

THE IMPACT OF ALGORITHMIC RISK ASSESSMENT TOOL LEGISLATION
ON RACIAL DISPARITIES IN CRIMINAL SENTENCING

by

Hannah Clare Brauch

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Applied Economics

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2023

©COPYRIGHT

by

Hannah Clare Brauch

2023

All Rights Reserved

ACKNOWLEDGEMENTS

I am deeply grateful to my thesis advisor, Dr. Wendy Stock, for her guidance and support throughout my entire undergraduate and graduate career. I would also like to extend my gratitude to Dr. Stoddard and Dr. Gilpin for their invaluable input. It has been an honor to work with the three of you.

I could not have done this without the immense support of my cohort. I love each and every one of you dearly. I am so grateful for every moment I spent learning and laughing with our little motley crew.

TABLE OF CONTENTS

1. INTRODUCTION	1
2. BACKGROUND AND PREVIOUS WORKS	4
Legislation on Sentencing Guidelines.....	4
Evidence of Racial Disparities in Sentencing.....	5
Algorithms	6
RAT Laws.....	8
RAT Intake Surveys.....	9
Impact of RATs on Racial Disparities in Sentencing.....	10
3. DATA	16
Final Dataset	17
Data Limitations	19
4. EMPIRICAL METHODOLOGY	23
Baseline Two-Stage Differences-in-Differences	23
Research Design	23
Assumptions	24
Heterogeneity by Law Characteristic.....	26
Heterogeneity by Race	27
Event Study	27
Assumptions	28
5. RESULTS	29
6. ROBUSTNESS CHECKS	36
Control Group Selection.....	36
Law Group Selection.....	36
Drug Crime Regressions	36
7. CONCLUSION	41
REFERENCES CITED.....	42
APPENDIX: Law Data Sources	47

ABSTRACT

The prevailing presence of racial disparities in criminal sentencing motivated the introduction of algorithmic risk assessment tools (RATs) in the U.S. judicial system. These tools provide judges with an algorithm-generated risk score and sentencing recommendation to consider in their decisions. Although this technology is well-intentioned, researchers find that RATs produce racial disparities in their outputs. My research examines the impact of state laws regulating the use of RATs on racial disparities in sentence length and likelihood of receiving probation. Utilizing Gardner’s (2021) two-stage differences-in-differences methodology, I exploit the natural experiment arising from 29 states passing some form of RAT law at different times. I find that the impact of RAT laws depends on the components of the state’s RAT law, and that the effect varies by racial group. My results suggest that RAT laws significantly decrease the racial sentencing disparity for Hispanics, but increase the disparity for Blacks. Although my results are somewhat sensitive to specification, they still bear critical policy implications regarding the use of RATs in the judicial system.

INTRODUCTION

Racial disparities and discrimination are pervasive in the United States judicial system (Alesina and Ferrara, 2014; Depew et al., 2017; Leslie and Pope, 2017). Blacks receive 20% longer sentences than Whites for similar crimes, and are 7.5 times more likely to be imprisoned than Whites (Bushway and Piehl, 2001; Mitchell 2004). Similarly, the Bureau of Justice Statistics reports that Hispanics are 3.1 times more likely to be incarcerated than Whites (United States, 2018). Although disparity does not necessarily indicate discrimination, the U.S. Department of Justice cites that “disparity implies that many offenders are not being sentenced in accordance with legally relevant factors” (Jareborg, 1987). Theoretically, technology could be used to address this issue, since it only considers “legally relevant factors.” This motivated the introduction of algorithmic risk assessment tools (RATs) into the judicial system to alleviate racial disparities in sentencing. However, this technology is shown to produce its own racial disparities in outcomes (Larson et al., 2016; Stevenson and Doleac, 2021). These findings illustrate the important distinction between RATs *affecting* racial disparities in sentencing and RATs *being embedded* with racial disparities. The question remains whether or not RATs, despite their flaws, can decrease racial disparities in sentencing. My research investigates the impact of RAT legislation on racial disparities in criminal sentencing.

Algorithms are extremely powerful technological tools that allow for nearly instantaneous analysis of massive datasets. They are increasingly used in online advertising, social media feeds, and financial markets. Using this technology, RATs were developed to objectively measure defendants’ risk of reoffending based on their background, history, and other socioeconomic factors. Their use has been promoted as a way to eliminate racial

disparities in the judicial system.¹ RATs produce several reports designed to aid judges’ decisions, including a risk score, sentence recommendation, leniency suggestion, bail and parole recommendations, and whether or not to detain the defendant pretrial.

In a process called machine learning, RAT algorithms are developed using data from past judicial records and recidivism rates, allowing them to recognize patterns in recidivism and sentencing and apply them to current cases.² However, research has demonstrated racial disparities in historical judicial decisions, with the implication that RATs incorporate these disparities into current sentencing decisions (Mayson, 2019). Starting in 2007, states began passing legislation regulating the use of RATs in the judicial system.³ As of 2023, 18 states have laws that require the use of RATs and mandate empirical validation of the unbiasedness of the tool.⁴ Table 1 lists each state with a RAT law, the year the law became effective, and the characteristics of the law. I exploit the natural experiment created by variation in the timing and characteristics of RAT laws to evaluate their impact on racial disparities in sentencing. Although previous literature has examined the effect of RATs on racial disparities, this paper is among the first to provide causal evidence of the effect of RAT *legislation* on racial disparities in sentencing. I utilize two-stage differences-in-differences (2sDiD) and event study methodologies to examine the impact of RAT legislation. I find that the effects of RAT laws depend on the particular characteristics of the law that are present, and the effects vary for Blacks and Hispanics. The results suggest that RAT laws requiring judges to use a validated RAT significantly decrease racial disparities

¹Although many RAT producers claim they will also help decrease incarceration, researchers find the opposite effect (Stevenson and Doleac, 2021). A potential mechanism for this is that RATs might deem a defendant ineligible for probation. Being deemed ineligible for alternative treatment during sentencing often turns into incarceration, possibly explaining the observed effect.

²Machine learning teaches an algorithm, or machine, to extract patterns from a “training dataset” and apply them to a new dataset. The machine is then able to make predictions for the new dataset without being explicitly programmed to do so. In the context of RATs, the algorithms learn how to determine risk and sentence defendants based on outcomes for similar defendants in the past. “What is Machine Learning?”)

³Virginia was the first state to implement a RAT law, in 2007.

⁴As of 2023, 29 states have some form of legislation regulating the use of RATs.

in sentencing for Hispanics, but increase disparities for Blacks. These findings have crucial policy implications for how legislation regulating the use of RATs can potentially reduce racial disparities in sentencing.

The remainder of the paper proceeds as follows. Section 2 provides a background on how algorithms are developed and a review of the literature. Section 3 describes the dataset. Section 4 describes the empirical methodology. Section 5 reports the results and Section 6 discusses various robustness checks. Section 7 concludes the paper.

BACKGROUND AND PREVIOUS WORKS

This section discusses how racial disparities in sentencing led to federal sentencing guidelines and eventually motivated the introduction of RATs in the judicial system. I describe how RAT algorithms are developed, why they have the potential to produce biased recommendations, the legislation regulating their use, and research demonstrating how RATs impact racial disparities in sentencing.

Legislation on Sentencing Guidelines

The first legislation regulating sentencing guidelines was the Sentencing Reform Act (SRA) of 1984.¹ The sentencing guidelines are based on offense severity and criminal history, rather than socioeconomic or demographic traits. However, judges can choose to deviate from the SRA guidelines, and research shows that judges are more likely to depart when sentencing Blacks, males, and individuals with low levels of education (Mustard, 2001).² Mustard (2001) finds that 55% of the Black-White sentence gap arises from judges departing from the federal guidelines, as opposed to judges sentencing differently within the guidelines.

Yang (2015) finds that Black defendants experienced a 4% increase in average sentence length compared to similar White defendants after the *United States v. Booker* Supreme Court case. This case withdrew a judge’s ability to enhance sentences using evidence not reviewed by the jury, which was found to be a violation of the Sixth Amendment (“United States v. Booker”).³ Fischman and Schanzenbach (2012) also find that racial disparities in sentencing increased after the *Booker* decision. This case restricted judges’ discretion in

¹The Sentencing Reform Act of 1984 created the United States Sentencing Commission (USSC), an independent body in the judicial system tasked with promoting federal sentencing guidelines (United States Sentencing Commission).

²The SRA allowed for federal appeals courts to review lower courts’ sentences that exceed federal guideline maximums.

³This ruling does not currently apply to the use of RAT recommendations to enhance sentences.

sentencing, and in turn increased the relevance of mandatory minimums. These minimums have a disproportionate impact on minority defendants, explaining the increase in racial disparities.

Evidence of Racial Disparities in Sentencing

Racial disparities in criminal case outcomes can manifest at several points in the judicial process. Researchers observe racial disparities in whether defendants face charges, are detained pretrial, and their options during the trial. In addition, several papers (described below) explain how unrelated events like judicial elections or college football games can produce racial disparities in sentencing.

Alesina and Ferrara (2014) test for racial bias in sentencing using patterns in capital sentence reversals, which is when a defendant no longer receives the death penalty due to a “prejudicial” error in court proceedings.⁴ In other words, the error “substantially undermines the reliability of the guilt finding or death sentence imposed at trial (Alesina and Ferrara, 2014). The results show 3-9 percentage point larger rates of reversals for minority defendants with White victims relative to those with non-White victims. The higher rate of reversals for minorities with White victims indicates a higher error rate in those cases, implying they were sentenced to the death penalty for a crime that did not warrant it. The researchers emphasize that these results can only be interpreted as bias against minorities with White victims, not all minorities.

Whether or not a defendant is detained before their trial can have a large impact on other judicial outcomes. Leslie and Pope (2017) show that being detained leads to a 13 percentage point increase in likelihood of conviction for felony offenses. In addition, because pretrial detention induces individuals to plead guilty when they would have pleaded not

⁴An error in court proceedings is considered “prejudicial” if there is a reasonable probability that the outcome would have been different had the error not occurred.

guilty had they been released, disparities in pretrial detention can exacerbate disparities in sentencing. Dobbie et al. (2018) find that pretrial detention increases the probability of conviction by 14 percentage points, with the main mechanism being an increase in guilty pleas. Minorities face higher pretrial detention rates, explaining 40% of the Black-White discrepancy in being sentenced to prison (Dobbie et al., 2018).

Researchers have also found that unrelated outside events impact court decisions and racial disparities in sentencing. Park (2017) finds that judges change their sentencing behavior towards minorities over the course of the judicial election cycle, reporting that incarceration rates increase by 2.4 percentage points in the last six months of the election cycle, but only for Black defendants. Park posits that judges' sentencing behaviors towards minorities change over the course of the judicial election cycle as concerns for reelection increase nearing the election date.

Further, Eren and Mocan (2018) show that unexpected outcomes in prominent colleges' football games can serve as emotional shocks that influence judges' decision-making. They find that unexpected football game losses increase sentence lengths by 13% if the judge earned their bachelor's degree from the losing college, with effects disproportionately felt by Black defendants. Phillipe and Ouss (2018) find that after media coverage of unrelated criminal justice events, sentence lengths increase by 2.4%. This increase occurs due to media exposure to crime, not crime itself.

Algorithms

In a process called machine learning, algorithms build models based on a sample dataset called "training data." The algorithm detects patterns in the training data and, without being explicitly programmed to do so, uses them to make predictions for new data. In the context of criminal sentencing, the algorithm learns how to predict a defendant's risk of recidivism and suggests a sentencing outcome based on outcomes for similar defendants in the past.

In theory, RAT technology has the potential to reduce, if not eliminate, racial disparities in sentencing outcomes by lowering judges’ ability to be discriminatory in their assessments of risk. At the same time, the training data used to develop RAT algorithms come from past judicial rulings, which have been proven to be racially biased (Alesina and Ferrara, 2014; Anwar et al., 2012; Depew et al., 2017). RATs are also developed using data on recidivism rates. If Blacks are historically more likely to be arrested, convicted, receive longer sentences, and have higher recidivism rates than Whites with otherwise similar characteristics, the training data used to develop a RAT will link factors correlated with being Black to higher risk of recidivism.⁵ In turn, the RAT will suggest longer sentences to individuals with those characteristics, who are disproportionately Black. As a result, the risk scores and sentencing recommendations produced by a RAT can perpetuate historical disparities.⁶

Opinions diverge regarding whether or not race should be included in RAT algorithms.⁷ Skeem and Lowenkamp (2020) and Cabrerros et al. (2022) maintain that including race, rather than “blinding” its effects, can improve prediction and minimize error rates. Even when defendant race and ethnicity are excluded from the training data, Arnold et al. (2021) find that algorithms recommend releasing White defendants pretrial 11% more often than their Black counterparts.⁸

⁵Skeem et al. (2020) study the interaction between risk assessment and wealth, which is highly correlated with race. They report that risk assessments reduce the probability of incarceration for relatively affluent individuals, but increase the likelihood of incarceration for relatively poor defendants. Whereas poor defendants were previously convicted less due to “mitigated perceived blameworthiness,” algorithms have transformed them into individuals with higher perceived risk.

⁶This is particularly important since Hartmann and Wenzelburger (2021) show that judges are apt to conform to the RAT’s recommendation.

⁷Public opinions diverge regarding a RAT’s ability to be unbiased. Scurich and Krauss (2020) gauge local public awareness of state government considering the adoption of an algorithmic approach. When polling public opinions on whether algorithms would worsen racial disparities in the criminal justice system, 71% of those who were not undecided believed it would “probably” or “definitely” increase disparities. The authors offer a potential mechanism for this response as a general aversion to algorithms, as human behavior is perceiving as being too complex to be captured by a machine.

⁸Rambachan et al. (2021) argue that equity preferences are not reason enough to alter how information is extracted, claiming that equity only matters in regards to how information is applied to decision-making. The authors argue that optimal regulation would involve algorithmic transparency, but no restriction on algorithmic development practices.

The public has also expressed an ethical dilemma in terms of the use of RATs in courts. Mayson (2019) studies the notion of “bias in, bias out” that pervades the use of algorithms and their tendency to generate disparate results. Mayson argues that, “in a racially stratified world, any method of prediction will project the inequalities of the past into the future.” Schwerzmann (2021) also argues against prediction in sentencing, because it makes an individual’s present life “inescapably doomed by its past.” She claims that the recursive process of prediction suppresses outliers and alters its predictions to match the forecasts. In this algorithmic framework, she argues, justice takes the form of maintaining a social homeostasis that systematically discriminates against minorities.

RAT Laws

Legislation regarding the use of RATs varies across state and time. Table 1 lists each state with a RAT law, the year the law became effective, and the characteristics of the law.⁹ Virginia passed the first RAT law in 2007, but most RAT laws were passed between 2017-2021. As of 2023, 29 states have some form of legislation regulating the use of RATs. Although RAT legislation varies across states, they regulate two fundamental components of the algorithmic tool.

First, some laws *require* judges to consider the algorithm’s suggestion when making sentencing decisions. Other policies allow (but do not *require*) the use of RATs. These policies are denoted as *May Use* in Table 1.¹⁰ The *require* component of RAT laws is motivated by the notion that mandating judges to at least consider a RAT’s recommendations might make their sentencing behaviors more consistent.

⁹Figure 1 reports the same information on a map. We can see somewhat of a geographic pattern in states’ implementation of RAT laws, with states in the middle of the country being less likely to pass RAT laws. This raises concerns regarding endogeneity of RAT laws.

¹⁰No state law requires the judge to completely adhere to the RAT’s sentencing recommendations, though research finds that judges tend to conform to the tool’s suggestion (Hartmann and Wenzelburger, 2021).

Second, many RAT laws mandate empirical studies to test the tool’s predictive validity and unbiasedness in terms of race, sex, and other protected classes.¹¹ The *validate* component of RAT laws is motivated by concerns that a biased algorithmic tool has the potential to perpetuate past racial disparities in judicial decisions.¹² Despite the potential for RATs to perpetuate historical racial disparities, Idaho is the only state that requires companies or courts to disclose the training data used in the formation of the algorithm. Idaho’s law also gives defendants the right to see how the RAT data is being used in their sentencing recommendations.¹³

RAT Intake Surveys

When a defendant is arrested in a judicial system that uses RATs, they complete a lengthy intake survey. Their answers are used to determine the RAT’s recommendations.¹⁴ The nature of some of the survey questions raises concerns regarding the objectivity of the RAT, and how this information is used to measure the defendant’s risk of recidivism. For example, in addition to basic questions regarding employment or criminal history, the survey may contain questions like, “How many of your friends/acquaintances are taking drugs illegally?”, “How often do you feel bored?”, or “Agree or disagree with the statement: A hungry person has a right to steal.” Some of the survey questions are prefaced with the statement “There are no ‘right or wrong’ answers,” but every answer is used in the RAT’s risk calculation. Figure 2 shows a sample page of the COMPAS survey.

¹¹Validation studies test whether a RAT’s estimated risk for an individual corresponds to their actual criminal behavior (“Risk Validation”).

¹²Some states combat this cycle by designating that a RAT’s results cannot be the only consideration in a judge’s sentencing decision.

¹³Idaho passed this transparency law in 2019.

¹⁴The widely used COMPAS survey is eight pages long with 137 questions (Angwin and Larson, 2016).

Impact of RATs on Racial Disparities in Sentencing

RATs have major implications for criminal justice. If a RAT is wrong in one way, someone receives an overly-harsh sentence. If a RAT is wrong in another way, a dangerous criminal walks free. Larson et al. (2016) compare RAT-generated risk scores to defendants' actual recidivist behavior. Figure 3 shows the distribution of risk scores they find for Whites and Blacks (Larson et al., 2016). Black defendants' risk scores were relatively uniformly distributed from low risk to high risk, but the risk scores for Whites were skewed toward the lowest risk category. The risk scores were "remarkably unreliable" in predicting violent crime. Only 20% of those predicted to commit violent crimes actually did so (Larson et al., 2016). In addition, the RAT was inaccurate for both White and Black defendants at a similar rate, but in different ways. Black defendants were twice as likely to be misclassified as being high risk, while Whites were frequently misclassified as low risk. Black defendants who did not reoffend were misclassified as high risk 45% of the time, compared to 24% of the time for Whites. Similarly, Black defendants who did recidivate were misclassified as low risk 28% of the time, compared to 48% of the time for Whites.

Stevenson and Doleac (2021) report similar racial disparities in RAT-generated risk scores, as well as disparities in the application of the risk scores. Judges are 43% more likely to follow leniency suggestions associated with a risk score for White defendants than Black defendants. Similarly, Albright (2019) reports that RAT-recommended default bail decisions were 25% more likely to be overridden in favor of harsher bonds for Black defendants than their White counterparts.

In summary, the literature finds that bias in criminal sentencing can stem from judges, election cycles, crime coverage in the media, and even college football games. Although algorithmic tools were introduced into the court system in an effort to reduce this disparity, studies find that the risk scores produced by the RATs can actually perpetuate bias.

In addition, research demonstrates that judges apply RAT-generated recommendations differentially by race/ethnicity.

State	Year Effective	Law
Virginia	2007	Require & Validate
Connecticut	2008	Require & Validate
Minnesota	2008	Require
Georgia	2012	Require & Validate
Kentucky	2012	Require & Validate
Hawaii	2013	Require & Validate
Arizona	2014	Require & Validate
Colorado	2014	Require & Validate
Delaware	2014	Require & Validate
Indiana	2014	Require & Validate
New Jersey	2014	Require & Validate
Vermont	2015	May Use
Maryland	2017	Require & Validate
Mississippi	2017	Require & Validate
New Mexico	2017	Require
Rhode Island	2017	Require & Validate
Utah	2017	Require
Alaska	2018	Require & Validate
Illinois	2018	May Use & Validate
Michigan	2018	Require & Validate
Pennsylvania	2018	May Use & Validate
West Virginia	2018	Require & Validate
Idaho	2019	May Use & Validate
Missouri	2019	Require & Validate
Montana	2019	Require & Validate
Nevada	2019	May Use & Validate
New York	2020	May Use & Validate
California	2021	May Use & Validate
Tennessee	2021	Require

Table 1: Table lists each state that has a RAT law, the year the law became effective, and the law’s characteristics. *Require* denotes laws that require courts to use a RAT. *Validate* denotes laws that mandate an empirical study to test the unbiasedness of the RAT. *May Use* denotes laws that allow but do not require the use of RATs.

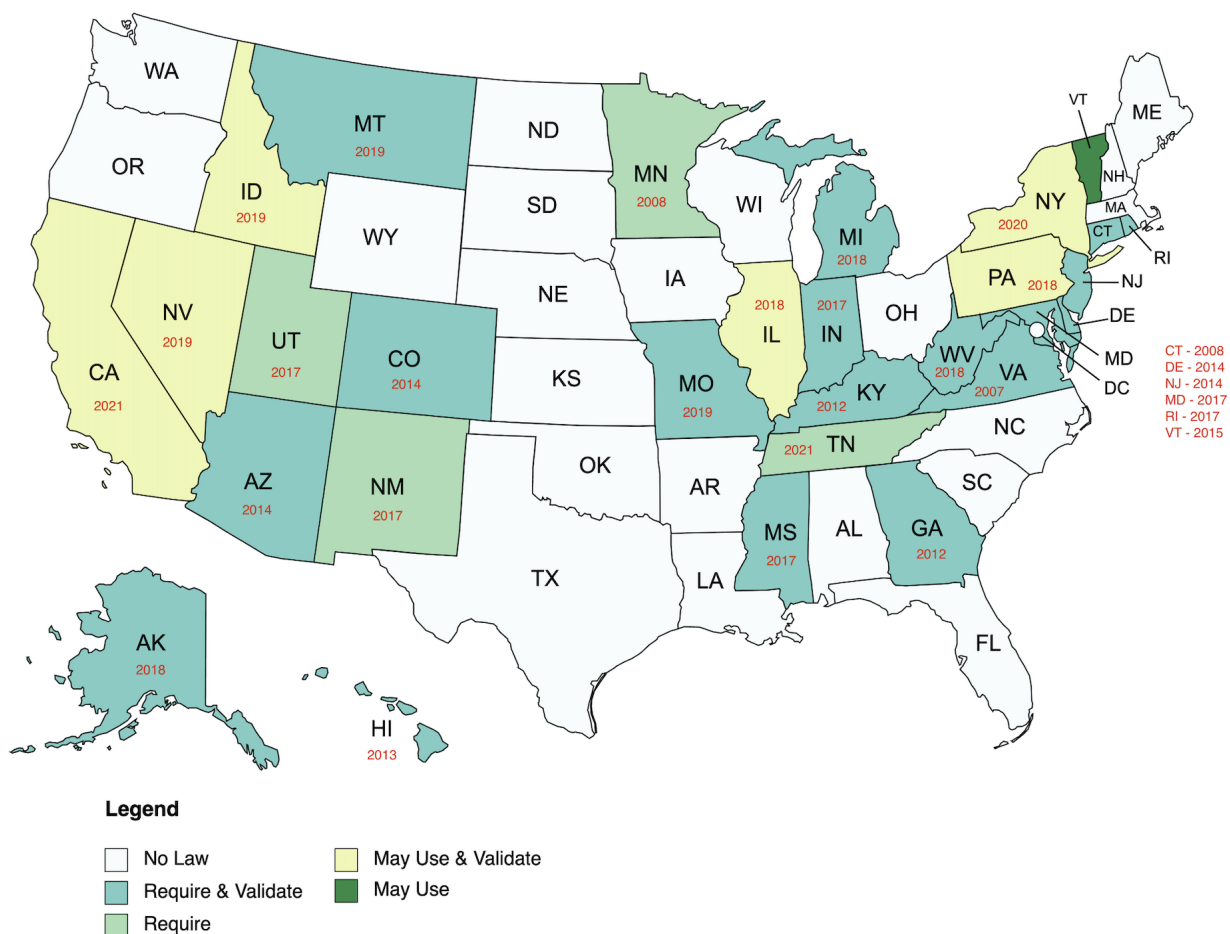


Figure 1: Map of states with a RAT law, the year the law became effective, and the law's characteristics.

125. "I have never intensely disliked anyone."
☐ Strongly Disagree ☒ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
126. "I have a short temper and can get angry quickly."
☐ Strongly Disagree ☒ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree

Criminal Attitudes

The next statements are about your feelings and beliefs about various things. Again, there are no 'right or wrong' answers. Just indicate how much you agree or disagree with each statement.

127. "A hungry person has a right to steal."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
128. "When people get into trouble with the law it's because they have no chance to get a decent job."
☐ Strongly Disagree ☒ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
129. "When people do minor offenses or use drugs they don't hurt anyone except themselves."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
130. "If someone insults my friends, family or group they are asking for trouble."
☐ Strongly Disagree ☐ Disagree ☒ Not Sure ☐ Agree ☐ Strongly Agree
131. "When things are stolen from rich people they won't miss the stuff because insurance will cover the loss."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
132. "I have felt very angry at someone or at something."
☐ Strongly Disagree ☐ Disagree ☐ Not Sure ☒ Agree ☐ Strongly Agree
133. "Some people must be treated roughly or beaten up just to send them a clear message."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
134. "I won't hesitate to hit or threaten people if they have done something to hurt my friends or family."
☐ Strongly Disagree ☐ Disagree ☒ Not Sure ☐ Agree ☐ Strongly Agree
135. "The law doesn't help average people."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
136. "Many people get into trouble or use drugs because society has given them no education, jobs or future."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree
137. "Some people just don't deserve any respect and should be treated like animals."
☒ Strongly Disagree ☐ Disagree ☐ Not Sure ☐ Agree ☐ Strongly Agree

Figure 2: Excerpt from survey used by COMPAS recidivism algorithm. Source: Angwin and Larson (2016).

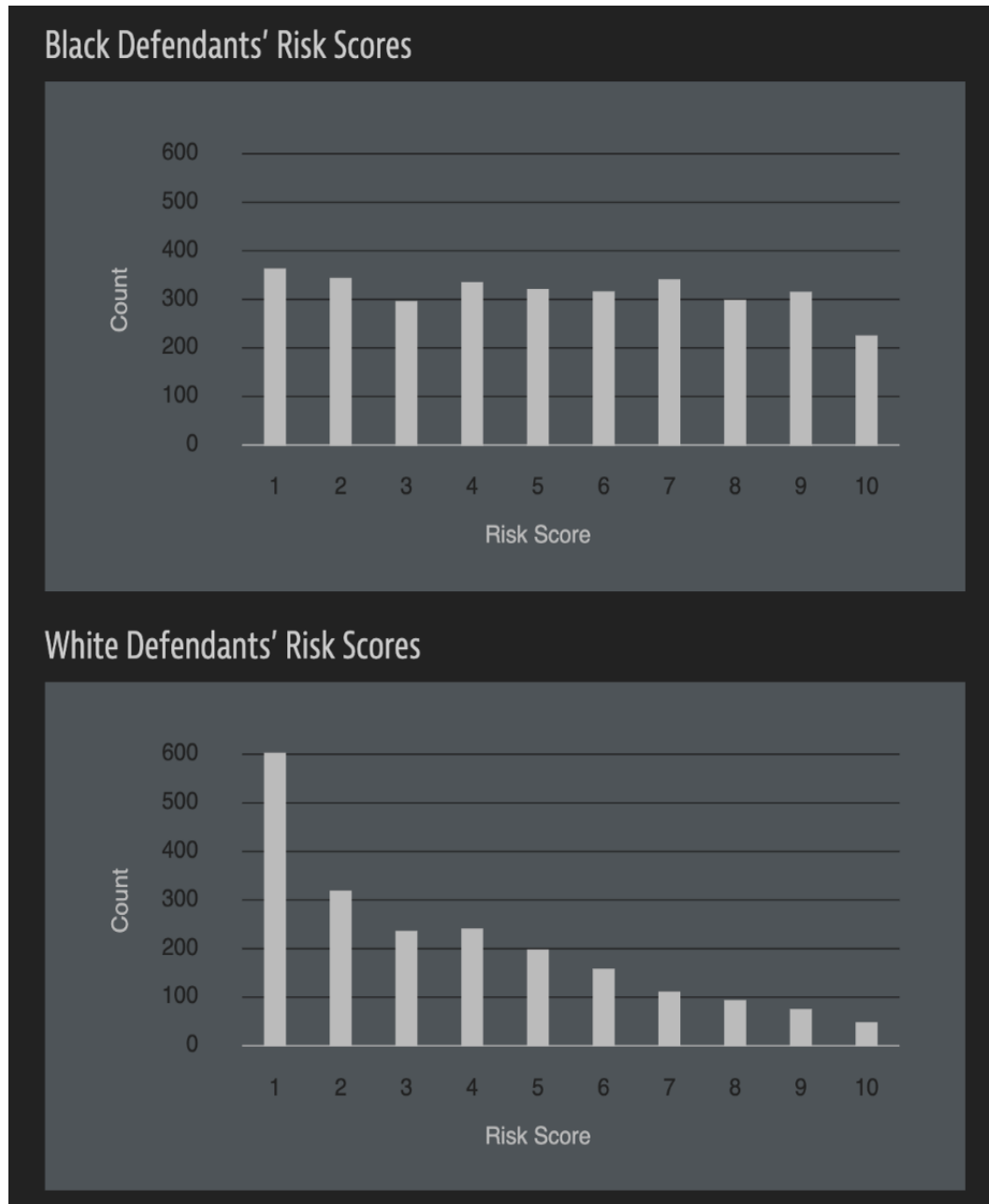


Figure 3: Figure displays the distribution of algorithm-generated risk scores for Black and White defendants in Broward County, Florida. Source: Larson et al. (2016).

DATA

In addition to the RAT law characteristics described earlier, I use publicly available incident-level information on sentence length, probation, and defendant characteristics from 2002-2021 (U.S. Sentencing Commission).¹ The USSC dataset documents crime and sentencing outcomes for every federally sentenced crime in each year. These outcomes include sentence length in months, and binary indicators for probation, community confinement, bail, and whether the case was settled by a plea agreement or trial. Crime information includes offense type, drug involvement, presence of vulnerable victims for hate crimes, and use of weapons. The USSC dataset also includes age, sex, race, ethnicity, education, citizenship status, number of dependents, and criminal history of the perpetrator. I limit the sample to White non-Hispanic, Black non-Hispanic, and Hispanic males in order to focus on racial disparities.

The USSC dataset documents only federally sentenced crimes, though it is possible to be prosecuted by both state and federal governments for a given crime.² Whether or not a case can be prosecuted by both state and federal courts usually depends on the offense committed. Crimes can include drug distribution and trafficking, kidnapping, bank robbery, and fraud. They can also include crimes committed on federal land, involving federal officials, or crimes committed across state lines.³ For these crimes, the USSC dataset’s sentence outcome variable aggregates the state and federal sentences. Since a federal judge is not subject to a state’s RAT laws, I limit the sample to observations where the defendant was also charged by the state. This ensures that the defendant’s sentence would be impacted by

¹Individual offender datafiles are available to commission online from the USSC at <https://www.ussc.gov/research/datafiles/commission-datafiles#individual> (“Individual”).

²The Double Jeopardy Clause of the Fifth Amendment prohibits trying an individual twice for the same offense. However, the sovereign exception allows a person to be tried in both state and federal courts since the federal and state governments are separate (United States, Dept. of Justice, 2017).

³Crimes that become punishable under federal law due to “interstate commerce” could present issues with my analysis if the defendant is crossing between treatment and control states.

the presence of a RAT law in that state. Unfortunately, the USSC dataset does not contain information on the length of the state sentence alone. However, federal judges sentence defendants according to strict federal sentencing guidelines, and must provide a reason for any departures from the guidelines. To account for variation in federal judges' sentencing patterns, I limit the sample to observations with no departures from the federal sentencing guidelines. Consistency in federal sentences makes the aggregated value a more accurate proxy for state sentence length.

Final Dataset

Though my RAT law data is at the state-level, my crime data contains information at the judicial district-level. Districts exist within states and do not cross state lines.⁴ I merge the sentencing data with the RAT law data, creating an incident-by-district-by-year panel. This leads to a total sample of 139,868 offenses committed between 2002 and 2021 by Black non-Hispanic, White non-Hispanic, and Hispanic males.

Table 2 shows the proportions of the sample covered by RAT laws. The baseline regressions define states with any form of RAT law as the treatment group. Subsequent regressions iterate through different treatment group selections to examine the effect of different law characteristics. Based on these estimates, RAT laws with both the *require* and *validate* components have the largest impact on sentencing outcomes. Because *require* and *validate* are the most salient characteristics of RAT laws, I use the 18 states with *Require* & *Validate* laws as the treatment group in subsequent regressions.⁵

Table 3 reports mean sentence length and mean likelihood of receiving probation in states that did and did not eventually pass RAT laws. I denote these as “ever-treated” and

⁴Most states contain one district, but states with multiple districts are generally divided into North/-South/East/West/Central regions.

⁵I examine whether my estimates change when I define the treatment group differently in Table 8.

“never-treated” states. I use data from 2002-2006 for the values in the table, and separate the values into four crime types: violent, property, drug, and financial. Violent crime includes murder, manslaughter, kidnapping/hostage taking, sexual abuse, and assault. Property crime includes robbery, burglary/breaking and entering, arson, automobile theft, and larceny. Drug crime includes trafficking/manufacturing/importing, communication facilities, and simple possession.⁶ Financial crime includes fraud/embezzlement, forgery/counterfeit/copyright infringement, bribery/corruption, tax offenses, money laundering, racketeering/extortion, and gambling/lottery offenses. The last column of Table 3 reports the differences in means within race, and whether or not the difference is statistically significant. The estimates for drug and property crimes appear to be statistically different for never-treated and ever-treated states. However, the magnitudes are small relative to mean sentence lengths and likelihood of receiving probation, mitigating concerns about endogeneity in which states choose to pass RAT laws.

Table 4 displays the means for each control variable, again separated by never-treated and ever-treated states. Black and Hispanic offenders account for 30% and 38% of the overall sample, respectively. On average, Black and Hispanic offenders are younger, have lower levels of education, and are more likely to have at least one dependent. Hispanic offenders are least likely to have a criminal history (58%), whereas 80% of Whites and 87% Blacks have criminal histories. Blacks and Whites have similar rates of U.S. citizenship, whereas Hispanics are about half as likely to be U.S. citizens, and ten times more likely to be in the country illegally.

Black offenders are more likely than Whites and Hispanics to have their case go to trial, as opposed to a plea agreement. In a plea agreement, defendants immediately plead guilty, usually in hopes of receiving a lesser sentence than if they were to go to trial and be found guilty. Hispanics are the most likely to be convicted for drugs involving cocaine, heroin,

⁶I do not separate the regressions into different crime groups. However, in a robustness check, I estimate regressions separately for different drug crimes in order to control for the type of substance involved.

and marijuana. Blacks are much more likely to be convicted of drug crimes involving crack. Drug crimes involving meth have similar proportions of Whites and Hispanics.

Data Limitations

Several data limitations exist that could bias my findings. First, it is difficult to ascertain whether RATs are utilized in states without RAT laws, since judges could be using RATs without regulation. A second issue is the lack of information on when judges began using algorithmic tools since it is possible that judges used RATs before any RAT legislation was passed. Mitigating these concerns are the high fixed costs associated with implementing a RAT. These costs make it unlikely that judges use RATs unless required by law. A related issue is that some judges may rely on RATs more heavily than others, which could bias the treatment effect estimates.

As previously discussed, the variable measuring sentence length aggregates state and federal sentences, posing a potential problem with measurement error in the dependent variable. I address this issue by limiting the sample to defendants that received sentences within the federal sentencing guidelines. Nonetheless, any systematic variation in within-guideline federal sentencing that is correlated with a state's propensity to have a RAT law could threaten the validity of my results.

Law	Incidents (%)	N	District-Years (%)	N
Any Law	2.908	4,095	14.369	270
May Use Only	0.010	14	0.373	7
Require Only	0.304	417	1.437	27
May Use & Validate	0.032	450	2.235	42
Require & Validate	2.574	3,617	10.325	194

Table 2: Table shows the proportion of observations covered by RAT laws with different characteristics. The values for *May Use Only* and *Require Only* represent the states that have only this component of the law and no *validate* component. The incidents column shows the percentage of cases that were committed after each type of the law was implemented. The district-years column shows the percentage of district-years covered by each type of RAT law.

Crime	Outcome	Never-Treated			Not-Yet-Treated			Difference	
		Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	N
Violent	Sentence								
	White	74.601	108.549	217	61.867	91.878	380	-12.733	597
	Black	104.394	107.537	100	106.682	154.012	167	28.461	267
	Hispanic	82.533	102.940	133	81.339	131.217	165	-1.193	298
	Probation								
	White	0.124	0.331	217	0.147	0.355	380	0.023	597
	Black	0.100	0.302	100	0.078	0.269	167	-0.222	267
	Hispanic	0.053	0.224	133	0.067	0.250	165	0.014	298
Property	Sentence								
	White	57.613	87.488	1,402	47.886	70.410	1,911	-9.727***	3,313
	Black	80.884	115.001	995	70.205	86.481	1,420	-10.680**	2,415
	Hispanic	37.430	51.620	283	41.037	56.454	374	3.607	657
	Probation								
	White	0.206	0.405	1,402	0.254	0.435	1,911	0.048***	3,313
	Black	0.177	0.382	995	0.199	0.399	1,420	0.022	2,415
	Hispanic	0.191	0.394	283	0.278	0.449	374	0.087**	657
Drugs	Sentence								
	White	89.775	92.760	4,544	76.556	75.267	5,017	-13.219***	9,561
	Black	134.426	106.42	5,980	123.404	101.884	7,251	-11.022***	13,231
	Hispanic	72.492	69.118	11,884	77.999	73.367	7,541	5.507***	19,425
	Probation								
	White	0.062	0.241	4,544	0.086	0.280	5,017	0.024***	9,561
	Black	0.016	0.126	5,980	0.020	0.141	7,251	0.004*	13,231
	Hispanic	0.010	0.102	11,884	0.016	0.124	7,541	0.005***	19,425
Finance	Sentence								
	White	25.154	59.631	3,404	25.688	50.572	5,246	0.534	8,650
	Black	27.474	51.300	1,991	29.821	55.476	2,875	2.346	4,866
	Hispanic	27.485	80.802	1,448	28.177	53.089	1,528	0.692	2,976
	Probation								
	White	0.250	0.433	3,404	0.254	0.436	5,246	0.004	8,650
	Black	0.215	0.411	1,991	0.216	0.412	2,875	0.001	4,866
	Hispanic	0.168	0.374	1,448	0.206	0.405	1,528	0.038**	2,976

Table 3: This table reports the mean sentence length in months and likelihood of receiving probation for Whites, Blacks, and Hispanics, separated by never- and not-yet-treated states. I calculate these values using data from the years 2002-2006, which is before any treatment takes place. The third column reports the difference between the values for control and treatment states, and reports if the estimate is statistically significant. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Variable	Never-Treated			Not-Yet-Treated			Difference		
	W	B	H	W	B	H	W	B	H
Age (years)	38.508	32.476	32.618	39.487	32.960	32.842	0.978***	0.484***	0.224*
Below HS	0.241	0.409	0.629	0.240	0.366	0.632	-0.001	-0.043***	0.004
HS grad	0.398	0.369	0.241	0.383	0.376	0.227	-0.014**	0.007	-0.014**
Some college	0.240	0.183	0.099	0.237	0.212	0.111	-0.004	0.030***	0.012***
Any children	0.523	0.684	0.708	0.499	0.643	0.682	-0.024***	-0.041***	-0.027***
Criminal history	0.736	0.874	0.566	0.934	0.868	0.609	-0.034***	-0.006	0.042***
US citizen	0.950	0.918	0.427	0.934	0.927	0.385	-0.15***	0.009**	-0.042***
Illegal alien	0.023	0.032	0.334	0.024	0.030	0.371	0.002	-0.002	0.038***
Trial	0.057	0.087	0.045	0.063	0.082	0.051	0.006*	-0.005	0.006**
Hate crime	0.009	0.003	0.001	0.011	0.003	0.002	0.002	0.000	0.001
Cocaine	0.063	0.149	0.259	0.048	0.120	0.233	-0.014***	-0.029***	-0.026***
Crack	0.012	0.418	0.021	0.023	0.384	0.058	0.005***	-0.034***	0.037***
Heroin	0.012	0.022	0.058	0.012	0.043	0.107	0.000	0.021***	0.049***
Marijuana	0.112	0.056	0.404	0.099	0.057	0.218	-0.013***	0.001	-0.185***
Meth	0.184	0.004	0.109	0.137	0.006	0.157	-0.047***	0.002**	0.048***
Number of obs.	10,458	9,092	13,830	13,609	11,757	9,658	24,067	20,849	23,488

Table 4: This table reports the means for each control variable separated by treatment for Whites (W), Blacks (B), and Hispanics (H). The level of observation is an incident. I calculate these values using data from the years 2002-2006, which is before any treatment takes place. The third column reports the differences in means by treatment status and if the estimate is statistically significant. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

EMPIRICAL METHODOLOGY

State RAT legislation can vary along several dimensions. Some states' RAT laws *require* (as opposed to just allow) the use of algorithmic tools, and some mandate an empirical study to *validate* the tool for unbiasedness. The baseline regressions use states with any form of RAT law (*anylaw*) as the treatment group. Subsequent regressions analyze heterogeneity across law characteristics.

Baseline Two-Stage Differences-in-DifferencesResearch Design

My goal is to investigate whether state RAT legislation decreases racial disparities in sentencing outcomes. I exploit the natural experiment arising from RAT legislation being implemented at different times in different states. Following Gardner (2021), I estimate the impact of RAT laws on sentence length and probation using a two-stage differences-in-differences (2sDiD) framework. The first stage of the 2sDiD framework identifies the district and year fixed effects by regressing the outcome on the sample of untreated observations. The second stage identifies the treatment effect after removing these fixed effects. Normal DiD estimation can show spurious violations of the parallel trends assumption, which 2sDiD addresses by only using the pre-treatment trends from untreated observations in the analysis. This method is well-suited for policies with staggered treatment timing as it is robust to treatment-effect heterogeneity (Gardner, 2021). The baseline estimating equation is:

$$(1) Y_{idt} = \beta_0 + \beta_1(\text{anylaw} * \text{post})_{dt} + \text{offense}_{idt} + \theta_{idt} + \text{districtFE} + \text{yearFE} + \epsilon_{idt}$$

where Y_{idt} can alternately measure sentence length for individual i in district d in year t or probation (a binary indicator for whether or not individual i received probation instead

of a prison sentence). The variable $(anylaw * post)_{dt}$ indicates states and times with any form of RAT law in place. In other words, *anylaw* aggregates states with any combination of *require*, *may use*, and *validate* components in the law. The variable $offense_{idt}$ controls for the specific offense committed. The vector θ_{idt} represents demographic variables for the individual and incident, including age, education, citizenship, criminal history, number of dependents, whether the defendant stood trial, and whether the offense had vulnerable victims.¹ Lastly, districtFE and yearFE serve as district and year fixed effects to control for district-specific differences that do not vary over time, and time-specific differences that affect all states, like a national event or enactment of a federal law. All models include standard errors that are clustered by state, which is the level of the treatment. The coefficient of interest, β_1 , measures the average effect of having any RAT law on sentence length or likelihood of receiving probation. I estimate the equation separately for White non-Hispanics, Black non-Hispanics, and Hispanics of any race.

Assumptions

There are several identifying assumptions that underlie the two-stage differences-in-differences methodology. First, the model assumes that there are no contemporaneous effects influencing the outcomes. For example, if a state passed its RAT law as part of a policy bundle, one of the other policies could be responsible for the estimated effect on sentencing outcomes. This would cause the regression to overestimate the impact of RAT legislation. However, this seems unlikely in this case as RAT laws specifically target racial disparities in sentencing and are not generally part of a policy package.

A second identifying assumption is that there are parallel pre-treatment trends in the outcome variable for the treatment and control groups. If the pre-treatment trends are non-

¹These variables aid in controlling for differential treatment in the judicial system. For instance, judges might sentence crimes with vulnerable victims more harshly (Phillipe, 2018).

parallel, the model could suggest a causal effect even if the actual treatment effect was zero. The crux of this assumption is that absent the treatment, the trends would remain parallel into the post-treatment period. It is impossible to observe the counterfactual, so we must rely on the pre-treatment trends being parallel to assess this assumption. I examine the trends for sentence length and probation with event study plots in Figures 4 and 5, which are discussed in the results section.²

The third identifying assumption is that assignment to the treatment group is exogenous, or that there is no selection into the treatment. It is possible that states implemented RAT laws due to concerns about discriminatory sentencing outcomes. This would indicate reverse causality in the data and render the model invalid. Judging from press statements and other media, it appears that states pass the legislation to be more equitable and set a precedent of unbiasedness in the judicial system, hopefully satisfying this assumption (Lipton, 2019; “Bipartisan,” 2019). If selection into treatment is an issue, the model will underestimate the treatment effect.

We must also assume that states did not take voluntary precautions before the law was enacted, or that people did not anticipate the law and adjust their behavior. This would entail states or districts implementing a RAT without a mandate. It seems unlikely a state or district would do this unless required, since it is costly to implement RATs.

Lastly, the model assumes that there are no spillover effects on individuals outside the treatment group. It is possible that an individual from a control state committed a crime in a treatment state, subjecting them to the use of an algorithmic risk assessment tool and potentially a different sentence.³

²The estimates indicate no significant differences in sentence length or probability of probation in the pre-treatment period.

³The sentencing dataset only contains information on where the defendant was sentenced, not where they reside, so I cannot define the exact number of observations with this issue.

Heterogeneity by Law Characteristic

The baseline regressions examine whether having any form of RAT law impacts sentencing outcomes. However, there is reason to suspect that different combinations of the *require*, *may use*, and *validate* components of RATs will have heterogeneous treatment effects. The presence of only one of the *require* and *validate* components can have ambiguous effects on racial disparities in sentencing. For example, requiring the use of RATs could feasibly worsen racial disparities if the algorithm used is biased. In addition, we cannot assume that states use a RAT unless the law has the *require* component.

To examine potential heterogeneous impacts of different types of RAT laws, I estimate a model with controls for each law characteristic. This shows the differential impact of the *require* and *may use* components, and whether the effects change when the *validate* component is also present.⁴ Economic theory suggests that *Require & Validate* laws should decrease racial disparities in sentencing, relative to a more ambiguous treatment effect when only one component is present. States that never have any form of RAT law serve as the control group. The estimating equation is:

$$(2) \ Y_{idt g} = \beta_0 + \beta_1(R * post)_{dt} + \beta_2(MV * post)_{dt} + \beta_3(RV * post)_{dt} + \beta_4 offense_{idt} + \theta_{idt} + districtFE + yearFE + \epsilon_{idt}$$

where $Y_{idt g}$ can alternately measure sentence length or whether the individual received probation. In these regressions, β_1 estimates the effect of laws with only the *require* component for individual i in district d and year t , by racial group g . β_2 and β_3 estimate the effect of *May Use & Validate* or *Require & Validate* laws, relative to no law. $Offense_{idt}$

⁴Vermont is the only state with a *May Use* law without the *validate* component. I group Vermont with the *May Use & Validate* states in the model, which I analyze in the robustness section. Unfortunately, this makes it so I cannot examine the impact of the *may use* component on its own.

and θ_{idt} represent the control variables as discussed above.⁵

Heterogeneity by Race

I estimate baseline regressions separately for Whites, Blacks, and Hispanics in order to identify differences in outcomes within racial groups. To identify potential differences in outcomes across racial groups, I estimate a regression that includes all of the races. The estimating equation is:

$$\begin{aligned} (3) \quad Y_{idt} = & \beta_0 + \beta_1(RV * post * black)_{idt} + \beta_2(RV * post * hispanic)_{idt} + \beta_3(RV * post)_{dt} + \\ & \beta_4(RV * black)_{idt} + \beta_5(post * black)_{idt} + \beta_6(RV * hispanic)_{idt} + \beta_7(post * hispanic)_{idt} + \\ & \beta_8black_{idt} + \beta_9hispanic_{idt} + offense_{idt} + \theta_{idt} + districtFE + yearFE + \epsilon_{idt} \end{aligned}$$

where Y_{idt} can alternately measure sentence length or whether the individual received probation. The equation includes three-way interactions between the treatment and each racial group. β_1 is the treatment effect of *Require & Validate* RAT laws for Blacks, β_2 represents the treatment effect for Hispanics, and β_3 represents the treatment effect for Whites.

Event Study

The event study methodology allows for heterogeneity in the treatment effect for each period in the sample. The treatment effect is interpreted as the deviation from the counterfactual, or what we would expect to happen to the outcome absent the treatment (passage of a RAT law).

⁵Because the results indicate that laws with *require* and *validate* components have the largest impact on sentencing outcomes, subsequent regressions use these 18 states as the treatment group. For these regressions, states that have other variations of the law are included in the control group. I discuss the impact of completely excluding these treatment states from the sample in the robustness section.

To avoid an unbalanced panel in the plot, I limit the sample to states that passed the law between 2007 and 2017. This allows me to observe outcomes for four years before and after the law is passed for these states. As with the other regressions, I estimate the equation separately for Whites, Blacks and Hispanics. The estimating equation is:

$$(4) \ Y_{idt} = \beta_0 + \sum_{\tau=-4}^4 \beta_{\tau} RV * \tau_t + offense_{idt} + \theta_{idt} + districtFE + \epsilon_{idt}$$

where Y_{idt} can alternately measure sentence length or whether the individual received probation.

Assumptions

The first identifying assumption of an event study is that the treatment is not anticipated, or that treatment states do not adjust their behavior due to the treatment while still in the pre-treatment period. Estimates will be biased if judges in treatment states change their sentencing behavior in anticipation of a RAT law. Most states that are required to adopt a RAT are allotted government funds to do so, and thus would not likely adopt a RAT before the funds are issued. The second assumption is the absence of contemporaneous trends, which the model addresses with district and time fixed effects.

RESULTS

Table 5 presents the baseline 2sDiD estimates of the effect of having any RAT law on sentence length and probation, separated for Whites, Blacks, and Hispanics. The mean sentence length and mean likelihood of probation are reported to provide context to the size of the estimated treatment effect. I find that having any RAT law significantly decreases sentence lengths by 4.35 months for Whites and 9.45 months for Hispanics. The mean sentence length is 63 months for Whites and 67 months for Hispanics, indicating that these coefficients are relatively large and bear economic significance. The estimate for Blacks is not statistically significant. None of the treatment estimates for the impact of having any RAT law on probation are statistically significant.

Table 6 reports the results from the regression that tests for heterogeneous impacts of RAT law characteristics. The results suggest that the effect of different law characteristics varies by race. *Require & Validate* (RV) laws decrease sentence length by 4.8 months for Whites and 12.4 months for Hispanics. I find no relative impact of RV laws for Blacks. *Require* laws significantly decrease sentence length by 11.7 months for Blacks, but increase sentence length by 7.2 months for Hispanics. Conversely, *May Use & Validate* laws increase sentence length by 37.6 months for Blacks, and decrease sentence length for Hispanics by 76.3 months. We see the sign of the effect flip from positive to negative for Hispanics when the *validate* component is included, and the opposite relationship for Blacks. The results for probation are mostly insignificant, except for the finding that *Require* laws significantly increase the likelihood of receiving probation for Hispanics by 0.8 percentage points.

Table 7 reports results from regressions testing for heterogeneity in the effect of *Require & Validate* (RV) laws by race. I find that RV laws significantly decrease sentence length for Whites by 18.7 months. For Hispanics, RV laws decrease sentence lengths an additional 14.2 months, implying an overall decrease of 32.9 months. This suggests that the law decreased

sentence disparities for Hispanics. For Blacks, RV laws significantly increase sentence length by 8.8 months relative to Whites, implying an overall decrease of 9.9 months for Blacks. This suggests that even though RV laws decrease sentence lengths for Blacks, the decrease is even larger for Whites, which increases overall disparities. Relative to the mean sentence length of 76 months, the coefficients are large and bear economic significance. The model does not find a significant impact of RV laws on probation.

Figures 4 and 5 display the event study plots for the impact of RV laws on sentence length and probation for Whites, Blacks, and Hispanics. Figure 4 shows that in the pre-period, all intervals contain zero, implying no treatment effect in the pre-period. The plot shows a decrease in sentence lengths for Blacks 3 years after having an RV law. The point estimates for Hispanics show a larger decrease for Hispanics, but the confidence intervals are also larger and sometimes contain zero. There does not appear to be a strong treatment effect for Whites. For probation, Figure 5 shows that most of the confidence intervals contain zero in the pre-treatment period, again indicating no differential pre-treatment trends. The plot shows a small but significant increase in the likelihood of receiving probation for Blacks. The estimates for Whites and Hispanics show no consistent impact of RV laws on likelihood of receiving probation.

	Sentence Length			Probation		
	W	B	H	W	B	H
Anylaw*post	-4.354*** (1.691)	-0.199 (2.667)	-9.450** (3.972)	-0.004 (0.008)	0.005 (0.003)	-0.001 (0.004)
Mean sentence length	63.146	100.413	67.325			
Mean probation likelihood				0.147	0.076	0.034
Number of observations	44,337	42,170	53,361	44,337	42,170	53,361

Table 5: Table reports 2sDiD treatment estimates of having any RAT law on sentence length and probation, separately for Whites, Blacks, and Hispanics. The first column reports the change in sentence length in months, and the second column reports the change in likelihood of receiving probation. The mean sentence length and mean likelihood of receiving probation are included to provide scale for the coefficients. Robust standard errors clustered by state are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

	Sentence Length			Probation		
	W	B	H	W	B	H
Require*post	-0.204 (2.467)	-11.716*** (1.619)	7.172*** (2.671)	-0.001 (0.004)	0.001 (0.003)	0.008** (0.004)
MayUse&Validate*post	-54.315 (36.524)	37.595*** (7.711)	-76.267*** (22.954)	-0.017 (0.048)	0.003 (0.054)	-0.007 (0.016)
Require&Validate*post	-4.757*** (1.567)	0.322 (2.431)	-12.419*** (2.801)	-0.005 (0.009)	0.005 (0.003)	-0.002 (0.004)
Mean sentence length	63.146	100.413	67.325			
Mean probation likelihood				0.147	0.076	0.034
Number of observations	44,337	42,170	53,361	44,337	42,170	53,361

Table 6: Table reports 2sDiD estimates separated by race when each characteristic of the law is controlled for separately. The characteristics are *Require*, *May Use&Validate*, and *Require&Validate*. Since there are no states that only have the *validate* component, this coefficient is not estimated in the regression. Vermont is the only state with a law that has the *may use* component without the *validate* component. I group it with the *May Use&Validate* states and examine the impact of this grouping in the robustness section. The first column reports the change in sentence length in months, and the second main column reports the change in likelihood of receiving probation. The mean sentence length and mean likelihood of receiving probation are included to provide scale for the coefficients. Robust standard errors clustered by state are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

	Sentence Length	Probation
RV*post*black	8.849*** (2.299)	-0.004 (0.006)
RV*post*hispanic	-14.160*** (1.769)	-0.001 (0.018)
RV*post (Whites)	-18.735*** (1.519)	-0.003 (0.003)
Mean sentence length	75.976	
Mean probation likelihood		0.082
Number of observations	139,868	139,868

Table 7: Table reports 2sDiD estimates of *Require & Validate* (RV) laws on sentence length and probation. The sample is no longer separated by race, so the estimates show the differential effect of the law for Whites, Blacks, and Hispanics. Robust standard errors clustered by state are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

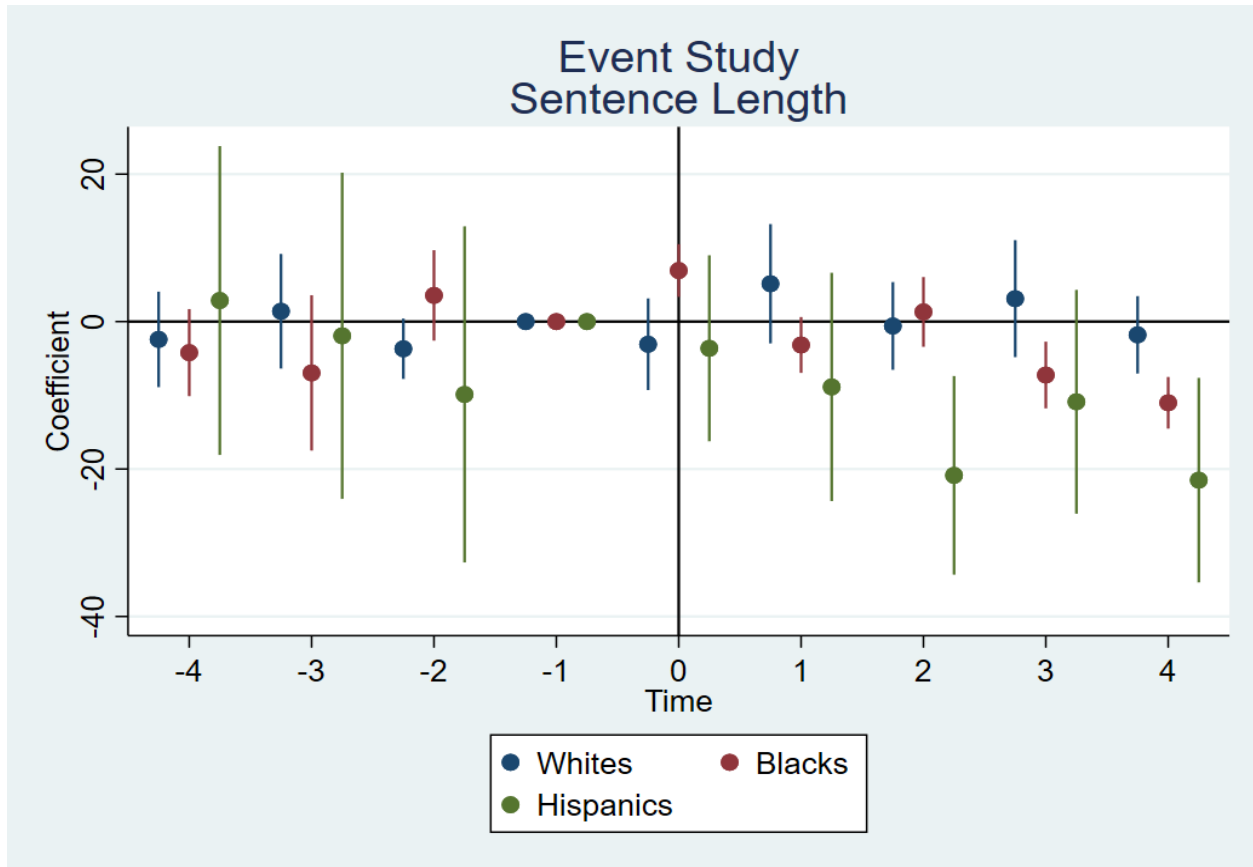


Figure 4: Figure displays event study plot for sentence length for Whites, Blacks and Hispanics when the treatment is *Require & Validate* laws. The year before the treatment is omitted, which is standard for event studies.

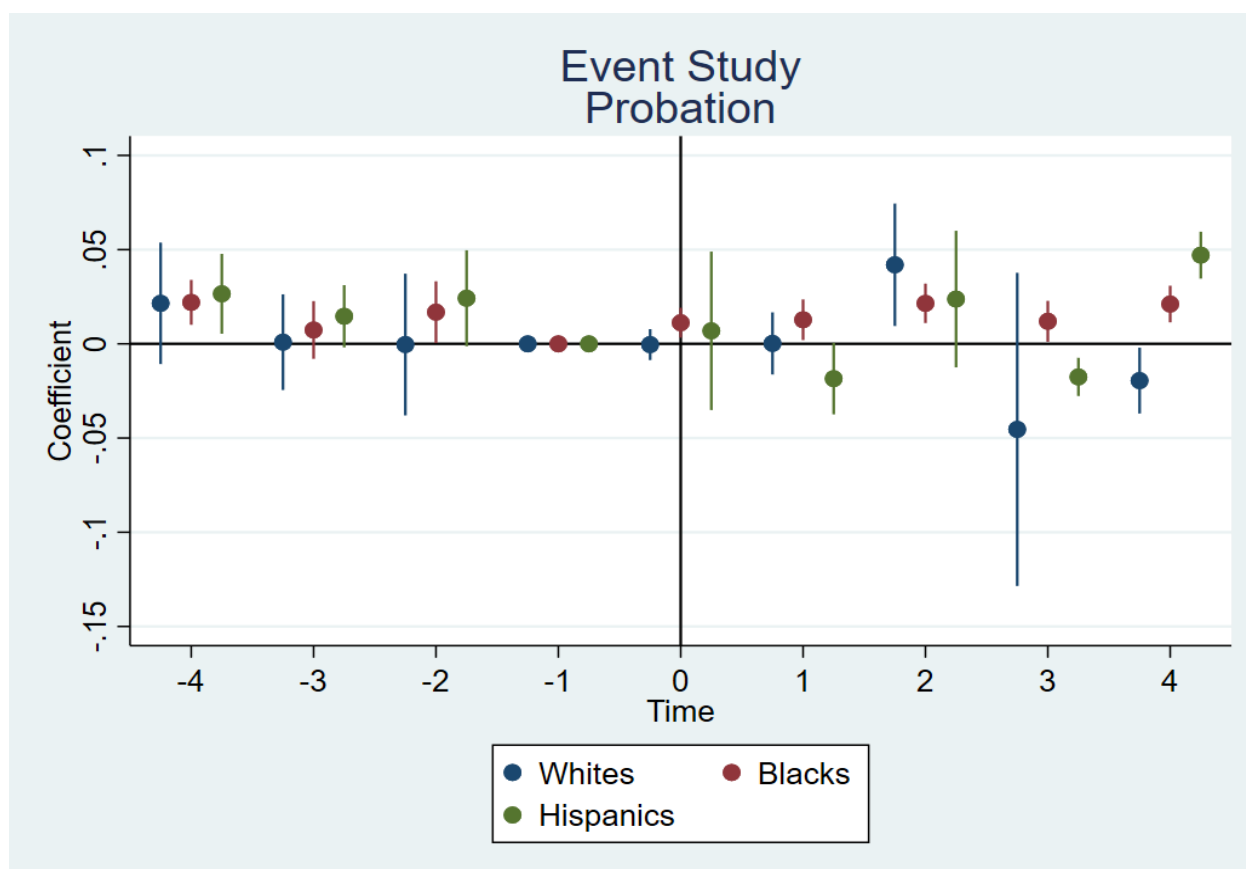


Figure 5: Figure displays event study plot for probation for Whites, Blacks and Hispanics when the treatment is *Require & Validate* laws. The year before the treatment is omitted, which is standard for event studies.

ROBUSTNESS CHECKS

Control Group Selection

In this section, I investigate the sensitivity of my estimates in several ways. Table 8 reports the 2sDiD estimates when states with RAT laws other than *Require & Validate* (RV) are dropped from the sample, as opposed to being included in the control group. These results imply that RV laws significantly increase sentence lengths for Blacks by 24.89 months, and significantly decrease sentence lengths for Whites by 2.95 months. The estimate for Hispanics loses statistical significance in this model. Interestingly, when non-RV treatment states are dropped, all probation estimates gain statistical significance. The results indicate that RV laws decrease the likelihood of receiving probation for Whites, Blacks, and Hispanics by 5-7 percentage points. This finding suggests that laws with different characteristics impact sentence length and probation differently.

Law Group Selection

Table 9 reports the 2sDiD estimates for RV laws when Vermont, a *May Use* state, is dropped from the sample as opposed to being grouped with *May Use & Validate* states. The results stay the same qualitatively, with slight decreases in the magnitudes of the coefficients. Since the results do not change when I drop Vermont, I choose to keep it in the sample in the baseline model.

Drug Crime Regressions

Table 10 reports the 2sDiD estimates when I regress drug crime sentences separately by the type of drug. This allows me to control for the substance involved, which include cocaine, crack, heroin, marijuana, and meth. The results suggest that the impact of RV laws

varies by race for defendants convicted of drug crimes. Looking at defendants sentenced for drug crimes involving marijuana, RV laws decrease sentence length by 15.4 months for Whites. Sentence lengths for Blacks and Hispanics increase by 8.4 and 14.1 months relative to Whites, indicating an overall decrease in sentence length. The regressions for the other drug types show no obvious pattern of the impact of the RV law. RV laws increase the likelihood of receiving probation for Black and Hispanic defendants sentenced for crimes involving cocaine and meth. Many of the estimates for probation gain statistical significance, but show no consistent pattern in the effect of the law.

	Sentence Length	Probation
RV*post*black	24.892*** (4.310)	-0.071*** (0.007)
RV*post*hispanic	-0.375 (5.868)	-0.054*** (0.009)
RV*post	-2.948* (1.688)	-0.056*** (0.010)
Mean sentence length	76.524	
Mean probation likelihood		0.076
Number of observations	115,775	115,775

Table 8: Table reports 2sDiD estimates of *Require & Validate* (RV) laws when states with *May Use & Validate* and *Require* laws are dropped from the sample as opposed to being included in the control group. Robust standard errors clustered by state are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

	Sentence Length	Probation
RV*post*black	8.334*** (2.490)	-0.006 (0.007)
RV*post*hispanic	-11.699*** (3.144)	0.005 (0.015)
RV*post	-16.067*** (3.220)	-0.003 (0.003)
Mean sentence length	76.012	
Mean probation likelihood		0.082
Number of observations	139,669	139,669

Table 9: Table reports 2sDiD estimates of *Require & Validate* (RV) laws when Vermont (*May Use* law) is dropped from the sample as opposed to being grouped with *May Use & Validate* states. Robust standard errors clustered by state are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

	Crack	Cocaine	Heroin	Marijuana	Meth
Panel A: Sentence Length					
RV*post*black	1.768 (2.575)	17.120** (7.542)	-12.091*** (2.703)	8.432*** (2.577)	-13.017 (39.226)
RV*post*hispanic	-13.764*** (2.711)	-26.201*** (3.232)	-36.187*** (2.548)	14.140*** (3.983)	7.063* (3.924)
RV*post	-26.951*** (2.474)	-4.027 (7.460)	-17.995 (17.780)	-15.415*** (4.203)	-8.566*** (2.862)
Mean sentence length	141.094	100.637	83.285	42.745	117.893
Number of observations	19,366	21,372	5,564	25,680	13,816
Panel B: Probation					
RV*post*black	-0.003** (0.001)	0.009** (0.004)	-0.007 (0.007)	-0.048** (0.024)	0.005** (0.002)
RV*post*hispanic	0.043*** (0.011)	0.019* (0.011)	-0.058*** (0.015)	-0.046 (0.034)	0.004** (0.002)
RV*post	0.030 (0.022)	0.008 (0.011)	-0.018 (0.016)	-0.029 (0.030)	0.001 (0.001)
Mean probation likelihood	0.004	0.013	0.012	0.045	0.010
Number of observations	19,366	21,372	5,564	25,680	13,816

Table 10: Table reports 2sDiD estimates for *Require & Validate* (RV) laws when I regress drug crimes separately by the substance that was involved. Panel A reports the results when sentence length is the outcome variable, and Panel B reports the results when probation is the outcome variable. Robust standard errors clustered by state are in parentheses. *p<0.1, **p<0.05, ***p<0.01.

CONCLUSION

This paper is among the first to examine the impact of RAT laws on racial disparities in sentencing outcomes. My baseline results show that having any form of RAT law significantly decreases sentence lengths for Whites and Hispanics, but not Blacks. When I test for heterogeneity across law characteristics, I find that the *require* component of RAT laws significantly decreases sentence length for Blacks, but increases sentence length for Hispanics. When the *validate* component is also present, RAT laws significantly increase sentence length for Blacks, but decrease sentence length for Hispanics. When I test for heterogeneity across race, I find that *Require & Validate* (RV) laws significantly decrease overall sentence lengths for all races, but only decrease the sentence disparity for Hispanics relative to Whites. Most of my baseline regressions do not find a statistically significant impact of RAT laws on the likelihood of receiving probation. Overall, my results are somewhat sensitive to changes in how I define the treatment and control groups.

Potential mechanisms for *Require & Validate* laws specifically worsening the Black-White sentencing disparity include lack of uniformity in RAT law validation requirements and their enforcement. If judges are required to consider outputs from a poorly-validated RAT, it could increase racial disparities in sentencing. However, this mechanism implies that poor validation requirements negatively impact only Blacks and not Hispanics.

In summary, my results imply that properly regulated RATs have the potential to decrease racial disparities in sentencing, particularly for Hispanics relative to Whites. At the same time, my estimates of the impact of RAT laws on Black-White sentencing disparities are sensitive to specification. Clearly, there is more work to be done in this area. My paper only addresses the impact of RAT laws on sentence length and likelihood of receiving probation. Future research could examine pretrial detention, plea deals, and other defendant outcomes.

REFERENCES CITED

- Abrams, D. S., Bertrand, M., & Mullainathan, S. (2012). Do Judges Vary in Their Treatment of Race? *Journal of Legal Studies*, 41(2), 347-384.
- Albright, Alex. 2019. If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions. Working paper.
- Alesina, A., & La Ferrara, E. (2014). A Test of Racial Bias in Capital Sentencing. *American Economic Review*, 104(11), 3397-3433.
- Angwin, Julia, and Jeff Larson. "Machine Bias." ProPublica, ProPublica, 23 May 2016.
- Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. *Quarterly Journal of Economics*, 127(2), 1017-1055.
- Arnold, D., Dobbie, W. S., & Hull, P. (2021). Measuring Racial Discrimination in Algorithms. National Bureau of Economic Research, Inc, NBER Working Papers: 28222.
- "Attorney General Eric Holder Speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference." United States Department of Justice, USDOJ, 1 Aug. 2014, Accessed 11 Mar. 2023.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Rahwan, I. (2018). The Moral Machine Experiment. *Nature*, 563(7729), 59-3,64A-64K.
- Bielen, S., Marneffe, W., & Mocan, N. H. (2018). Racial Bias and In-group Bias in Judicial Decisions: Evidence from Virtual Reality Courtrooms National Bureau of Economic Research, Inc, NBER Working Papers: 25355.
- "Bipartisan Legislators Introduce Legislation in Idaho to Prevent Non-Transparent and Potentially Biased Pretrial Algorithms." American Bail Coalition, American Bail Coalition, 13 Feb. 2019, Accessed 2 Mar. 2023.
- Bushway, S. D., & Piehl, A. M. (2001). Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing. *Law & Society Review*, 35(4), 733-764.
- Cabreros, Irineo, et al. "Predicting Race And Ethnicity To Ensure Equitable Algorithms For Health Care Decision Making." *Health Affairs Web Exclusive*, vol. 41, no. 8, 2022, pp. 1153-59.
- Cohen, A., & Yang, C. S. (2019). Judicial Politics and Sentencing Decisions. *American Economic Journal: Economic Policy*, 11(1), 160-191.

- Depew, B., Eren, O., & Mocan, N. (2017). Judges, Juveniles, and In-group Bias. *Journal of Law and Economics*, 60(2), 209-239.
- Dobbie, W., Goldin, J., & Yang, C. S. (2018). The Effects of Pretrial Detention on Conviction, Future crime, and Employment: Evidence from Randomly Assigned Judges. *American Economic Review*, 108(2), 201-240.
- Eren, O., & Mocan, N. (2018). Emotional Judges and Unlucky Juveniles. *American Economic Journal. Applied Economics*, 10(3), 171-205.
- Feigenberg, B., & Miller, C. (2021). Racial Divisions and Criminal Justice: Evidence from Southern State Courts. *American Economic Journal: Economic Policy*, 13(2), 207-240.
- Fischman, J. B., & Schanzenbach, M. M. (2012). Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums. *Journal of Empirical Legal Studies*, 9(4), 729-764.
- Gardner, John. "Two-Stage Differences in Differences." IDEAS Working Paper Series from RePEc, 2021.
- Hartmann, K., Wenzelburger, G. "Uncertainty, Risk and the Use of Algorithms in Policy Decisions: A Case Study on Criminal Justice in the USA." *Policy Sci* 54, 269–287. 9 Oct. 2021.
- "Individual Offender Datafiles." United States Sentencing Commission, USSC, 16 Feb. 2023.
- Jareborg, N. (1987). "Council of Europe's Eighth Criminological Colloquium." Disparity in Sentencing: Some General Observations. United States Department of Justice, Accessed 21 Mar. 2023.
- Kleinberg, J., & Mullainathan, S. (2019). Simplicity Creates Inequity: Implications for Fairness, Stereotypes, and Interpretability. National Bureau of Economic Research, Inc, NBER Working Papers: 25854.
- Larson, Jeff, et al. "How We Analyzed the COMPAS Recidivism Algorithm." ProPublica, ProPublica, 23 May 2016.
- Leslie, E., & Pope, N. G. (2017). The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments. *Journal of Law and Economics*, 60(3), 529-557.
- Lipton, Beryl. "Idaho Legislators Approve Law Requiring Transparency for Risk Assessment Tools." MuckRock, 26 Mar. 2019, Accessed 26 Feb. 2023.

- Mayson, S. (2019). Bias in, Bias Out. *The Yale Law Journal.*, 128(8), 2218.
- Mustard, D. B. (2001). Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *Journal of Law and Economics*, 44(1), 285-314.
- Park, K. H. (2017). The Impact of Judicial Elections in the Sentencing of Black Crime. *Journal of Human Resources*, 52(4), 998-1031.
- Pettit, B., & Gutierrez, C. (2018). Mass Incarceration and Racial Inequality. *American Journal of Economics and Sociology*, 77(3-4), 1153-1182.
- Philippe, A., & Ouss, A. (2018). "No Hatred or Malice, Fear or Affection: Media and Sentencing." *Journal of Political Economy*, 126(5), 2134-2178.
- Rambachan, A., Kleinberg, J., Mullainathan, S., & Ludwig, J. (2020). An Economic Approach to Regulating Algorithms National Bureau of Economic Research, Inc, NBER Working Papers: 27111.
- Rehavi, M. M., & Starr, S. B. (2014). Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy*, 122(6), 1320-1354.
- "Risk Validation." Public Safety Risk Assessment Clearinghouse, Bureau of Justice Assistance, U.S. Department of Justice.
- Schwerzmann, K. (2021). Abolish! Against the Use of Risk Assessment Algorithms at Sentencing in the US Criminal Justice System. *Philosophy & Technology*, 34(4), 1883-1904.
- Scurich, N., & Krauss, D. A. (2020). Public's Views of Risk Assessment Algorithms and Pretrial Decision Making. *Psychology, Public Policy, and Law*, 26(1), 1-9.
- Sen, M. (2015). Is Justice Really Blind? Race and Reversal in US Courts. *Journal of Legal Studies*, 44(1), S187-S229.
- Skeem, J., & Lowenkamp, C. (2020). Using Algorithms to Address Tradeoffs Inherent in Predicting Recidivism. *Behavioral Sciences & the Law*, 38(3), 259-278.
- Skeem, J., Scurich, N., & Monahan, J. (2020). Impact of Risk Assessment on Judges' Fairness in Sentencing Relatively Poor Defendants. *Law and Human Behavior*, 44(1), 51-59.
- Sloan, C., Naufaul, G., & Caspers, H. (2018). The Effect of Risk Assessment Scores on Judicial Behavior and Defendant Outcomes IZA Discussion Paper Series: No. 11948.

- Stevenson, Megan, and Jennifer L. Doleac. 2021. Algorithmic Risk Assessment in the Hands of Humans. Working paper.
- Starr, S. B., & Rehavi, M. M. (2013). Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of Booker. *Yale Law Journal*, 123(1), 2-80.
- United States, Department of Justice, Bureau of Justice Statistics. "Prisoners in 2016" Jan. 2018.
- United States, Department of Justice, Office of Justice Programs. "Double Jeopardy Clause of the Fifth Amendment." Apr. 2017.
- United States Sentencing Commission. *Federal Sentencing Guidelines Manual*. St. Paul, Minn. West Pub. Co.
- "United States v. Booker." Oyez, www.oyez.org/cases/2004/04-104. Accessed 27 Mar. 2023.
- "What Is Machine Learning?" IBM, IBM, <https://www.ibm.com/topics/machine-learning>. Accessed 14 Mar. 2023.
- Yang, C. S. (2015). Free at Last? Judicial Discretion and Racial Disparities in Federal Sentencing. *Journal of Legal Studies*, 44(1), 75-111.

APPENDIX

LAW DATA SOURCES

State	Source
Virginia	https://tinyurl.com/82nax8uf
Connecticut	https://tinyurl.com/2p8jvwyx
Minnesota	https://tinyurl.com/pzmmh78c4
Georgia	https://tinyurl.com/2s3jh7j9
Kentucky	https://tinyurl.com/33ae8cty
Hawaii	https://tinyurl.com/ypfw837j
Arizona	https://tinyurl.com/y3jpc9rf
Colorado	https://tinyurl.com/4c8r3r9n
Delaware	https://tinyurl.com/3xh3hdhk
Indiana	https://tinyurl.com/3ejfuf4y
New Jersey	https://tinyurl.com/bdzdj45j
Vermont	https://tinyurl.com/e96b8xh5
Maryland	https://tinyurl.com/bddhk596
Mississippi	https://tinyurl.com/23u534ma
New Mexico	https://tinyurl.com/2p859vvt
Rhode Island	https://tinyurl.com/bde7ep6h
Utah	https://tinyurl.com/2bkjpywn
Alaska	https://tinyurl.com/yck3c5st
Illinois	https://tinyurl.com/4rkymztk
Michigan	https://tinyurl.com/r7bkk9r3
Pennsylvania	https://tinyurl.com/3sb9knmz
West Virginia	https://tinyurl.com/ume8et6n
Idaho	https://tinyurl.com/2p88rr6u
Missouri	https://tinyurl.com/kbtvmfnj
Montana	https://tinyurl.com/yd8byk8t
Nevada	https://tinyurl.com/2cmjj5cc
New York	https://tinyurl.com/crcwps8s
California	https://tinyurl.com/tvcf9s4h
Tennessee	https://tinyurl.com/2vfnetdk

Table 11: This table lists the sources for each state’s law data.