HCEO WORKING PAPER SERIES

Algorithmic Risk Assessment in the Hands of Humans

Megan T. Stevenson
Jennifer L. Doleac

Working Paper

2020-055

07/2020

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Algorithmic Risk Assessment in the Hands of Humans*

Megan T. Stevenson & Jennifer L. Doleac†

November 18, 2019

Abstract

We evaluate the impacts of adopting algorithmic predictions of future offending (risk assessments) as an aid to judicial discretion in felony sentencing. We find that judges' decisions are influenced by the risk score, leading to longer sentences for defendants with higher scores and shorter sentences for those with lower scores. However, we find no robust evidence that this reshuffling led to a decline in recidivism, and, over time, judges appeared to use the risk scores less. Risk assessment's failure to reduce recidivism is at least partially explained by judicial discretion in its use. Judges systematically grant leniency to young defendants, despite their high risk of reoffending. This is in line with a long standing practice of treating youth as a mitigator in sentencing, due to lower perceived culpability. Such a conflict in goals may have led prior studies to overestimate the extent to which judges make prediction errors. Since one of the most important inputs to the risk score is effectively off-limits, risk assessment's expected benefits are curtailed. We find no evidence that risk assessment affected racial disparities statewide, although there was a relative increase in sentences for black defendants in courts that appeared to use risk assessment most. We conduct simulations to evaluate how race and age disparities would have changed if judges had fully complied with the sentencing recommendations associated with the algorithm. Racial disparities might have increased slightly, but the largest change would have been higher relative incarceration rates for defendants under the age of 23. In the context of contentious public discussions about algorithms, our results highlight the importance of thinking about how man and machine interact.

*We are grateful to Kelsey Pukelis and Arjun Ravi for excellent research assistance. We also thank Alex Albright, Patrick Bayer, Meredith Farrar-Owens, Brandon Garrett, Sandra Mayson, John Monahan, Michael Mueller-Smith, Derek Neal, Aurelie Ouss, Ben Schoenfield, Sonja Starr; the staff at the Virginia Criminal Sentencing commission; and seminar participants at CELS, NBER Summer Institute, University of Michigan Law & Economics Colloquium, UT Austin Law & Economics Colloquium, Northwestern Law & Economics Colloquium, the IADB, the CCC Economics and Fairness Workshop, UCLA Workshop on AI and Law Enforcement, Manne Junior Faculty Forum, GMU Schar School, UT Austin Economics, the University of Oregon, the 2019 IRP Summer Research Workshop, the 2019 Public Choice Society meetings, Williams College, the SDSU Center for Health Economics & Policy Studies, the ASHE-TAMU conference, Baylor University, the University of Missouri, the Harvard Kennedy School, the University of Pittsburgh, Ohio State University, the University of Laussane, and the Minnesota Population Center, as well as student reviewers from the University of Michigan graduate Labor Economics class for helpful comments and suggestions. Stevenson: Assistant Professor of Law, George Mason University, msteven@gmu.edu. Doleac: Associate Professor of Economics, Texas A&M University, jdoleac@tamu.edu.

†Both authors contributed to idea development and writing. Stevenson did all the data analysis.

1 Introduction

Algorithmic predictions of future offending, known as risk assessments, are proliferating in criminal justice. They are used to help inform decision-making at almost every stage of the criminal proceedings: setting bail, sentencing, determining probation supervision levels, offender placement within the prison system, and parole. Developed by statistically analyzing court and police records to identify which factors best predict recidivism, risk assessment tools summarize relevant data in a standardized way. Recent research has highlighted the errors that judges make in predicting reoffending (Berk et al., 2016; Jung et al., 2017; Kleinberg et al., 2018; Arnold et al., 2018). If humans are systematically making mistakes, prediction algorithms may lead to efficiency gains. By incarcerating those who pose a high risk of reoffending, and releasing those who don't, we should be able to lower crime rates without increasing incarceration, or vice versa. A policy simulation conducted by Kleinberg et al. (2018) suggests that using an algorithm to determine pretrial detention could reduce crime by up to 24% without increasing the jail population.

However, there are increasing concerns about risk assessment's impact on racial disparities (Mayson, 2019; Albright, 2019). Algorithms don't include race as a predictor, but they often include direct socio-economic markers like employment, housing or education status, which are correlated with race (Starr, 2014). Even criminal history variables such as the number of prior convictions are likely to reflect racially disparate policing and prosecution patterns as much as actual criminal activity. While concerns about racial bias are well-founded, its impact on risk assessment depends on a comparison to the status quo (Doleac and Stevenson, 2016; Cowgill and Tucker, 2017). Given the substantial evidence of racial bias in human predictions (Arnold et al., 2018), as well as statistical discrimination in the absence of objective information (Agan and Starr, 2017; Doleac and Hansen, 2018), a risk algorithm might alleviate some of the disadvantage black defendants face (Cowgill and Tucker, 2019; Kleinberg et al., 2019). Some initial evidence supports this: policy simulations in Kleinberg et al. (2018) suggest that risk assessment use can lower racial disparities while simultaneously increasing efficiency.

While nascent economics research has thus far largely supported the use of risk assessment, most studies have implicitly assumed that algorithms would take the place of human discretion. Generally, this is not the case. Risk assessments are used as a supplement to human discretion; people retain final decision-making power. Thus, while the conversation about risk assessment tools is often framed as a question of man versus machine, the relevant question is 'how do the two interact?' The impact of risk assessment depends not only on the properties of the algorithm, but on how the algorithm enters into the objective function of its user. This can lead to unexpected

results.

This study is about what happens when risk assessments are placed in the hands of humans. In the early 2000s, Virginia courts began using risk assessments at sentencing for nonviolent felony offenders and sex offenders. They were incorporated into the sentence guidelines in order to (1) divert a large share of low-risk nonviolent offenders from jail or prison, and (2) enable longer sentences for high-risk sex offenders. Other offenders were not affected by the policy change, and judges retained final authority over sentencing. Using several different research methodologies, we find that risk assessment was influential in sentencing. Although the recommendations associated with the risk score were not uniformly followed, sentences increased for those with higher risk scores and decreased for those with lower risk scores. But there is no robust evidence that this reshuffling led to a reduction in recidivism. We explore how people use risk assessments to better understand why not. Consistent with a deep-rooted tradition of treating youth as a mitigating factor in sentencing, judges/prosecutors are systematically more likely to be lenient with young defendants despite their higher risk scores (Stevenson and Slobogin, 2018). Although there was a relative increase in sentences for young defendants after risk assessment adoption, this increase was nowhere near as large as it would have been had there been full compliance with the algorithm. Age is one of the most important predictors of criminal activity, and, accordingly, has large weight in almost every risk assessment currently in use. A reluctance to incarcerate young people puts a ceiling on risk assessment's ability to reduce crime by incarcerating those at highest risk of reoffending.

Our results on racial disparities are mixed. We find no evidence that risk assessment affected racial disparities in sentencing statewide. Although black defendants have substantially higher risk scores than white defendants, judges sentence in a racially disparate manner even without risk assessment. However, we find that racial disparities increased in the subset of courts where risk assessment appears most influential. This is partly driven by the risk score recommendations, and partly because judges were more likely to sentence leniently for white defendants with high risk scores than for black defendants with the same score.

These results were obtained using multiple different research strategies, as described in the next few paragraphs. Our primary method of evaluating whether judges pay attention to the risk tool takes advantage of the fact that scoring right above or below various cutoffs in the risk score triggers different sentencing recommendations. Using regression discontinuity, we find that nonviolent offenders whose risk score is right below the cutoff (who just barely received a diversion recommendation) are 6 percentage points less likely to be incarcerated and have sentences that are approximately 23% percent shorter than those whose risk score is right above the cutoff. Sex offenders

whose risk score is right above the relevant cutoff (who just barely received an expanded upper bound of the guidelines-recommended sentence) are 6 percentage points more likely to be incarcerated and have sentences that are 34% longer than those whose score is right below. While these results show that risk assessment is influential, judicial discretion still plays a large role. Less than half of those who were recommended for diversion by the nonviolent risk assessment actually get diverted.

We next evaluate how risk assessment's use affected outcomes relative to the status quo (i.e judicial decision-making without risk assessment). We use a difference-in-differences research design, which compares risk-assessment-eligible/-ineligible defendants before/after risk assessment was adopted, to evaluate risk assessment's net impact on the probability of incarceration, sentence length and recidivism. The net effect on both sentencing outcomes for nonviolent offenders was a precisely estimated null. In other words, prison and jail terms appear to have been reallocated from defendants rated low-risk to those rated high-risk so that overall incarceration for nonviolent offenders remained the same. Under the efficiency hypothesis, if the risk assessment helped the judge to evaluate risk more accurately, this would have led to a reduction in crime. However, we do not find robust evidence that the adoption of risk assessment led to a decline in recidivism. And, over time, use of risk assessment appeared to decline, suggesting that judges did not find it to be useful.

Next, we explore the nonviolent risk assessment's impact by age and race. Even after differences in the guidelines-recommended sentence are accounted for, young defendants have substantially higher risk scores than older defendants, and black defendants have substantially higher risk scores than non-black defendants. (Age is an explicit factor in the calculation of the risk score; race is not.) While these gaps are concerning, the mere fact of higher risk scores does not necessarily mean that the use of risk assessment will lead to worse average outcomes compared to the status quo. This depends on a) the extent to which these factors influence sentencing in the absence of formal risk evaluation and b) whether the information provided by the risk score is interpreted differently for defendants of a different race or age (Agan and Starr, 2017; Doleac and Hansen, 2018; Albright, 2019; Skeem et al., 2019). On net, we find that risk assessment use adversely impacted the young. Using a triple-differences research design we find that risk assessment use led to a 4 percentage point increase in the probability of incarceration for young defendants (relative to older defendants), and a 12% relative increase in the sentence length. Using simulations, we benchmark this increase against what would have occurred if judges had fully complied with the sentencing recommendations associated with the algorithm. We find that full compliance would have entailed a much larger increase in age disparities: a 15 percentage point relative increase in the probability of incarceration, and a 53% relative increase in the

sentence length. Judges used their discretion to divert young defendants despite their high risk scores, thus minimizing the adverse impacts on young people.

We find no evidence that risk assessment use increased racial disparities statewide, although our standard errors prevent us from ruling out a moderate increase in sentence lengths for black defendants (relative to white). However, judges vary in the extent to which risk assessment influenced decision-making. Using regression discontinuity at the judge level, we identify a subset of courts in which the low-risk designation appeared to lead to the largest increase in leniency for defendants at the margin. In this 'most responsive' group of judges, we find that risk assessment adoption led to an increase in racial disparities relative to judicial discretion alone. This is partially explained by the fact that black defendants have higher risk scores, and partially because black defendants are sentenced more harshly than white defendants with the same risk score. (Simulations provide suggestive evidence that full compliance would have increased racial disparities, although this is not robust.) Our other results – the net impact on sentencing and recidivism, as well as differential effects on young defendants – look similar in this 'most responsive' subgroup as they do in the full sample.

We consider several hypotheses for why the nonviolent risk assessment did not lead to the decline in recidivism that researchers had anticipated. Our first hypothesis, alluded to above, is that judges are making fewer prediction errors than previously believed. Instead, they are pursuing objectives that may conflict with the goal of incapacitating those at highest risk of reoffending. A reluctance to incarcerate young people despite their high risk scores suggests lower potential gains from adopting prediction tools. Our second hypothesis is that the criminogenic effects of incarceration for higher risk defendants may have effectively canceled out its incapacitative effects. Using discontinuities in the risk score as an instrument for incarceration, we find no evidence to support the criminogenic channel: incarceration has a net negative effect on recidivism for up to seven years after sentencing. Finally, we consider whether the Virginia risk assessment simply does not predict recidivism very well, and whether we can build a more predictive instrument. We find that even under ideal conditions, predictions of future offending are unable to explain more than a tiny fraction of the variation in recidivism (R^2 <0.03). Future criminal activity is simply hard to predict. However, despite the low R^2, both the real risk score and our home-built one are able to successfully sort many defendants by recidivism risk. The extent to which this provides information that judges did not already have is unclear.

We conduct only a limited number of analyses on the sex offender risk assessment, since the number of convicted sex offenders are relatively small. The sex offender risk assessment was incorporated into the sentence guidelines in such a way as to only authorize an *increase* in sentences for higher risk offenders. Despite this, we find a

net *decrease* in incarceration rates and sentence lengths after the risk assessment was adopted, suggesting that even though sentences increased for higher risk defendants, this was more than offset by a decrease in sentences for lower risk defendants. It's possible that the risk assessment changed judges' beliefs so that sex offenders were, on average, perceived as less risky than they had previously believed. However, the risk assessment tool provides no statistical information about recidivism risk, only an ordinal ranking, making this explanation less than satisfying. It's also possible that the risk assessment provided a sort of political shield that, at least for those with low risk scores, empowered judges to be more lenient. In our interviews with judges, many mentioned a skew in decision-making: releasing someone who goes on to reoffend has larger consequences for the judge than failing to release someone who would not have reoffended. By acting as a second opinion, the risk assessment may protect judges from political backlash if someone who was classified as low risk goes on to reoffend, counteracting the aforementioned skew. (We have nicknamed this mechanism the 'Willie Horton hypothesis' after the infamous offender used in political attack ads in the 1980s.) We do not evaluate recidivism effects for sex offenders due to their lengthy sentences.[1]

There is a large literature, both in criminal justice and beyond, that has predicted substantial gains from adopting predictive algorithms.[2] However, there are few studies that evaluate their real world impacts. This small literature includes studies of algorithms in prison placement (Berk et al., 2002), parole (Berk, 2017) and pretrial detention decisions (Stevenson, 2019; Sloan et al., 2018; Albright, 2019; Cowgill, 2019) – and in some non-criminal-justice contexts such as hiring (Cowgill, 2018). We add to this literature by providing the first evaluation of how risk assessment at sentencing affects outcomes relative to the status quo.[3] This setting is important both because of its high stakes – months, years, or even decades of a person's liberty – as well as the rapid expansion of prediction algorithms in this domain. Our results highlight the importance of considering human incentives in the use of algorithms. When decision-makers follow objectives that differ from, or even conflict with, the goals that the algorithm is supposed to advance (i.e. reducing incarceration/crime by reallocating jail beds towards those at the highest risk of reoffending), algorithm adoption may lead to unexpected results. Virginia's nonviolent risk assessment reduced neither incarceration nor recidivism; its use disadvantaged a vulnerable group (the young); and failed to reduce racial disparities. Virginia's sex offender risk assessment lowered sentences for

[1]Most sex offenders are incarcerated throughout the available follow-up periods; recidivism rates thus do not tell us about offenders' actual risk.

[2]See, e.g., (Chetty et al., 2014; Berk et al., 2016; Goel et al., 2016; Chalfin et al., 2016; Jung et al., 2017; Kleinberg et al., 2018).

[3]This complements descriptive studies, both qualitative and quantitative, of how judges use risk assessments at sentencing, e.g. Ostrom et al. (1999); Garrett et al. (2019); Garrett and Monahan (2018a).

those convicted of rape: a group that the Sentencing Commission had targeted for *increased* sentences. Such outcomes would be hard to predict with policy simulations, and underline the importance of empirical evaluation.

We expect at least some our findings to be generalizable, both to other jurisdictions and to other criminal justice uses of risk assessment. For instance, an aversion to putting teenagers and young adults in jail is likely to influence decisions in many settings, including those where the sole justification for incarceration is purportedly preventive, such as bail. This suggests that prior research may have overestimated the extent to which judges are making prediction errors, and thus the potential gains from adopting them (Jung et al., 2017; Kleinberg et al., 2018). We also expect risk assessment will generally be an ineffective way of trying to control incarceration or implement policy. Judges have their own sets of priorities. Attempts to nudge them towards particular policy goals via the risk assessment could backfire: judges may ignore the risk assessment altogether or respond strategically, using it to advance their own agenda (Cowgill and Stevenson, 2019). Career aspirations and the pressures of re-election/reappointment add an extra layer of complexity (Lim et al., 2015; Berdej and Yuchtman, 2013). These incentives could interact with the risk score in a counterintuitive fashion, as our Willie-Horton hypothesis demonstrates. Unless there is a meaningful penalty that discourages judges from deviating from the action-recommendations associated with each risk score (e.g. release for low risk defendants), risk assessment is unlikely to be an effective way of implementing reform.

The paper proceeds as follows: Section 2 provides more background on risk assessment algorithms and the Virginia policy context. Section 3 presents our data and describes race/age disparities in risk assessment. Section 4 explains our empirical strategies and presents results for each of the following: (1) effect of risk assessment on judges' decisions, (2) net effect on incarceration rates, sentence lengths and recidivism, (4) effect on race/age disparities. Section 5 explores how judicial discretion to follow/ignore the sentencing recommendations associated with the algorithm impacted our findings and Section 6 evaluates two other potential explanations for why risk assessment use did not lower recidivism. Section 7 discusses and concludes.

2 Background

2.1 Algorithmic risk assessment in criminal justice

Judges' decisions are, in part, predictions about risk (e.g. the risk of reoffending, or the risk of not appearing for trial). Tools that improve the accuracy of these predictions may therefore enable better decisions. Risk assessment tools aim to do just that, and

are now used broadly in criminal justice: in determining bail and pretrial custody, probation supervision levels, offender placement within the prison system, release on parole, and in sentencing.[4] Sentencing is one of the most controversial applications of risk assessment. This is both because of the high stakes of criminal sentencing and because sentencing decisions are informed by many other goals beyond simply incapacitating those at high risk of reoffending, such as retribution or rehabilitation. Nonetheless, desire to make sentencing more efficient and consistent has prompted rapid expansion of risk assessment into the sentencing decision, particularly in the last five to ten years. Risk assessment tools are used at sentencing in 28 states; at least 7 additional states use risk assessment at sentencing in at least some counties. (See Appendix A.1 for a full list.)

Risk assessment tools are built by statistically analyzing court and police records to evaluate how various inputs correlate to measures of future offending, such as rearrest or re-conviction. The most common inputs include age, gender, and criminal history (prior convictions, arrests, periods of incarceration, etc.). Many tools also include socio-economic markers such as employment or housing status, and some also include indicators for mental health, attitudes, substance abuse, peer and family relationships. Some risk assessment tools are developed by private companies, who contract with local jurisdictions to provide their proprietary algorithms. These black-box tools are particularly controversial, since defendants cannot contest or even understand the assessments they provide. They have faced legal challenge, but courts have placed little restriction on their use.[5] Other risk assessments were developed by local jurisdictions, universities, or foundations. These algorithms are usually public knowledge.

Almost all risk assessment tools currently used today take the form of a 'weighted checklist', where inputs are assigned point values based on their statistical correlation with future offending, as measured by re-arrest or reconviction. The risk score is the sum of the points. This type of simple linear model is not able to exploit nonlinear or interactive relationships in the same way that more sophisticated machine learning algorithms do. However, evidence suggests that simple tools, with integer weights and only a handful of inputs, can rival the accuracy of complex prediction models in a wide variety of domains. In the criminal justice context, Jung et al. (2017) show that "simple rules that consider only two features – age and prior FTAs [failureto-appear in court] – perform nearly identically to state-of-the-art machine learning models (random forest and lasso regression) that incorporate all 64 available features".[6] Other authors have shown that it is possible to construct simple two-feature prediction tools that

[4]Predictive algorithms are also used in policing, but the tools and applications differ somewhat from the court context discussed in this paper.

[5]State v. Loomis, 881 N.W.2d 749 (Wisc. 2016)

[6]This tool was designed to predict future failures-to-appear in court for pretrial defendants.

perform as well as the well-known risk assessment tool COMPAS, which has access to 136 input variables, or a non-linear support vector machine trained on 7 input variables (Angelino et al., 2017; Dressel and Farid, 2018).

Defendants who score above/below certain cutoffs in the risk score will receive classifications such as low, moderate or high risk. Usually, jurisdictions will have a policy that states which intervention is recommended for defendants with different risk classifications. For instance, Virginia recommends diversion for nonviolent defendants with the low-risk designation. However, judges or other criminal justice decision-makers almost always retain discretion to deviate from the recommended actions.

2.2 Felony sentencing in Virginia

Virginia uses a voluntary sentence guidelines regime in which judges are recommended, but not required, to sentence within a particular range.[7] This system has been in place since the 1980s. In 1994, Virginia adopted a major 'truth-in-sentencing' reform act that abolished parole, mandated that offenders serve at least 85% of their sentence, and increased sentences for violent offenders. In order to free up state prison beds for violent offenders who were expected to serve longer terms as a result of the reform, Virginia also set the goal of diverting 25% of nonviolent offenders from jail or prison. Risk assessment was the proposed method of achieving this. It was to be designed and implemented by the newly founded Virginia Criminal Sentencing Commission (VCSC, or Sentencing Commission).

While most of the 'truth-in-sentencing' reforms were implemented immediately after the bill passed, risk assessment came later. The VCSC designed a new risk tool for nonviolent offenders in the late 1990s, and piloted it in 6 judicial circuits. They launched a revised risk nonviolent risk assessment statewide on July 1, 2002. This tool was in use until 2013, when several further revisions were made.[8] Our analysis is focused around the time that risk assessment was adopted statewide, and uses data from judicial circuits that did not participate in the pilot.[9]

The nonviolent risk score was developed by analyzing a randomly selected sample of 1500 nonviolent offenders who either had received a non-carceral sentence or had re-

[7]The information in this section was derived from interviews with the director of the Virginia Criminal Sentencing Commission, Meredith Farrar-Owens, as well as Ostrom et al. (1999); VCSC (2001); Ostrom et al. (2002).

[8]The revised instrument no longer includes socioeconomic markers. We find no evidence that incarceration rates, sentence lengths, racial disparities, or recidivism changed when this revised risk tool was adopted (results not shown).

[9]We have also evaluated the impacts of risk assessment in the pilot circuits and find results that are qualitatively similar to those shown here. We do not make use of the pilot period in a panel difference-in-difference design because the instrument implemented then was meaningfully different from the one later implemented statewide.

cently been released from jail or prison. It is designed to predict which defendants would be reconvicted of a felony within three years of return to the community. Its inputs include indicators for whether the charge of conviction was drug, larceny or fraud; an indicator for whether there were any additional offenses; gender; age; employment and marital status; recent arrests or confinement; prior felony convictions or adjudications; and prior adult incarcerations.[10] Based on the mandate from Virginia's legislature, VCSC chose a cutoff in the risk score that identified the 25% lowest risk defendants. (This cutoff was increased by three points on July 1, 2004.)[11] Those who score below the cutoff are recommended for diversion from jail or prison. For defendants whose guidelines-recommended sentence is prison (more than 12 months), diversion means probation or a shorter jail sentence. For defendants whose guidelines-recommended sentence is jail (less than or equal to 12 months), diversion means probation or some other non-carceral sentence.

The nonviolent risk assessment is only administered on eligible cases: a defendant must be convicted of a drug, larceny or fraud charge, and cannot have a prior or current conviction for a violent crime. Since the tool is designed for diversion, only those who were recommended for jail or prison by the sentence guidelines are eligible. The final two eligibility requirements are that defendants must not have sold large quantities of cocaine or be convicted of an offense that includes a mandatory term of incarceration.

Virginia uses a separate risk assessment tool for sex offenders. The VCSC began to develop this tool in 1999, launched it statewide on July 1, 2001, and it is still in use today. The sex offender risk score was developed by analyzing a sample of 579 felony sex offenders. A risk algorithm was trained to predict rearrest for a new sex offense or any other crime against a person within 5 years after release. The input factors include age; education; employment; relationship with the victim; aggravated sexual battery; location of offense; history of arrest for sex or other against-person crimes; prior incarceration; and prior mental health/substance abuse treatment. The sex offender risk score has three cutoffs; defendants whose score surpasses each cutoff will have the upper bound of the sentence guidelines increased by 50%, 100% or 300% respectively. The sex offender risk assessment is administered exclusively for defendants convicted of rape or sexual assault.

Both the sentence guidelines and the risk scores are calculated using a set of worksheets that are filled out by a probation officer or a prosecutor after conviction. (The risk assessment results are therefore available during plea negotiations.) These worksheets are then provided to the judge for sentencing. The cover page of these worksheets prominently displays the sentence range that is recommended by the guidelines.

[10] The pilot risk score included indicators for lack of accomplices, prior felony drug convictions, and prior juvenile incarcerations, and did not include indicators for the charge of conviction.

[11] Unfortunately, this change affected too few defendants to be able to evaluate its impacts.

A sentence is considered 'guidelines-compliant' if it falls within this range, and the midpoint of the range is what we refer to as the 'guidelines-recommended sentence'. If a judge chooses a sentence that is not guidelines-compliant, they must provide a written justification.

The risk assessment is incorporated into the guidelines by expanding the set of sentences that are considered guidelines-compliant: diversion for low-risk nonviolent offenders, and longer sentences for higher risk sex offenders. Right beneath the guidelines-recommended sentence on the nonviolent offenders' cover sheet is a checkbox with the label 'Recommended for Alternative Punishment'. This box is checked if the defendant scores below the low-risk cutoff. Sex offenders have three boxes right beneath the guidelines-recommended sentence. These boxes state that the upper bound of the sentence guidelines is increased by 300%, 100% and 50% respectively. If the person scores above one of the respective cutoffs in the sex offender risk score, one of these boxes will be checked. (Both risk assessments, and their respective cover sheets, are provided at the end of the Appendix.) If the judge is interested in the exact risk score, as opposed to merely the risk classification and its associated sentence-recommendation, they can find it by flipping through the pages of the worksheets. (As in most jurisdictions, the risk score simply consists of a number denoting an ordinal ranking among the defendants. No statistical information about the risk of reoffending is provided.) After sentencing, the final sentence is written on a separate worksheet. All sheets are mailed to the Sentencing Commission, who maintains a database on all felony sentences in Virginia.

In addition to housing records of felony sentencing, the VCSC maintains detailed records of every change to the sentence guidelines and every change to sentencing policy going back at least to 1995. This meticulous record-keeping helps to ensure that there are no important policy changes concurrent to the adoption of risk assessment that would confound our analysis. We find only trivial changes to sentencing practices for all offense categories during the time period of our analysis: fiscal years 2000-2004. The only change concurrent with the adoption of the nonviolent risk assessment is a policy shift that makes it easier for prosecutors and probation officers to access juvenile records. However, this change applies to all cases, and we have no reason to believe that this will impact nonviolent risk-assessment-eligible defendants more than anyone else. There are no changes to sentencing for sex offenders that are concurrent with the adoption of the sex offender risk assessment. However, there is a technical change in how the guidelines-recommended sentence is calculated for defendants convicted of sexual assault: the risk score is used as an input into the algorithm that calculates the guidelines-recommended sentence, corrupting one of our most important control variables. (This does not affect those convicted of rape.) For this reason we omit sexual

assault from our difference-in-differences analyses.

In this paper, we refer frequently to judges as if they are the sole decision-makers in sentencing. Of course, this is not strictly true. For one, juries determine sentences for all cases that resolved through jury trial.[12] These, however, constitute only about 2% of felony convictions. Formally, judges are in charge of sentencing for all remaining cases. They directly set the sentence in bench trials, which account for approximately 10% of felony convictions (VCSC, 2003). But the remaining felony convictions come from guilty pleas. If this entails an agreement between the prosecution and defense, then the judge's influence is less direct. A plea agreement can stipulate three things: that other charges will be dropped, that the prosecutor will make a specific (non-binding) recommendation for the sentence, or that the defendant should receive a specific sentence.[13] We don't have good data on how many cases are resolved each way, but anecdotally, stipulated sentences are common. Other actors, particularly the prosecutor, are influential in negotiated sentences. However, all negotiations occur in the 'shadow of the judge'. In other words, since the sentence must be approved by the judge, the plea-bargaining process is influenced by expectations about what type of sentence will be approved.[14] For the purposes of concision in language, and since the judge is both formally and practically influential in sentencing, we often refer to judges as the primary actors – for instance, when describing judicial discretion in risk assessment use. We acknowledge that this is not ideal, and hope that the reader will understand that prosecutorial discretion plays a large role as well.

Judges are appointed by majority vote of the Virginia General Assembly. They are up for reappointment every eight years. The reappointment procedure includes an interview with the legislative committee and then a vote in the General Assembly. Compliance with sentencing guidelines is monitored by the Sentencing Commission, reported (at the state level) in an annual report, and became formally included in the reappointment evaluation in 2009. Compliance rates have hovered around 80% since fiscal year 2000, with deviations evenly split between aggravation and mitigation.[15] Prosecutors in Virginia are called Commonwealth Attorneys, who set policy and manage operations in each individual jurisdiction. They are elected officials who serve four-year terms and can appoint Assistant Commonwealth Attorneys to carry out the day-to-day work of prosecution.

[12]The judge is authorized to deviate downward from the jury's sentence but they cannot deviate upward.

[13]Virginia Rule 3A.8.

[14]The judge can reject the plea agreement, but she must then recuse herself from the case and it will be passed on to another judge.

[15]See the Annual Reports put out by the VCSC: http://www.vcsc.virginia.gov/reports.html

3 Data

3.1 General descriptive statistics

The data used in this paper comes from two sources. The primary data source is the VCSC, which collects and maintains records on all felony sentences in the state of Virginia since 1995. This data was acquired through a public records request and contains lots of information relevant to the case, such as the charges, sentences, sentence guidelines, and risk score. In addition, it contains dozens of variables that are used to calculate the risk score and the guidelines-recommended sentence, including those pertaining to the current offense, the criminal record, and personal characteristics of the defendant. Unfortunately, neither the risk score nor the exact variables used to calculate it are available for defendants who did not receive a risk score. Therefore we do not have risk score information for eligible defendants sentenced before risk assessment was adopted, nor for defendants whose charge/criminal history makes them ineligible for risk assessment. VCSC data also does not contain the race or gender of the defendants.

Race, gender, and an alternative recidivism measure (new felony charges) were obtained by matching the Sentencing Commission data with bulk court records scraped from a public online site. For most counties, court records from as far back as the year 2000 can be found online.[16] Alexandria and Fairfax counties are an exception; we were unable to match VCSC data back to the original court records for cases in these counties. The match was conducted using the fastLink package in R, which conducts probabilistic matching across multiple variables (Enamorado et al., 2018). The variables used for the match include the first name, last name, offense date, birth month and county of arrest. With a probabilistic match threshold of 0.92, 96% of the post-2000 cases in the Sentencing Commission data (excluding Alexandria and Fairfax) found a match in the court records.

Descriptive statistics for defendants sentenced in fiscal year 2001 (which extends from July 1, 2000 through June 30, 2001, and is the last year before the sex offender risk assessment was adopted) can be found in Table 1. The first column shows data for nonviolent risk-assessment-eligible offenders. This includes all defendants who were convicted of a drug, larceny, or fraud offense; whose guidelines-recommended sentence was jail or prison; and who did not have a current or prior violent conviction.[17] The

[16] Many thanks to Ben Schoefield for the public service of scraping the data and making it available to researchers.

[17] This group is expected to be slightly over-inclusive, since we can't identify offenders who would have been ineligible due to a mandatory minimum sentence or large quantities of cocaine unless they were sentenced during the post-risk-assessment period. However, in cases for which data is available, these ineligibility criteria rule out only 2% and 1% of cases, respectively.

second column shows nonviolent offenders who were not risk-assessment-eligible, due to either a non-carceral guidelines-recommended sentence or a violent conviction. The third column corresponds to those convicted of sex offenses and the final column includes all other ineligible offense categories.

Roughly 2/3 of the sample is convicted of a nonviolent offense, and a little more than half of these are risk-assessment-eligible. Drug convictions make up the largest share of nonviolent cases, followed by larceny and then fraud. Both actual sentences and guidelines-recommended sentences are longer for risk-assessment-eligible nonviolent cases than ineligible nonviolent cases. This is because the primary eligibility criteria is whether or not the defendant was recommended for a carceral sentence by the sentence guidelines.

Sex offenders account for only 3% of the entire sample. This smaller sample imposes some restrictions on our analysis for the sex offender risk score; for instance, we do not do heterogeneity analysis for the sex offender risk score, nor do we conduct auxiliary analyses in order to better understand the results. Almost all of those convicted of sex offenses receive a carceral sentence, and many are quite long. The mean carceral sentence is 83 months, and the median is 32 months.

About 31% of felony convictions are for offense categories that are ineligible for any sort of risk assessment. Assault, burglary and traffic each constitute roughly 1/5 of these cases. Robbery accounts for an additional 10%, and the remaining convictions are either weapons or other miscellaneous offenses. The average incarceration rate for this group is 81% and the average sentence is 74 months.

Our primary measure of recidivism is an indicator for having been convicted of a new felony offense in the state of Virginia within three years of the original sentencing date. We focus on this measure because it is very similar to the target variable that Virginia's nonviolent risk assessment was trained to predict. The main difference is that our time-counter for recidivism begins at sentencing, and Virginia's time-counter for recidivism begins at release. In our recidivism measure, incarcerated defendants will be incapacitated (at least for crimes in which the victims are outside of prison walls) during a portion of the recidivism time window. This is intentional: we are curious about the extent to which risk assessment use lowers recidivism by incapacitating those at highest risk of reoffending. According to this definition, 12-16% of nonviolent offenders and 9% of offenders in the 'other' category recidivate. The three year recidivism rate for sex offenders is only 2%, due at least in part to their longer sentences. Since the incapacitation period is so long for sex offenders, and recidivism rates are so low, we do not evaluate risk assessment's impact on recidivism for this group.

The bottom panel shows descriptive statistics for the variables that were acquired from matching to the Circuit Court data set. As above, these descriptive statistics

are for cases sentenced in fiscal year 2001. 65% of nonviolent risk-assessment-eligible defendants are black. In contrast, only 39% of sex offenders are black. Almost all of the remaining defendants are labeled as white in our data. Women composed 23-29% of nonviolent offenders, 2% of the sex offenders, and 9% of other offenses. Our other recidivism measure, being charged with a new felony offense within three years of sentencing, shows higher rates than the previous measure. (Most of the felony charges that don't result in a conviction are either dismissed or are bargained down to a misdemeanor conviction.) Nonviolent offenders have a 33-40% recidivism rate; sex offenders have a 16% recidivism rate, and those charged with other offense categories have a 26% recidivism rate.

Virtually everyone convicted of a sex offense has a sex offender risk score administered once they become available. However, only about 70% of those who appear eligible for the nonviolent risk assessment actually has one administered. This is partially due to the fact that we choose a definition of risk-assessment-eligible that is consistent across the pre and post-period, and data limitations force us to be over-inclusive. However, it appears that court actors are failing to administer the nonviolent risk assessment as required in some cases. Our contacts at the VCSC report that a risk assessment is less likely to be administered when the prosecutor has negotiated a plea agreement with a stipulated sentence.

3.2 Race and age disparities in the nonviolent risk score

In this subsection we present some descriptive statistics pertaining to race and age disparities in the nonviolent risk score. While race does not specifically factor into the risk score, many of its inputs – such as the criminal record, employment, and marital status – are correlated with race. Age, on the other hand, is a direct input to the risk score. Being under the age of 30 results in 13 additional points to the risk score. As a reference point, having five or more prior adult incarcerations adds only nine points to the risk score.

Sentencing law and practice has often entailed leniency for young offenders due to the belief that their immaturity and impulsivity makes them less culpable. This leniency, however, does not extend to defendants in their late 20s. For this reason, when we are exploring risk assessments' impact on the young, we define youth as being under the age of 23. This was also the youngest age bracket in Virginia's pilot risk assessment.

Figure 1a plots the likelihood of receiving a low risk classification against the guidelines-recommended sentence, shown separately by race. The sample includes all defendants who received a risk assessment and whose guidelines-recommended sentence

is prison.[18] At all points in the guidelines-recommended sentence, white defendants are more likely to be recommended for diversion by the risk assessment than black defendants. The gap is particularly large at high guidelines-recommended sentences. Using regression, we find that white defendants are 26% (11 percentage points) more likely to be classified as low risk then black defendants with the same guidelines-recommended sentence.[19]

Figure 1b is similar, except that the sample is divided by age. Individuals under the age of 23 are substantially less likely to receive the low risk designation, particularly at higher guidelines-recommended sentences. Using regression, we find that older defendants are 43% (15 percentage points) more likely to be classified as low risk than young defendants with the same guidelines-recommended sentence.

Given the large weight that age holds in calculating the risk score, disparities across age are not surprising. But disparities across races stems solely from disparities across the inputs to the risk score. Which risk assessment factors contribute most to the race gaps in scores? We evaluate this using the decomposition method proposed in Gelbach (2016). This technique uses the omitted variables bias formula to measure how much the race gap in risk assessment score can be explained by the different inputs to the risk score, conceptualized here as 'omitted variables'. Results are presented graphically in Figure 1c, in the order of their contribution to the black-white gap in risk scores. (This figure presents the decomposition after controlling for the guidelines-recommended sentence; results are similar if controls for the recommended sentence are not included.) The top four contributing factors are prior incarceration, employment status, gender, and marriage status. It is worth noting that two of the top four factors – employment and marriage status – are socio-economic markers that do not reflect any culpable behavior. These are among the most controversial inputs to risk assessment tools, precisely because they are expected to exacerbate class- and race-based disadvantage (Starr, 2014).

4 Main results

4.1 Do judges use the risk assessment tools?

We begin by evaluating whether the risk assessment affects sentencing decisions. While judges and prosecutors have access to the raw risk score, the most salient in-

[18] Results are similar for defendants whose guidelines-recommended sentence is jail. However, the guidelines-recommended sentence is not a continuous variable for these individuals. Rather, the guidelines-recommended sentence comes in bins such as '1 day-6 months jail time', making it harder to graph.

[19] We regressed a low risk indicator on a race indicator, with a fully saturated set of fixed effects for the exact guidelines-recommended sentence.

formation is simply whether or not the defendant's risk score triggers a change to the guidelines-recommended sentence. Nonviolent offenders just below the cutoff receive a recommendation for diversion; those above the cutoff do not. Sex offenders right above the cutoff receive an expanded upper bound of the sentence guidelines; those right below the cutoff do not. (For sex offenders, we focus on the first cutoff – the one that expands the upper bound of the guidelines-recommended sentence by 50% – due to limited observations at the higher cutoffs.) Our strategy is to test for discontinuous changes in sentencing around the risk score cutoffs using a regression discontinuity design with risk score as the running variable. Our specification is shown in Equation 1, where $riskscore$ is the raw risk score, $high\_risk$ is an indicator for being above the cutoff and X includes the guidelines-recommended sentence, age, recent prior convictions, and offense. Observations i are at the case level.

$$Y_i = \alpha + riskscore_i * \beta_1 + high\_risk_i * \beta_2 + riskscore_i * high\_risk_i * \beta_3 + X_i * \beta_4 + \epsilon_i \quad (1)$$

The regression discontinuity design relies on the assumption that, with the exception of the risk classification, people who score right below the cutoff are very similar to people who score right above. This assumption is negated if there is 'gaming' in the calculation of the risk score: that is, if the person who administers it strategically assigns points to ensure that someone is rated low- or high-risk. Such strategic manipulation is often observable in an uneven bunching of scores right below or right above the cutoff, or in discontinuous changes in personal characteristics such as age, criminal history, or race.

The distribution of the risk scores is shown in Appendix Figures A.1a and b. The risk assessment has been normalized so that scores below zero fall into the lowest risk category. Nonviolent offenders who score below zero are recommended for diversion; sex offenders who score below zero do not have an expanded guidelines-recommended sentence. There is no obvious bunching around the cutoffs that would indicate strategic manipulation of the score. However, the distribution of the risk scores is lumpy, which complicates both visual evaluation as well as formal distribution tests. The lumpiness is due to the fact that integer weights are used to calculate the risk score; integers that are multiples of three are particularly common in the nonviolent risk score. Thus, we rely primarily on tests for discontinuities in defendant characteristics around the risk score cutoff to assuage concerns about strategic manipulation.

Figure 2 shows no evidence of discontinuous changes in either the guidelines-recommended sentence, the defendant's age, or recent prior convictions.[20] Panel A of Table 2 shows this more formally: using regression discontinuity, we find no evidence of a discontinuous change in gender, race, age, guidelines-recommended sentence, recent prior

[20]Note: the guidelines-recommended sentence is not altered by the risk assessment.

convictions, or the conviction offense across either the nonviolent cutoff or the first sex offender cutoff. (As noted above, we don't evaluate the higher sex offender cutoffs due to small sample size in this part of the distribution). Unless otherwise indicated, we use the optimal mean squared error bandwidth selection method proposed in Calonico et al. (2014) for all regression discontinuity estimations in this paper. None of the estimates are statistically significant, and the coefficients are small compared to their means (shown in Table 1).

We do, however, find that the cutoff delineating a change in risk classification corresponds to a discontinuous change in both the probability of incarceration and in the sentence length. We use an inverse hyperbolic sine (arcsinh) transform on sentence length both to reduce the influence of extreme outliers and to allow coefficients to be interpreted approximately as a percent-change in the outcome. These results are shown graphically (Figure 3) and in regression discontinuity (Panel B of Table 2). Defendants above the nonviolent cutoff are 6 percentage points ($p < 0.05$) more likely to be incarcerated, and receive sentences that are approximately 23% longer ($p < 0.01$). Defendants above the sex offender cutoff are 6 percentage points ($p < 0.01$) more likely to be incarcerated, and receive sentences that are approximately 34% longer ($p < 0.001$). All together, these results imply that judges and/or prosecutors do pay attention to the recommendations associated with the risk assessment, and adjust their sentences accordingly.

Of course, if the risk assessment only changed sentencing for defendants at the margins of the different risk classifications then there is no reason to expect large gains. The efficiency gains that policy simulations predicted come from releasing low risk defendants who otherwise would have been incarcerated, or incarcerating high risk defendants who otherwise would have gone free. Testing changes in sentencing for defendants at the tails of the risk score distribution would be straightforward if we had risk assessment information for defendants sentenced before risk assessment was adopted. Unfortunately this data is not available. We do, however, have access to much of the same information that is used to calculate the risk score: age, gender (for most defendants), offense and criminal history. Using this information we are able to calculate risk score predictions for nonviolent risk-assessment-eligible defendants. We use a triple-differences specification, to evaluate how risk assessment affected sentencing at various quintiles of the predicted risk score distribution. More details about this process, as well as the table of results, are available in Appendix Section A.2 and Table A1. In sum, we find that risk assessment adoption also influenced sentencing at the tails of the risk score distribution. Sentence lengths increased by about 10% for defendants in the highest risk score quintile and declined by approximately 7% for defendants in the lowest risk score quintile after risk assessment was adopted

While there is ample evidence that judges used the risk scores, compliance was far from perfect. In fact, judges followed the recommendation for diversion in less than half (44%) of the nonviolent cases for which diversion was recommended.[21] Even though the risk score is influential, judicial discretion still plays a substantial role.

4.2 Did risk assessment reduce incarceration and/or re-offending?

In this section we evaluate how the adoption of risk assessment affects average incarceration rates, sentence lengths and recidivism compared to the status quo (i.e. sentencing without risk assessment). Our primary method of analysis is a difference-in-difference research design that compares outcomes before/after the adoption of risk assessment for defendants who are/are not eligible for risk assessment. Thus we are comparing outcomes during the time period in which risk assessment is used to outcomes during the period right before its adoption, controlling for any changes over time that would impact risk-assessment-eligible and -ineligible defendants similarly. We estimate the impacts of the sex offender risk assessment and the nonviolent risk assessment jointly as shown in Equation 2:

$$\begin{aligned} Y_i = \alpha &+ Post02_i * NV\_eligible_i * \beta_0 + Post01_i * Sex\_offender_i * \beta_1 \\ &+ NV\_eligible_i * \beta_2 + Sex\_offender_i * \beta_3 + Post02_i * \beta_4 + Post01_i * \beta_5 \\ &+ X_i * \beta_6 + \psi_i \end{aligned} \tag{2}$$

Each observation i consists of a single case. $Post02$ refers to the period after nonviolent risk assessment is adopted and $NV\_eligible$ refers to nonviolent risk-assessment-eligible cases. $Post01$ refers to the time period after the sex offender risk assessment is adopted, and $Sex\_offender$ is an indicator for sex offender cases. Covariates (X) include the guidelines-recommended sentence, offense, age, recent prior charges, judge fixed effects, and fixed effects for the month, year and week-day of sentencing. Observations i are at the case level.

As discussed in Section 2, we drop defendants convicted of sexual assault; our difference-in-differences estimates for the sex offender risk assessment pertain only to people convicted of rape. In addition, we drop circuits that participated in the pilot program and thus were already using a (different) risk assessment tool. Our primary specification includes cases that were sentenced within two years before and two years after risk assessment was adopted. Since the two risk assessments were adopted in subsequent years, our sample includes the five years encompassing fiscal years 2000-2004. Motivated in part by the design-based inference methods proposed in Abadie

[21] Research by Garrett et al. (2019) suggests that judges are more likely to divert low risk defendants if some sort of alternative sanction, such as mental health treatment, is available in that county.

et al. (2017) we cluster standard errors at the level of the judge. Judges vary in the extent to which they use the risk assessment tool, meaning that defendants seen by different judges have varying levels of 'treatment'. Just as a medical trial that assigned varying intensities of treatment to different clusters of people (e.g. different hospitals) would cluster at the level of treatment assignment, we cluster at the level of judge assignment. With the exception of large counties in which there are multiple judges, this is effectively clustering at the county level.

The difference-in-differences estimates are shown in Table 3. The odd columns do not include controls; the even columns do. Columns 1 – 4 show risk assessment's impact on sentencing: both the probability of incarceration and the sentence length with an arcsinh transform. We find no evidence that the nonviolent risk assessment led to a net change in either the probability of incarceration or the length of the sentence. The coefficients are small, precisely estimated, and statistically insignificant. We find, however, that the sex offender risk assessment led to a five percentage point decline in the probability of being incarcerated ($p < 0.01$) and an approximate 24% decline in the sentence length ($p < 0.05$).

Interestingly, the net sentencing effects are quite different from what one might have expected based on how the risk assessment was incorporated into sentence recommendations. Both the nonviolent risk assessment and the sex offender risk assessment were incorporated in a unidirectional manner: to lower sentences for nonviolent offenders and increase sentences for sex offenders. Yet sentences remain the same for nonviolent offenders and went in the opposite direction for those convicted of rape. What is going on here?

Subsection 4.1 showed us that, at least around the margins, the sentencing recommendations associated with the risk classification affected sentencing decisions. But apparently judges responded as much to the absence of the recommendation as they did to the recommendation itself. To whatever extent risk assessment increased the diversion rate for nonviolent offenders who received a recommendation for diversion – those who scored below the cutoff – it must have decreased the diversion rate for those who did not. To whatever extent it increased sentences for sex offenders who were recommended for a longer sentence – those who scored above the cutoff – it must have decreased sentences even more for those who were not.

Note that neither risk assessment provides any absolute information about the statistical likelihood of recidivism. The only information it provides is relative information: information about how a defendant scores relative to the cohort. Thus the decline in the likelihood of incarceration for sex offenders is particularly interesting. It is possible that this is a simple information story: the sex offender risk assessment taught judges that defendants were less risky on average than they had previously believed. However,

this hypothesis is hard to reconcile with the fact that the risk assessment conveyed no information about the statistical likelihood of reoffending. We believe this evidence is most consistent with the 'Willie Horton' hypothesis: that judges feel more comfortable releasing defendants when they can point to the low risk-assessment score as a second opinion to support their decision. This is a theory that was expressed to us by multiple judges during the course of our research. And concern about public backlash when granting lenient sentences to those convicted of sex offenses is not just hypothetical. In 2016, Judge Aaron Persky gave only a six month jail sentence to a Stanford student who had been convicted of sexual assault. This resulted in huge public backlash. Aaron Persky became a household name, and activists mounted a successful recall movement that cost him his job. If the Stanford student had received a low risk score, this would almost certainly be cited to help defend the low sentence he received. It may not have changed the fate of Judge Persky, but it may change outcomes in more marginal situations, or in cases where an apparently low-risk offender went on to commit another crime.

Although one of the policy goals for nonviolent risk assessment – lowering jail and prison populations – was not met, the policy may still have led to a more efficient use of criminal justice resources: lowering average recidivism rates by incarcerating those at higher risk to reoffend and releasing those at a lower risk. Unfortunately, there is little evidence that this happened. Our main recidivism measure is the likelihood of a new felony conviction within three years of the initial conviction. The point estimates on recidivism outcomes, shown in Columns 5 and 6 of Table 3, are positive and statistically significant, albeit small. They suggest that the nonviolent risk assessment led to an increase of about one percentage point in the likelihood that offenders would be convicted of a new felony charge within three years. We explore potential explanations for why risk assessment use did not lead to a reduction in recidivism in Sections 5 and 6. (We do not report recidivism results for sex offenders. Their sentences are very long and their recidivism rates are very low, making it hard to detect any change.)

4.3 Did risk assessment use affect race/age disparities?

We next evaluate whether the adoption of the nonviolent risk assessment affected defendants differently by race or age. Due to small sample size we do not test for heterogeneity in the impacts of the sex offender risk score.

As Figures 1a and b show, the added information provided by the risk assessment score directs judges towards relatively harsher sentences for young and black defendants, at least compared to older and white defendants. If the sentence guidelines are a set of instructions from the state to the judge, and the risk assessments are an amendment to those instructions, this amendment is unfavorable to black people and

to young people. However, the simple fact that black and young defendants tend to have higher risk scores does not automatically mean that adoption of the risk score will lead to harsher sentences for these two groups. If judges are already sentencing in a racially biased manner, it's unclear whether or not risk assessment will make this worse. Statistical discrimination adds an extra layer of complication. Presumably, the judges are conducting some sort of intuitive evaluation of the risk a defendant poses. Lacking more detailed information about risk, they may make generalizations – either accurate or not – about offending rates based on age and race. In the presence of statistical discrimination, changing the amount of information available can have counterintuitive effects. In the hiring context, scholars have found that removing information about the criminal record lowers hiring rates for black applicants since it increases the likelihood that employers will use race as a proxy for criminal history (Doleac and Hansen, 2018; Agan and Starr, 2017). Since the risk assessment increases the amount of information available to the judge, it could lower statistical discrimination and prove net-beneficial to young and black defendants.

Theory provides no clear answers about risk assessment's expected impact on racial disparities. The little empirical research currently available in the pretrial context has documented an increase in race disparities after risk assessment was adopted, although this is largely due to regional variation in take-up (Stevenson, 2019; Albright, 2019). A vignette study conducted on a sample of practicing judges shows that risk assessment use increased socioeconomic disparities in sentencing, potentially by triggering negative stereotypes about poor people (Skeem et al., 2019). We are unaware of any prior empirical work on risk assessments impact on age disparities.

We test whether risk assessment affects defendants differently by race or age using a triple differences research design, as shown in Equation 3. Depending on the specification, Z is an indicator for being black or being under the age of 23. The rest of the variables are as described above.

$$\begin{aligned} Y_i = \alpha + Post02_i * NV\_eligible_i * \beta_0 + Post02_i * NV\_eligible_i * Z * \beta_1 + NV\_eligible_i \\ * \beta_2 + Post02_i * \beta_3 + Z_i * \beta_4 + NV\_eligible_i * Z_i * \beta_5 + Post02_i * Z_i * \beta_6 + X_i * \beta_7 + \gamma_i \end{aligned} \tag{3}$$

Our sample for heterogeneity analysis is restricted to fiscal years 2001-2004: two years before and two years after the nonviolent risk assessment was adopted. Alexandria and Fairfax counties are omitted from the racial disparities specification, since race information is not available for these jurisdictions. We drop sex offenders in order to focus on the impact of the nonviolent risk assessment. Results are shown in Table 4. We see no evidence that racial disparities in sentencing changed after the adoption of the nonviolent risk assessment. The point estimates (β_1) are small and statistically

insignificant, although the standard errors are large enough that we can't rule out a moderate (12%) increase in the sentence length. We do, however, see a relative increase in sentence severity for young defendants. Defendants under the age of 23 are four percentage points more likely to be incarcerated and receive sentences that are approximately 12% longer.

Recidivism results by race and age are reported in Appendix Table A2. As expected, standard errors on these outcomes are substantially larger than in the difference-in-differences specification. The point estimates suggest a small relative decline in recidivism for young offenders, consistent with an incapacitation effect due to increased incarceration. However, the estimates are too noisy to draw clear inference.

4.4 Robustness tests and event-study graphical analysis

The 'parallel trends' assumption under which the difference-in-differences estimates can be interpreted causally is that trends in outcomes between treated and control group would have been parallel after the reform had it not been for the adoption of risk assessment. One potential challenge to this assumption is that prosecutors may have changed charging practices in response to the adoption of risk assessment, thus changing both the frequency and composition of risk-assessment-eligible cases at the same time that risk assessment was adopted. As demonstrated in Appendix Figure A.2, we find no evidence that the frequency of either nonviolent risk-assessment-eligible cases or rape cases changed when their respective risk assessments were adopted. Furthermore, by comparing even and odd columns in Table 3 we find that the estimates are stable to the inclusion of covariates, even though these covariates increase the R^2 substantially. While not dispositive, this eases concerns about changes in the composition of cases occurring at the time of risk assessment adoption.

In order to evaluate whether outcome trends were parallel before risk assessment adoption, and to verify that the decline in sentencing for sex offenders was coincident with the adoption of risk assessment, we add lag and lead indicators to our main specification. In particular, we generate dummy indicators for each fiscal year 1999-2006, as well as interactions between these dummies and risk assessment eligibility. This specification is shown in Equation 4, where *laglead* indicates the eight year dummies and the other variables are as defined above.

$$\begin{aligned} Y_i = \alpha + laglead_i * NV\_eligible_i * \beta_0 + laglead_i * Sex\_eligible_i * \beta_1 \\ + NV\_eligible_i * \beta_2 + Sex\_eligible_i * \beta_4 + laglead_i * \beta_5 + \gamma * X_i + \epsilon_i \end{aligned} \tag{4}$$

Figures 4a-e shows coefficient plots for β_0 and β_1, where the dummy for the year before risk assessment adoption is dropped to serve as a baseline. While there is some year-on-year noise, there is no evidence of divergent trends between those who are/are

not eligible for the risk assessment before the reform. Nor do we see any evidence that average sentences or recidivism for nonviolent offenders changed after the reform, consistent with the small, mostly statistically insignificant coefficients shown in Table 3. For sex offenders, we see a decline in both the probability of incarceration and the sentence length in the years immediately after the reform.

We also conduct an event-study style graphical analysis of the triple-differences specifications, with the sentence length as the outcome. In particular, we substitute *post*02 in Equation 3 with a set of lag/lead dummy variables in order to get a better sense of the timing of the changes. The age specification includes eight year-dummies, from fiscal year 1999 to 2006; in the race specification we are only able to document trends from 2001 to 2006. Figures 5a and b present the results. There is no visual evidence of diverging outcomes in the years prior to risk assessment adoption. However, there is an increase in the sentence length for defendants under the age of 23, coincident with the adoption of risk assessment.

We conduct a variety of tests to ensure that our estimates for risk assessment's impact on average sentencing and recidivism are robust to various ways of defining the sample. These results are shown in Appendix Tables A3 and A4. First, we show that our results are robust to varying the sample time window. In our main specification, we used 5 years of data. Results are generally similar using 3, 7, and 9 year samples as well. We then showed that our results are robust to different ways of defining the control groups. We select two alternative control groups for nonviolent risk assessment eligible defendants: those convicted of a nonviolent offense but who were risk-assessment-ineligible, and those who were convicted of an offense type for which there was no risk assessment. We select one alternative control group for the sex offenders: defendants who were convicted of violent offenses. Results are qualitatively similar using these alternative control groups.

We also use alternate measures of recidivism to evaluate whether our estimates of risk assessment's impact on recidivism are robust. In particular, we test to see if risk assessment led to a change in new felony convictions for crimes committed within various time windows after sentencing (6 months and 1, 3, 5, or 7 years). We also use an alternative measure of recidivism: having new felony charges for offenses committed after sentencing within the same five time windows listed above. This outcome is only available for the cases in which we were able to match to the court data, thus we drop fiscal year 2000 as well as Alexandria and Fairfax counties. Results are shown in Appendix Table A5. The top panel uses the full sample and the new felony conviction measure but with varying time windows. The middle panel continues to use the new felony conviction measure with varying time windows, but the sample is restricted to exclude Alexandria and Fairfax. The outcome measure in the bottom panel is the

likelihood of receiving a new felony charge within varying time windows; Alexandria and Fairfax are dropped because data is unavailable. Across these 15 specifications, few estimates are statistically significant. There is some evidence that risk assessment use may have led to a decline in the likelihood of receiving a new felony charge within five or seven years, although the estimates are only statistically significant at the 10% level. However, the coefficient on the sentence length is positive in this subsample, suggesting that sentence lengths may have increased by about 4%, again significant at the 10% level. If sentences increase slightly then recidivism may decline mechanically due to an incapacitation effect.

Across a variety of specifications, there is little evidence that the nonviolent risk assessment led to a more efficient use of jail and prison beds. The specification most favorable to risk assessment suggests that it led to a slight decline in the likelihood of receiving new charges within 5-7 years. However, average sentences increased in this subsample, so it would be difficult to conclude that this is due to improved information about risk. The evidence does not support the type of dramatic gains that policy simulations of risk assessment had predicted.

5 The role of discretion

In the previous section, we saw that the real-world impacts of risk assessment often differed from what was expected. In order to better understand the results, this section explores how judges/prosecutors use their discretion to follow or ignore the recommendations associated with the risk assessment. Our analysis focuses on the nonviolent risk assessment due to the small sample of sex offenders. First, we test to see what factors predict deviation from the algorithm. Second, we conduct simulations to see how race/age disparities would have changed if judges had fully complied with the sentencing recommendations associated with the algorithm. Third, we conduct an exploratory analysis to see how sentencing and recidivism changed in the subset of courts where risk assessment appeared most influential. Finally, we evaluate risk assessment's influence over time.

5.1 What factors predict deviation from the algorithm-recommended sentence?

While Subsection 4.1 showed that sentencing was influenced by the risk assessment, judges still made ample use of their discretion to ignore the sentencing recommendations associated with the algorithm. In this section we consider several hypotheses that may explain why judges deviate. First, judges may ignore the risk assessment if they

believe a defendant to be higher or lower risk than the risk score says. If their beliefs are correct, these deviations would be a beneficial complement to the algorithm. If they are incorrect, they would reduce the efficacy of the tool.

Judges may also deviate from the risk score if they believe, either consciously or unconsciously, that it is mis-calibrated for people of different races. For instance they may be less likely to trust a low-risk classification for black defendants than for white defendants. Alternatively, they may believe that the risk assessment is biased against black defendants, and may try to compensate for this by treating black defendants more leniently than white defendants with the same risk score.

Finally, judges may deviate from the risk score when they are considering factors besides risk when making the sentencing decision. For instance, judges may be considering culpability, external effects on family/community, or whether some other characteristic of the case evokes mercy.

We explore these different hypotheses by testing which factors predict deviation from the risk assessment. In particular, we regress a dummy that is equal to one if the judge grants the defendant a diversion from jail or prison on each factor, controlling for the exact risk score and the exact guidelines-recommended sentence. The top panel shows results for defendants with a low risk score: those who received a recommendation for diversion. The bottom panel shows results for defendants who did not receive a diversion recommendation. The first factor we consider is an alternative risk score that we build ourselves using a random forest model on a sample of nonviolent offenders sentenced during the two years before risk assessment was adopted statewide.[22] If judges are deviating because they have risk information not captured by the real risk assessment tool, this might be detectable with our alternative measure. However, as can be seen in Column 1 of Table 5 we find no evidence that recidivism risk, as measured by our alternative prediction tool, is correlated with the decision to deviate.

We next test to see how race affects the likelihood of deviating. Column 2 shows that judges are about three percentage points less likely to grant diversion to black defendants in the highest risk category than to white defendants with similar risk scores and guidelines-recommended sentences. The difference is not statistically significant in the low risk sample.

We then test to see how employment status, young age and gender affect judges' decision-making, since these factors may evoke more mercy-based rationales in sentenc-

[22]Like the real risk assessment, our target variable is the likelihood of a new felony conviction within three years. Our input variables include age, gender, offense, the total score for each of the three sentence-guidelines worksheets, and all the criminal history variables from the first page of the worksheets. This includes additional current offenses, prior against-person felonies, prior drug felonies, prior property felonies, prior convictions, prior incarcerations, recent legal restraints, prior juvenile incarcerations, prior misdemeanor convictions, and prior weapons offenses. We restrict the training sample to those who received a non-carceral sentence since recidivism is unobserved during the time period of incarceration.

ing. Judges may feel more sympathy for a defendant who engages in illegal commerce due to poverty. Judges may choose leniency for young defendants either due to their more limited culpability, or the adverse consequences that incarceration may have on the impressionable. Finally, judges may hesitate to incarcerate women due to concerns about separating them from their children. Consistent with the mercy hypothesis, we find that judges are more likely to divert defendants who are unemployed, young or female. The effects are particularly pronounced for young defendants: defendants under the age of 23 are 6-7 percentage points ($P < 0.001$) more likely to be diverted than older defendants. This is true both in the low risk sample, where the mean diversion rate is 44%, and the high risk sample, where the mean diversion rate is only 16%.

Our hypotheses are exploratory, but the regression results demonstrate that deviation from the risk score is non-random. Judges are choosing to follow/ignore the risk assessment because other factors are at play in their decision, with disparate impact on different demographic groups. The risk score itself embeds disparities. But judicial discretion in its use can both exacerbate them (black defendants) and reduce them (young defendants).

5.2 How would race/age disparities have changed if there was full compliance with the algorithm?

In this subsection we conduct a simulation to see how race/age disparities would have changed if judges had fully complied with the sentencing recommendations associated with the nonviolent risk assessment. Since the thought experiment we are interested in is the comparison between sentencing done by human intuition and sentencing done by algorithm, we develop the following guidelines for our simulations. First, 'sentencing by algorithm' entails a uniformity requirement, meaning that differences across defendants and cases only impact sentencing through their impact on the risk score and the guidelines-recommended sentence. Thus, a person with the same risk score and guidelines-recommended sentence would receive the same sentence, regardless of race, age, gender, or any other personal characteristic. 'Sentencing by algorithm' also requires a policy that maps the risk score to the sentence. Here we follow Virginia policy: for nonviolent offenders whose score is below the low-risk cutoff, our simulations assign a non-carceral sentence. (In alternative specifications, we revise this so that full compliance is defined as a 6 or 12 month jail sentence for those whose guidelines-recommended sentence is prison.) For defendants whose score is above the low-risk cutoff, we simply impose the uniformity requirement, and assign them a predicted sentence based on their risk score and guidelines-recommended sentence.[23]

[23]Since these predictions are generated by training a model on actual sentences, we are deferring to judicial discretion in the extent to which risk scores should influence sentencing in the absence of specific

We use the same triple-differences specifications described in Sections 4.2 and 4.3 to estimate the effect of full compliance with the algorithm. In these simulations, the sentence of defendants who received a risk score is replaced with the sentence they would have received if judges had fully complied with the algorithm. Defendants who did not receive a risk score – either because they were sentenced before risk assessment was adopted, or because they were not risk-assessment-eligible – do not have their sentences altered. The simulated sentences are summarized below:

- Nonviolent offenders whose risk scores are below the low-risk cutoff are assigned non-carceral sentences (sentence length = 0).
- Nonviolent offenders whose risk scores are above the low-risk cutoff are assigned sentences based on the predictions of a random forest algorithm whose only inputs are the guidelines-recommended sentence and the risk score.[24]
- Defendants who did not receive a risk score were assigned their actual sentence. This includes all defendants sentenced before the nonviolent risk assessment was adopted as well as those who did not receive one because they were ineligible.

Using the triple-differences specification shown in Equation 3, with our simulated sentences as the outcome variable, we estimate how full compliance with the algorithm would have affected race/age disparities in sentencing. These results are shown in Table 6. First, we see that the relative probability of incarceration for black defendants would have gone up by 3.7 percentage points relative to white defendants. The coefficient on the sentence length is not statistically significant, but corresponds to an approximate 8% increase in black sentences relative to white.

The estimates for age disparities are striking. The relative probability of incarceration for young defendants would have increased by 15 percentage points, and relative sentence lengths for young defendants would have increased by approximately 45%.

These simulations suggest that, even though age disparities increased after risk assessment was adopted, judicial discretion minimized the full impact on young people. Young age is one of the most important predictors of future offending and, accordingly, is given large weight in virtually every risk assessment tool (Stevenson and Slobogin, 2018). If the goal at sentencing is to prevent future crime by incarcerating those who pose the highest risk of committing it, then jails and prisons should be full of young people. Sentencing by algorithm would achieve just that. Yet many would argue that this is undesirable. There is a long history of leniency in criminal justice for teenagers and young adults. Every state has a separate justice system for people below a certain

recommendations. However, since the only predictors are the risk score and the recommended sentence, we are imposing the uniformity requirement described above.

[24]This model is trained on all defendants who received a nonviolent risk score in fiscal years 2003-2004.

age, usually 18 years. Recent movements have pushed to increase the maximum age for juvenile justice even further. Vermont will raise the age for juvenile justice courts to 20 years by 2022, and Massachusetts, Illinois, and Connecticut have proposals to raise the age to 21.[25] A number of states have youthful offender' provisions that call for more lenient treatment of offenders tried in adult court who are under a certain age (e.g., 21 or 25).[26] Leniency for youthful offenders is often motivated by the idea that young people are less culpable. Their brains are still developing; they are more impulsive, more susceptible to peer influence. Many will simply grow out of the lawbreaking phase. Given this history, it seems unlikely that judges are simply making prediction errors when they divert young people. Rather, they are pursuing goals that are in conflict with risk-based sentencing. If one of the most important inputs to risk assessment is effectively 'off-limits', then the risk assessment is handicapped. Expectations as to the extent to which its use will increase efficiency should be revised downward.

The results for age disparities are very robust. Regardless of how 'full compliance' is defined, it results in a large relative increase in sentences for young defendants. The results for race disparities are less robust. If full compliance is defined so that defendants are assigned shorter jail sentences (6 or 12 months) if they are recommended for diversion but have a guidelines-recommended sentence of prison, then the coefficients on race disparities shrink and lose statistical significance. This is because racial disparities in the algorithm are amplified when policies map low risk scores onto sentences that are very different from high risk scores.

Why do simulations suggest that full compliance with the risk algorithm would have disadvantaged young people so much more than it would have disadvantaged black people? Both groups have high risk scores. The key is 'compared to what'. If judges sentence in a racially disparate manner without a formal risk assessment, risk assessment tools may not increase the race gap. Age is different. When judges have discretion, they do not sentence young offenders as severely as risk assessments recommend.

5.3 Analyzing effects for the 'most responsive' judges

In this subsection, we conduct an exploratory analysis to evaluate how risk assessment affected sentencing and recidivism among the subset of judges that appeared to use risk assessment most. In some senses, this is similar in spirit to our previous exercise. We are engaging in a thought experiment to evaluate how sentencing would change if the algorithm played a greater role. There are two important differences, however.

[25] https://chronicleofsocialchange.org/youth-services-insider/juvenile-justice-raise-the-age-vermont-missouri-state-legislation/31430

[26] See, e.g., FLA. STAT. 958.04; GA. CODE. ANN. 42-7-2; N.J. STAT. ANN. 30:4-148 (West 2018).

First, we are testing what judges actually do, not just a simulated hypothetical. Judges still have, and utilize, discretion to deviate from the risk assessment at will. Second, we can evaluate recidivism effects in courts that appear to use risk assessment most. We refrain from doing this in our simulations due to the highly speculative nature of inferring counterfactual criminal activity for defendants who were incarcerated.

Qualitative research has found that Virginia judges vary significantly in their views of risk assessment. In a statewide survey only half "always" or "almost always" consider the results of the nonviolent risk assessment in sentencing. In contrast, 38% rely "primarily on judicial experience" when making decisions (Monahan et al., 2018). As one particularly dry Virginia judge puts it "I also don't go to psychics" (Garrett and Monahan, 2018b). Our strategy entails measuring judge-responsiveness by the magnitude of the discontinuity around the low risk cutoff. In particular, we generate regression discontinuity estimates at the judge level: judges for whom the RD estimate is large are classified as more responsive to the risk score. We then redo both the difference-in-differences and the triple-differences analyses from Section 4 on the subsample of 'most responsive' courts.

Table 7 shows results from the subset of judges that appear to use risk assessment most. This subset includes all judges whose RD estimate is greater than the median in a regression discontinuity with a bandwidth of seven and probability of incarceration as the outcome.[27] Most of the results look similar to what was shown with the full sample. There is no evidence that the nonviolent risk assessment affected average sentencing or that it reduced recidivism. Sentences declined for sex offenders and increased for the young. Note, however, that the increase in sentences for the young is of similar estimated magnitude to the full sample, and substantially less than would be seen in full compliance. Even though judges are using the risk assessment more, they still use their discretion to minimize its impact on young people. This could help explain why risk assessment use did not lead to a detectable decline in recidivism.

The most striking difference between this subset analysis and the full sample is that here we see an increase in racial disparities. The probability of incarceration for black defendants increased by about four percentage points ($P < 0.10$) relative to white defendants, and the length of the sentence increased by approximately 17% ($P < 0.05$). This increase is partially due to racial disparities in the risk score, but partially due to the fact that, as in the full sample, judges are more likely to deviate downward for white defendants with high risk scores than black.

How should we interpret these results? In particular, how should we interpret the provocative finding that racial disparities increased after risk assessment was adopted

[27]In alternative specifications, we vary between bandwidths of 4, 7, and 10, and use the sentence length as the outcome in the RD regressions. The full set of results for these different specifications is shown in Appendix Table A6 and is qualitatively similar to the chosen specification.

among the 'most-responsive' judges? We acknowledge that our measure of risk-assessment responsiveness is noisy. We are partially selecting our subsample based on risk assessments' impact and partially based on idiosyncratic factors that led to higher sentences for defendants just above the threshold. This noise is likely to increase type II errors in the difference-in-differences and triple differences specifications. However, with the finding of increased racial disparity, we are more concerned about type I errors. Does the fact that we are selecting partially on noise around the cutoff increase the likelihood of false positives? We don't believe so. We see little reason why this noise should be correlated with a change in sentencing trends across races for nonviolent offenders at the time of risk assessments' adoption. If we are selecting on a factor that is orthogonal to treatment it should not increase the likelihood that we falsely reject the null.

That being said, what sort of external validity does this result have? Judges who respond more to the risk assessment algorithm may differ in other ways from the less responsive judges.[28] Thus, while we consider the increase in racial disparities documented in this subsection to be a red flag on a very important issue, we hesitate to draw more general conclusions about how intensity of risk assessment use would impact people of color.

5.4 How does risk assessment use change over time?

Finally, we inquire as to how risk assessment use changes over time. Evaluating use over time provides indirect evidence on how useful judges find the tool. It also provides evidence about the extent to which our difference-in-differences and triple-differences results generalize to more recent time periods in Virginia. If risk assessment use increases, then our estimates from Section 4 might understate the impact of risk assessment. If risk assessment use decreases, then our results might be thought of as an upper bound on risk assessment's impact.

Our primary metric for evaluating the extent to which risk assessment was influential is the magnitude of the sentencing discontinuity around the high risk cutoff. Thus, to examine risk assessment's use over time, we analyze changes in the magnitude of the discontinuity over time. Specifically, we conduct regression discontinuity estimates as described in Equation 1 on rolling two-year samples: first fiscal years 2003-2004; then fiscal years 2004-2005; and so forth. Figure 6a shows results with sentence length as the outcome and using a bandwidth of 7; we find similar results with bandwidths of 4 and 10, as well as with using the probability of incarceration as the outcome. Discontinuities are greatest in the first two years of risk assessment use: fiscal years 2003-2004,

[28] We have investigated to see whether judge responsiveness to the risk assessment correlates with the judges' experience, sentencing severity, or level of sentencing disparity across races. In general, the results are not robust enough to warrant inclusion.

the leftmost coefficient shown in the plot shown in the plot. By fiscal years 2004-2005 the discontinuity has already declined. While there is some year-on-year variation, the 95% confidence interval includes zero in most of the subsequent time periods.

We provide an alternative metric for evaluating the extent to which judges use the risk assessment: the fraction of low-risk defendants who receive a diversion. Figure 6b plots this by year from 1998 (the adoption of the pilot risk assessment) through 2017. There is a gentle downward trend in the fraction of low risk defendants granted diversion: 45-50% in the earlier years down to 35-40% in the later years

Judges appear to respond most to the tool when it was first introduced. This could be because the tool was particularly salient in the early years: new things generate attention, and trainings brought reminders and encouraged use. The decline may also be because judges tried it but did not find it to be useful.

6 Other challenges to risk assessment's efficacy

If the risk assessment had worked as intended, the change in the distribution of those incarcerated (shifting jail beds towards those at the highest risk of reoffending) should have led to a net decline in reoffending. There is no evidence this happened. In the previous section we proposed one potential explanation: a reluctance to incarcerate young defendants despite their high statistical risk of reoffending. In this section we explore two alternative hypotheses for why we didn't see a decline in recidivism. First, we inquire as to whether the criminogenic effects of incarceration effectively canceled out the incapacitative effects for higher risk defendants. Second, we ask whether the risk assessment tool simply wasn't very good - and whether it's possible to build a better one.

6.1 Did criminogenic effects cancel out incapacitation?

Some scholars have argued that incarceration is criminogenic, i.e. that the experience of incarceration makes someone more likely to commit crime after release. This could be due to peer effects during incarceration, the decay of human capital, or difficulties in finding employment after release (Bayer et al., 2009; Stevenson, 2017). If incarceration is criminogenic, then the net effects of incarcerating more high-risk people are ambiguous, and could even lead to an increase in future offending. Thus one potential explanation for why risk assessment led to no detectable reduction in recidivism is that the criminogenic effects of incarceration for high-risk-score defendants effectively canceled out the incapacitative effects. We explore this hypothesis in this subsection.

Prior research on the criminogenic effects of incarceration has come to varying conclusions. Most papers that have attempted to identify the causal impact of incar-

ceration on post-release offending have exploited random assignment to judges with varying propensities to incarcerate. Several studies using this method have found evidence supporting the criminogenic channel (Mueller-Smith, 2015; Aizer and Doyle, 2009), but others have not (Kling, 2006; Loeffler, 2013). Another recent study, using regression discontinuity with a sentence guidelines score as the running variable, found no evidence for a criminogenic effect (Rose and Tov, 2018).

We return to our discontinuity-in-risk-score specification described in the Equation 1 to provide new evidence on how incarceration affects recidivism. If the criminogenic effect dominates the incapacitative effect then we would expect to see increased recidivism rates for defendants right above the low-risk cutoff, since these defendants were more likely to be incarcerated and had longer sentences. Table 8 presents reduced form results as well as instrumental variables estimates for the impact that sentence length has on recidivism, using sentencing discontinuities around the low-risk threshold as an instrument. We use new felony charges as our recidivism measure. Even though this means we need to drop Alexandria and Fairfax counties, the higher recidivism rates using this measure increases our power to detect effects.[29] Our sample includes nonviolent offenders with a risk score in fiscal years 2003-2004. Columns 1-2 shows how the high-risk classification increases the probability of sentence and the length of the sentence (with an arcsinh transform) for defendants at the margin. Columns 3-5 show reduced form estimates where the outcome in Equation 1 is recidivism within a one, three, and seven year window. Columns 6-8 shows results from the fuzzy RD estimation, where the endogenous variable is the sentence length, i.e. Column 2 is the first stage.

If incarceration reduces criminal activity through incapacitation but then subsequently increases it through its criminogenic effects, we expect to see coefficients that are negative for shorter recidivism time windows and positive for longer ones. Instead, we see only negative coefficients. The reduced form estimates show that defendants whose score is right above the low-risk cutoff are 8 percentage points less likely to have been charged with a new felony 3 years after sentencing ($P < 0.05$), and 6 percentage points less likely to be charged with a new felony 7 years after sentencing ($P < 0.05$). The five-year outcome is shown graphically in Figure 6a.

The IV estimates suggest that a doubling of the sentence length (from a mean of about 11 months) leads to a 35 percentage point decline ($P < 0.10$) in the likelihood of being charged with a new felony within three years, and a 26 percentage point decline ($P < 0.10$) within seven years.

In Appendix Table 8 we conduct the same analysis with our sample limited to defendants under the age of 23. Our sample size is substantially smaller. The results are

[29] Regression discontinuity results with new felony convictions as an outcome are statistically insignificant.

not statistically significant, but the coefficients are negative and of a similar magnitude to what was shown in Panel B.

In sum, incarceration lowers recidivism for compliers in our sample. If incarceration has a criminogenic effect, it is not large enough to cancel out the crime-reducing channels of incapacitation and/or specific deterrence. Of course it's possible that incarceration is criminogenic only for the type of high-risk-score defendants who saw their sentences increase after risk assessment was adopted. We can't rule this out, but we find no evidence to support this theory.

6.2 Is the risk assessment tool just bad at predicting recidivism?

Perhaps the risk assessment did not bring about the desired results because it provided poor information about recidivism. This could be for several reasons. First it was trained on data that suffers from sample selection issues (Bushway and Smith, 2007). Since recidivism data is not available during the time period when offenders are incarcerated, researchers need to either drop them from the training model or include them after they were released - in which case they a) are older, and b) have different life circumstances, networks, and noncognitive traits due to incarceration. We refer to this as the 'missing data' problem. Second, the risk score uses integer weights in a simple algorithm developed through logistic regression, and therefore would not be able to pick up any interactions or nonlinearities that might help improve prediction. Third, even without problems of missing data or model constraints, recidivism may simply be hard to predict.

The traditional way of evaluating a risk score, known as a 'validation study' in the risk assessment literature, is to test how well it predicts recidivism among released defendants. Since risk assessments are trained on released defendants, it is expected to perform better on this group. However, risk assessments are expected to be less accurate for the type of defendant who is unlikely to be released, since the missing-data problem is more relevant to this group. If the predictive power of the tool declines substantially in the probability of incarceration, then the risk assessment tool may be only one-way informative: it provides good information about which released defendants have the highest recidivism risk, but bad information about which incarcerated defendants pose a low risk. Since the nonviolent risk tool is designed to identify the latter - to identify low risk defendants for diversion from jail or prison - this concern is of particular importance.

We conduct several exercises to evaluate how well Virginia's nonviolent risk tool predicts recidivism. As a benchmark, we use an alternative risk assessment tool that

we built using a random forest model. While we can do nothing about selection issues in the data, the random forest model is better equipped to identify nonlinear and interactive relationships between the predictor variables and recidivism. Our goal is to evaluate how well the real risk assessment tool sorts defendants by crime propensity as compared to our alternative risk tool.

The alternative risk score is trained on nonviolent offenders who received a non-carceral sentence during the two years before risk assessment was adopted statewide. Like the real risk assessment, our target variable is the likelihood of a new felony conviction within three years. Our input variables include age, gender, offense, the total score for each of the three sentence-guidelines worksheets, and all the criminal history variables from the first page of the worksheets.[30] Using this model, we generate out-of-bag alternative risk scores for all nonviolent offenders sentenced in fiscal years 2003-2013.

The first thing worth noting is that, even in ideal circumstances, recidivism is hard to predict. When we regress the recidivism dummy on indicators for each decile of our alternative risk score, the adjusted R^2 is less than 0.03. *This is using our training data, in which the risk prediction is optimized to perform well.* The same regression applied to the test sample of defendants with a non-carceral sentence in fiscal years 2003-2004 yields an adjusted R^2 of less than 0.02. If we run a similar regression in fiscal years 2003-2004, with the deciles of the real risk score instead of the deciles of our alternative risk score, we also find an adjusted R^2 of less than 0.02. If it were possible to know the counterfactual crime rate for incarcerated defendants, and include them in our regression, the R^2 would almost certainly go down even further due to the sample-selection issues outlined above.

Despite the low R^2, both the real risk score and our alternative risk score successfully sort defendants by recidivism risk, at least among the sample of defendants who receive a non-carceral sentence. The average recidivism rate for defendants in the highest risk decile is about 14 percentage points higher than recidivism in the lowest risk decile ($P < 0.001$, results not shown). This is true for both the real and our alternative risk score.

The real question is how accurately the risk score predicts recidivism among the type of defendants who receive carceral sentences. We cannot answer this question fully. But we can provide an estimate for risk assessment's accuracy among a group of defendants who are unlikely to receive completely non-carceral sentence: those who are on the margins of being sent to prison. (By definition, prison entails a sentence of at least 12 months.)

[30] This includes additional current offenses, prior against-person felonies, prior drug felonies, prior property felonies, prior convictions, prior incarcerations, recent legal restraints, prior juvenile incarcerations, prior misdemeanor convictions, and prior weapons offenses.

Our method uses an instrumental variables technique. In particular, we exploit discontinuities in a sentence guidelines score that determines whether or not the defendant will be recommended for a prison sentence, known colloquially as the 'in/out' score. Defendants who score above a certain cutoff in the in/out score will be recommended for prison, defendants who score below that cutoff will not. We use this discontinuity as an instrument to develop estimates of recidivism risk for defendants around the margins of being recommended for prison. Here, recidivism risk is measured as the new felony charges that would be averted if a group of defendants were incapacitated from committing crime by incarceration. Incarceration averts more crimes for defendants who pose a higher recidivism risk and fewer crimes for those that pose a low recidivism risk.

To evaluate how well Virginia's risk assessment sorts defendants at different levels of recidivism risk, we first sort them into three terciles of the risk score. We use regression discontinuity with the 'in/out' score as the running variable to develop estimates of recidivism risk (i.e. incarceration's impact on the likelihood of receiving new felony charges) for defendants in each tercile. We then evaluate the gap between the recidivism risk of defendants in the lowest and highest terciles. If the risk score successfully sorts defendants, recidivism risk for the lowest tercile should be substantially lower than recidivism risk for the highest tercile. We then repeat the analysis for our alternative risk score: we divide the sample into terciles according to this alternative metric, run the RD on each tercile, gather estimates of recidivism risk, and compare the recidivism risk of the highest tercile to that of lowest.

This method is a little complex and we leave the details to Appendix A.3. Using Virginia's risk assessment we find that recidivism risk for defendants in the highest tercile is about 13 percentage points higher than for defendants in the lowest tercile. Using our risk score, recidivism risk is about 14 percentage points higher in the highest tercile compared to the lowest. The results indicate that both the real risk score and our alternative risk score successfully sort defendants by recidivism risk, even among a sample that is likely to receive a carceral sentence.

To summarize: we find no evidence that Virginia's nonviolent risk assessment is impaired by its use of integer weights in a simple algorithm. In all tests, the real risk assessment performed similarly to our alternative risk assessment generated with a random forest algorithm. Both tools were able to successfully sort defendants with a non-carceral sentence by recidivism risk. Both tools were able to successfully sort defendants at the margins of a prison sentence by recidivism risk. But neither tool explained more than a tiny fraction of the variation in recidivism. This is due to structural issues, not to sample selection issues in the training data. Demographic data, employment/marital status, and the criminal record provide only limited information

about the likelihood someone will offend in the future.

7 Conclusion

In this paper, we document how the use of risk assessment affected judicial decision-making and recidivism. Despite being adopted with an express decarceral purpose, Virginia's nonviolent risk assessment did not result in lower incarceration rates or sentencing. Instead, incarceration was reallocated from those with lower risk scores to those with higher risk scores. Theoretically, this should have reduced recidivism rates. We find no evidence that this was the case. This can be at least partially explained by the use of judicial discretion. Although risk assessment use did lead to a relative increase in sentencing for young defendants, this increase was nowhere near as large as it would have been if judges had fully complied with the sentencing recommendations associated with the algorithm. Judges used their discretion to systematically divert young offenders, despite their higher risk of reoffending. This is likely due to long-standing practices of treating youth as a mitigating factor, as opposed to an error in prediction. Since young age is one of the most important predictors of reoffending, a preference for leniency for this group will curtail risk assessment's expected benefits.

The disappointing recidivism results may also be partially explained by structural challenges in predicting offending. Even in optimal circumstances, risk predictions explain only a tiny fraction (low R^2) of the variation in reoffending. Furthermore, the data used to train risk assessment algorithms suffer from a missing data problem: recidivism rates for incarcerated defendants are unobserved during the period of incarceration. While we find that Virginia's risk assessment successfully sorts at least some defendants by recidivism risk, it's unclear the extent to which this provides information that wasn't already known by judges. Over time, the use of risk assessment appeared to decline, suggesting that judges did not find it useful.

We find mixed results on racial disparities. We find no evidence that risk assessment affected racial disparities in sentencing statewide, although we cannot rule out moderate increases in the sentence length. However, we find that racial disparities increased in the courts that appear to use risk assessment most.

The sex offender risk assessment was incorporated into the sentence guidelines to authorize an *increase* in sentences. But its use led to a net *decrease* in sentences. Even though judges increased punitiveness for sex offenders with higher risk scores, the increased leniency for sex offenders with low risk scores proved dominant.

In sum, we find that the real-world impacts of risk assessment differs from what many had anticipated, in large part because the incentives of human decision-makers were not taken into account. Policy simulations and debates about hypotheticals pro-

vide no replacement for on-the-ground evaluation.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge, "When Should You Adjust Standard Errors for Clustering?," Working Paper 24003, National Bureau of Economic Research November 2017.

Agan, Amanda and Sonja Starr, "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 191–235.

Aizer, Anna and Joseph J. Jr. Doyle, "Juvenile Incarceration, Human Caoital and Future Crime: Evidence from Randomly-Assigned Judges," *National Bureau of Economic Research Working Paper*, 2009.

Albright, Alex, "If you give a judge a risk score: evidence from Kentucky bail decisions," Working Paper 2019.

Angelino, Elaine, Nicholas Larus-Stone, Daniel Alabi, Margo Seltzer, and Cynthia Rudin, "Learning Certifiably Optimal Rule Lists for Categorical Data," *arXiv e-prints*, April 2017, p. arXiv:1704.01701.

Arnold, David, Will Dobbie, and Crystal S Yang, "Racial Bias in Bail Decisions*," *The Quarterly Journal of Economics*, 05 2018, *133* (4), 1885–1932.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen, "Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections*," *The Quarterly Journal of Economics*, 02 2009, *124* (1), 105–147.

Berdej, Carlos and Noam Yuchtman, "Crime, Punishment, and Politics: An Analysis of Political Cycles in Criminal Sentencing," *The Review of Economics and Statistics*, 2013, *95* (3), 741–756.

Berk, Richard, "An impact assessment of machine learning risk forecasts on parole board decisions and recidivism," *Journal of Experimental Criminology*, Jun 2017, *13* (2), 193–216.

Berk, Richard A., Susan B. Sorenson, and Geoffrey Barnes, "Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions," *Journal of Empirical Legal Studies*, 2016, *13* (1), 94–115.

Berk, Richard, Heather Ladd, Heidi Graziano, and Jong-Ho Baek, "A Randomized Experiment Testing Inmate Classification Systems Inmate Classification," *Criminology & Public Policy*, 2002, *2*, 215.

Bushway, Shawn and Jeffrey Smith, "Sentencing Using Statistical Treatment Rules: What We Don't Know Can Hurt Us," *Journal of Quantitative Criminology*, Dec 2007, *23* (4), 377–387.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan, "Productivity and Selection of Human Capital with Machine Learning," *American Economic Review*, May 2016, *106* (5), 124–27.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, September 2014, *104* (9), 2633–79.

Cowgill, Bo, "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening," *Unpublished Working Paper*, August 2018.

_ , "The Impact of Algorithms on Judicial Discretion: Evidence from Regression Discontinuities," Working Paper 2019.

_ **and Catherine Tucker**, "Algorithmic Bias: A Counterfactual Perspective," Working Paper: NSF Trustworthy Algorithms December 2017.

_ **and** _ , "Economics, Fairness, and Algorithmic Bias," In preparation for the Journal of Economics Perspectives March 2019.

_ **and Megan Stevenson**, "Algorithmic Social Engineering," Working Paper 2019.

Doleac, Jennifer L and Benjamin Hansen, "The unintended consequences of ban the box: Statistical discrimination and employment outcomes when criminal histories are hidden," *Journal of Labor Economics*, 2018, *forthcoming.*

Doleac, Jennifer L. and Megan T. Stevenson, "Are Criminal Risk Assessment Scores Racist," *Brookings.edu*, August 2016.

Dressel, Julia and Hany Farid, "The accuracy, fairness, and limits of predicting recidivism," *Science Advances*, 2018, *4* (1).

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai, "Using a Probabilistic Model to Assist Merging of Large Scale Administrative Records," Technical Report, Unpublished working paper May 2018.

Garrett, Brandon L., Alexander Jakubow, and John Monahan, "Judicial Reliance on Risk Assessment in Sentencing Drug and Property Offenders: A Test of the Treatment Resource Hypothesis," *Criminal Justice and Behavior*, 2019, *46* (6), 799–810.

_ **and John Monahan**, "Judging Risk," *California Law Review*, 2018, *forthcoming.*

_ **and** _ , "Judging Risk," Virginia Public Law and Legal Theory Research Paper Series 2018-27 May 2018.

Gelbach, Jonah B., "When Do Covariates Matter? And Which Ones, and How Much?," *Journal of Labor Economics*, 2016, *34* (2), 509–543.

Glaeser, Edward L., Andrew Hillis, Scott Duke Kominers, and Michael Luca, "Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy," *American Economic Review*, May 2016, *106* (5), 114–18.

Goel, Sharad, Justin M. Rao, and Ravi Shroff, "Personalized Risk Assessments in the Criminal Justice System," *American Economic Review*, May 2016, *106* (5), 119–23.

Jung, J., C. Concannon, R. Shroff, S. Goel, and D. G. Goldstein, "Simple rules for complex decisions," *ArXiv e-prints*, February 2017.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*, 2018, *133* (1), 237–293.

_ , Jens Ludwig, Sendhil Mullainathan, and Cass R Sunstein, "Discrimination in the Age of Algorithms," *Journal of Legal Analysis*, 04 2019, *10*, 113–174.

Kling, Jeffrey R., "Incarceration Length, Employment, and Earnings," *American Economic Review*, June 2006, *96* (3), 863–876.

Lim, Claire S. H., Jr. Snyder James M., and David Strmberg, "The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems," *American Economic Journal: Applied Economics*, October 2015, *7* (4), 103–35.

Loeffler, Charles E., "Does Imprisonment Alter the Life Course? Evidence on Crime and Employment From a Natural Experiment," *Criminology*, 2013, *51* (1), 137–166.

Mayson, Sandra, "Bias In, Bias Out," *Yale Law Journal*, June 2019, *128* (8).

Monahan, John, Anne L. Metz, and Brandon L. Garrett, "Judicial Appraisals of Risk Assessment in Sentencing," Virginia Public Law and Legal Theory Research Paper Series 2018-27 April 2018.

Mueller-Smith, Michael, "The Criminal and Labor Market Impacts of Incarceration: Identifying Mechanisms and Estimating Household Spillovers," Working Paper August 2015.

Ostrom, Brian, Fred Cheesman, Ann M. Jones, Meredith Peterson, and Neal B. Kauder, "Truth in Sentencing in Virginia: Evaluating the Process and Impact of Sentencing Reform," Technical Report, National Center for State Courts 1999.

_ , Mathew Kleiman, Fred Cheesman, Randall M. Hansen, and Neal B. Kauder, "Offender Risk Assessment in Virginia: A Three Stage Evaluation," Technical Report, National Center for State Courts 2002.

Rose, Evan and Yotam Shem Tov, "Does Incarceration Produce Crime?," Working Paper November 2018.

Skeem, Jennifer Lynne, Nicholas Scurich, and John Monahan, "Impact of Risk Assessment on Judges Fairness in Sentencing Relatively Poor Defendants," Working Paper 2019.

Sloan, Carly Will, George S Naufal, and Heather Caspers, "The Effect of Risk Assessment Scores on Judicial Behavior and Defendant Outcomes," IZA DP No. 11948 November 2018.

Starr, Sonja, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination," *Stanford Law Review*, 2014, *66.*

Stevenson, Megan, "Breaking Bad: Mechanisms of Social Influence and the Path to Criminality in Juvenile Jails," *The Review of Economics and Statistics*, 2017, *99* (5), 824–838.

_ , "Assessing Risk Assessment in Action," *Minnesota Law Review*, 2019, *103.*

_ **and Christopher Slobogin**, "Algoirithmic Risk Assessment and the Double Edged Sword of Youth," *Washington University Law Review*, 2018, *96.*

VCSC, "Assessing Risk Among Sex Offenders in Virginia," Technical Report, Virginia Criminal Sentencing Commission January 2001.

_ , "Virginia Criminal Sentencing Commission 2003 Annual Report," Technical Report December 2003.

Table 1: Summary statistics from fiscal year 2001

	Nonviolent RA-eligible	Nonviolent RA-ineligible	Sex offenses	Other ineligible offenses
Variables from sentencing commission data:				
Age	31.39	31.41	33.34	31.63
Age<23	0.26	0.29	0.25	0.29
Recent felony conviction	0.09	0.04	0.02	0.05
Drug	0.53	0.48	0.00	0.00
Larceny	0.30	0.32	0.00	0.00
Fraud	0.17	0.19	0.00	0.00
Rape	0.00	0.00	0.35	0.00
Sex assault	0.00	0.00	0.65	0.00
Burglary	0.00	0.00	0.00	0.21
Robbery	0.00	0.00	0.00	0.12
Assault	0.00	0.00	0.00	0.19
Traffic	0.00	0.00	0.00	0.33
Guidelines-recommended sentence	13.00	8.81	90.15	41.96
Pr(incarceration)	0.78	0.33	0.81	0.83
Sentence length (months)	9.95	6.66	73.82	37.81
New felony conviction - 3yr	0.16	0.12	0.02	0.09
Observations	7479	6439	700	6065
Variables from court data:				
Black	0.65	0.55	0.39	0.52
Female	0.23	0.29	0.02	0.09
New felony charge - 3yr	0.40	0.33	0.16	0.26
Observations	6678	5570	594	5522

Note: This table provides summary statistics from fiscal year 2001, the year before the sex offender risk assessment was adopted. All variables except for age, guidelines-recommended sentence and months of incarceration are dummy variables. Age is measured in years and both the guidelines-recommended sentence and the actual sentence is measured in months. The likelihood of receiving a new felony charge or conviction within three years is measured from the time of sentencing, not the time of release. From left to right, the columns show statistics for those who are nonviolent risk-assessment-eligible, those who were convicted of a violent offense but are ineligible for risk assessment, those convicted of a sex offense, and those convicted of an offense category that has no risk assessment.

Table 2: Regression discontinuity around the low-risk cutoffs

	Nonviolent RA		Sex offender RA	
	(1)	(2)	(3)	(4)
Panel A				
Female	0.034		-0.005	
	(0.033)		(0.013)	
Black	-0.004		0.006	
	(0.030)		(0.035)	
Age	-0.395		-0.610	
	(0.745)		(0.739)	
Recent conviction	0.031		-0.002	
	(0.019)		(0.010)	
Guidelines-rec. sentence	0.534		12.473	
	(0.706)		(7.672)	
Drug	-0.008		0.000	
	(0.030)		(.)	
Larceny	0.017		0.000	
	(0.022)		(.)	
Fraud	-0.011		0.000	
	(0.027)		(.)	
Rape	0.000		0.014	
	(.)		(0.031)	
Panel B				
Pr(incarceration)	0.062***	0.059**	0.067***	0.060***
	(0.024)	(0.024)	(0.022)	(0.020)
Months sentence (arcsinh)	0.238***	0.230***	0.406***	0.343****
	(0.085)	(0.072)	(0.126)	(0.093)
Observations	12850	12850	9436	9436
Mean pr(incarceration)	0.820	0.820	0.870	0.870
Mean sentence (arcsinh)	2.297	2.297	3.838	3.838
Covariates	N	Y	N	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table uses regression discontinuity to test for changes around the low-risk cutoff for both the nonviolent risk score and the sex offender risk score. The top part tests for discontinuous changes in defendant characteristics and the bottom part tests for discontinuous changes in the probability of incarceration and sentence length, with an arcsinh transform. The estimates are generated using optimal mean-squared error bandwidths. Mean incarceration rates and sentence lengths (with an arcsinh transform) are shown at the bottom.

Table 3: Risk assessment's net impact on sentencing and recidivism

	Pr(Incarceration)		Sentence (arcsinh)		Recidivism (3yr)	
	(1)	(2)	(3)	(4)	(5)	(6)
NV eligible x post-02	-0.00838	0.00284	0.0208	0.0327	0.00861	0.0102**
	(0.00839)	(0.00799)	(0.0301)	(0.0215)	(0.00526)	(0.00496)
Rape x post-01	-0.0504***	-0.0495***	-0.262**	-0.245**		
	(0.0163)	(0.0170)	(0.116)	(0.104)		
Observations	77065	77065	77065	77065	77065	77065
R^2	0.223	0.439	0.211	0.637	0.00760	0.0377
Mean DV, NV	0.791	0.791	2.119	2.119	0.162	0.162
Mean DV, Rape	0.962	0.962	5.311	5.311	0.0157	0.0157
Covariates	N	Y	N	Y	N	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents difference-in-differences estimates in which outcomes are compared across eligible/ineligible cases before/after risk assessment is adopted. The outcomes are the probability of incarceration, the sentence length (with an arcsinh transform) and the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). Recidivism estimates for sex offenders are omitted to avoid drawing focus towards tests for which we are underpowered. The sample includes all defendants convicted of a felony in fiscal years 2000-2004.

Table 4: Risk assessment's impact by race and age

	Pr(Incarceration)		Sentence (arcsinh)	
	(1)	(2)	(3)	(4)
NV elig. x post-02 x black	-0.0141	0.0324		
	(0.0165)	(0.0466)		
NV elig. x post-02 x young			0.0416**	0.120**
			(0.0172)	(0.0526)
Observations	58744	58744	63700	63700
R^2	0.431	0.622	0.430	0.622
Mean DV, NV, black	0.820	2.277		
Mean DV, NV, young			0.733	1.984
Covariates	Y	Y	Y	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This figure shows triple difference estimates of risk assessment's impact by race and age. The outcomes are the probability of incarceration and the sentence length (with an arcsinh transform). Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). The samples are limited to defendants sentenced in fiscal years 2001-2004; Alexandria and Fairfax counties are dropped from race regressions since data is unavailable.

Table 5: What factors predict whether judges will follow the recommendations of the risk assessment?

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Diverted | risk = low						
Alternative risk score	0.013					0.010
	(0.010)					(0.010)
Black		-0.015				-0.014
		(0.015)				(0.016)
Unemployed			0.025			0.009
			(0.017)			(0.018)
Female				0.040**		0.038**
				(0.016)		(0.017)
Age<23					0.069****	0.065***
					(0.020)	(0.020)
Observations	3943	3943	3943	3943	3943	3943
R^2	0.204	0.204	0.204	0.205	0.206	0.280
Mean DV	0.44	0.44	0.44	0.44	0.44	0.44
Panel B: Diverted | risk = high						
Alternative risk score	-0.004					-0.007
	(0.005)					(0.005)
Black		-0.029***				-0.045****
		(0.010)				(0.012)
Unemployed			0.043****			0.018
			(0.012)			(0.012)
Female				0.038***		0.040***
				(0.013)		(0.014)
Age<23					0.065****	0.058****
					(0.011)	(0.011)
Observations	7598	7598	7598	7598	7598	7598
R^2	0.142	0.143	0.144	0.143	0.146	0.197
Mean DV	0.16	0.16	0.16	0.16	0.16	0.16

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: The outcome variable in all regressions is a dummy equal to one if the defendant is diverted from jail or prison. The top panel includes defendants who received a recommendation for diversion from jail or prison due to the low risk score. The bottom panel includes defendants with a high risk score who were not recommended for diversion. All regressions include controls for the exact nonviolent risk score and the exact guidelines-recommended sentence. The alternative risk score is a risk score we built using a random forest model and various demographic, criminal history, and offense variables.

Table 6: Simulating how full compliance with the algorithm would have affected race/age disparities

	(1) Pr(Inc.)	(2) Sent.	(3) Pr(Inc.)	(4) Sent.
NV elig. x post-02 x black	0.0377**	0.0851		
	(0.0187)	(0.0539)		
NV elig. x post-02 x young			0.156****	0.447****
			(0.0180)	(0.0520)
Observations	58744	58744	63700	63700
R^2	0.387	0.614	0.384	0.613
Mean DV, NV, black	0.826	2.277		
Mean DV, NV, young			0.747	1.984
Covariates	Y	Y	Y	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table provides simulation estimates of how risk assessment would have impacted race and age disparities in sentencing if judges had fully complied with the sentencing recommendations associated with the algorithm. The outcomes are the probability of incarceration and the sentence length (with an arcsinh transform). Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). The samples are limited to defendants sentenced in fiscal years 2001-2004; Alexandria and Fairfax counties are dropped from race regressions since data is unavailable.

Table 7: Impacts of risk assessment among the 'most responsive' judges

Panel A: Differences-in-differences						
	Pr(Incarceration)		Sentence (arcsinh)		Recidivism (3yr)	
	(1)	(2)	(3)	(4)	(5)	(6)
NV eligible x post-02	0.0000247	0.00475	0.0207	0.0189	0.0143*	0.0133*
	(0.0127)	(0.0106)	(0.0472)	(0.0325)	(0.00727)	(0.00702)
Rape x post-01	-0.0710***	-0.0767***	-0.337*	-0.309**		
	(0.0237)	(0.0227)	(0.188)	(0.143)		
Observations	32490	32490	32490	32490	32490	32490
R^2	0.209	0.429	0.201	0.631	0.00739	0.0331
Mean DV, NV	0.784	0.784	2.160	2.160	0.154	0.154
Mean DV, sex	0.983	0.983	5.375	5.375	0.0252	0.0252
Covariates	N	Y	N	Y	N	Y
Panel B: Triple-differences						
	Pr(Incarceration)		Sentence (arcsinh)			
	(1)	(2)	(3)	(4)		
NV elig. x post-02 x black	0.0361*	0.166**				
	(0.0191)	(0.0640)				
NV elig. x post-02 x young			0.0545**	0.138		
			(0.0268)	(0.0855)		
Observations	25532	27933	25532	27933		
R^2	0.428	0.424	0.619	0.618		
Mean DV, NV, black	0.795	2.207				
Mean DV, NV, young			0.713	1.961		
Covariates	Y	Y	Y	Y		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents estimates from the subset of judges who appear most responsive to the risk assessment, as measured by the magnitude of sentencing discontinuities around the low risk cutoff. The top panel presents difference-in-differences estimates in which outcomes are compared across eligible/ineligible cases before/after risk assessment is adopted. The bottom panel shows triple difference estimates of risk assessments' impact by race and age. The outcomes are the probability of incarceration, the sentence length (with an arcsinh transform) and the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). The sample for difference-in-differences includes all defendants convicted of a felony between fiscal years 2000-2004. The sample for triple-differences is limited to defendants sentenced in fiscal years 2001-2004; Alexandria and Fairfax counties are dropped from race regressions since data is unavailable.

Table 8: Effect of incarceration on recidivism: discontinuity-in-risk-score estimates

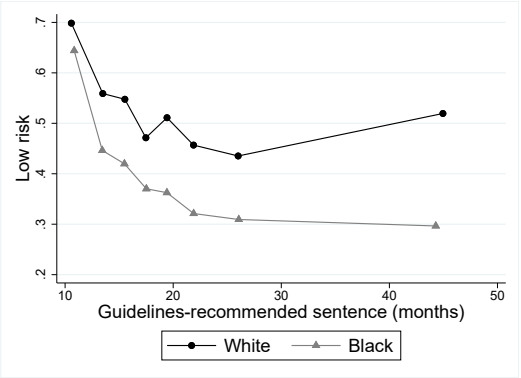
	Impact on sentencing		Reduced form impacts on recidivism			IV impacts on recidivism		
	Pr(inc.)	Sent.	1yr	3yr	7yr	1yr	3yr	7yr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD_Estimate	0.065***	0.217**	-0.025	-0.077**	-0.058**	-0.110	-0.354*	-0.264*
	(0.025)	(0.093)	(0.024)	(0.033)	(0.029)	(0.104)	(0.197)	(0.154)
Mean DV	0.766	2.032	0.182	0.352	0.423	0.182	0.352	0.423
Observations	11304	11304	11304	11304	11304	11304	11304	11304

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

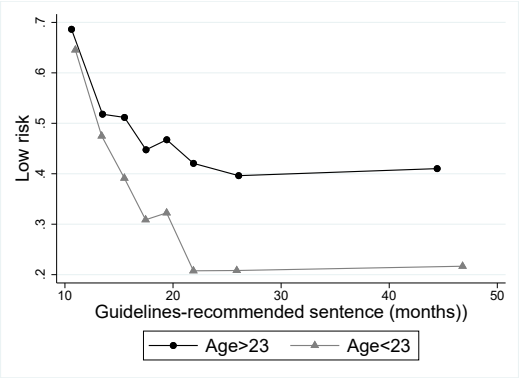
Note: All columns show RD estimates that exploit discontinuities in risk classification at a cutoff in the risk score. The outcome variables in Columns 1 and 2 are the probability of incarceration and the sentence length (arcsinh transform). The outome variables in Columns 3-5 are recidivism within different time windows. The outcome variables in Columns 6-8 are also recidivism in different time windows, but these estimates are from a fuzzy RD regression in which the length of the sentence (with an arcsinh transform) is the endogenous variable (in other words, Column 2 is the first stage). Recidivism is defined here as the likelihood of receiving a new felony charge within X years of sentencing. The mean dependent variable is shown for defendants whose risk score is within -3 to -1. Alexandria and Fairfax counties are dropped; the sample includes defendants who received a nonviolent risk assessment in fiscal years 2003-2004.

Figure 1: Nonviolent risk score and race/age disparities

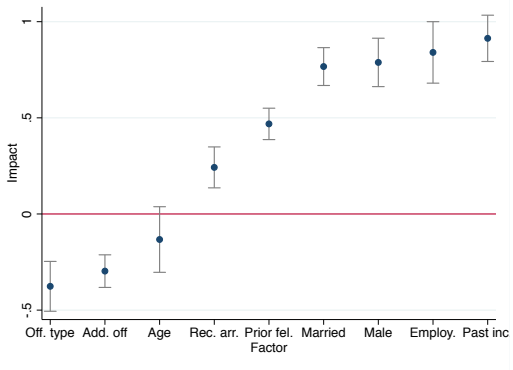
(a) Racial disparities in low-risk classification by recommended sentence



(b) Age disparities in low-risk classification by recommended sentence



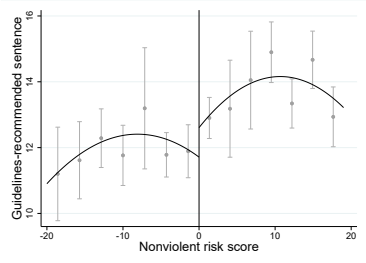
(c) Gelbach decomposition of racial disparities in the NV risk score



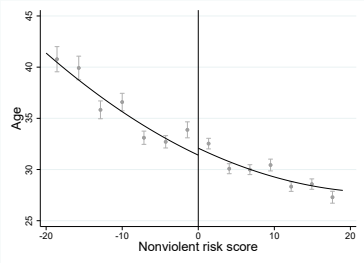
Note: The top two graphs plot the likelihood of receiving a low risk classification against the guidelines-recommended sentence, shown separately by race and age. The sample includes all defendants who received a risk score and whose guidelines-recommended sentence was prison. The bottom graph is a coefficient plot that shows how various nonviolent risk score factors contribute to the race gap in the nonviolent risk score using a Gelbach demposition. A positive coefficient means that that factor increases risk scores for black defendants relative to white defendants. The factors include offense type, additional offenses, age, recent arrest or confinement, prior felonies, marital status, gender, employment, prior incarcerations.

Figure 2: Covariate balance across risk score cutoffs

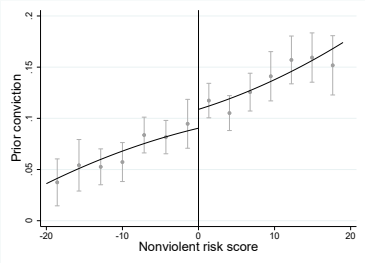
(a) Nonviolent risk score and the guidelines-recommended sentence



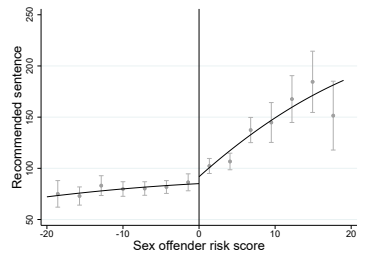
(b) Nonviolent risk score and age



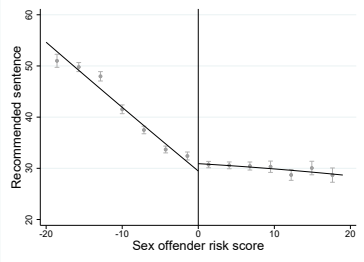
(c) Nonviolent risk score and prior convictions



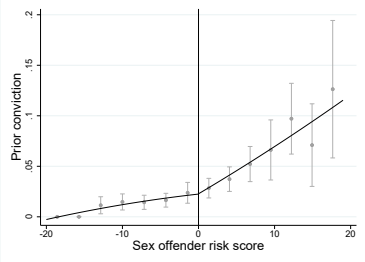
(d) Sex offender risk score and the guidelines-recommended sentence



(e) Sex offender risk score and age



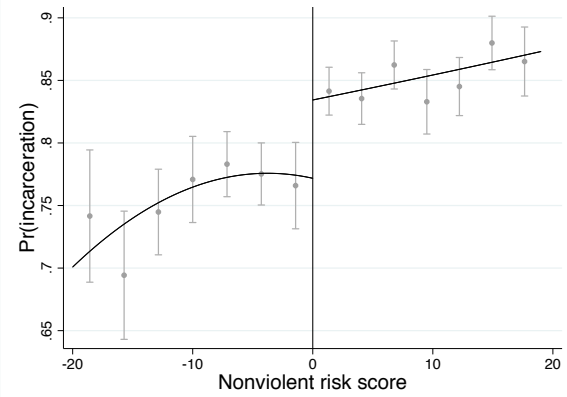
(f) Sex offender risk score and prior convictions



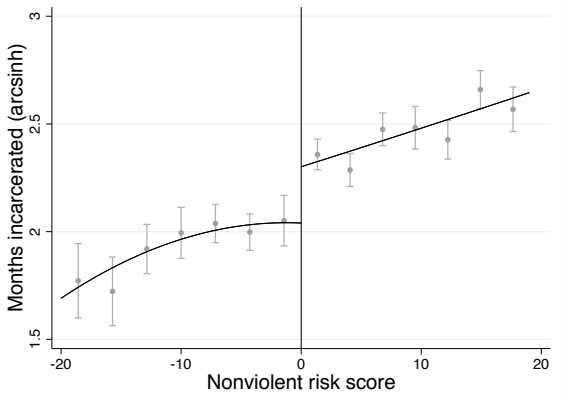
Note: The vertical axes of these figures show the guidelines-recommended sentence (which is not altered by the risk score), age, and a dummy for recent prior felony convictions. The horizontal axis in the top two figures is the nonviolent risk score, and the horizontal axis in the bottom figures is the sex offender risk score, normalized so that scores below 0 are in the lowest risk classification. Each dot shows the mean for a bin of three risk scores, and the whiskers show the 95% confidence interval for that mean. The lines represent fitted polynomial trends of degree 2.

Figure 3: Does the risk classification affect defendants' sentences at the margin?

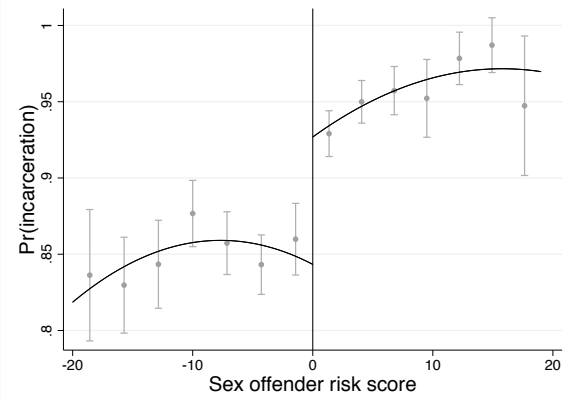
(a) Nonviolent risk score and probability of incarceration



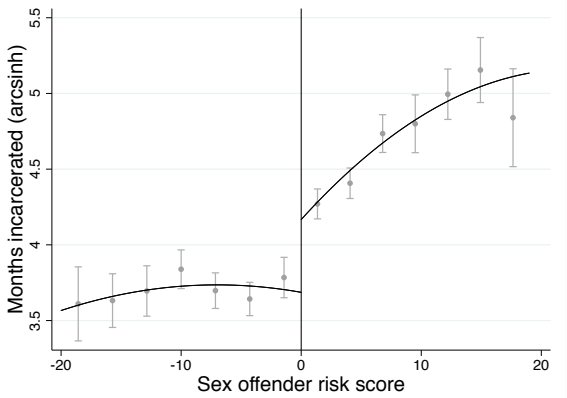
(b) Nonviolent risk score and the sentence length



(c) Sex offender risk score and probability of incarceration



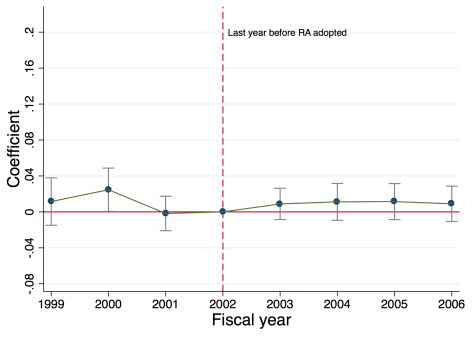
(d) Sex offender risk score and the sentence length



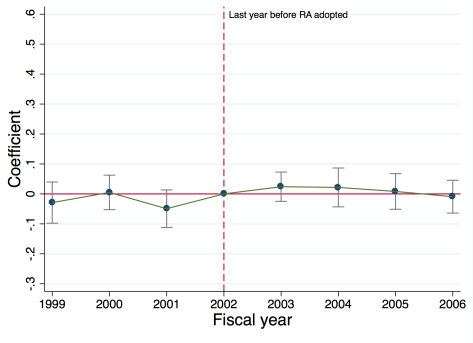
Note: The vertical axes of these figures show the probability of incarceration and the sentence length (with an arcsinh transform). The horizontal axis in the top two figures is the nonviolent risk score, and the horizontal axis in the bottom figures is the sex offender risk score. Each dot shows the mean for a bin of three risk scores, and the whiskers show the 95% confidence interval for that mean. The risk scores are normalized so that defendants with a score of 0 or above are in the higher risk classification. The lines represent fitted polynomial trends of degree 2.

Figure 4: Event-study graphical analysis of risk assessment's impact on sentencing and recidivism

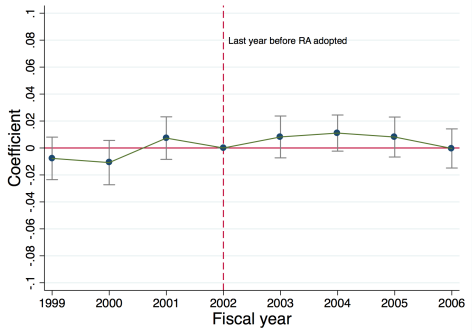
(a) Nonviolent risk score and probability of incarceration



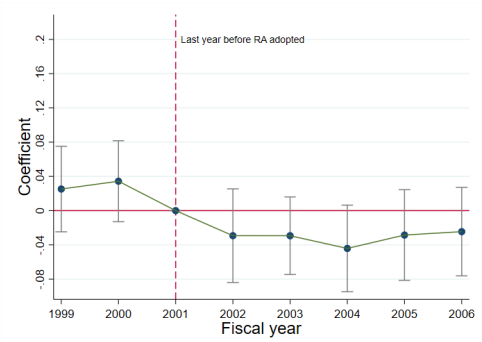
(b) Nonviolent risk score and the sentence length (arcsinh)



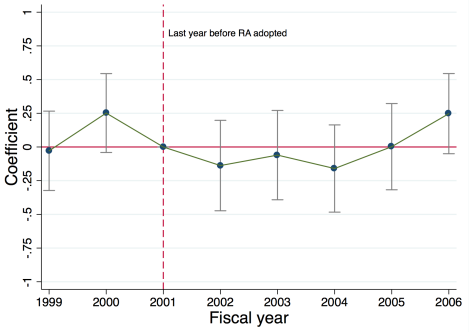
(c) Nonviolent risk score and recidivism (3yr)



(d) Sex offender risk score and probability incarceration



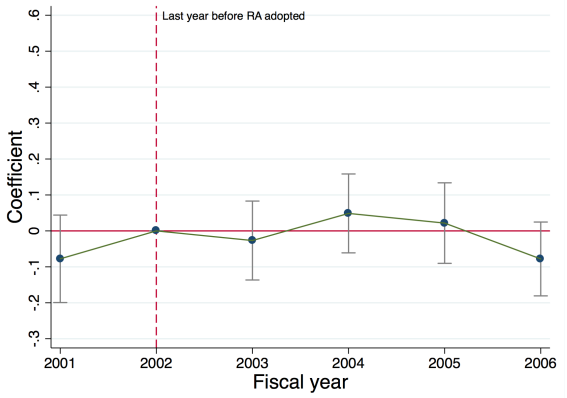
(e) Sex offender risk score and the sentence length (arcsinh)



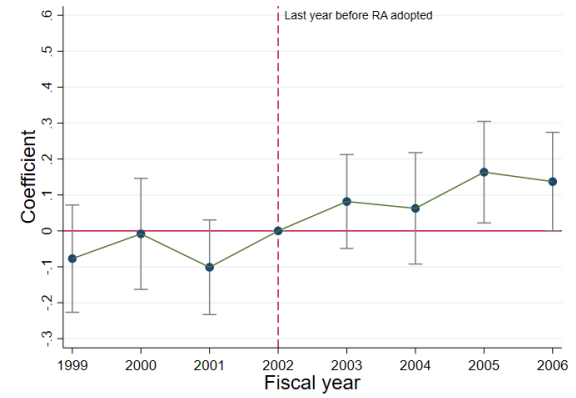
Note: These figures describe the net impact of risk assessment. Each dot represents a lead/lag treatment coefficient as described in Equation 4. The sample includes fiscal years 1999-2006. The year prior to each risk assessment adoption is omitted as the baseline and is indicated by the dashed vertical line. The outcomes are the probability of incarceration, the sentence length (with an arcsinh transform) and the likelihood of being convicted of a new felony within 3 years (for nonviolent offenders). 95% confidence intervals are shown.

Figure 5: Event-study graphical analysis of risk assessment's actual and simulated impact on race and age disparities

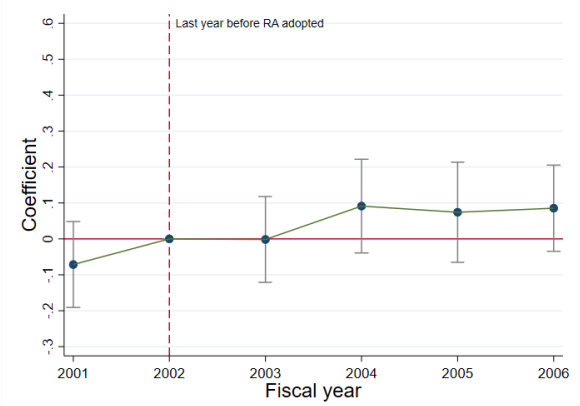
(a) Risk assesment's **actual** impact on sentencing for black defendants



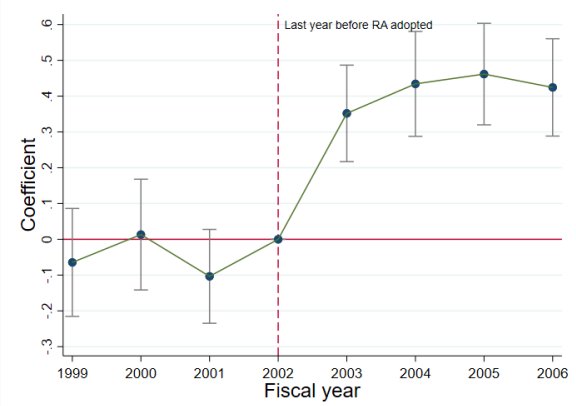
(b) Risk assesment's **actual** impact on sentencing for young defendants



(c) Risk assesment's **simulated** impact on sentencing for black defendants



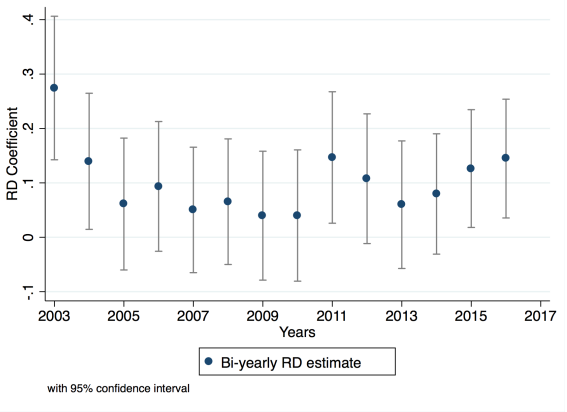
(d) Risk assesment's **simulated** impact on sentencing for young defendants



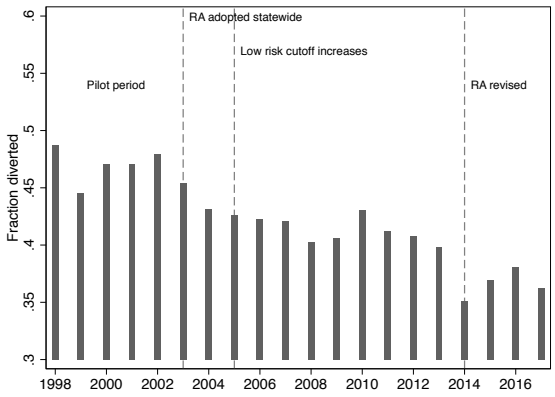
Note: The top two figures present lag/lead coefficient plots for the triple differences estimations of risk assessment's impact on the sentence length (arcsinh transform) for black defendants and young defendants. Each dot is the magnitude of a lead/lag treatment coefficient as described in Equation 4. The bottom figures are similar except they estimate how risk assessment would have impacted race and age disparities in sentence length if judges had fully complied with the sentencing recommendations associated with the algorithm. The year prior to risk assessment adoption is omitted as the baseline and is indicated by the dashed vertical line. 95% confidence intervals are shown.

Figure 6:

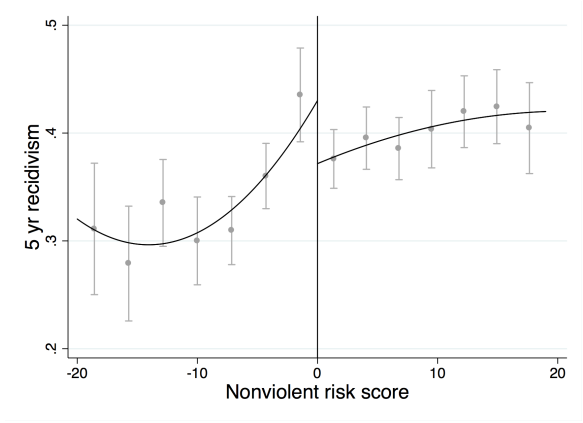
(a) Influence of risk assessment over time



(b) Fraction of nonviolent low risk diverted over time



(c) RD in risk score: recidivism (5yr)



Note: The top left figure is a coefficient plot designed to show how sentencing responds to the risk classification in different time periods. Each dot represents an RD estimate of the magnitude of sentencing discontinuity at the low-risk cutoff for nonviolent cases sentenced during a rolling two year period as indicated in the horizontal axis. The top right figure shows the yearly fraction of low-risk nonviolent defendants who were diverted from jail or prison. The bottom figure plots the likelihood of being charged with a new felony within 5 years of sentencing against the nonviolent risk score. The risk score has been normalized; scores below 0 receive a recommendation for diversion from jail or prison.

A Appendix

A.1 Use of risk assessment at sentencing

The following states use risk assessment at sentencing:

- AL (Ala. Stat. 12-25-33(6));
- CO (Colorado Revised Statues Title 16-11-102 (1)(b)(II));
- ID, NE, OR (Elek, J. K., Warren, R. K., & Casey, P. M. (2015). Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions. National Center for State Courts.);
- HI, IL (Howell, T. (n.d.). LSI-R, LS/RNR and LS/CMI Documentation. Public Safety Division, Multi-Health Systems, Inc. Retrieved from https://www.scstatehouse.gov/Archives/Ci
- KS (Kansas Judicial Branch. (2014, June 26). Court Services Officer Assessment of Adult Offenders (Rule 110B). Retrieved from http://www.kscourts.org/rules/District˙Rules/Rule%20
- CA, FL, WI (Kehl, D. L., & Kessler, S. A. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Retrieved from https://dash.harvard.edu/handle/1/33746041);
- PA (Meyers, R. S. (2018). Introducing Risk Assessment at Sentencing in Pennsylvania. Presented at the CJAB Annual Conference. Retrieved from https://www.pccd.pa.gov/training/I
- ND (North Dakota Department of Corrections and Rehabilitation. (2013). 2011-2013 Biennial Report. Retrieved from https://docr.nd.gov/biennial-report-archive);
- NY (Office of Probation and Correctional Alternatives. (2015). New York Correctional Offender Management Profiling for Alternative Sanctions (NYCOMPAS) Risk and Needs Assessment Instrument: Practitioner Guidance for Probation and Community Corrections Agencies. New York State Division of Criminal Justice Services. Retrieved from http://www.criminaljustice.ny.gov/opca/pdfs/2015-5-NYCOMPAS-Guidance-August-4-2015.pdf);
- IA (Presentation to the Iowa Board of Corrections: Risk Assessments in Presentence Investigations. (2011). Retrieved from http://justicereformconsortium.org/wp-content/uploads/2011/11/BOC˙LSIR1.pdf);
- AZ, IN, KY, MI, MO, OH, OK, UT, VA, WA, WV (Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. Stan. L. Rev., 66, 803?872.);
- VT (State of Vermont Department of Corrections. (2011, December 20). Presentence Investigation (PSI) Reports (Directive #342.01). VT. Retrieved from http://doc.vermont.gov/about/policies/rpd/correctional-services-301-550/335-350-district-offices-general/copy˙of˙342-01-pre-sentence-investigation-psi-reports);
- TN (TENN. CODE ANN. 41-1-412).

There are 7 more states in which at least one county uses risk assessments in sentencing:

- NC (Howell, T. (n.d.). LSI-R, LS/RNR and LS/CMI Documentation. Public Safety Division, Multi-Health Systems, Inc. Retrieved from https://www.scstatehouse.gov/Archives/Ci
- LA (Louisiana Stat. Ann. 15:326(A));

- MN (Minnesota Department of Corrections. (2011). Study of Evidence-Based Practices in Minnesota: 2011 Report to the Legislature. St. Paul, MN. Retrieved from https://mn.gov/doc/assets/12-10EBPreport˙tcm1089-271698.pdf);
- NM (New Mexico 2nd Judicial District Criminal Justice Strategic Plan. (2012). Bernalillo County. Retrieved from https://www.bernco.gov/uploads/FileLinks/33e0766212d24ba7a61e
- ME (Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. Stan. L. Rev., 66, 803?872.);
- TX (Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. Stan. L. Rev., 66, 803?872. Elek, J. K., Warren, R. K., & Casey, P. M. (2015). Using Risk and Needs Assessment Information at Sentencing: Observations from Ten Jurisdictions. National Center for State Courts.);
- AR (Warren, R. K. (2013). State Judicial Branch Leadership In Sentencing and Corrections Reforms. Center for Sentencing Initiatives, Research Division, National Center for State Courts. Retrieved from https://www.ncsc.org/ /media/Microsites/Files/CSI/State%20Judicial%20Branch%20Leadership%20Brief%20csi.ashx).

A.2 Changes at the tails of the risk distribution

In this section, we evaluate whether the adoption of the risk assessment affected sentencing at the tails of the risk distribution. Unfortunately, we don't have risk score information for defendants who did not receive a risk score. However, the sentence-guidelines worksheets contain much of the same information – albeit weighted differently. We use this data to develop a prediction model for the nonviolent risk score. (Note: we are predicting the risk score, not recidivism.) This model was trained using a random forest algorithm on the set of defendants who received a nonviolent risk score within fiscal years 2003-2004. The correlation coefficient between the actual risk score and the predicted risk score (generated using 'out of bag' predictions, or predictions in which that particular observation was not used) is 0.73. While not a perfect prediction, the two are strongly correlated. Predictions were then generated for all drug, larceny, and fraud defendants who were sentenced during fiscal years 2001-2004. We divide this predicted risk score into quintiles and generate dummies for each. We then use the triple differences specification described in Equation 3, where Z are the quintile dummies, to evaluate how risk assessment affected sentencing at various quintiles of the predicted risk score distribution. Appendix Table A1 shows results, with the third quintile dropped to serve as a baseline. (The cutoff for the low risk classification lies at the 38th percentile of the risk score distribution: near the top of the second quintile.) Sentencing for defendants in the second and third quintile at the margins of the low risk classification is fairly similar. Sentencing changes the most for defendants in the highest risk quintile: an increase of three percentage points in the likelihood of incarceration and approximately 10% in the sentence length relative to the middle quintile. The probability of incarceration declined by 1.5 percentage points (not statistically significant) and the sentence length declined by approximately 7% for the lowest risk quintile relative to the middle quintile.

Since we use the predicted risk score instead of the actual risk score we are introducing measurement error into the specification which should both bias estimates

towards zero and increase the standard errors. Despite these caveats, the evidence suggests that risk assessment affected decision-making at the tails of the distribution as well as at the margins of the low risk cutoff.

A.3 Calibrating the risk assessment with an instrumental variables method

In this section, we describe a novel method of evaluating how well the risk assessment sorts defendants by recidivism risk. Our method uses an instrumental variables technique to estimate recidivism risk for defendants in different terciles of the risk distribution. The instrumental variables method exploits discontinuities in a sentence guidelines score that determines whether or not the defendant will be recommended for a prison sentence, known colloquially as the 'in/out' score. Defendants who score above a certain cutoff in the in/out score will be recommended for prison, defendants who score below that cutoff will not. This creates sharp discontinuities in the incarceration length. Using regression discontinuity with the sentence guidelines as the running variable, we develop estimates of recidivism risk for different groups. Here, recidivism risk is measured as the new felony charges that would be averted if a group of defendants were incapacitated from committing crime by incarceration. Incarceration averts more crimes (i.e., the IV estimated impact of incarceration on recidivism is larger) for defendants who pose a higher recidivism risk and fewer crimes for those who pose a low recidivism risk.

As a benchmark, we build an alternative risk assessment tool using a random forest model. While we can do nothing about selection issues in the data, the random forest model is better equipped to identify nonlinear and interactive relationships between the predictor variables and recidivism. Our goal is to evaluate how well the real risk assessment tool sorts defendants by crime propensity as compared to our alternative risk tool.

The alternative risk score (described in more detail in Section 6.2) is trained on nonviolent offenders who received a non-carceral sentence during the two years before risk assessment was adopted statewide. Like the real risk assessment, our target variable is the likelihood of a new felony conviction within three years. The correlation coefficient between the alternative risk score and the real risk score is 0.29.

We use all defendants who have a nonviolent risk score between fiscal years 2003-2013, whose case matches to the court data, and whose offense does not automatically guarantee a recommendation for prison. (More serious offenses do not rely on the in/out score to ascertain whether the guidelines recommended sentence entails prison time.) First, we partition the sample into terciles of the real risk score. We exploit discontinuities in the in/out score to build IV estimates of recidivism risk for these three groups. Then, using the same sample, we partition the sample again according to terciles of our alternative risk assessment. Using the same IV method we build estimates of recidivism risk for these three groups. A tool that is accurate for defendants at the margins of the prison sentence should be able to successfully sort defendants; in other words, there should be a large gap in recidivism risk between higher- and lower-risk-score groups.

Appendix Figures A.3b and c provide motivation for the RD specification. Figure A.3b shows a sharp jump up in the fraction of the first year spent incarcerated for

defendants who are right above the prison cutoff. Figure A.3c shows a corresponding decline in the likelihood of having a new felony charge within one year of sentencing.

Table A8 shows the IV results. We use a bandwidth of 7 for the RD but find similar results with bandwidths of 4 and 10. The endogenous variable is equal to one if the sentence is 12 months or greater, or it is equal to the months of the sentence divided by 12 if the sentence is less than a year. The outcome variable is the likelihood of having a new felony charge within one year of sentencing; Alexandria and Fairfax counties are dropped. (We focus on a relatively short time window post-incarceration to increase the likelihood that what we are capturing is an incapacitation effect.) The first three columns show the real risk score and the final three columns show the alternative risk score. The top panel shows defendants who are in the lowest terciles in each respective risk score, the middle panel shows defendants in the middle terciles, and the bottom panel shows the highest terciles. The first stage, shown in Column 1 for the real risk score and in Column 4 for the alternative risk score, is very strong for all subgroups, and suggests that scoring right above the cutoff leads to an increase of 0.31-0.37 in the fraction of the first year spent incarcerated. Columns 2 and 5 show the RD results without controls, Columns 3 and 6 show the RD results with controls for age, gender, offense and recent prior convictions.

The IV estimates are all negative and all highly statistically significant: being incarcerated decreases the likelihood of having new felony charges within the first year after sentencing. Furthermore, the estimates remain pretty stable to the inclusion of controls, easing concerns about omitted variable bias or manipulation of the running variable. Both risk scores effectively sort defendants by recidivism risk. For the real risk score, the recidivism risk is 14% for the lowest-scoring tercile, 23% for the middle-scoring tercile, and 27% for the highest-scoring tercile. The lowest, middle and highest terciles of the alternative risk score have an estimated recidivism risk of 13%, 21% and 27% respectively. By these metrics, the alternative risk score might outperform the real risk score slightly: the gap in recidivism risk between the bottom and top terciles is 13 percentage points for the real risk score and 14 percentage points for the alternative one. However, this difference is small and the standard errors preclude drawing clear inference.

This test demonstrates that Virginia's risk assessment tool can successfully sort defendants on the margins of being imprisoned. In addition, we find no evidence that one can build a substantially better risk assessment with a more complicated algorithm.

Table A1: How risk assessment affected sentencing for defendants in different quintiles of the risk distribution

	Pr(Incarceration)	Sentence (arcsinh)
	(1)	(2)
1st quintile risk	-0.0196	-0.0773*
	(0.0151)	(0.0408)
2nd quintile risk	-0.000435	0.00331
	(0.0144)	(0.0396)
4th quintile risk	0.0148	0.0334
	(0.0121)	(0.0345)
5th quintile risk	0.0287**	0.101***
	(0.0113)	(0.0385)
Observations	64431	64431
R^2	0.432	0.633
Covariates	Y	Y
Mean DV, NV	0.791	2.119

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents estimates from a triple differences specification, which compares outcomes pre-/post-nonviolent risk assessment adoption across risk-assessment-eligible/-ineligible defendants. The dummy for being risk-assessment-eligible is interacted with quintiles of the predicted risk score to ascertain how adoption of the risk assessment affects outcomes for defendants at different risk levels. The outcomes are the probability of incarceration and the sentence length (with an arcsinh transform). Standard errors are clustered at the judge level. The bottom row shows the mean of the outcome variable for nonviolent risk-assessment-eligible cases during the two years before risk assessment was adopted. The sample includes all defendants convicted of a felony in fiscal years 2001-2004.

Table A2: Risk assessment's impact on recidivism by race and age

	Recidivism (3yr)	
	(1)	(2)
NV elig. x post-02 x black	0.0111	
	(0.0118)	
NV elig. x post-02 x young		-0.000684
		(0.0124)
Observations	58744	63700
R^2	0.0391	0.0371
Mean DV, NV, black	0.164	
Mean DV, NV, young		0.163
Covariates	Y	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows triple difference estimates of risk assessment's impact by race and age on the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). The samples are limited to defendants sentenced in fiscal years 2001-2004; Alexandria and Fairfax counties are dropped from race regressions since data is unavailable.

Table A3: Robustness tests for main results: varying time windows

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pr(incarceration)			Months sentence			Recidivism(3yr)		
	3yr	7yr	9yr	3yr	7yr	9yr	3yr	7yr	9yr
NV eligible x post-02	0.00827	0.00258	0.00577	0.0407*	0.0329*	0.0261	0.00534	0.0111**	0.00994**
	(0.00884)	(0.00759)	(0.00799)	(0.0239)	(0.0190)	(0.0181)	(0.00669)	(0.00453)	(0.00399)
Rape x post-01	-0.0338	-0.0480***	-0.0378***	-0.124	-0.166*	0.00000123	0.00353		
	(0.0219)	(0.0145)	(0.0135)	(0.145)	(0.0885)	(0.0804)	(0.00853)		
Observations	48163	107429	138766	48163	107429	138766	48163	107429	138766
R^2	0.427	0.438	0.433	0.625	0.635	0.634	0.0394	0.0355	0.0337
Mean DV, NV	0.802	0.809	0.809	2.197	2.234	2.249	0.148	0.152	0.148
Mean DV, Rape	0.952	0.955	0.952	5.176	5.221	5.213	0.00708	0.0127	0.0128

Standard errors in parentheses

Clustered Standard Errors

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the main difference-in-differences results shown in Table 3. The specification is the same as that shown in Equation 2 but instead of using five years of data we use three, seven, and nine years, centered around fiscal year 2002. The outcomes are the probability of incarceration, the sentence length (with an arcsinh transform) and the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). Recidivism estimates for sex offenders are omitted to avoid drawing focus towards tests for which we are underpowered.

Table A4: Robustness tests for main results: varying control groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Only nonviolent offenses			No NV-ineligible or sex offense			Only violent offenses	
	Pr(inc.)	Sent.	Recid.	Pr(inc.)	Sent.	Recid.	Pr(inc.)	Sent.
NV eligible x post-02	0.00318	0.0403	0.0123**	0.00358	0.0295	0.00635		
	(0.0115)	(0.0257)	(0.00602)	(0.00654)	(0.0258)	(0.00562)		
Rape x post-01							-0.0409**	-0.202*
							(0.0163)	(0.106)
Observations	52388	52388	52388	52155	52155	52155	14554	14554
R^2	0.436	0.584	0.0352	0.171	0.498	0.0453	0.372	0.633
Mean DV, NV	0.807	2.206	0.153	0.807	2.206	0.153		
Mean DV, Rape							0.954	5.208

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table presents robustness tests for the main difference-in-differences results shown in Table 3. The specification is the same as that shown in Equation 2 but our control groups vary. The control groups for nonviolent risk-assessment-eligible cases are as follows: Columns 1-3 use nonviolent cases that are not risk-assessment-eligible as a control and Columns 4-6 use cases whose offense type makes them ineligible for any risk assessment (e.g. assault, robbery, burglary, traffic, etc.) as a control. In Columns 7-8, the control group for sex offenders consists only of those convicted of violent offenses, such as assault and robbery. The outcomes are the probability of incarceration, the sentence length (with an arcsinh transform) and the likelihood of being convicted of a new felony within 3 years. Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). Recidivism estimates for sex offenders are omitted to avoid drawing focus towards tests for which we are underpowered. The sample includes all defendants convicted of a felony within two years of the adoption of each respective risk assessment.

Table A5: Varying measures of recidivism

	Sentence		Recidivism				
	Inc.	Months	6mo	1yr	3yr	5yr	7yr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main sample; recidivism=felony conviction							
NV eligible x post-02	0.00284	0.0327	0.000553	0.00143	0.0102**	0.00959	0.0106
	(0.00799)	(0.0215)	(0.00234)	(0.00327)	(0.00496)	(0.00607)	(0.00681)
Observations	77065	77065	77065	77065	77065	77065	77065
R^2	0.439	0.637	0.0187	0.0255	0.0377	0.0484	0.0573
Mean DV	0.608	2.093	0.0159	0.0320	0.0983	0.159	0.207
Panel B: Subsample that matches to court data; recidivism=felony conviction							
NV eligible x post-02	0.00956	0.0417*	-0.00135	-0.00299	0.00436	0.00526	0.00860
	(0.00874)	(0.0245)	(0.00270)	(0.00373)	(0.00557)	(0.00690)	(0.00777)
Observations	57963	57963	57963	57963	57963	57963	57963
R^2	0.430	0.621	0.0169	0.0253	0.0383	0.0479	0.0552
Mean DV	0.605	2.057	0.0165	0.0333	0.103	0.167	0.217
Panel C: Subsample that matches to court data; recidivism=any new charge							
NV eligible x post-02	0.00956	0.0417*	0.00664	0.00130	-0.0117	-0.0141*	-0.0138*
	(0.00874)	(0.0245)	(0.00500)	(0.00605)	(0.00719)	(0.00754)	(0.00796)
Observations	57963	57963	57963	57963	57963	57963	57963
R^2	0.430	0.621	0.0483	0.0821	0.0994	0.0889	0.0801
Mean DV	0.605	2.057	0.0871	0.149	0.289	0.343	0.371

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows how nonviolent risk assessment use affects recidivism, using a variety of different recidivism measures. In Panel A, the recidivism measure is the likelihood of receiving a new felony conviction within varying time windows: 6 months, or 1, 3, 5 and 7 years. In Panel B, we use the same recidivism measures but restrict our sample to cases that we are able to match to court data; Alexandria and Fairfax are dropped. In Panel C, our recidivism measure is the likelihood of receiving new felony charges within varying time windows; again, Alexandria and Fairfax are dropped. The first two columns show how risk assessment affects the probability of incarceration and the months of the sentence with an arcsinh transform in these various samples. Standard errors are clustered at the judge level. The mean dependent variables during the pre-risk assessment period are shown in the bottom row (NV= nonviolent risk-assessment-eligible). The sample includes all defendants convicted of a felony within two years of the adoption of the nonviolent risk assessment.

Table A6: Robustness tests for 'most responsive' judges

	(1) h=7 Sentence	(2) h=7 Pr(Inc.)	(3) h=4 Sentence	(4) h=4 Pr(Inc.)	(5) h=10 Sentence	(6) h=10 Pr(Inc.)
Panel A: Dependent variable=Pr(incarceration)						
NV eligible x post-02	0.008	0.005	-0.005	0.000	0.004	-0.003
	(0.012)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
Rape x post-01	-0.052*	-0.077***	-0.077***	-0.077***	-0.059**	-0.046*
	(0.026)	(0.023)	(0.024)	(0.023)	(0.023)	(0.024)
NV elig. x post-02 x black	0.032	0.036*	0.026	0.023	0.011	0.031
	(0.022)	(0.019)	(0.020)	(0.020)	(0.021)	(0.019)
NV elig. x post-02 x young	0.029	0.055**	0.031	0.034	0.026	0.040
	(0.031)	(0.027)	(0.029)	(0.029)	(0.027)	(0.027)
Panel B: Dependent variable=months sentence (arcsinh)						
NV eligible x post-02	0.051	0.019	0.030	0.017	0.038	0.013
	(0.035)	(0.032)	(0.034)	(0.034)	(0.033)	(0.030)
Rape x post-01	-0.307*	-0.309**	-0.324**	-0.366**	-0.241*	-0.161
	(0.159)	(0.143)	(0.154)	(0.147)	(0.139)	(0.144)
NV elig. x post-02 x black	0.159**	0.166**	0.162**	0.136**	0.065	0.140**
	(0.063)	(0.064)	(0.061)	(0.062)	(0.062)	(0.058)
NV elig. x post-02 x young	0.095	0.138	0.078	0.075	0.074	0.089
	(0.089)	(0.085)	(0.088)	(0.092)	(0.081)	(0.085)
Panel C: Dependent variable=new fel. conviction w/in 3yrs						
NV eligible x post-02	0.008	0.013*	0.012	0.017**	0.011	0.015**
	(0.009)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)
Rape x post-01	0.008	0.000	0.003	-0.004	0.021*	0.002
	(0.018)	(0.020)	(0.018)	(0.019)	(0.011)	(0.018)
NV elig. x post-02 x black	0.034**	0.023	0.027	0.032*	0.016	0.025
	(0.016)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)
NV elig. x post-02 x young	-0.027	-0.013	-0.021	-0.018	-0.018	0.001
	(0.017)	(0.019)	(0.019)	(0.019)	(0.018)	(0.016)

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: This table shows robustness tests for our 'most responsive' judges analysis. We use six different methods of defining the sample of 'most responsive' judges. Each subsample is defined by using regression discontinuity to identify judges for whom there is a large sentencing discontinuity around the low-risk cut off for the nonviolent risk assessment. However, the exact specification varies: the bandwidth, h, is chosen from 4,7,10 and the outcome is chosen from sentence length (arcsinh), probability of incarceration. We include all cases seen by a judge who has an RD coefficient above the median in the subsample. The bandwidth and outcome that defines the specification are shown as column headers. For instance, Column 1 shows estimates for the subsample defined by a bandwidth of seven and the sentence length as the RD outcome. The dependent variables in Panel A-C respectively are the probability of incarceration, the sentence length (arcsinh), and the likelihood of being convicted of a new felony within three years of sentencing. The top part of each panel shows difference-in-differences results; the bottom two parts of each panel show triple differences results for race and age disparities.

Table A7: Effect of incarceration on recidivism for young defendants: discontinuity-in-risk-score estimates

			Reduced form			IV		
	Pr(inc.)	Sent.	1yr	3yr	7yr	1yr	3yr	7yr
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD_Estimate	0.072	0.208	-0.005	-0.058	-0.051	0.011	-0.234	-0.215
	(0.062)	(0.205)	(0.049)	(0.062)	(0.063)	(0.193)	(0.288)	(0.293)
Mean DV	0.710	1.817	0.155	0.343	0.415	0.155	0.343	0.415
Observations	2837	2837	2837	2837	2837	2837	2837	2837

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Note: All columns show RD estimates that exploit discontinuities in risk classification at a cutoff in the risk score. The outcome variables in Columns 1 and 2 are the probability of incarceration and the sentence length (arcsinh transform). The outome variables in Columns 3-5 are recidivism within different time windows. The outcome variables in Columns 6-8 are also recidivism in different time windows, but these estimates are from a fuzzy RD regression in which the length of the sentence (with an arcsinh transform) is the endogenous variable (in other words, Column 2 is the first stage). The sample includes only defendants under the age of 23; recidivism is defined here as the likelihood of receiving a new felony charge within X years of sentencing. The mean dependent variable is shown for defendants whose risk score is within -3 to -1.

Table A8: Evaluating the risk scores' ability to sort defendants by recidivism risk using discontinuities in the in/out score

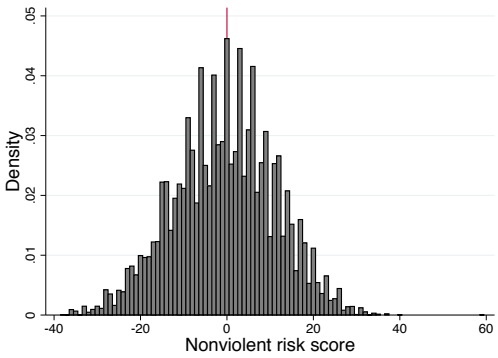
	Real risk score			Alt. risk score		
	1st stage	2nd stage		1st stage	2nd stage	
	1yr incarc.	1yr recidivism		1yr incarc.	1yr recidivism	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Lowest tercile in risk score						
RD_Estimate	0.315****	-0.158****	-0.144****	0.369****	-0.122***	-0.127***
	(0.0116)	(0.0449)	(0.0425)	(0.0119)	(0.0430)	(0.0439)
Observations	15586	15586	15586	15731	15731	15731
Mean DV	0.358	0.358	0.358	0.431	0.431	0.431
Panel B: Middle tercile in risk score						
RD_Estimate	0.339****	-0.235****	-0.233****	0.361****	-0.181****	-0.208****
	(0.0111)	(0.0433)	(0.0428)	(0.0107)	(0.0397)	(0.0410)
Observations	17019	17019	17019	16723	16723	16723
Mean DV	0.465	0.465	0.465	0.503	0.503	0.503
Panel C: Highest tercile in risk score						
RD_Estimate	0.351****	-0.267****	-0.268****	0.339****	-0.282****	-0.268****
	(0.0105)	(0.0414)	(0.0413)	(0.0110)	(0.0408)	(0.0403)
Observations	17070	17070	17070	17221	17221	17221
Mean DV	0.591	0.591	0.591	0.487	0.487	0.487
Covariates	N	N	Y	N	N	Y

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

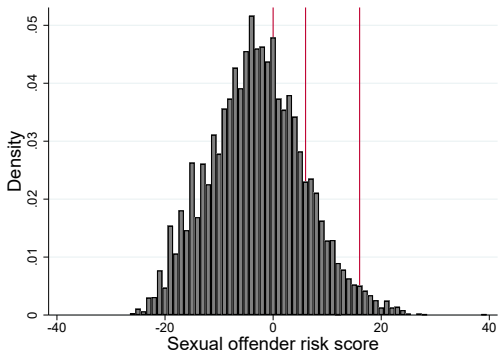
Note: This table presents RD estimates of the impact of incarceration on recidivism for defendants who are in different terciles of both the real risk scores and the alternative risk scores. The real risk scores are those used on nonviolent offenders in Virginia. The alternative risk scores were built by us using a random forest model trained on data from released defendants during the two years prior to risk assessment adoption. Columns 1 and 4 show the first stage of the instrumental variables regression; the endogenous variable is equal to one if the sentence is at least one year and is equal to the number of months of the sentence divided by 12 if the sentence is less than one year. The outcome is whether or not the defendant is charged with a new felony offense within one year of sentencing.

Figure A.1: Distribution of risk scores

(a) Distribution of nonviolent risk score



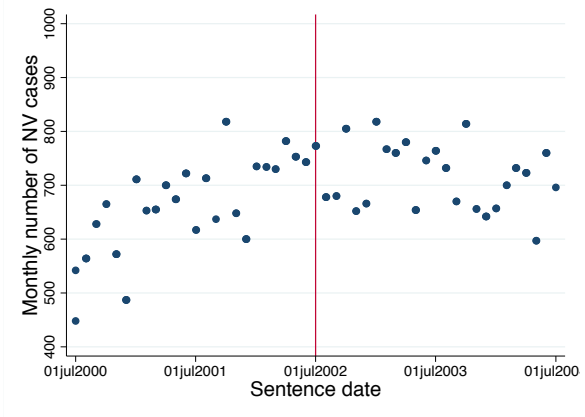
(b) Distribution of sex offender risk score



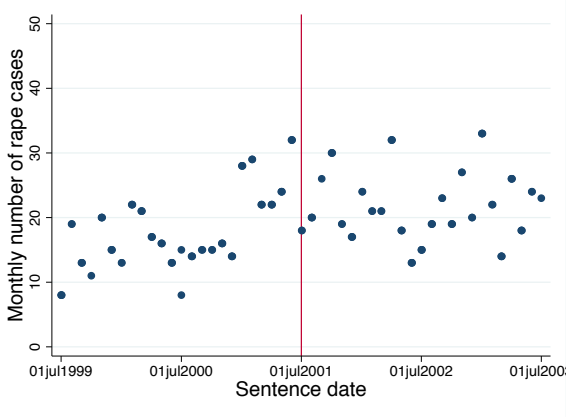
Note: These figures show the distribution of the nonviolent risk score and the sex offender risk score. The vertical lines indicate the cutoffs that delineate the different risk classifications and trigger a change in the sentence recommendations: diversion for nonviolent offenders below the cutoff, and an increased upper bound of the guidelines-recommended sentence for sex offenders above the various cutoffs. The risk scores have been normalized so that defendants with a score of 0 and above are in the higher risk classification.

Figure A.2: Case frequency around time of risk assessment adoption

(a) Case frequency of nonviolent risk-assessment-eligible cases around time of risk assessment adoption



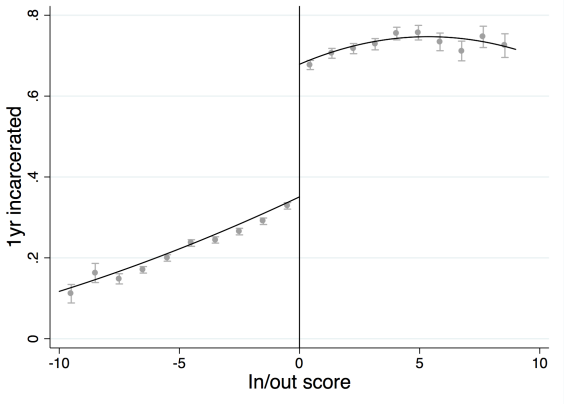
(b) Case frequency of rape cases around time of risk assessment adoption



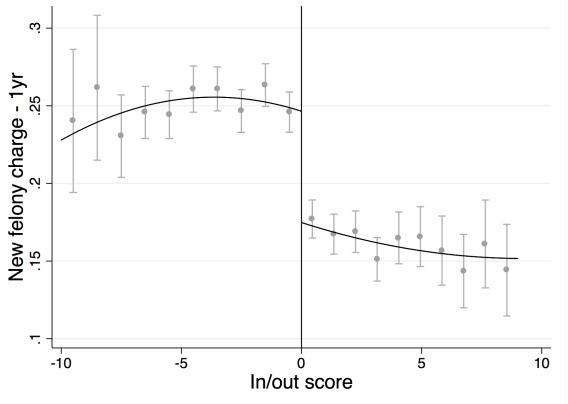
Note: These two scatterplots show case frequency over time for nonviolent risk-assessment-eligible offenders and sex offenders. Each dot represents a one month average. The vertical line indicates the date that each risk assessment is adopted.

Figure A.3:

(a) RD in the in/out score: incarceration (1yr)



(b) RD in the in/out score: recidivism (1yr)



Note: The left-hand figure plots the fraction of the first post-sentencing year spent incarcerated against the in/out score score (used to determine whether the guidelines-recommended sentence is prison). The in/out score score has been normalized; scores above 0 are recommended for prison. The right figure plots the one year recidivism rate (defined as being charged with a new felony offense within one year of sentencing) against the in/out score score.

Nonviolent Risk Assessment — Section D

Offender Name:

◆ Ineligibility Conditions

A. Was the offender recommended for **Probation/No Incarceration** on Section B? ☐ Yes ☐ No

B. Do any of the offenses at sentencing involve the sale, distribution, or possession with intent, etc. of cocaine of a combined quantity of 28.35 grams (1 ounce) or more? ☐ Yes ☐ No

C. Are any prior record offenses violent (Category I/II listed in Table A of the Guidelines Manual)? ☐ Yes ☐ No

D. Are any of the offenses at sentencing violent (Category I/II listed in Table A of the Guidelines Manual)? ☐ Yes ☐ No

If answered YES to ANY, go to "Nonviolent Risk Assessment Recommendations" on cover sheet and check Not Applicable. If answered NO to ALL, complete remainder of Section D worksheet.

◆ Offense Type *Select the type of primary offense*

Drug	3
Fraud	3
Larceny	11

◆ Additional Offense(s) — If YES, add 5 → 0

◆ Offender Score factors A to D and enter the total score

A.	Offender is a male	8
B.	Offender's age at time of offense	
	Younger than 30 years	13
	30 - 40 years	8
	41 - 46 years	1
	Older than 46 years	0
C.	Offender not regularly employed	9
D.	Offender at least 26 years of age & never married	6

\+ \+ \+ = Enter A to D Total: 0

◆ Arrest or Confinement Within Past 18 Months (prior to instant offenses) — If YES, add 6 → 0

◆ Prior Felony Convictions and Adjudications *Select the combination of adult and juvenile felony convictions/adjudications that characterizes the offender's prior record.*

Adult felony convictions only	3
Juvenile felony convictions or adjudications only	6
Both adult and juvenile felony convictions/adjudications	9

0

◆ Prior Adult Incarcerations

Number:	1 - 2	3
	3 - 4	6
	5 or more	9

0

Total Score → 0

☐ 38 or less, check Recommended for Alternative Punishment.
☐ 39 or more, check NOT Recommended for Alternative Punishment.

Go to Cover Sheet and fill out Nonviolent Risk Assessment Recommendations.

Go to Cover Sheet | Clear Form

Drug Schedule I or II/ Section D Eff. 7-1-04

Sentencing Guidelines Cover Sheet

Complete this form ONLY for applicable felonies sentenced on or after **July 1, 2004**.

Clear Form

◆ OFFENDER

First: _____ Middle: _____

Last: _____ Suffix: _____

Date of Birth: ____ (Month) / ____ (Day) / ____ (Year) Social Security Number: _____

CCRE: V A _____ PSI Number: _____ *For Use by Probation Officer*

◆ COURT

Judicial Circuit : _____ City/County: _____ FIPS Code: _____

Judge's Name: _____ *Office Use Only*

Preparer Name: _____ Preparer Title: ☐ Commonwealth's Attorney ☐ Probation Officer

Prosecuting Commonwealth's Attorney: _____ Defense Attorney: _____

◆ CONVICTIONS

Offense	Counts	VCC	Offense Date (Month / Day / Year)
Primary Offense:			
Additional Offenses:			

More

Primary Offense Code Section : § _____ Docket Number: _____

◆ METHOD OF ADJUDICATION

☐ Jury Trial → **Sentence Set by Jury:** ☐ Life Sentence (Enter Sentence: Years / Months / Days)

☐ Bench Trial ☐ Guilty Plea ☐ Alford Plea/Nolo contendere

◆ SENTENCING GUIDELINES RECOMMENDATIONS

Section B

- ☐ Probation / No Incarceration
- ☐ Incarceration 1 Day to 3 Months
- ☐ Incarceration 1 Day to 6 Months
- ☐ Incarceration 3 to 6 Months
- ☐ Probation / No Incarceration or Incarceration to 6 Months

Mandatory Minimum _____

Section C

- ☐ Life Sentence
- ☐ Incarceration *(Enter Midpoint and Range Below)*

Range Midpoint ____ Years ____ Months

Sentence Range ____ Years ____ Months TO ____ Years ____ Months

☐ Recommendation Adjusted for **Mandatory Minimum**

◆ NONVIOLENT RISK ASSESSMENT Section D of Drug, Fraud, and Larceny Worksheets

- ☐ Recommended for Alternative Punishment
- ☐ NOT Recommended for Alternative Punishment
- ☐ Not Applicable

Eff. 7-1-04

Rape — Section A

Offender Name:

◆ **Offender's Age at Time of Offense**

Younger than 35 years	12
35 to 46 years	4
Older than 46 years	0

◆ **Less than 9th Grade Education** — If YES, add 4 → 0

◆ **Not Regularly Employed** — If YES, add 5 → 0

◆ **Offender's Relationship with Victim**

Victim Under Age 10	Relative	0
	Known to victim (not relative or step-parent)	4
	Stranger	4
	Step-parent	9
Victim Age 10 or more	Relative	2
	Known to victim (not relative or step-parent)	3
	Stranger	8
	Step-parent	2

0

◆ **Location of Offense**

Place of employment	0
Shared victim/offender residence	3
Outdoors	3
Motor Vehicle	4
Victim's residence (not offender's)	5
Offender's residence or other residence	9
Location other than listed	3

0

◆ **Prior Adult Felony/Misdemeanor Arrests for Crimes Against Person**

Number:			
	0 Felonies	1 - 3 Misdemeanors	1
		4+ Misdemeanors	8
	1 Felony	0 - 2 Misdemeanors	5
		3+ Misdemeanors	8
	2+ Felonies	0 - 3 Misdemeanors	8
		4+ Misdemeanors	15

◆ **Prior Incarcerations/Commitments** — If YES, add 3 → 0

◆ **Prior Treatment**

Prior mental health commitment	0
Prior mental health treatment	2
Prior alcohol or drug treatment	3
No prior treatment	4

0

Risk Score → 0

Risk Level (Risk Score)

☐ 44 or more	Level 1
☐ 34 - 43	Level 2
☐ 28 - 33	Level 3
☐ up to 27	No Adjustment

Go to **Section C**

Rape Sentencing Guidelines Cover Sheet

Complete this form ONLY for applicable felonies sentenced on or after **July 1, 2002**.

Clear Form

◆ OFFENDER

First: _____ Middle: _____

Last: _____ Suffix: _____

Date of Birth: ____ (Month) / ____ (Day) / ____ (Year) Social Security Number: _____

CCRE: V A _____ PSI Number: _____ *For Use by Probation Officer*

◆ COURT

Judicial Circuit : _____ City/County: _____ FIPS Code: _____

Judge's Name: _____ *Office Use Only*

Preparer Name: _____ Preparer Title: ☐ Commonwealth's Attorney ☐ Probation Officer

Prosecuting Commonwealth's Attorney: _____ Defense Attorney: _____

◆ CONVICTIONS

Offense	Counts	VCC	Offense Date
Primary Offense:			Month / Day / Year
Additional Offenses:			

Primary Offense Code Section : § _____ Docket Number: _____

◆ METHOD OF ADJUDICATION

☐ Jury Trial → **Sentence Set by Jury:** ☐ Life Sentence Enter Sentence _____ Years _____ Months _____ Days

☐ Bench Trial ☐ Guilty Plea ☐ Alford Plea/Nolo contendere

◆ SENTENCING GUIDELINES RECOMMENDATIONS

Section C

☐ Incarceration *(Enter Midpoint and Range Below)* ☐ Life Sentence

Range Midpoint _____ Years _____ Months

Sentence Range _____ Years _____ Months TO _____ Years _____ Months

☐ Recommendation Adjusted for **Mandatory Minimum**

Modifications Based on Risk Assessment

The upper end of the sentence range can be adjusted based on the risk assessment level.

Characteristics of the offender and the circumstances of the offense may have correlated with a significant risk of recidivism among other sex offenders. If so, the upper end of the recommended sentence range has been increased by:

Check one

☐ 300% - Level 1
☐ 100% - Level 2
☐ 50% - Level 3
☐ No Adjustment

Adjusted High End _____ Years _____ Months

Eff. 7-1-02