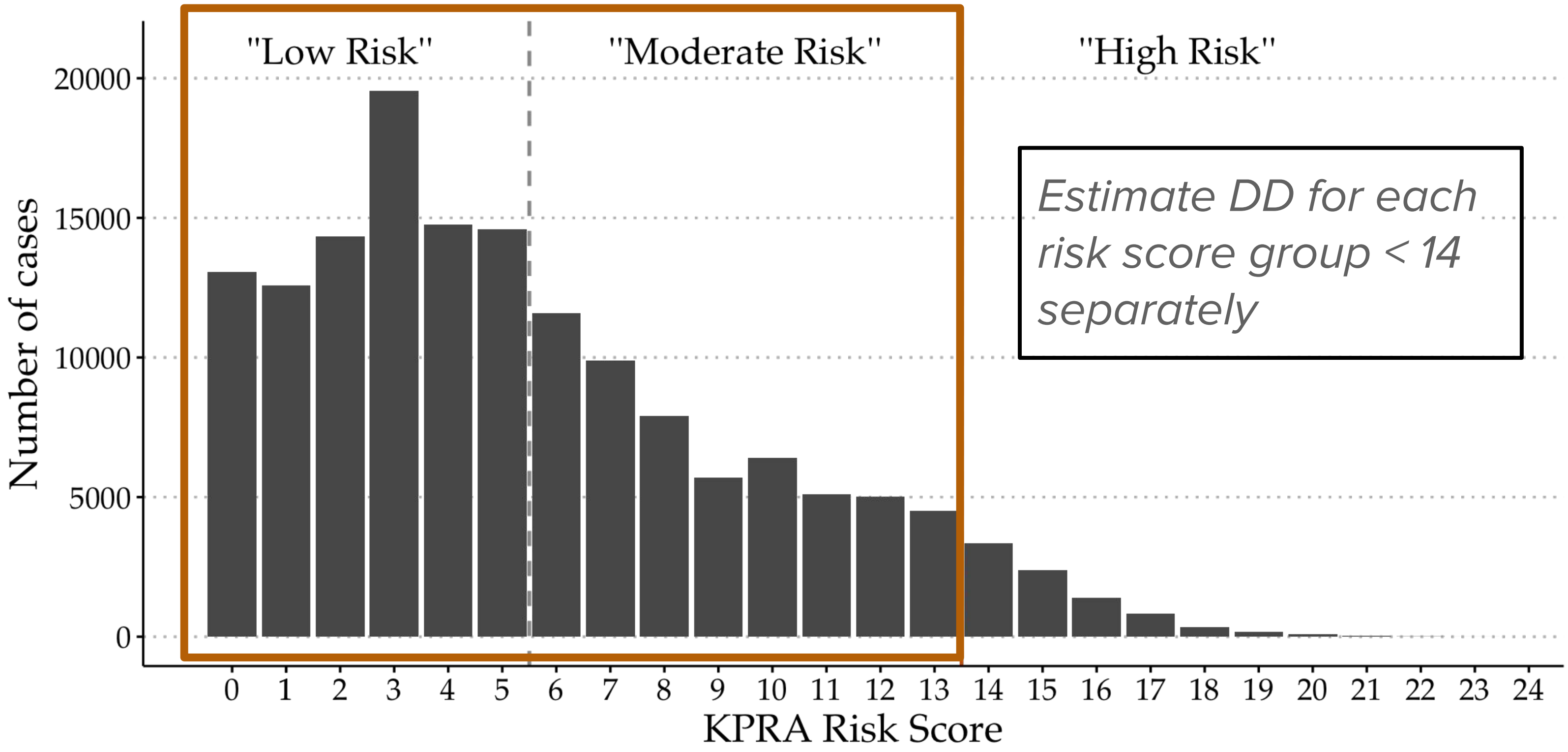


<u>Pooled DD:</u> **15 pp increase / 50% increase** (off the 30% baseline)

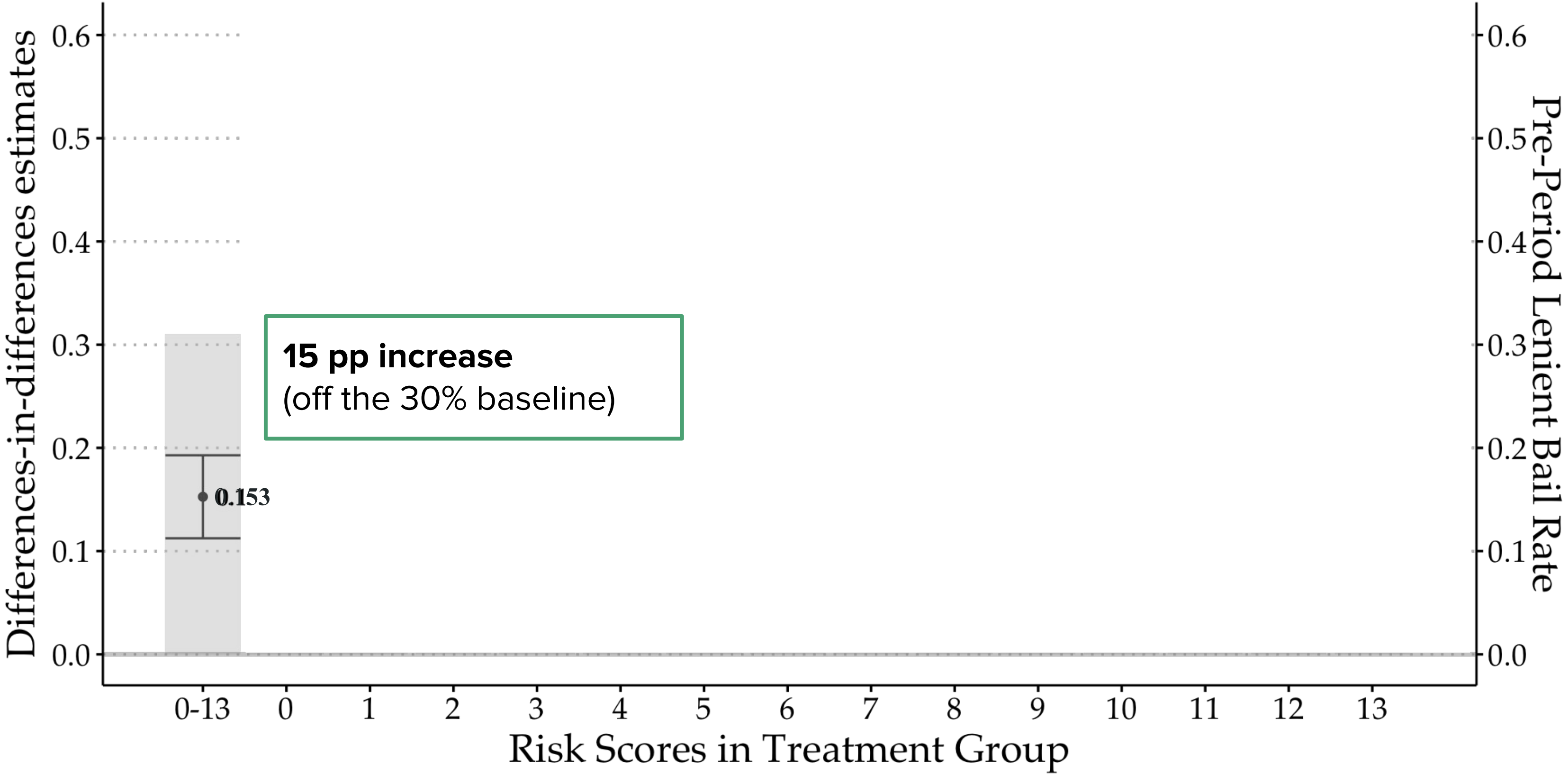
Heterogeneity in effects across the risk score distribution



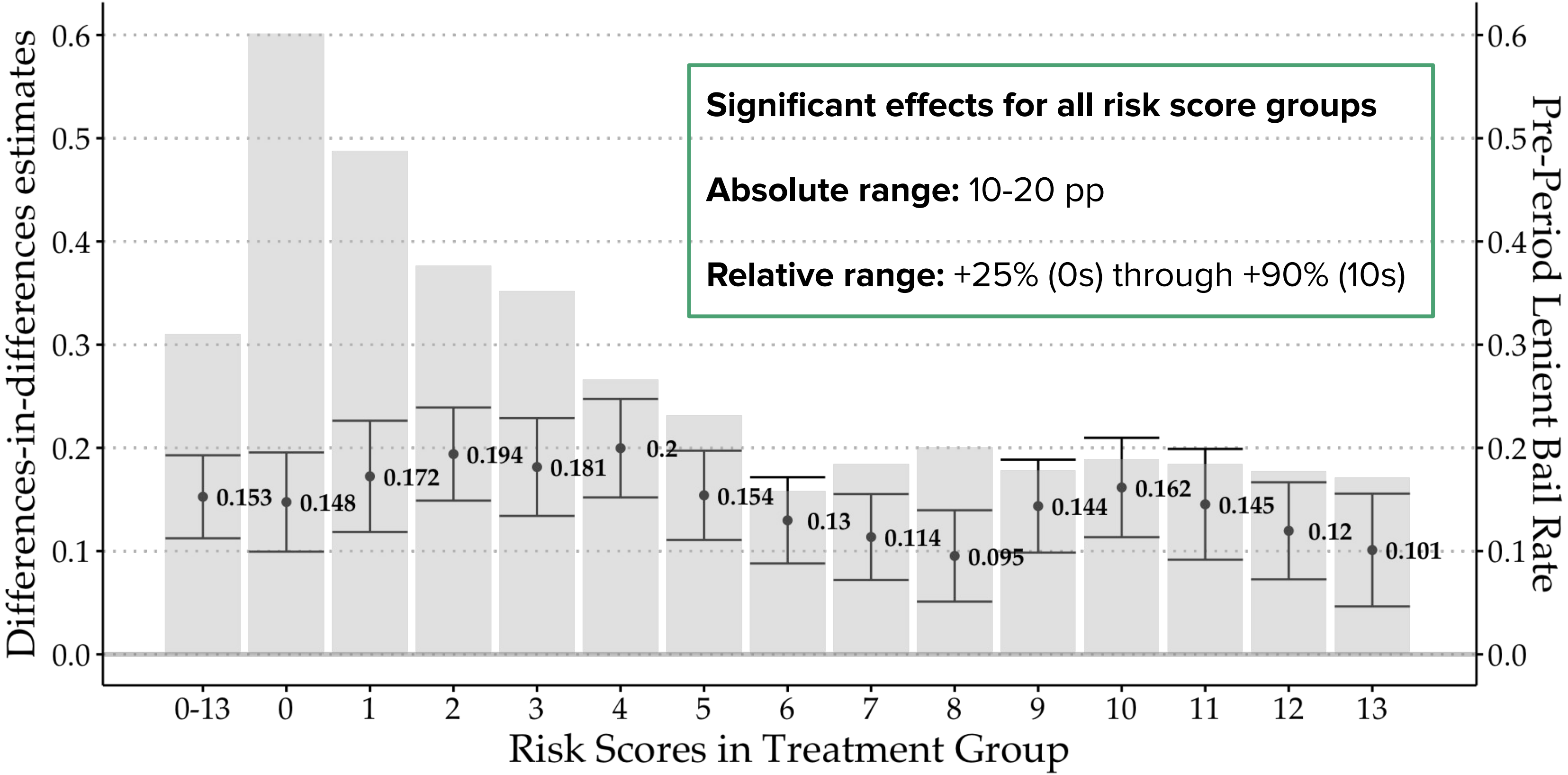
Heterogeneity in effects across the risk score distribution



Heterogeneity in effects across the risk score distribution

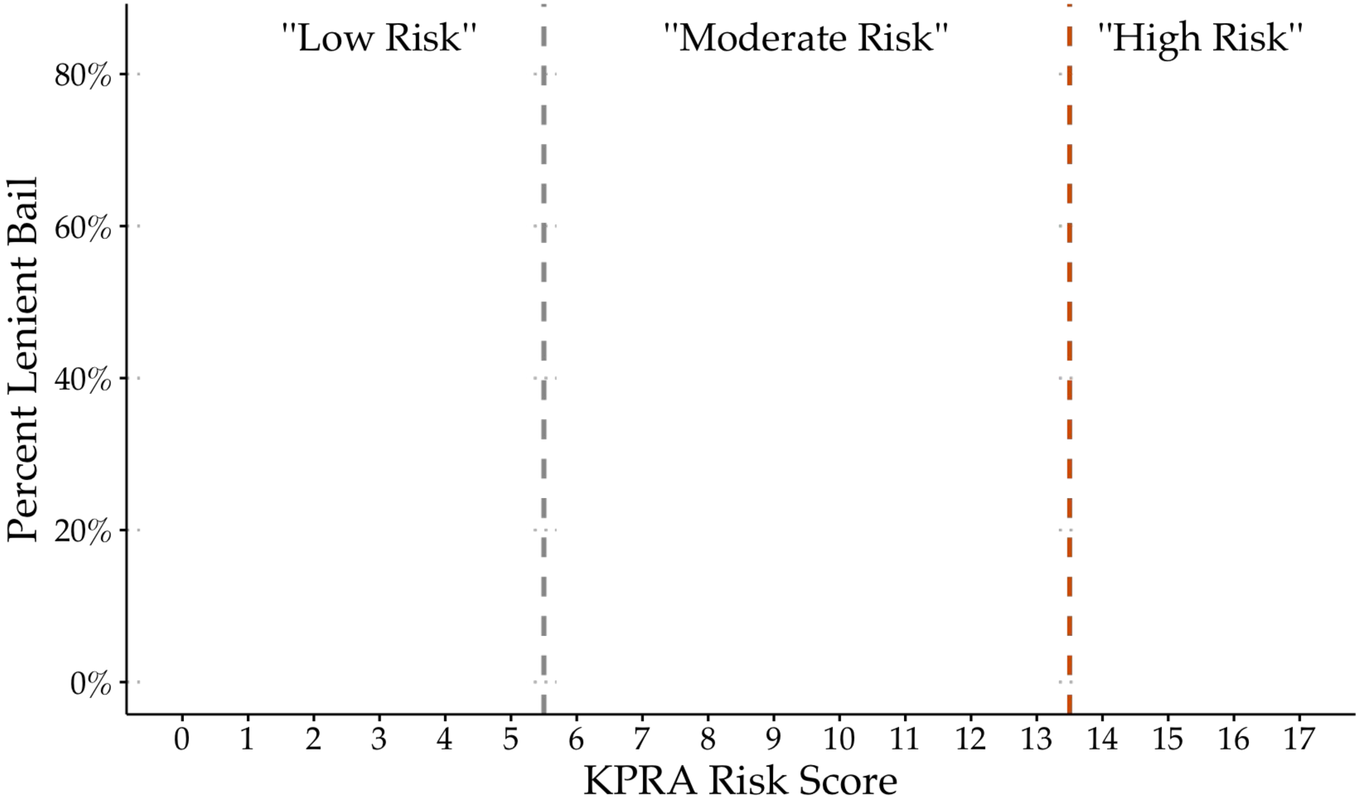


Heterogeneity in effects across the risk score distribution

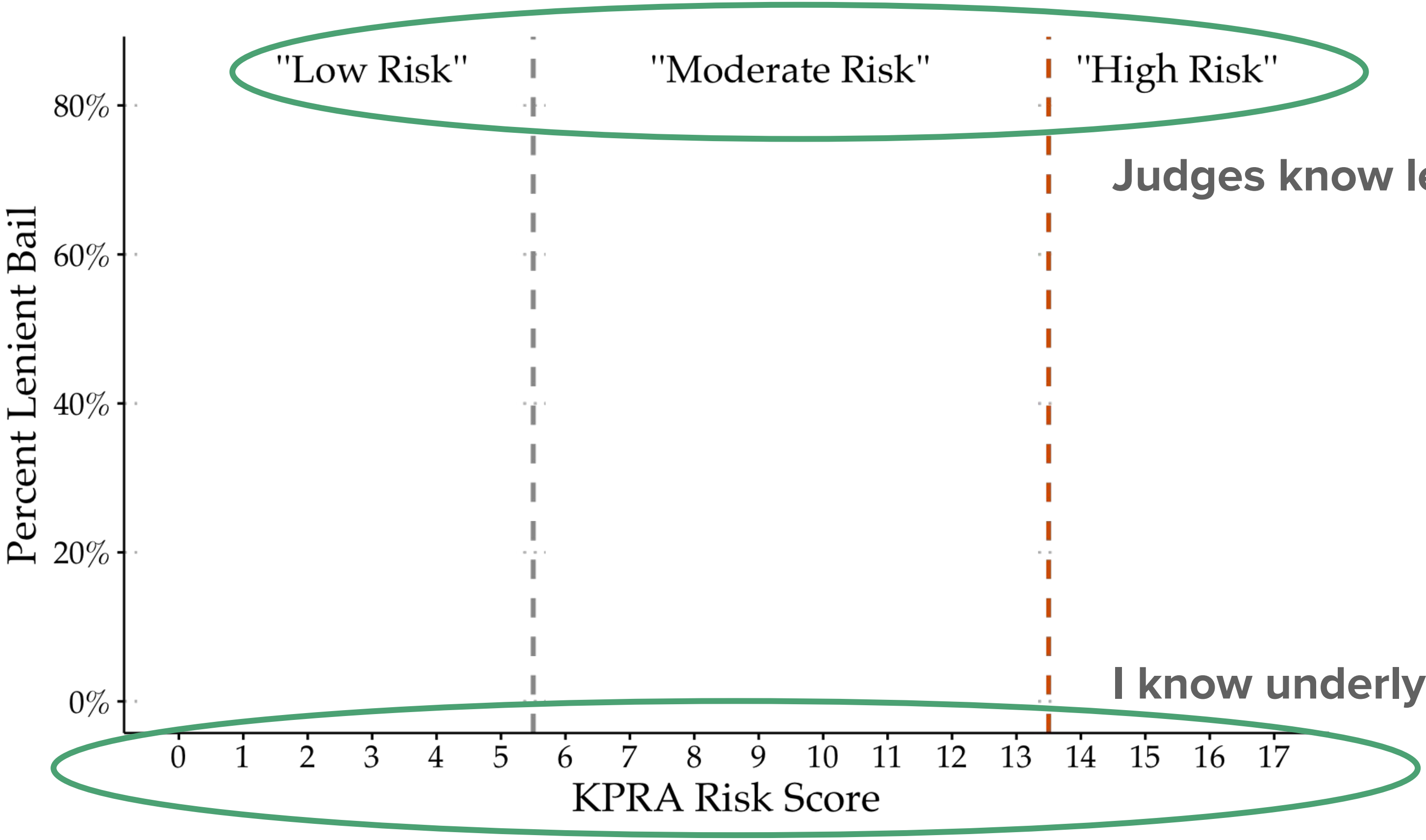


Another approach: leverage discontinuities

Another approach: leverage discontinuities



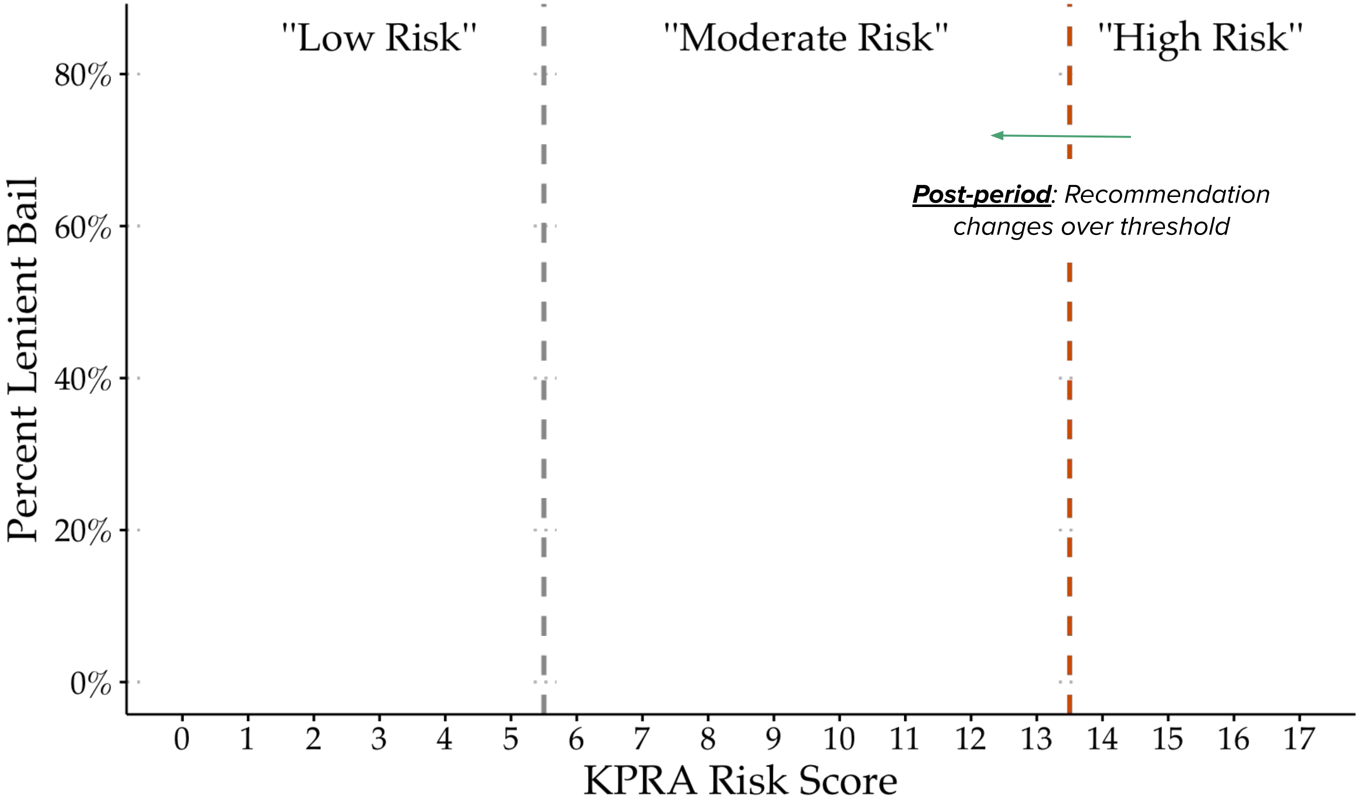
Another approach: leverage discontinuities



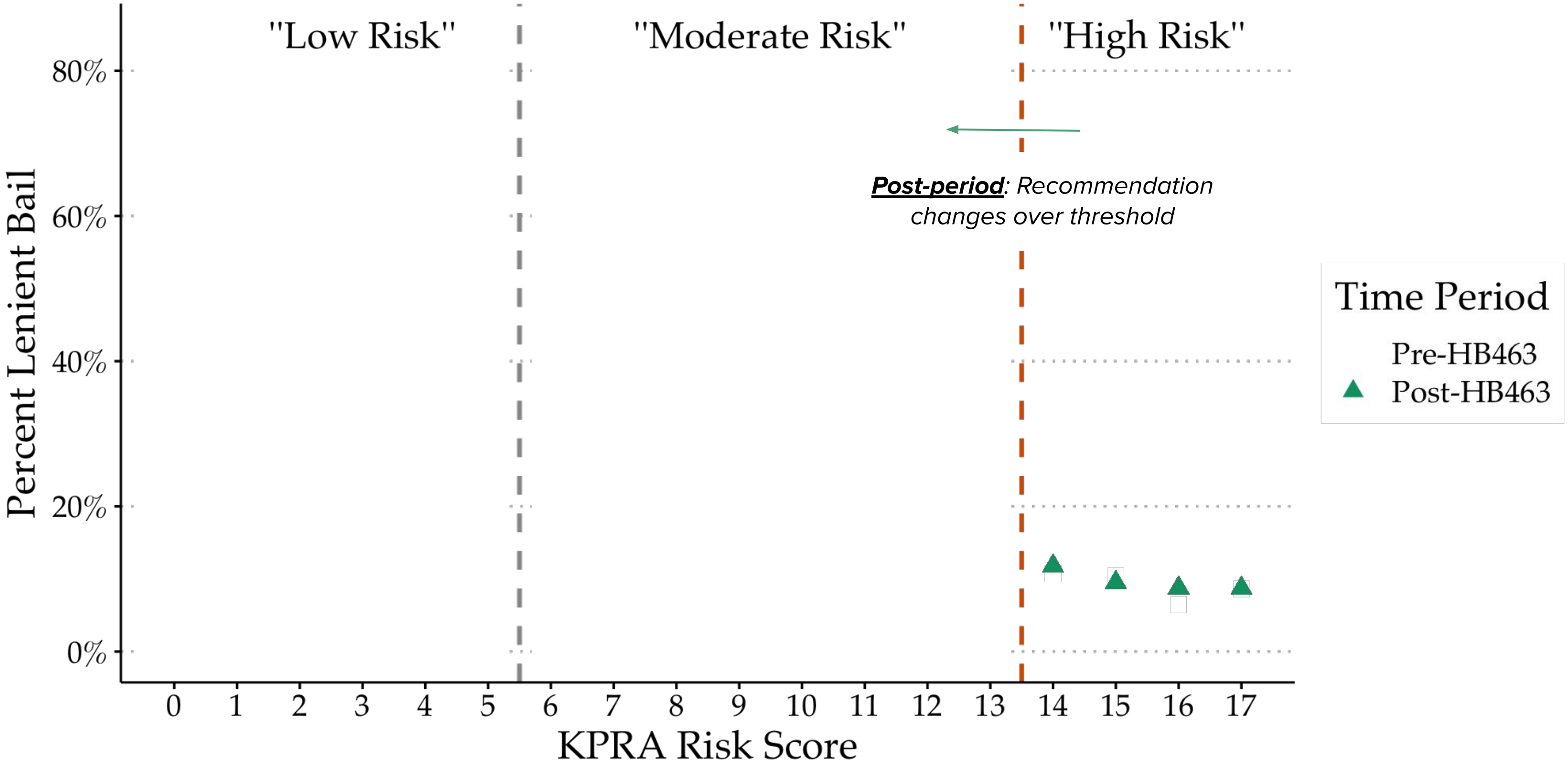
Judges know levels

I know underlying scores

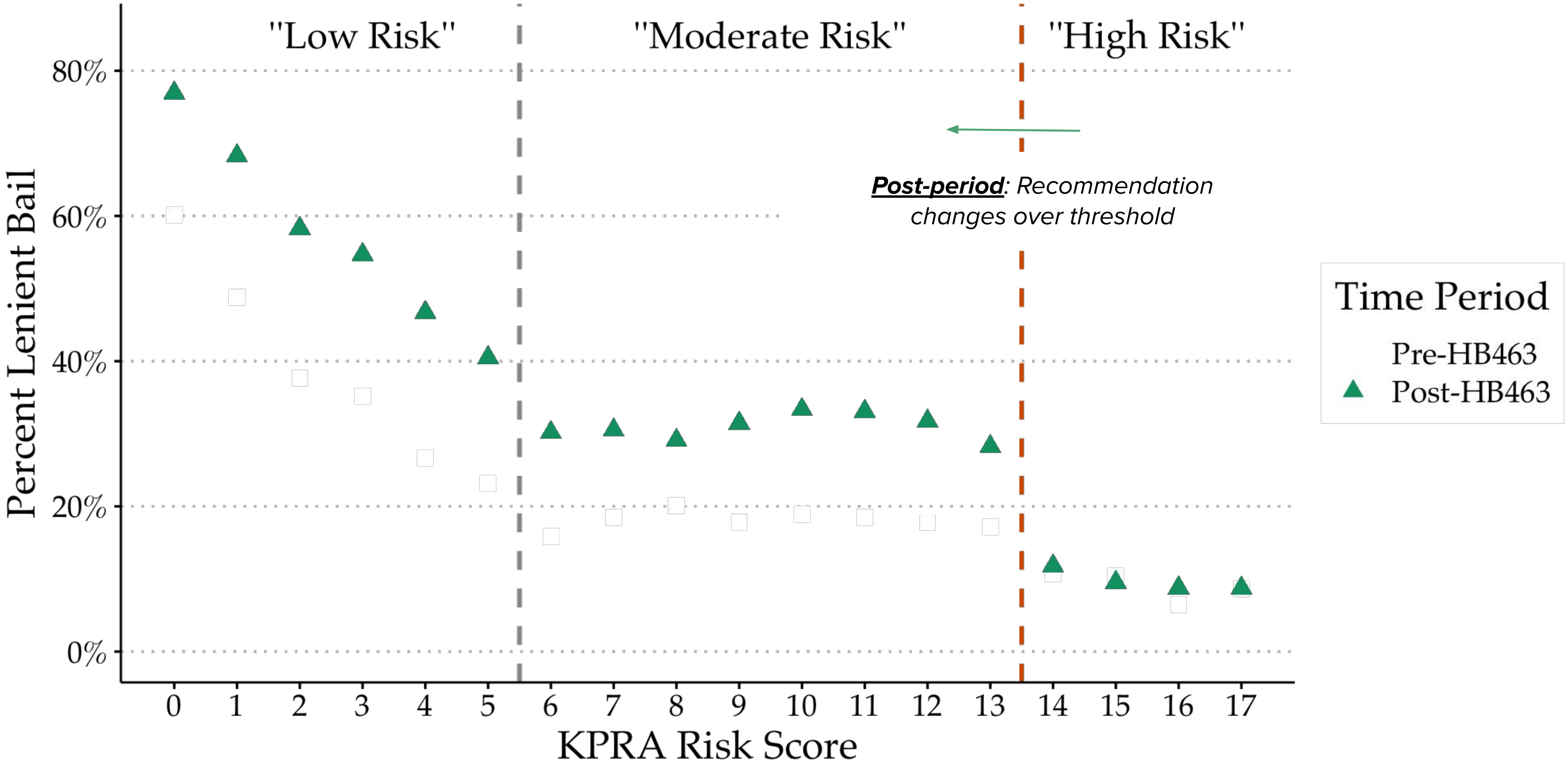
Regression discontinuity after recommendations



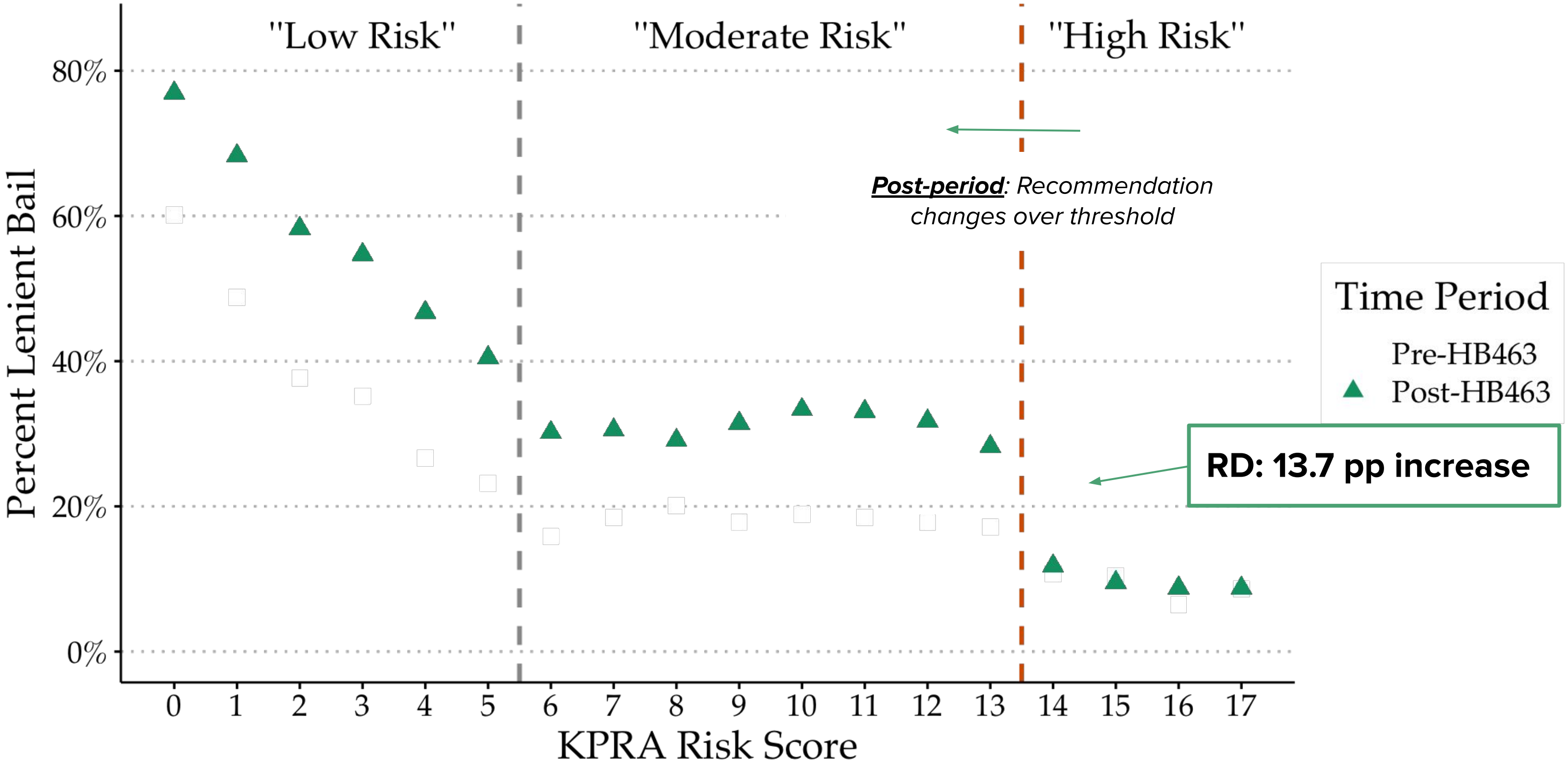
Regression discontinuity after recommendations



Regression discontinuity after recommendations



Regression discontinuity after recommendations



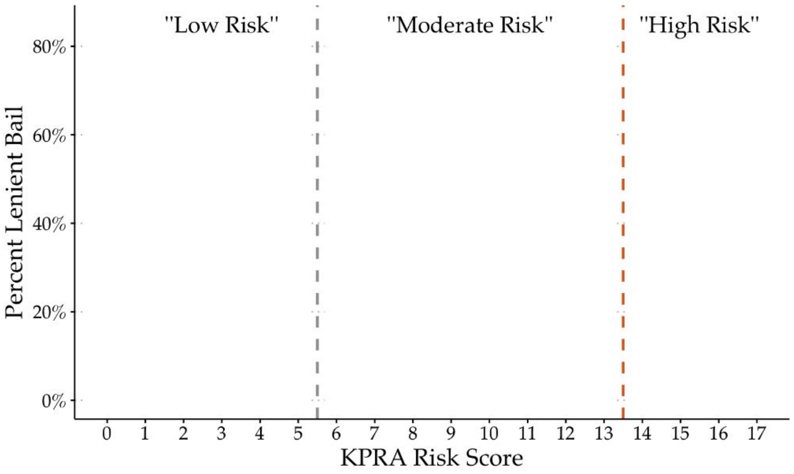
Regression discontinuity after recommendations =/= recommendation effect of interest

Two other factors change discontinuously over threshold

Regression discontinuity after recommendations =/= recommendation effect of interest

Two other factors change discontinuously over threshold

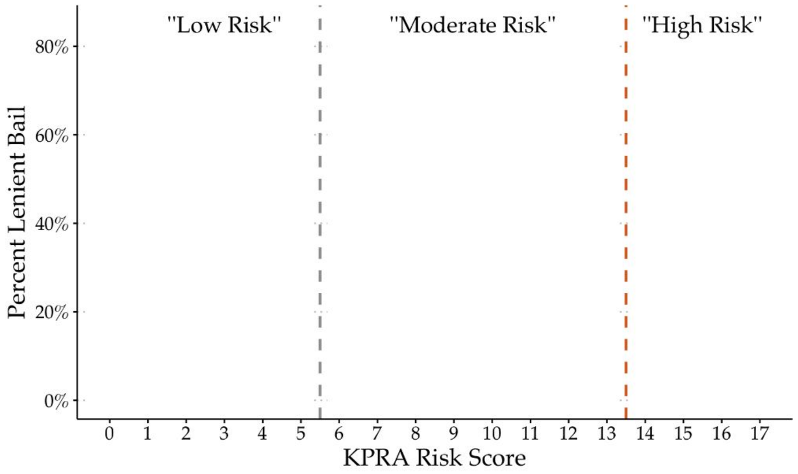
1. Risk level label



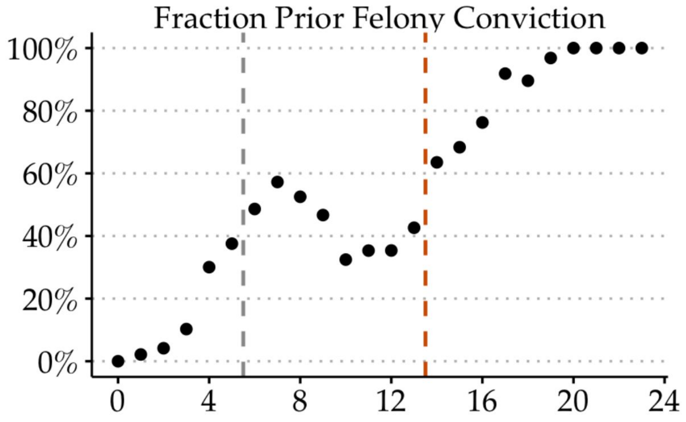
Regression discontinuity after recommendations =/= recommendation effect of interest

Two other factors change discontinuously over threshold

1. Risk level label

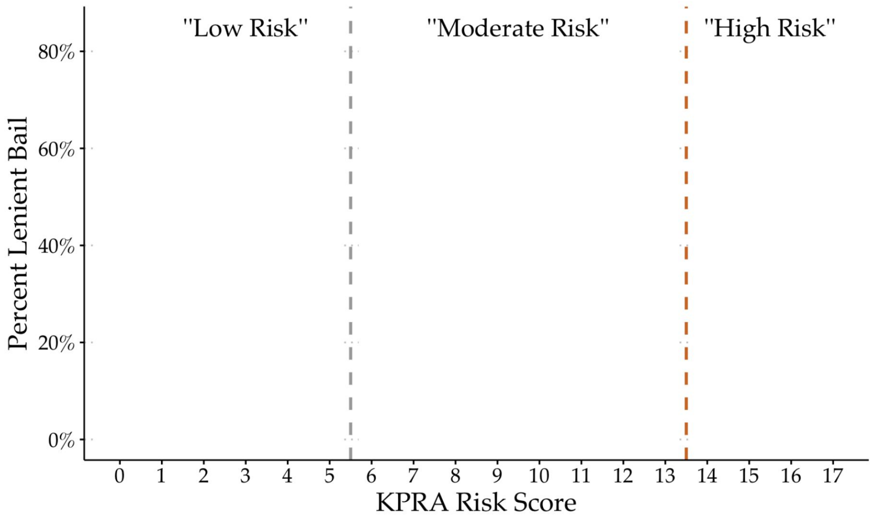


2. Prior felony conviction rate

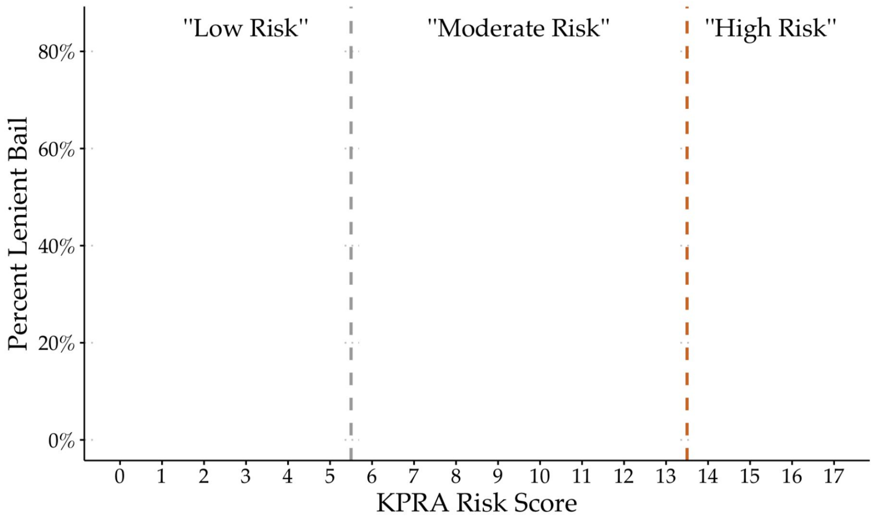


Solution: leverage discontinuities *across time periods*

PRE-PERIOD

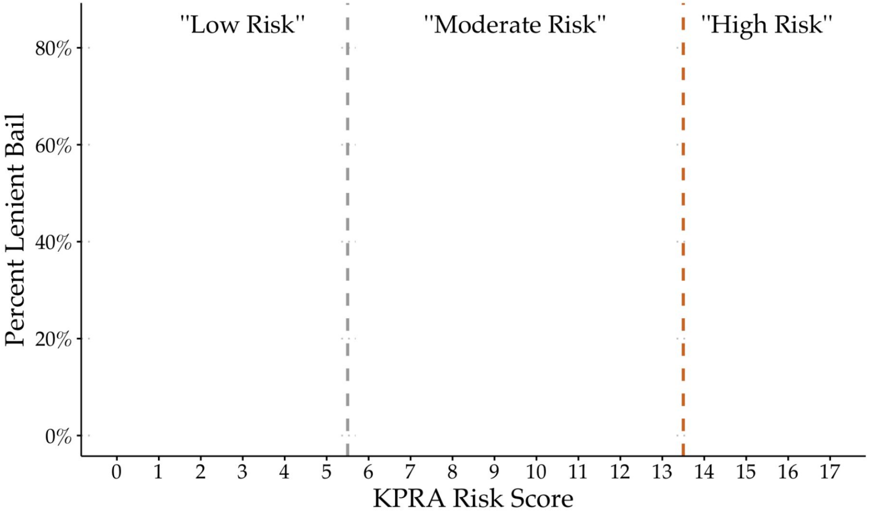


POST PERIOD

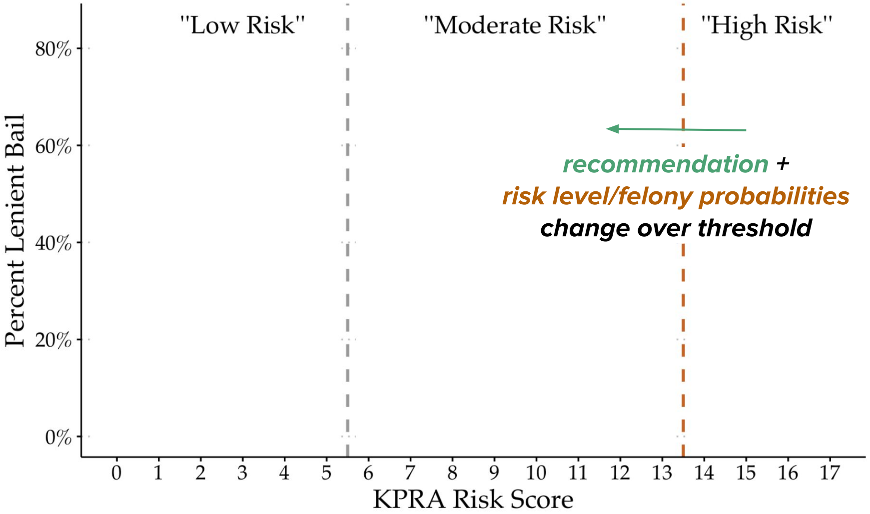


Solution: leverage discontinuities *across time periods*

PRE-PERIOD

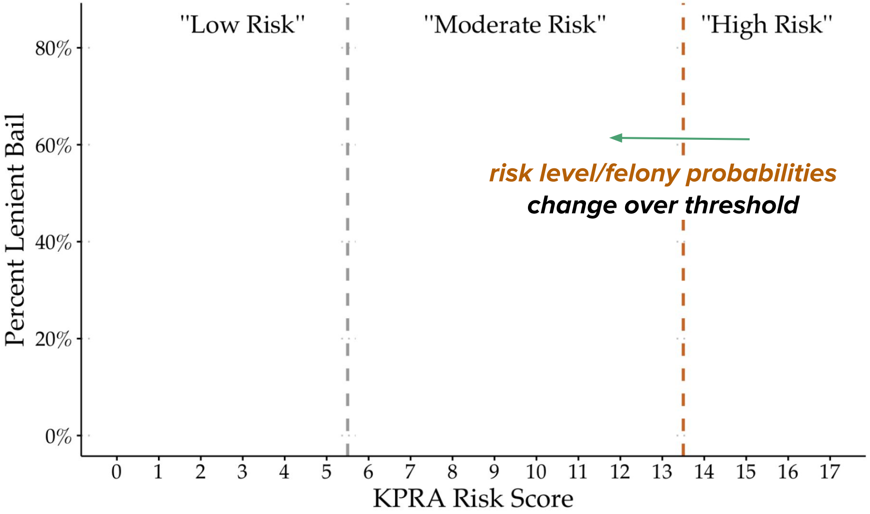


POST PERIOD

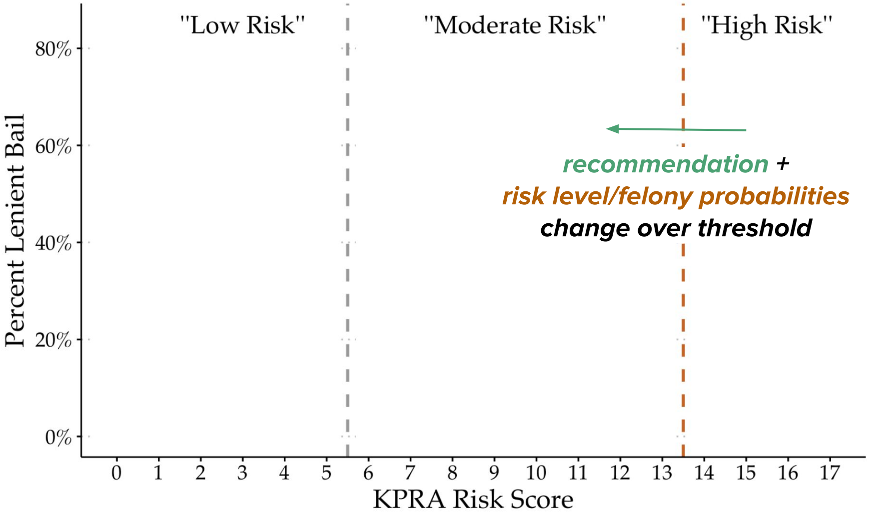


Solution: leverage discontinuities *across time periods*

PRE-PERIOD

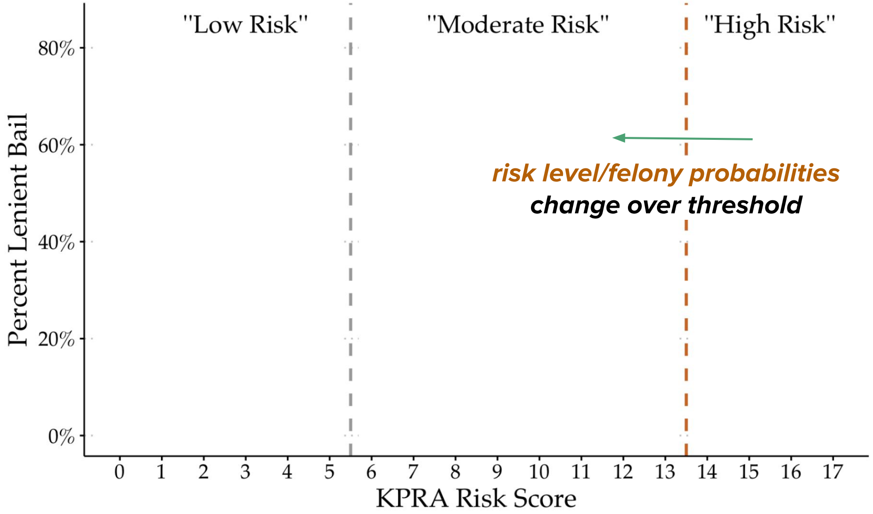


POST PERIOD

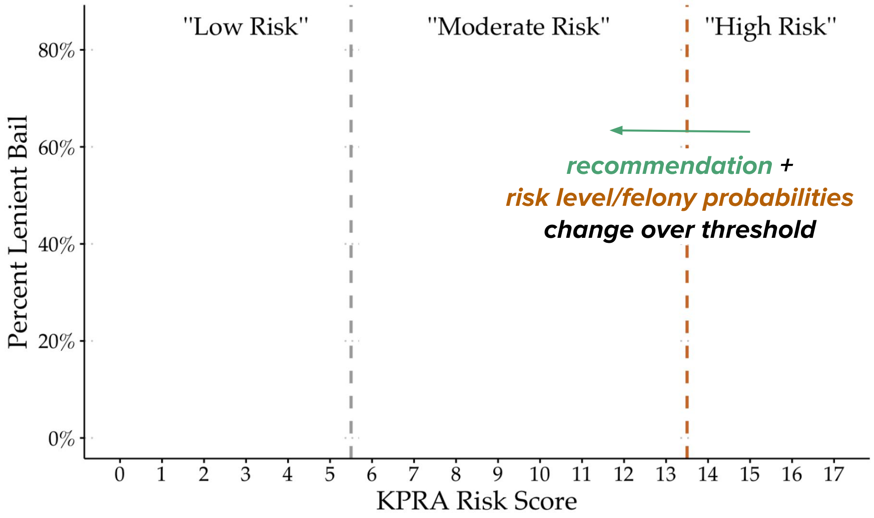


Solution: leverage discontinuities *across time periods*

PRE-PERIOD



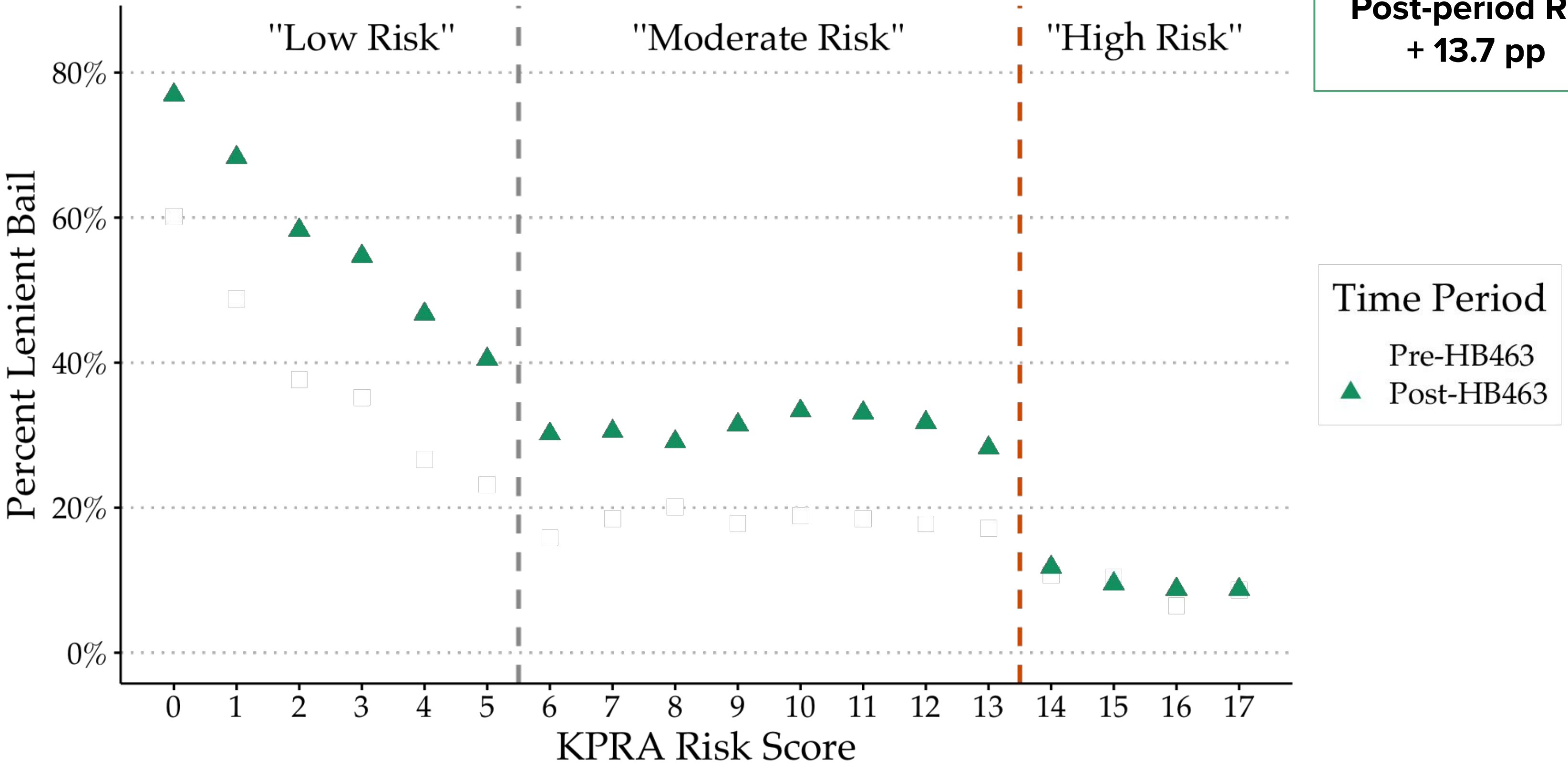
POST PERIOD



Difference-in-discontinuity (diff-in-disc)= RD(post)-RD(pre)

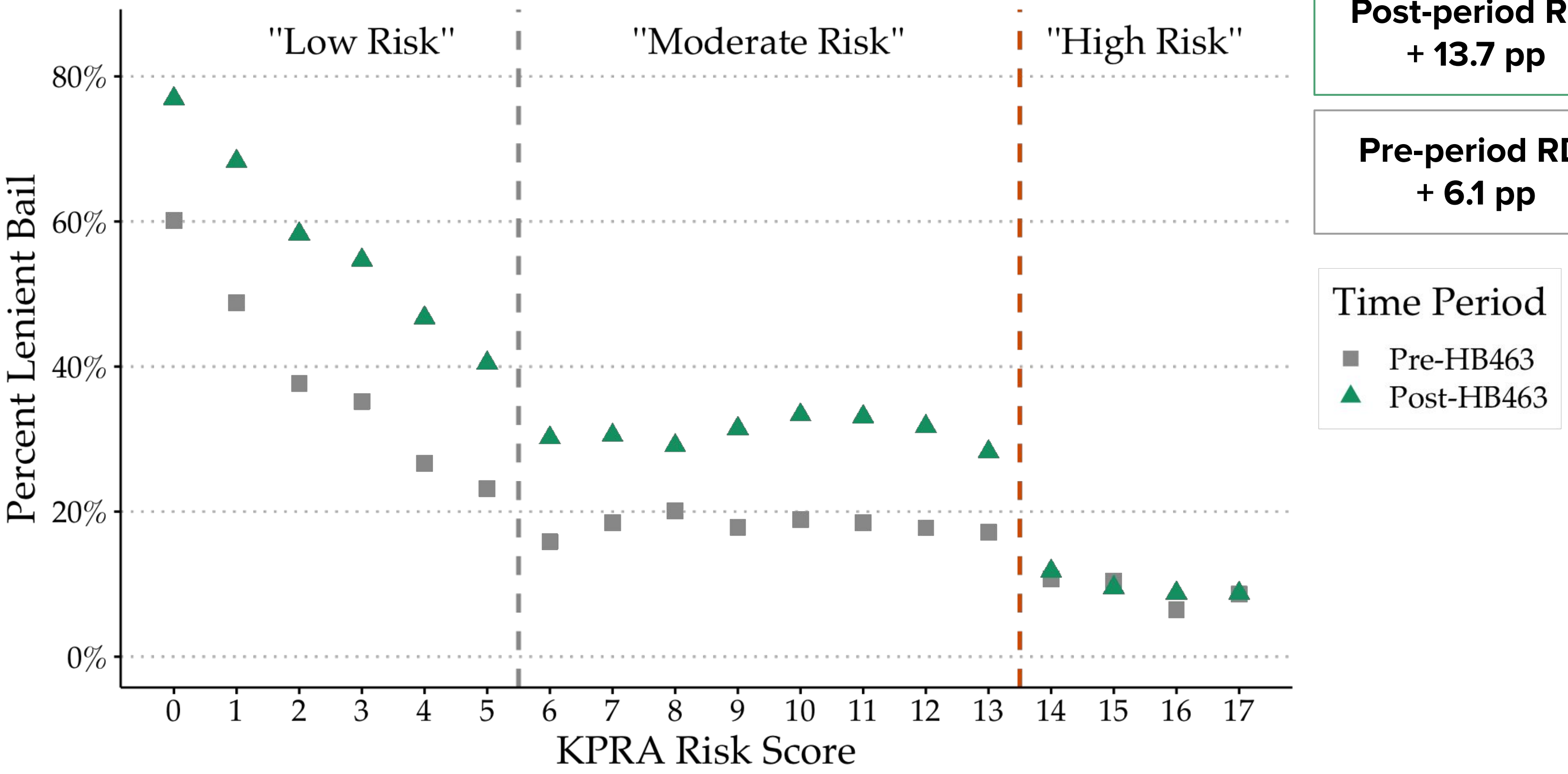
=> to isolate ***recommendation effect***

Differences-in-discontinuities



Post-period RD: + 13.7 pp

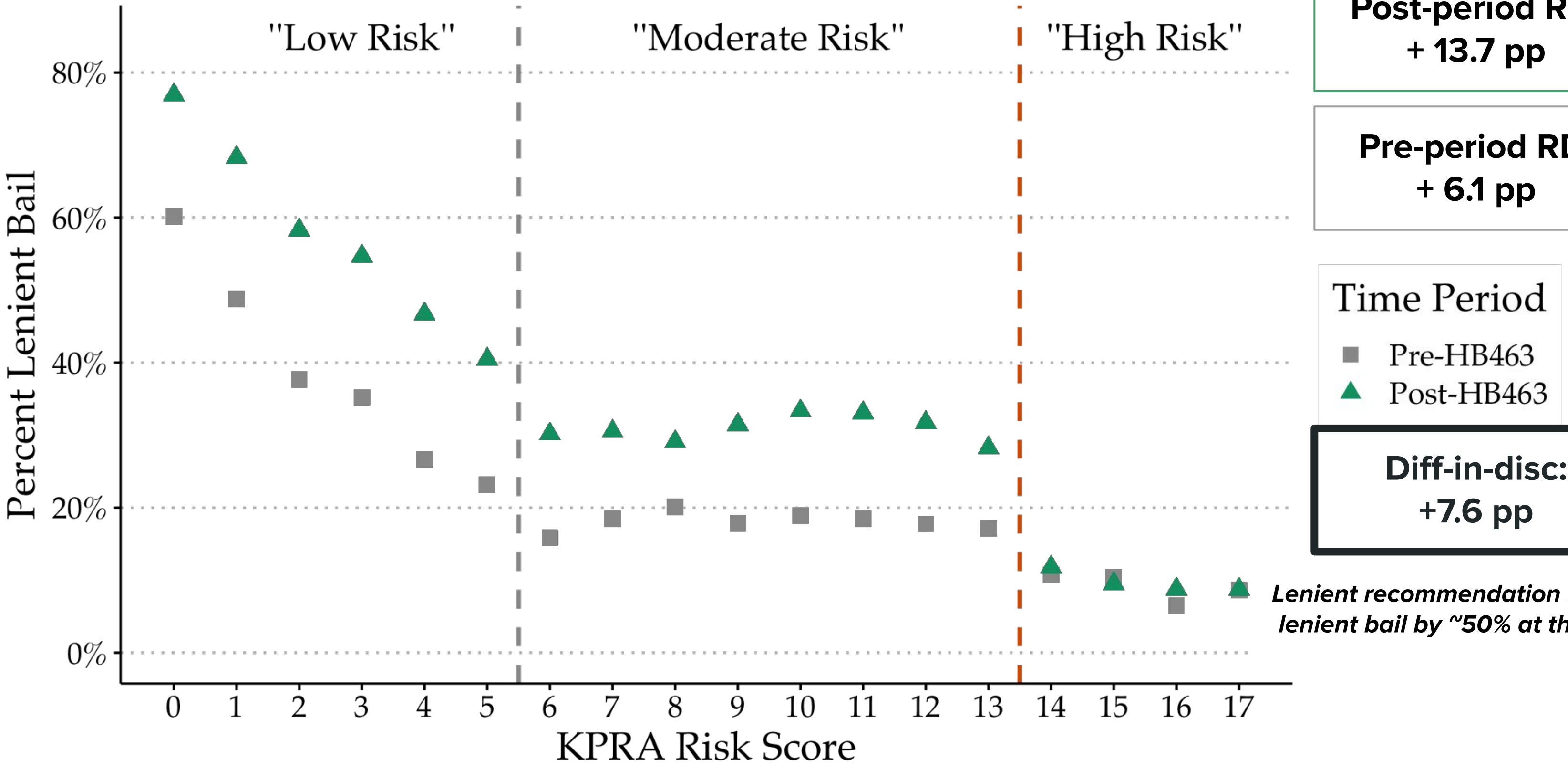
Differences-in-discontinuities



Post-period RD:
+ 13.7 pp

Pre-period RD:
+ 6.1 pp

Differences-in-discontinuities



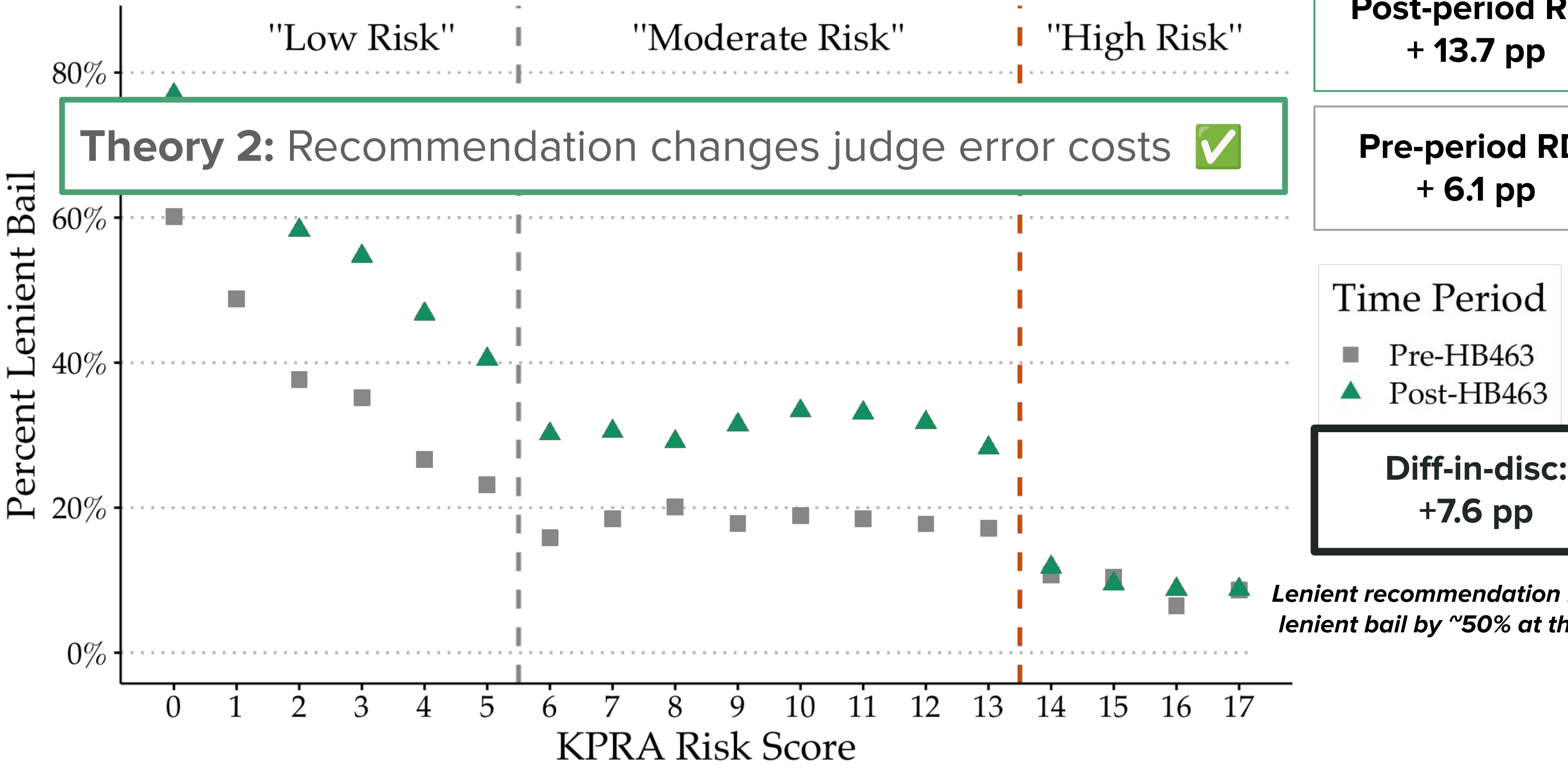
Post-period RD:
+ 13.7 pp

Pre-period RD:
+ 6.1 pp

Diff-in-disc:
+7.6 pp

Lenient recommendation increased lenient bail by ~50% at the margin

Differences-in-discontinuities



Theory 2: Recommendation changes judge error costs ✅

Post-period RD:
+ 13.7 pp

Pre-period RD:
+ 6.1 pp

Diff-in-disc:
+7.6 pp

Lenient recommendation increased lenient bail by ~50% at the margin

Addressing identification concerns

Changes over time + implications for estimates

same risk levels *available*

- No recommendation

June 2011

- Recommendation *(lenient bail for low/moderate cases)*

Changes over time + implications for estimates

same risk levels *available*

- No recommendation
- **Levels optional**

June 2011

- Recommendation *(lenient bail for low/moderate cases)*
- **Levels mandated**

Identification concern: risk levels not consulted in some cases in pre-period...

Changes over time + implications for estimates

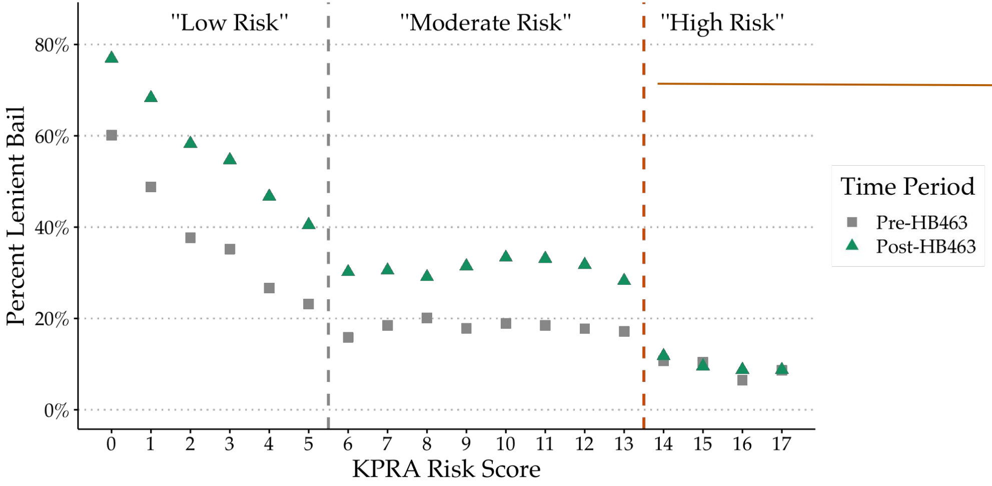
same risk levels *available*

- No recommendation
- **Levels optional**

June 2011

- Recommendation *(lenient bail for low/moderate cases)*
- **Levels mandated**

Identification concern: risk levels not consulted in some cases in pre-period...



Assuming levels used before:

<u>Post-period RD:</u>
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]

<u>Pre-period RD:</u>
[level eff$_{MH}$] + [prior felony eff]

<u>Diff-in-disc:</u>
[recommendation effect]

Changes over time + implications for estimates

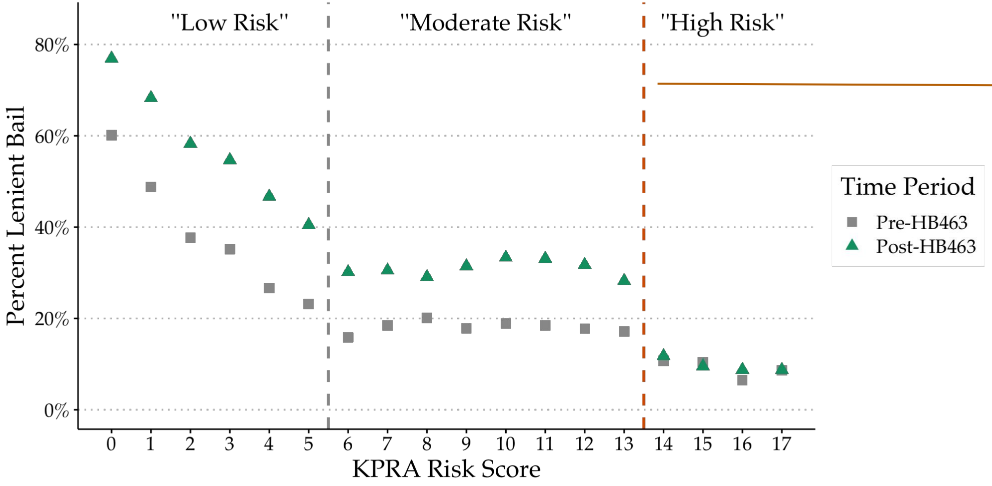
same risk levels *available*

- No recommendation
- **Levels optional**

June 2011

- Recommendation *(lenient bail for low/moderate cases)*
- **Levels mandated**

Identification concern: risk levels not consulted in some cases in pre-period…



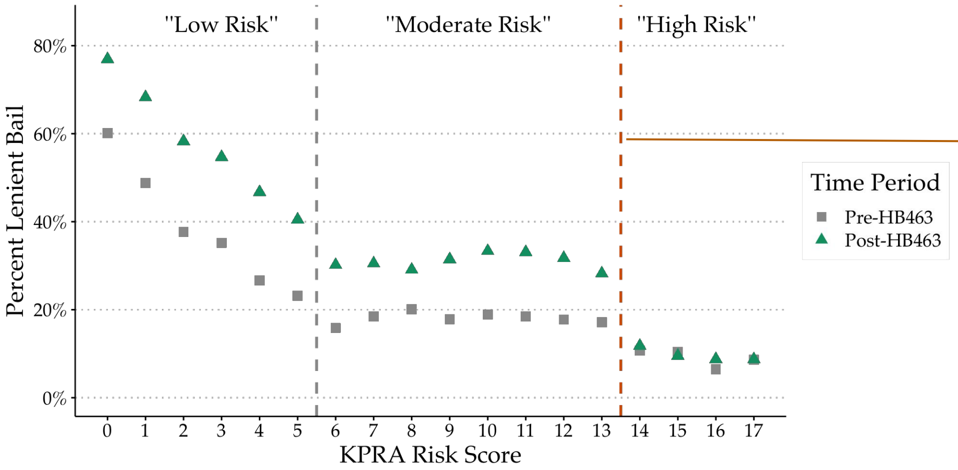
Beforehand, levels consulted in ω cases (in [0,1]) :

Post-period RD:
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]

Pre-period RD:
ω[level eff$_{MH}$] + [prior felony eff]

Diff-in-disc:
[recommendation effect] + **(1-ω)[level eff$_{MH}$]**

Method 1: Estimating ω



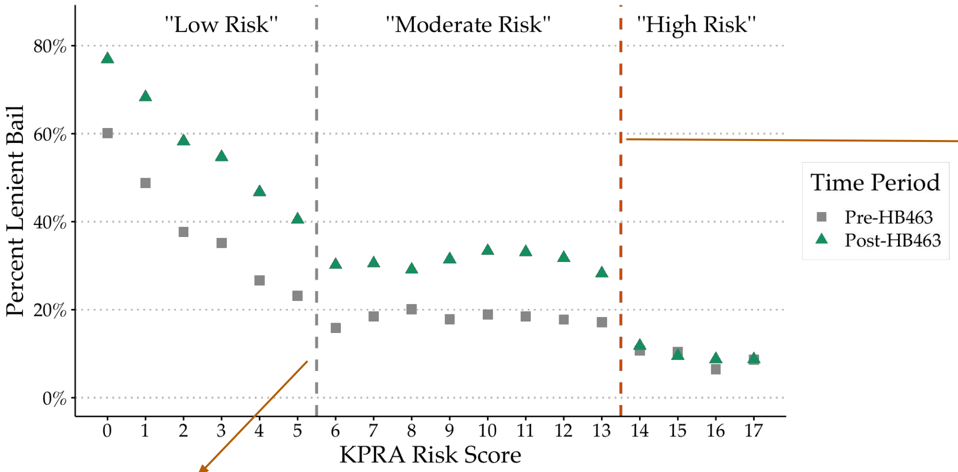
Beforehand, levels consulted in ω cases (in [0,1]) :

<u>Post-period RD:</u>
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]

<u>Pre-period RD:</u>
ω[level eff$_{MH}$] + [prior felony eff]

<u>Diff-in-disc:</u>
[recommendation effect] + **(1-ω)[level eff$_{MH}$]**

Method 1: Estimating ω



Beforehand, levels consulted in ω cases (in [0,1]) :

Post-period RD:
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]

Pre-period RD:
ω[level eff$_{MH}$] + [prior felony eff]

Diff-in-disc:
[recommendation effect] + **(1-ω)[level eff$_{MH}$]**

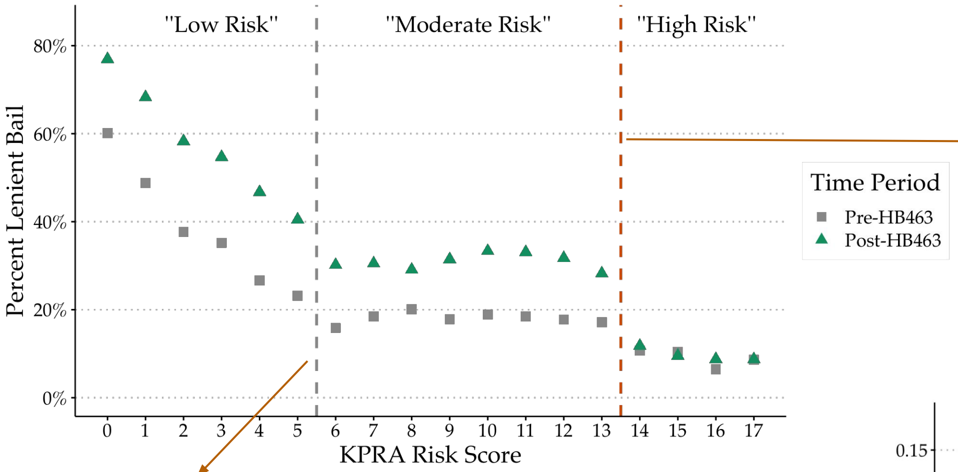
Using the low/moderate discontinuity:

Post-period RD:
[level eff$_{LM}$]

Pre-period RD:
ω[level eff$_{LM}$]

Diff-in-disc:
(1-ω)[level effect$_{LM}$]

Method 1: Estimating ω



Beforehand, levels consulted in ω cases (in [0,1]) :

<u>Post-period RD:</u>
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]

<u>Pre-period RD:</u>
ω[level eff$_{MH}$] + [prior felony eff]

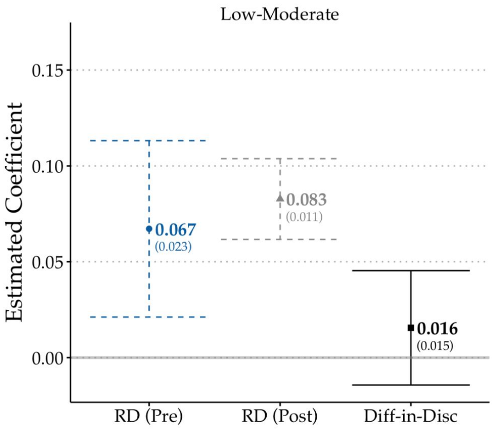
<u>Diff-in-disc:</u>
[recommendation effect] + **(1-ω)[level eff$_{MH}$]**

Using the low/moderate discontinuity:

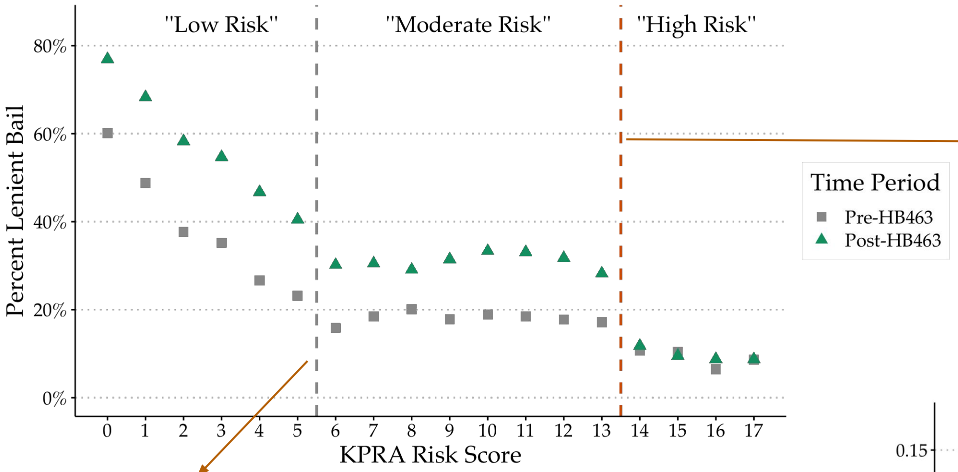
<u>Post-period RD:</u> **8.3pp**
[level eff$_{LM}$]

<u>Pre-period RD:</u> **6.7 pp**
ω[level eff$_{LM}$]

<u>Diff-in-disc:</u> **1.6 pp**
(1-ω)[level effect$_{LM}$]



Method 1: Estimating ω



Beforehand, levels consulted in ω cases (in [0,1]) :

Post-period RD:
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]

Pre-period RD:
ω[level eff$_{MH}$] + [prior felony eff]

Diff-in-disc:
[recommendation effect] + **(1-ω)[level eff$_{MH}$]**

Using the low/moderate discontinuity:

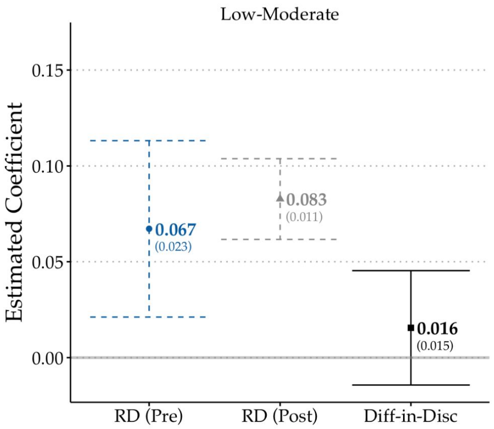
Post-period RD: **8.3pp**
[level eff$_{LM}$]

Pre-period RD: **6.7 pp**
ω[level eff$_{LM}$]

Diff-in-disc: **1.6 pp**
(1-ω)[level effect$_{LM}$]

ω=0.81

Levels consulted in 81% of cases



Method 1: Updating estimates with ω = 0.81

Parameter	Original Estimate ($\omega = 1$)
[recommendation eff] + [level eff_{MH}] + [prior felony eff]	13.7
[level eff_{MH}] + [prior felony eff]	6.1
[recommendation eff]	7.6

Method 1: Updating estimates with ω = 0.81

Parameter	Original Estimate ($\omega = 1$)	Adjusted Estimate ($\omega = 0.81$)
[recommendation eff] + [level eff_{MH}] + [prior felony eff]	13.7	
[level eff_{MH}] + [prior felony eff]	6.1	
[recommendation eff]	7.6	

Method 1: Updating estimates with ω = 0.81

Parameter	Original Estimate ($\omega = 1$)	Adjusted Estimate ($\omega = 0.81$)
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]	13.7	13.7
[level eff$_{MH}$] + [prior felony eff]	6.1	
[recommendation eff]	7.6	

Method 1: Updating estimates with ω = 0.81

Parameter	Original Estimate ($\omega = 1$)	Adjusted Estimate ($\omega = 0.81$)
[recommendation eff] + [level eff$_{MH}$] + [prior felony eff]	13.7	13.7
[level eff$_{MH}$] + [prior felony eff]	6.1	7.5
[recommendation eff]	7.6	6.2

Method 2: Intuitive subsetting

DD estimates **[recommendation effect]** + *(1-ω)[level effect]*

Method 2: Intuitive subsetting

DD estimates **[recommendation effect]** + $\underbrace{\textit{(1-ω)[level effect]}}_{0}$

Focus on cases where risk level does not provide new info, so we think level effect should be close to 0

Method 2: Intuitive subsetting

DD estimates **[recommendation effect]** + $\underbrace{\textit{(1-}\omega\textit{)[level effect]}}_{0}$

Focus on cases where risk level does not provide new info, so we think level effect should be close to 0

→ Misdemeanors + no risk factors / scores of 0: no convictions, no prior FTAs

→ 7% of the data

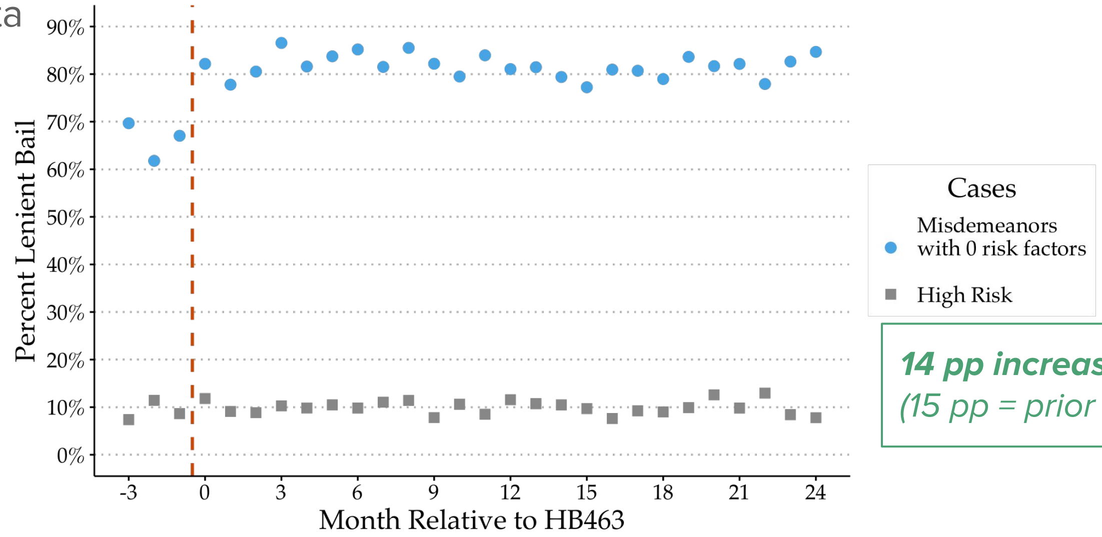
Method 2: Intuitive subsetting

DD estimates **[recommendation effect]** + $\underbrace{\textit{(1-}\omega\textit{)[level effect]}}_{0}$

Focus on cases where risk level does not provide new info, so we think level effect should be close to 0

➔ Misdemeanors + no risk factors / scores of 0: no convictions, no prior FTAs

➔ 7% of the data



14 pp increase in lenient bail

(15 pp = prior DD estimates)

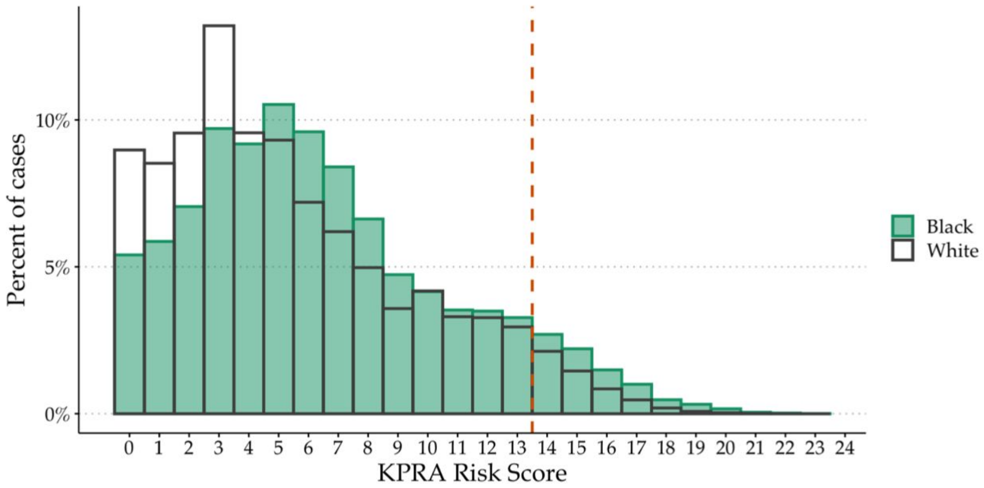
Heterogeneous effects by defendant race

Racial disparities in risk scores, recommendations, and outcomes

Concern that use of algorithms may widen racial disparities

Racial disparities in risk scores, recommendations, and outcomes

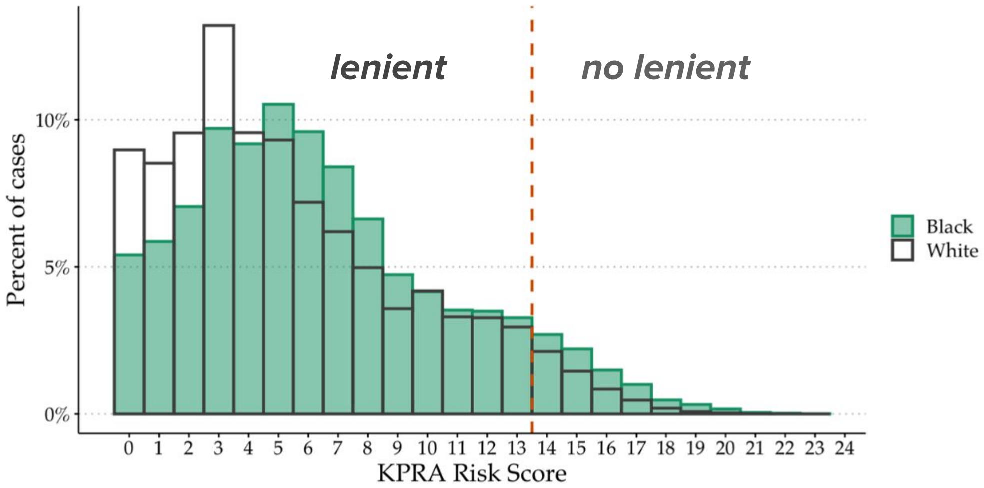
Concern that use of algorithms may widen racial disparities



Differences primarily due to:
FTA, prior felony conviction, prior violent conviction weights

Racial disparities in risk scores, recommendations, and outcomes

Concern that use of algorithms may widen racial disparities

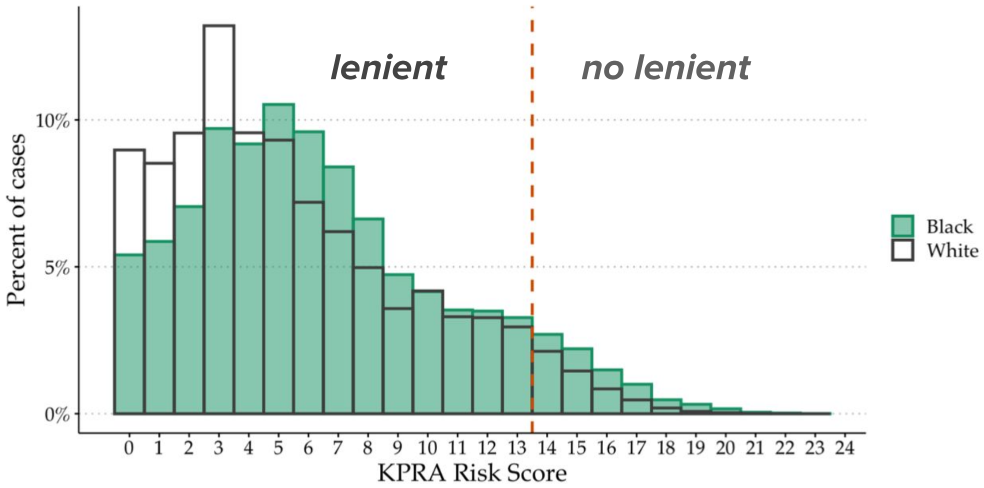


<u>If bail automatically set by recommendations</u> (low/mod => lenient; high => no lenient),

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 that use of algorithms may widen racial disparities



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

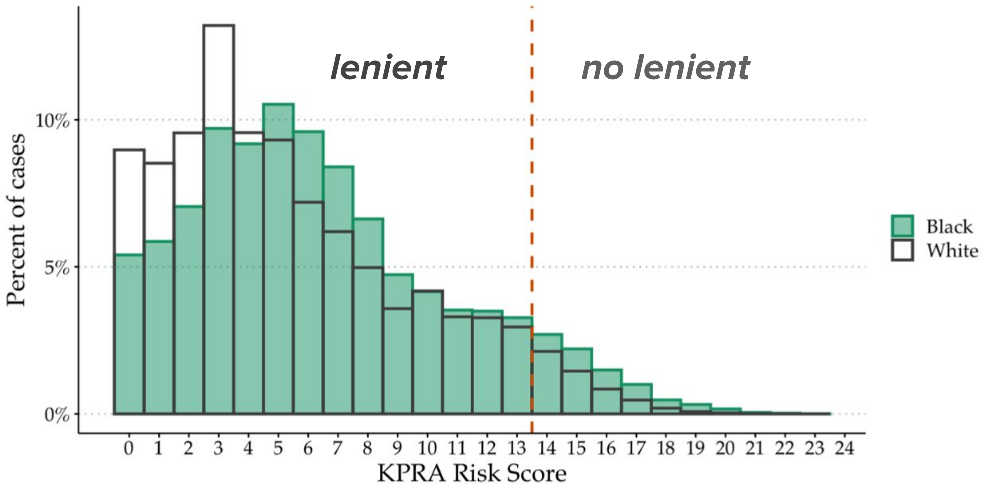
Black people would've been **3.3 pp less likely** to get lenient bail (91.5% vs. 94.8%) than white people

<u>After the recommendations implemented,</u>

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 that use of algorithms may widen racial disparities



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

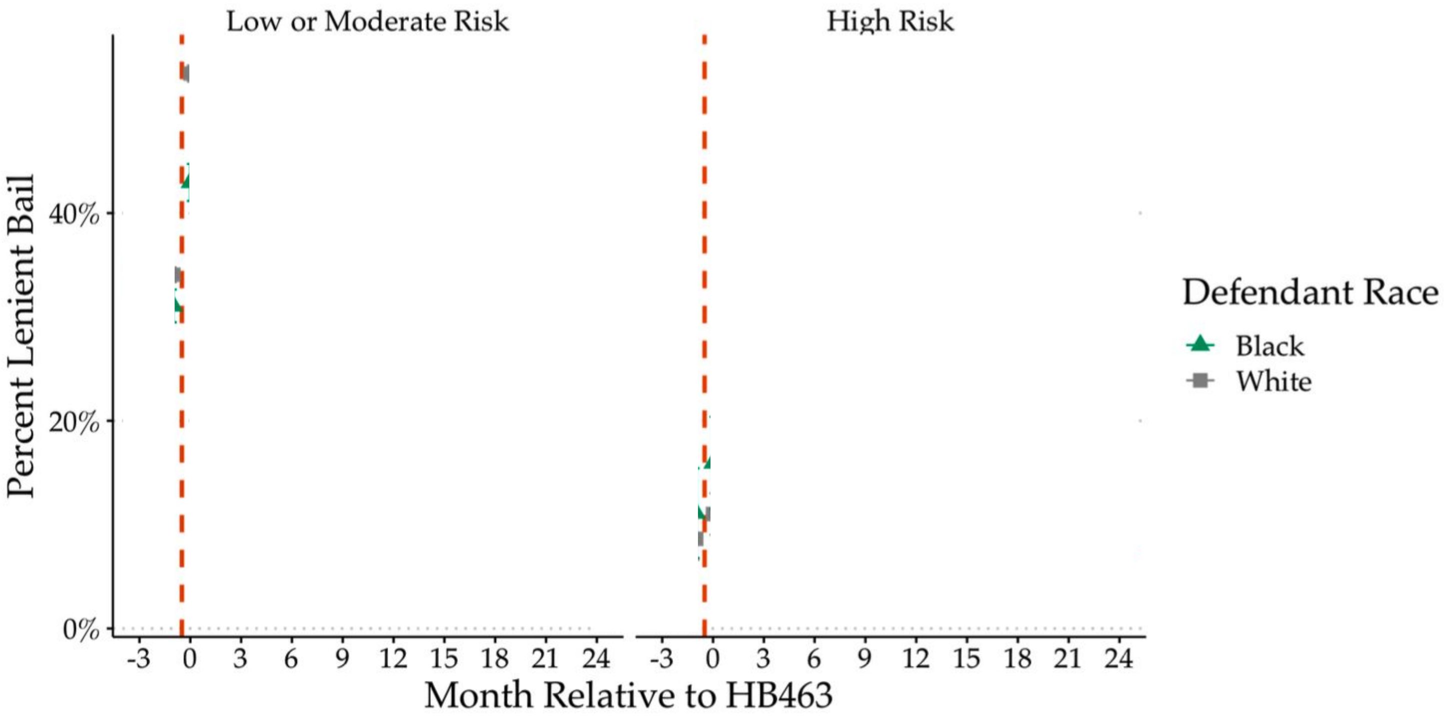
Black people would've been **3.3 pp less likely** to get lenient bail (91.5% vs. 94.8%) than white people

<u>After the recommendations implemented,</u>

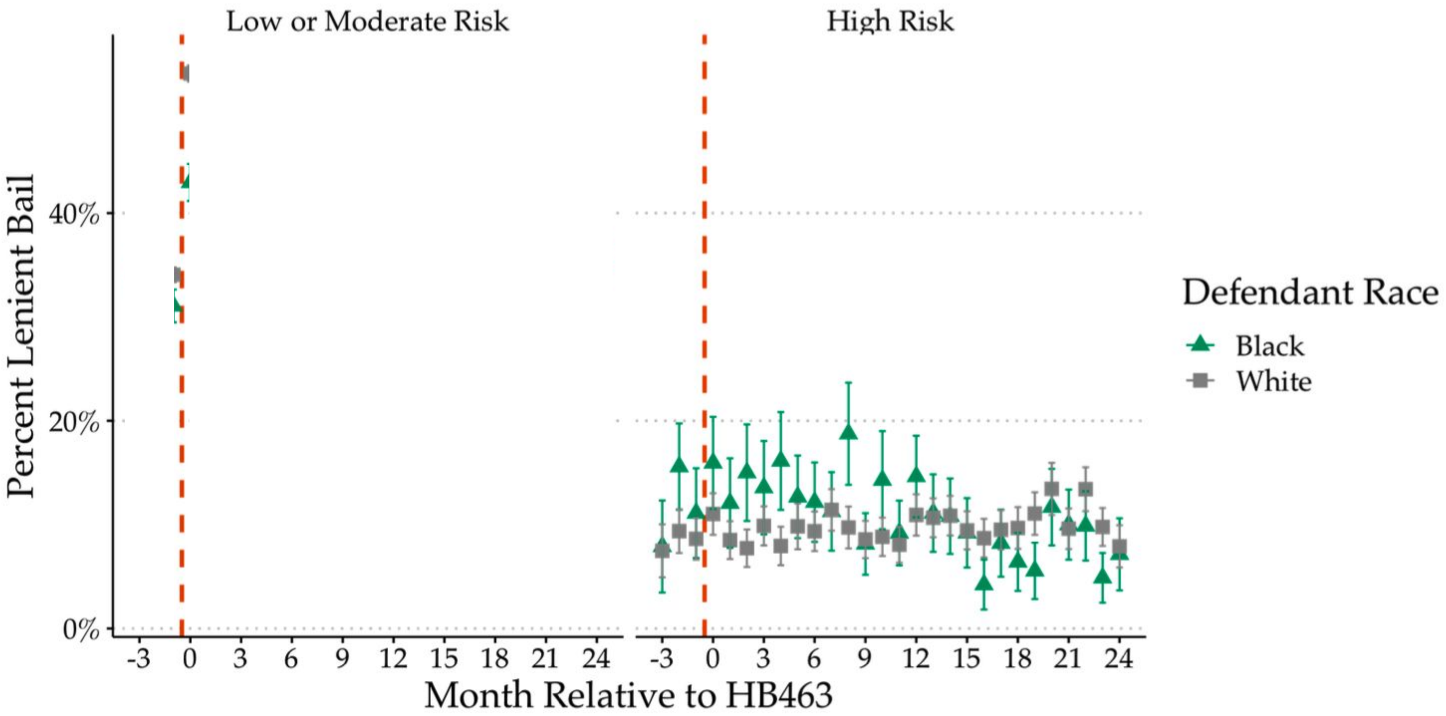
Black people were **9.3 pp less likely** to get lenient bail (36.7% vs. 46%) than white people

suggests: lenient recommendation effects vary by defendant race

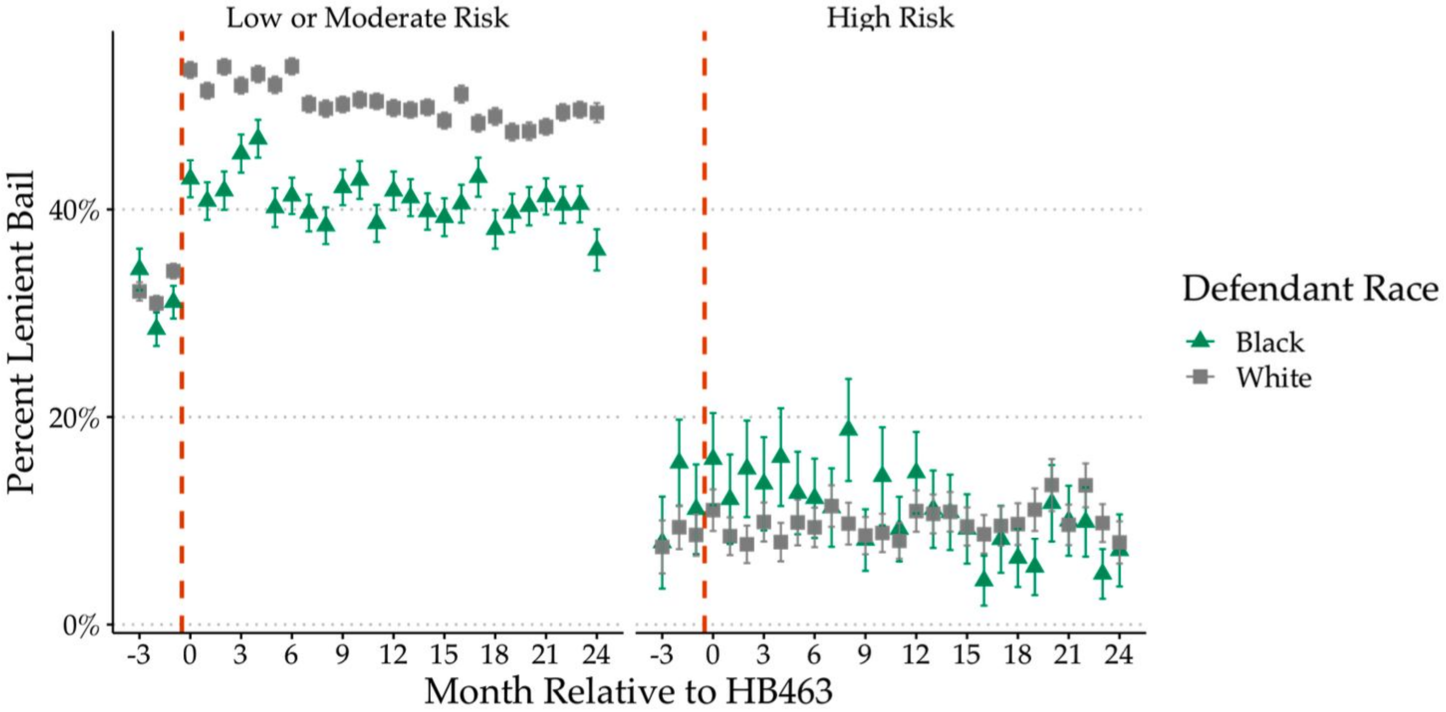
Split the original diff-in-diff approach by defendant race



Split the original diff-in-diff approach by defendant race



Split the original diff-in-diff approach by defendant race

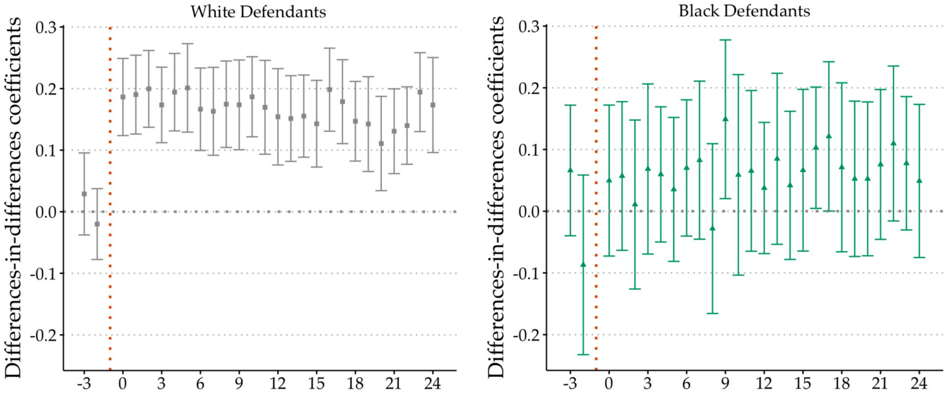


Effects of algorithmic recommendations differ by defendant race

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

Effects of algorithmic recommendations differ by defendant race

	Dependent variable: I(lenient bail)		
	DD *(White)*	DD *(Black)*	
	(1)	(2)	
I(score<14) x Post	0.175***	0.094**	
	(0.021)	(0.037)	



Mean Dep. Var. *(Pre-HB463)*	0.312	0.298	0.310

Effects of algorithmic recommendations differ by defendant race

	Dependent variable: I(lenient bail)		
	DD *(White)*	DD *(Black)*	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.175***	0.094**	
	(0.021)	(0.037)	
Mean Dep. Var. *(Pre-HB463)*	0.312	0.298	0.310

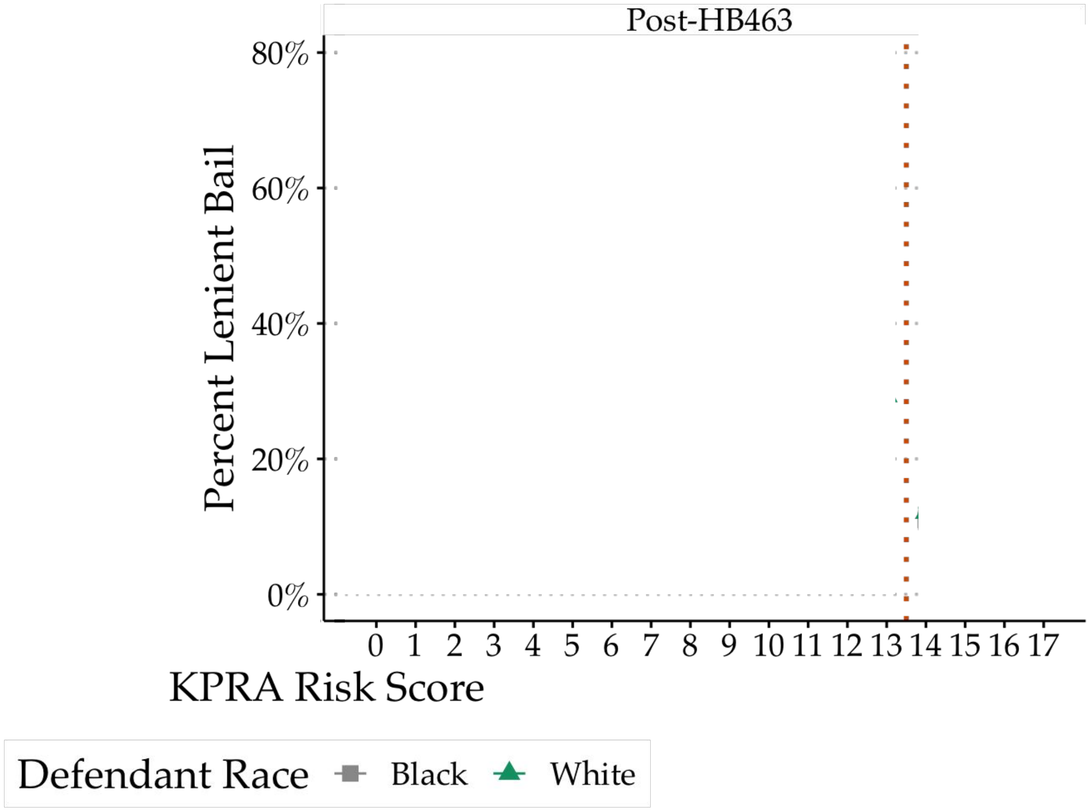
$$lenient_{itj} = \beta_1[I(score_i < 14) \times Post_t] + \beta_2[I(score_i < 14) \times Black_i] +$$

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

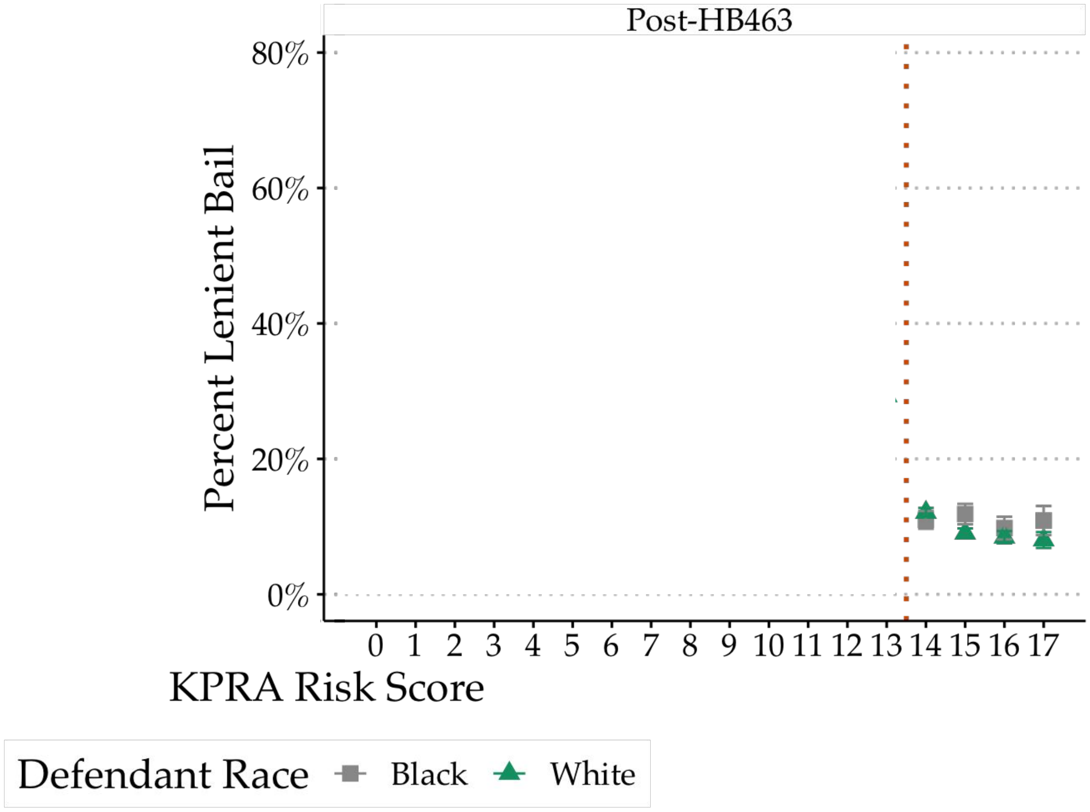
Effects of algorithmic recommendations differ by defendant race

	Dependent variable: I(lenient bail)		
	DD *(White)*	DD *(Black)*	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.175***	0.094**	0.174***
	(0.021)	(0.037)	(0.021)
I(score<14) x Black			0.026
			(0.031)
Post x Black			−0.0004
			(0.033)
I(score<14) x Post x Black			−0.080**
			(0.035)
Mean Dep. Var. *(Pre-HB463)*	0.312	0.298	0.310

Evidence across the risk score distribution

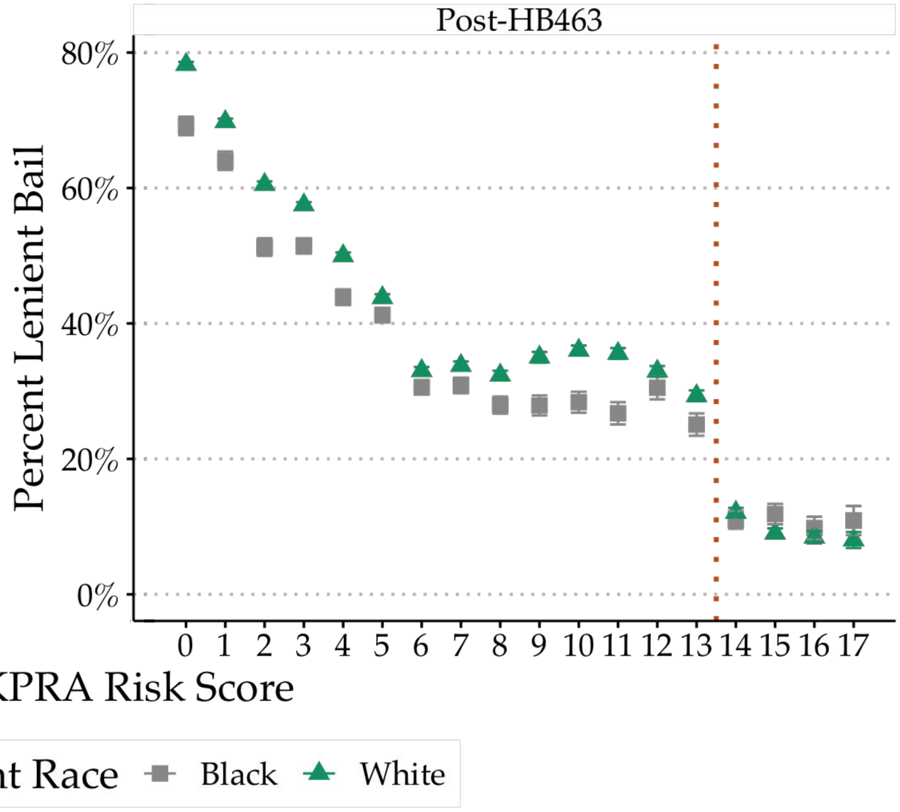


Evidence across the risk score distribution

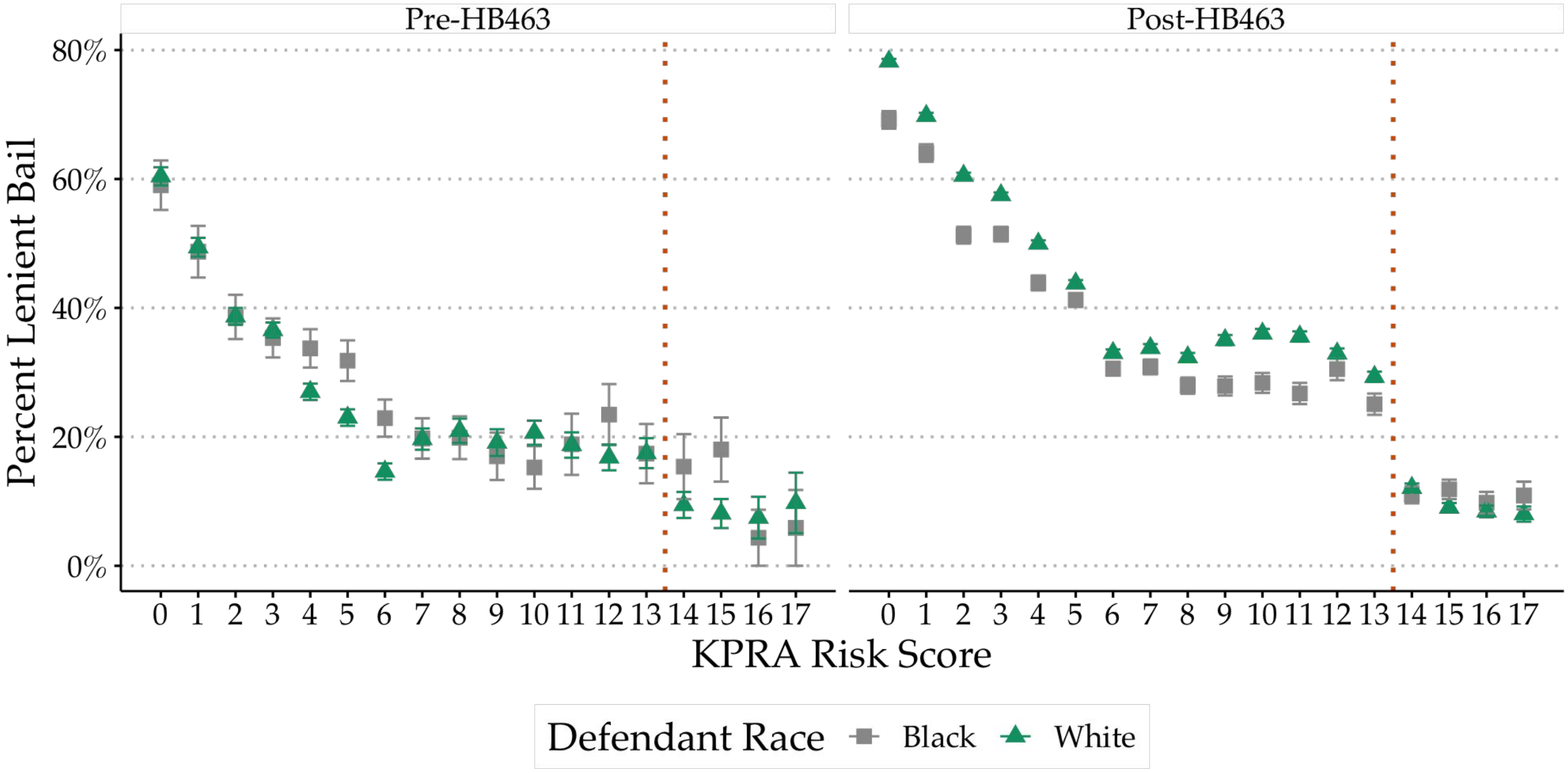


Evidence across the risk score distribution

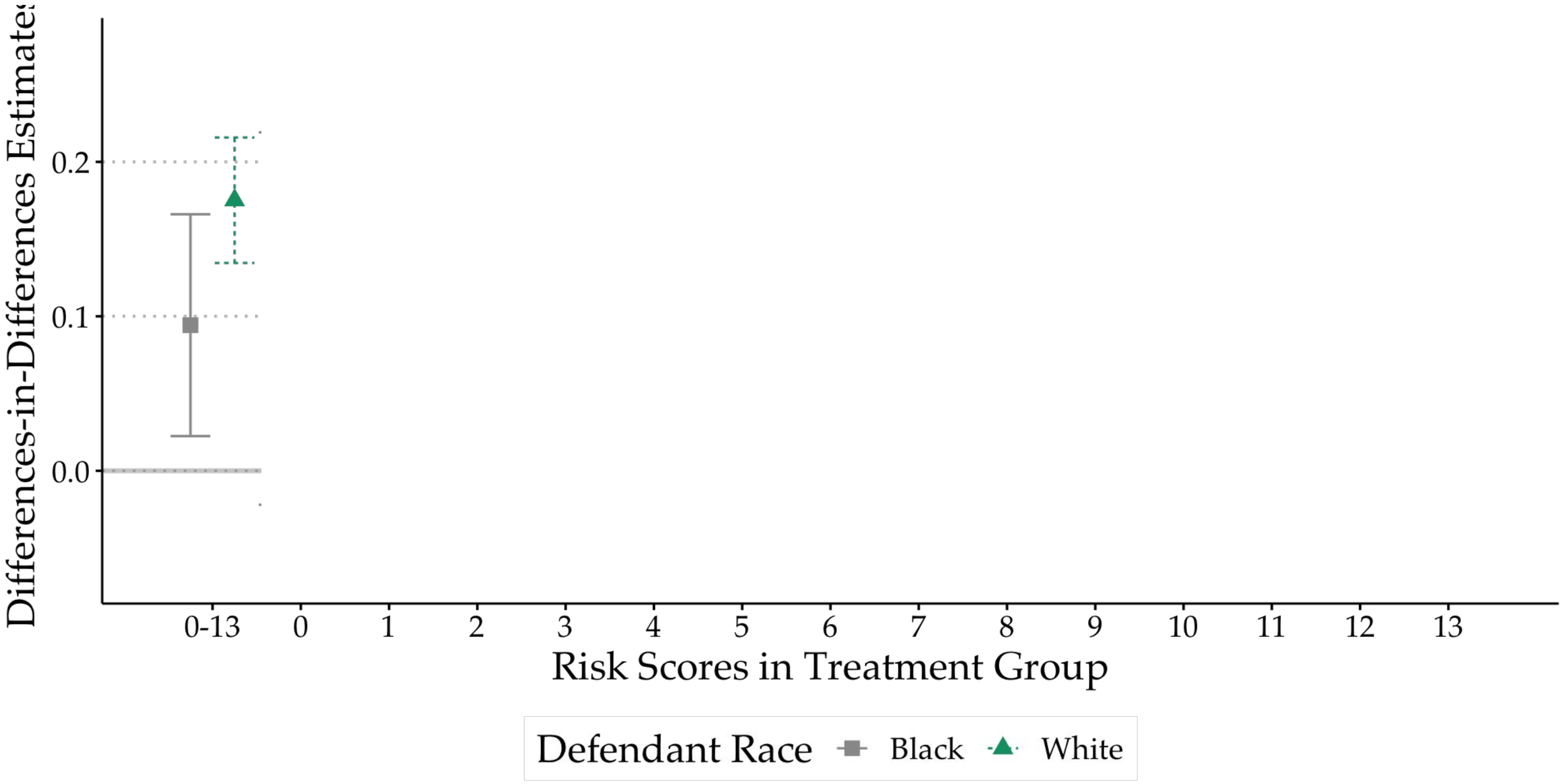
Black defendants are less likely to receive lenient bail than white defendants with identical risk scores



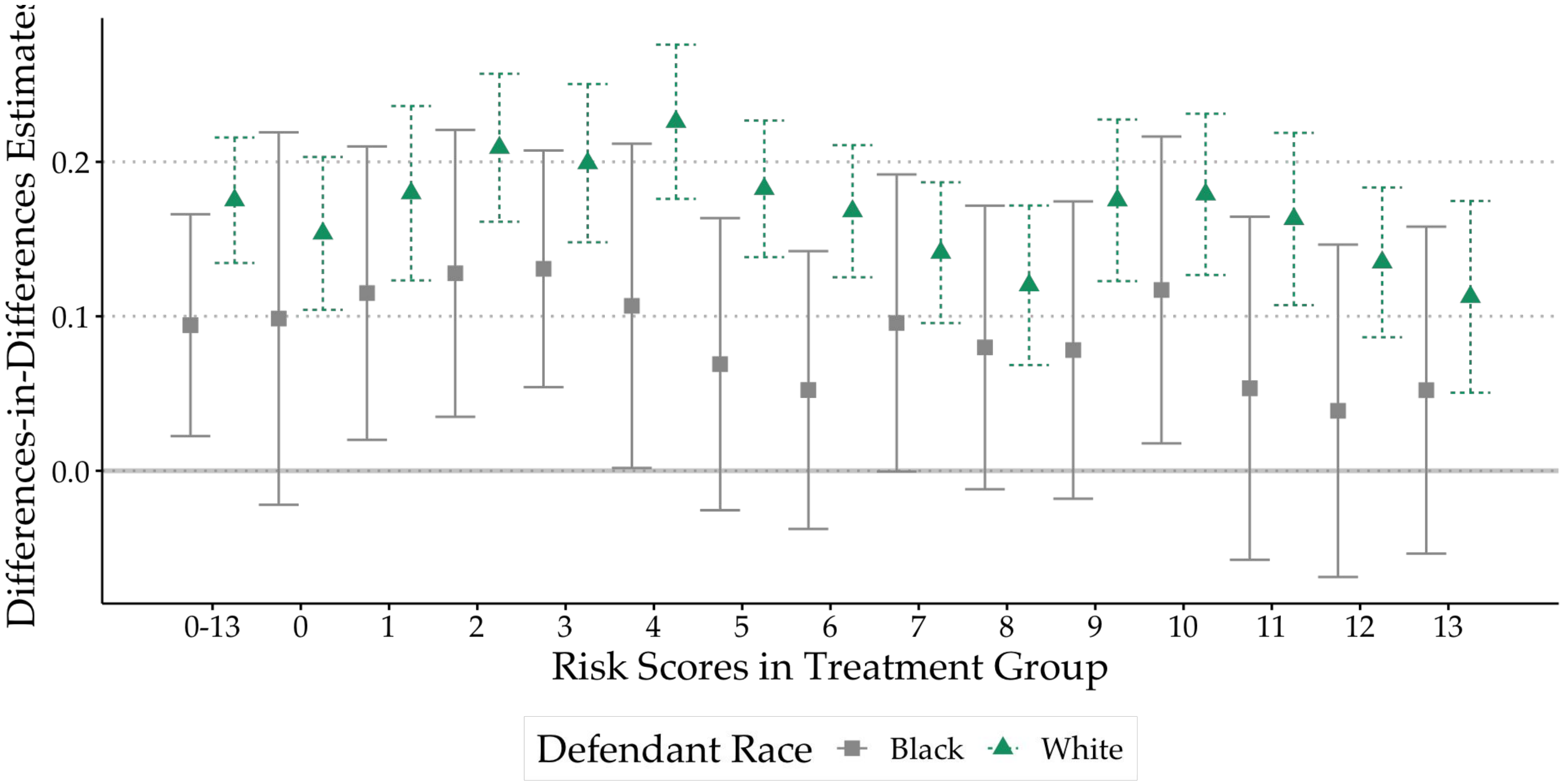
Evidence across the risk score distribution



Effects over the distribution, split by defendant race



Effects over the distribution, split by defendant race



What explains this heterogeneity?

	Dependent variable: I(lenient bail)
	DDD
	(1)
I(score<14) x Post	0.174***
	(0.021)
I(score<14) x Black	0.026
	(0.031)
Post x Black	−0.0004
	(0.033)
I(score<14) x Post x Black	−0.080**
	(0.035)
Mean Dep. Var. *(Pre-HB463)*	0.310
Additional Controls	-

What explains this heterogeneity?

	Dependent variable: I(lenient bail)
	DDD
	(1)
I(score<14) x Post	0.174***
	(0.021)
I(score<14) x Black	0.026
	(0.031)
Post x Black	−0.0004
	(0.033)
I(score<14) x Post x Black	−0.080**
	(0.035)
Mean Dep. Var. *(Pre-HB463)*	0.310
Additional Controls	-

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

What explains this heterogeneity?

	Dependent variable: I(lenient bail)	
	DDD	
	(1)	
I(score<14) x Post	0.174***	
	(0.021)	
I(score<14) x Black	0.026	
	(0.031)	
Post x Black	−0.0004	
	(0.033)	
I(score<14) x Post x Black	−0.080**	
	(0.035)	
Mean Dep. Var. *(Pre-HB463)*	0.310	0.310
Additional Controls	-	*judge-level-time varying FE's*

What explains this heterogeneity?

	Dependent variable: I(lenient bail)	
	DDD	DDD
	(1)	(2)
I(score<14) x Post	0.174***	
	(0.021)	
I(score<14) x Black	0.026	−0.013
	(0.031)	(0.036)
Post x Black	−0.0004	−0.003
	(0.033)	(0.031)
I(score<14) x Post x Black	−0.080**	−0.017
	(0.035)	(0.035)
Mean Dep. Var. *(Pre-HB463)*	0.310	0.310
Additional Controls	-	*judge-level-time varying FE's*

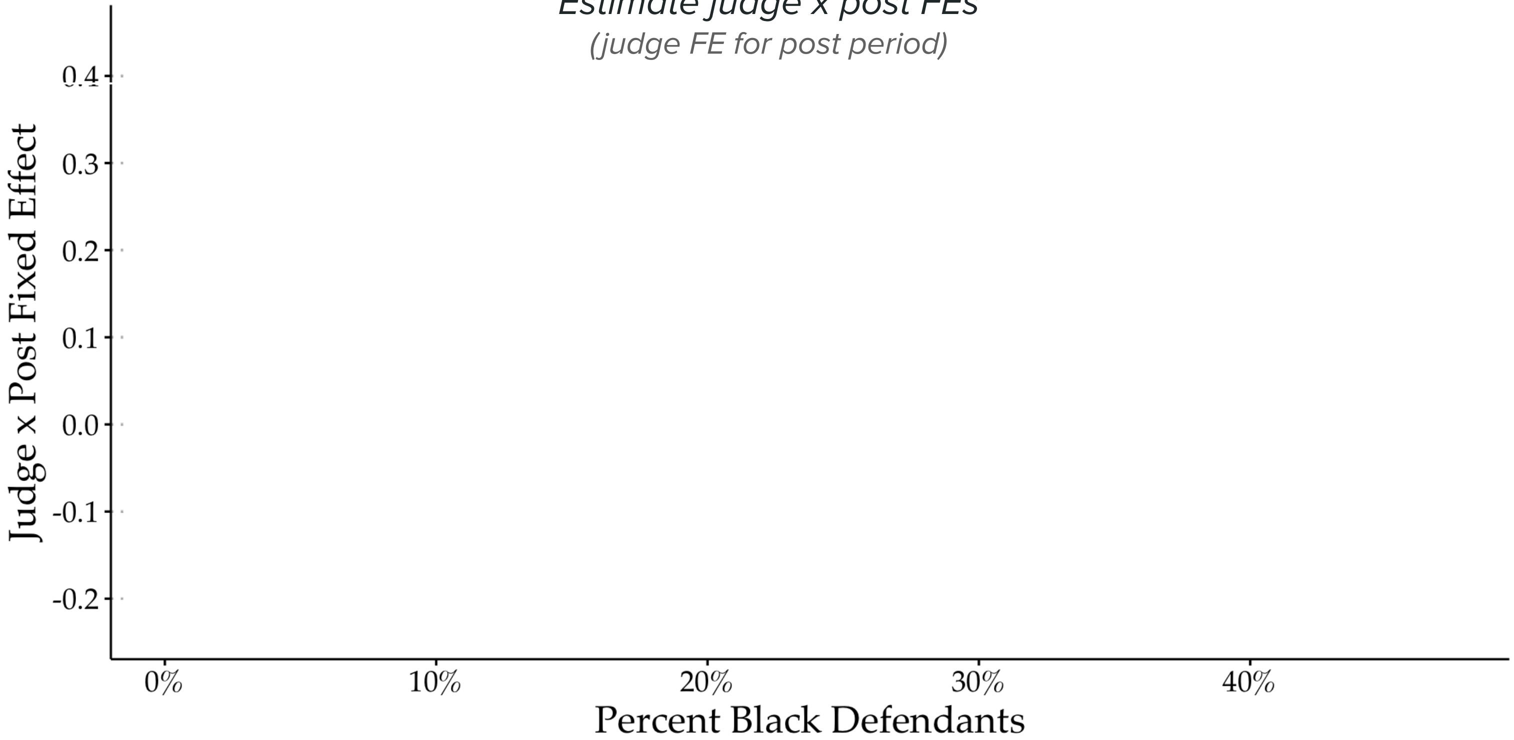
What explains this heterogeneity?

	Dependent variable: I(lenient bail)		
	DDD	DDD	DDD
	(1)	(2)	(3)
I(score<14) x Post	0.174***		
	(0.021)		
I(score<14) x Black	0.026	−0.013	−0.013
	(0.031)	(0.036)	(0.028)
Post x Black	−0.0004	−0.003	0.001
	(0.033)	(0.031)	(0.025)
I(score<14) x Post x Black	−0.080**	−0.017	−0.024
	(0.035)	(0.035)	(0.029)
Mean Dep. Var. *(Pre-HB463)*	0.310	0.310	0.310
Additional Controls	-	*judge-level-time varying FE's*	*county-level-time varying FE's*

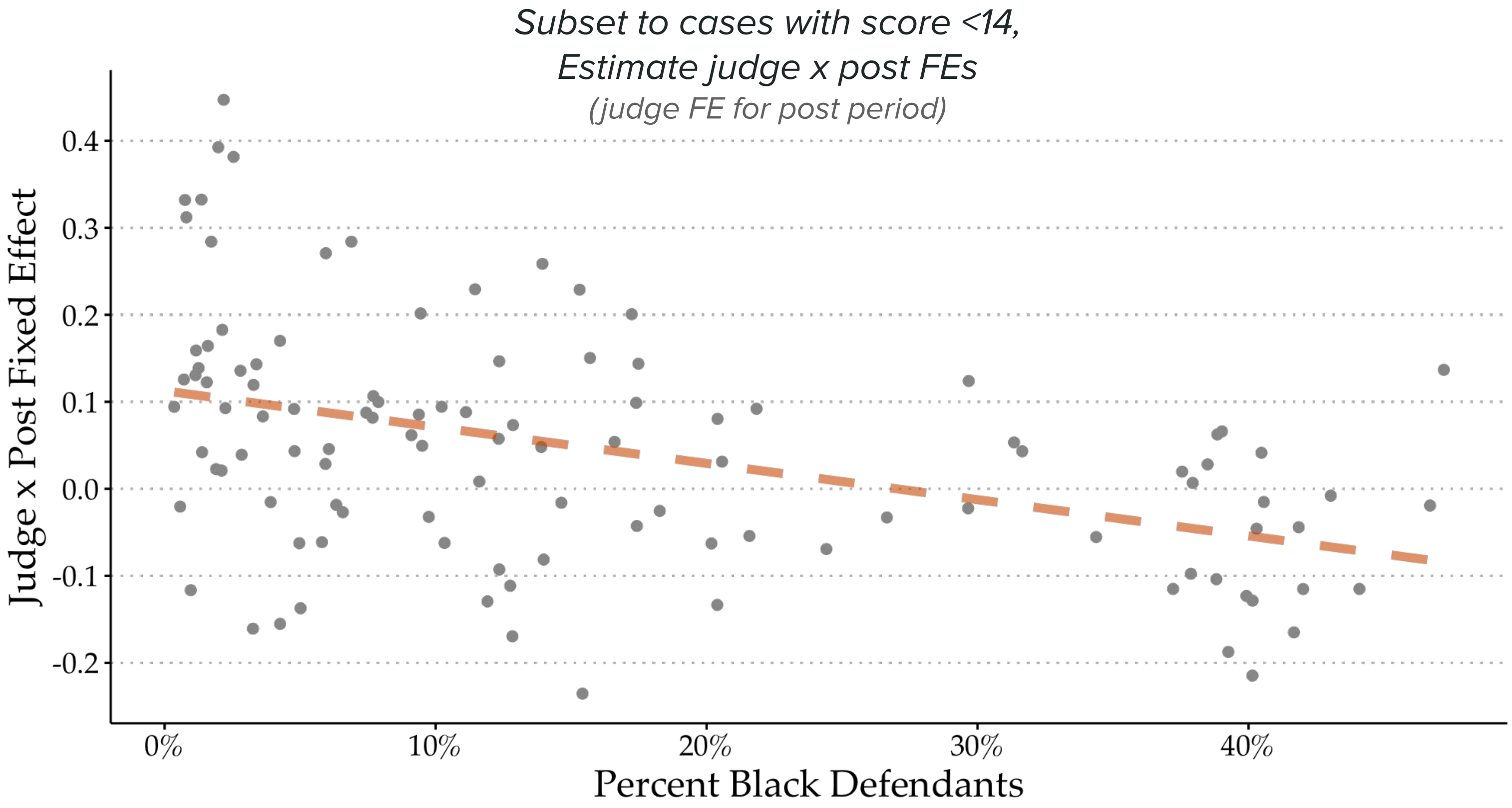
Judges with more Black defendants respond less to lenient recommendations

Judges with more Black defendants respond less to lenient recommendations

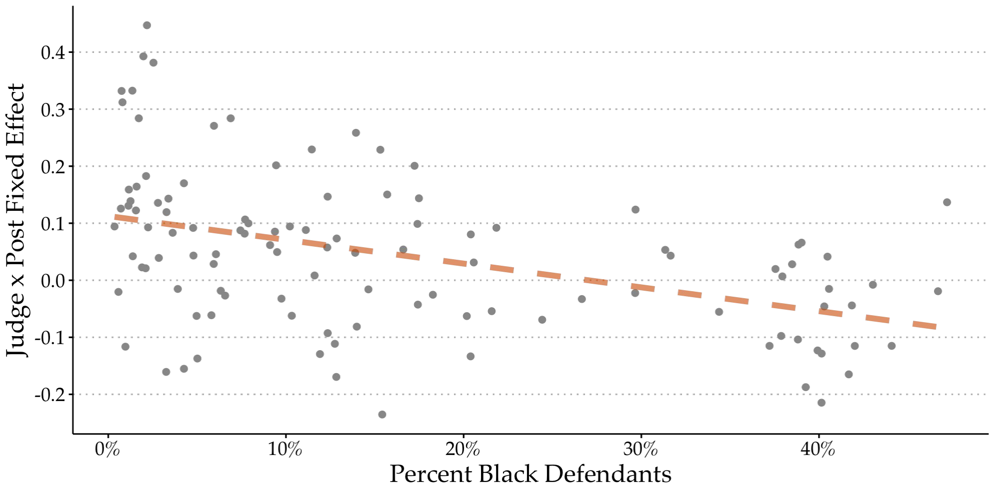
Subset to cases with score <14,
Estimate judge x post FEs
(judge FE for post period)



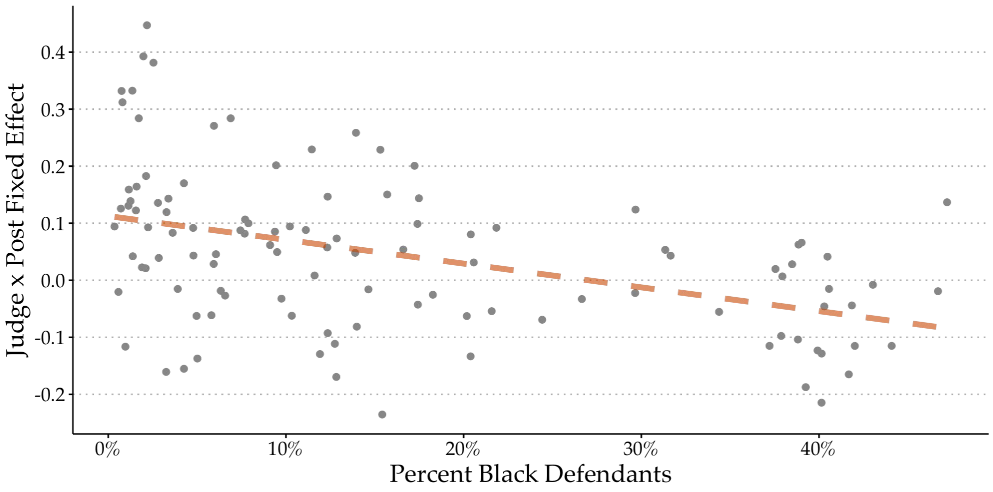
Judges with more Black defendants respond less to lenient recommendations



Why do they respond less?



Why do they respond less?



Could this relationship be explained by...

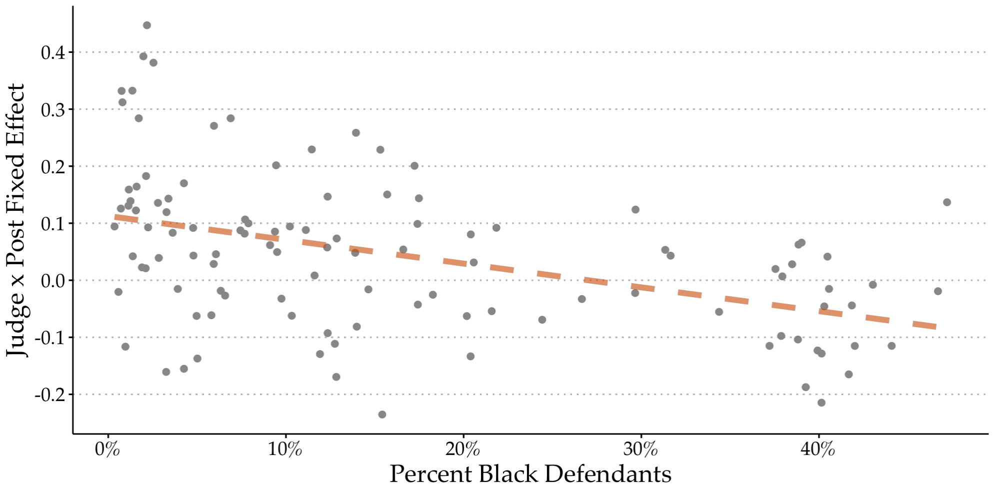
Judge characteristics?

- *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	−0.383***					
	(0.084)					

Judges who see 10 pp more Black defendants respond to the recommendation 3.8 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	−0.383***					
	(0.084)					
		+ 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 3.8 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	−0.383***	−0.391***	−0.378**	−0.320**	−0.275	−0.345*
	(0.084)	(0.088)	(0.151)	(0.157)	(0.178)	(0.187)
		+ 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 3.8 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	−0.383***	−0.391***	−0.378**	−0.320**	−0.275	−0.345*
	(0.084)	(0.088)	(0.151)	(0.157)	(0.178)	(0.187)
		+ 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 3.8 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

Conclusion

Summary of key results

1. **Algorithmic recommendations are common + they have independent effects on human decisions**
    - *Setting algorithmic recommendations =/= solving a prediction problem*
    - *Lenient recommendations increase lenient bail by 50%*

2. Why? **Recommendations can change private costs of errors**
    - *Making mistakes is less costly when decision consistent with recommendation (lenient recommendations provide "cover" for judges)*
    - *Algorithms can impact: - decision-maker incentives (rather than just predictions)*
      - composition of decisions (rather than just allocation)

3. Heterogeneity: **Recommendations can have unintended effects**
    - *Judges deviate from lenient recommendation more for Black defendants than for white defendants with the same algorithmic risk*

\end{talk}

Thanks for coming!

Email:

alex@albrightalex.com

Website:

albrightalex.com