ASSESSING RISK ASSESSMENT IN ACTION

Megan Stevenson[*][†]

December 8, 2017

ABSTRACT

Recent years have seen a rush towards evidence-based tools in criminal justice. As part of this movement, many jurisdictions have adopted actuarial risk assessment to supplement or replace the ad-hoc decisions of judges. Proponents of risk assessment tools claim that they can dramatically reduce incarceration without harming public safety. Critics claim that risk assessment will exacerbate racial disparities. Despite extensive and heated rhetoric, there is virtually no evidence on how use of this "evidence-based" tool affects key outcomes such as incarceration rates, crime, or racial disparities. The research discussing what "should" happen as a result of risk assessment is hypothetical and largely ignores the complexities of implementation. This Article is one of the first studies to document the impacts of risk assessment in practice. It evaluates pretrial risk assessment in Kentucky, a state that was an early adopter of risk assessment and is often cited as an example of best-practices in the pretrial area. Using rich data on more than one million criminal cases, the paper shows that a 2011 law making risk assessment a mandatory part of the bail decision led to a significant change in bail setting practice, but only a small increase in pretrial release. These changes eroded over time as judges returned to their previous habits. Furthermore, the increase in releases was not cost-free: failures-to-appear and pretrial crime increased as well. Risk assessment had no effect on racial disparities in pretrial detention once differing regional trends were accounted for.

Kentucky's experience does not mean we should abandon risk assessment, but it should temper the hyperbolic hopes (and fears) about its effects. Risk assessment in practice is different from risk assessment in the abstract, and its impacts depend on context and details of implementation. If indeed risk assessment is capable of producing large benefits, it will take research and experimentation to learn how to achieve them. Such a process would be evidence-based criminal justice at its best: not a flocking towards methods that bear the glossy veneer of science, but a careful and iterative evaluation of what works and what does not.

---

[*] Assistant Professor of Law, George Mason University, msteven@gmu.edu

TABLE OF CONTENTS

INTRODUCTION

In recent years, criminal justice has been marked by a surge of popularity for evidence-based practices. The evidence-based criminal justice movement promises to lower incarceration rates and taxpayer costs without affecting public safety through the use of methods that have been proven effective by scientific research. This idea has broad appeal across the political spectrum and has had a large impact on law and policy, particularly since the budgetary crises of the recent recession.[1] Correspondingly, it has attracted considerable attention from legal scholars. Most legal scholarship that discusses evidence-based practices explores the relationship between a specific legal or normative question and a specific practice, not the evidence-based movement in broad strokes.[2] In recent years, however, some scholars have begun to turn a critical eye towards the set of ideas behind evidence-based criminal justice. They express concerns about whether the evidence-based criminal justice movement will expand state correctional control, distort perceptions of justice and disproportionately harm minorities.[3]

At the forefront of the evidence-based criminal justice movement are algorithmic risk assessment tools.[4] Risk assessment tools have been adopted by dozens of jurisdictions around the country and are used throughout the criminal proceedings: for purposes of bail, sentencing, probation, parole, and in juvenile

---

[1] *See infra* Part I.a (providing examples of evidence-based practices and ideas integrated into law); Cecilia Klingele, *The Promises and the Perils of Evidence-Based Corrections,* 91 NOTRE DAME L. REV. 537, 566-567 (2016) (discussing the extent to which evidence-based practices have been incorporated into law).

[2] *See e.g.,* Christopher Slobogin, *Reconceptualizing Due Process in Juvenile Justice: Contributions from Law and Social Science*, 57 HASTINGS L.J. 955, 984-988 (2006) (discussing how social science research can help determine effective due process protections for juveniles); John Monahan & Jennifer Skeem, *Risk Assessment in Criminal Sentencing,* 12 ANN. REV. CLIN. PSYCH. 489, (2016) (generally discussing the various roles that risk assessment may play in the sentencing process, and the challenges therein).

[3] *See* Klingele, *supra* note 1 (arguing that evidence-based criminal justice could result in an expansion of state correctional control); Jessica Eaglin, *Against Neorehabilitation,* (arguing that the ideas associated with evidence-based criminal justice do not provide a good platform for reform).

[4] Risk assessments are listed as number one on multiple lists of evidence-based practices in criminal justice. *See, e.g.,* National Institute of Corrections, *A Framework for Evidence Based Decision Making in Local Criminal Justice Systems, Third Edition*, 13 (Apr. 16 2010), http://www.pretrial.org/download/performance-measures/The%20EBDM%20Framework%202010.pdf; Crime & Justice Institute, *Implementing Evidence-Based Practice in Community Corrections: The Principles of Effective Intervention,* 3 (April 30, 2004), https://s3.amazonaws.com/static.nicic.gov/Library/019342.pdf.

justice.[5] As Professor Sonja Starr puts it: "It is an understatement to refer to risk assessment as a criminal justice trend. Rather we are already in the risk assessment era."[6] Proponents of risk assessment argue that by replacing the subjective, error-prone and ad-hoc assessments of judges with "scientifically validated" prediction tools it is possible to dramatically lower both incarceration rates and crime.[7] In one of the most carefully executed instances of the literature, the authors state that "crime can be reduced by up to 24.8% with no change in jailing rates, or jail populations can be reduced by 42.0% with no increase in crime rates" as the result of making pretrial custody decisions on the basis of a risk assessment algorithm.[8] Critics of risk assessment raise a number of issues, but the question that has perhaps received the most attention is the extent to which risk assessment tools are racist themselves. This concern was voiced by former Attorney General Eric Holder[9] and reflected in a widely-read study by ProPublica that claimed that black defendants who did not reoffend were more than twice as likely to be wrongly classified as high risk than white defendants.[10]

Despite the heated rhetoric on both sides of the aisle, virtually nothing is known about how the implementation of risk assessment affects key outcomes: incarceration rates, crime, misconduct, or racial disparities. The empirical research evaluating whether outcomes are improved by incorporating algorithmic risk assessment into the decision-making framework is beyond thin; it is close to non-existent.[11] Many of the "facts" that are cited about the impacts

---

[5] Angele Christin et al., *Courts and Predictive Algorithms,* Data and Civil Rights: A New Era of Policing and Justice, 2-3 (October 27, 2015), http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf (describing how risk assessment is used around the country).

[6] Sonja B. Starr, *The Risk Assessment Era: An Overdue Debate*, 27 FEDERAL SENTENCING REPORTER 205, 205 (2015).

[7] *See, e.g.,* Shima Baradaran & Frank L. McIntyre, *Predicting Violence*, 90 TEX. L. REV. 497, 553 (2012); Jon Kleinberg et al., *Human Decisions and Machine Predictions,* 1 (Nat'l Bureau of Econ. Research, Working Paper No. 23180, 2017), http://www.nber.org/papers/w23180.

[8] *Id.* at 1.

[9] Massimo Calibresi, *Exclusive: Attorney General Eric Holder to Oppose Data-Driven Sentencing*, TIME.COM (July 31, 2014), http://time.com/3061893/holder-to-oppose-data-driven-sentencing/ (discussing Holder's concerns that risk assessment will disadvantage the poor and minorities).

[10] Jeff Larson et al. *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[11] *See infra* Part II.b (an overview of the risk assessment evaluation literature); Richard Berk, *An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism,* 13 J. EXP. CRIMINOLOGY 193, 193 (2017) (stating that debates around risk assessment "have unfolded with scant information about how actuarial risk assessments have affected practices and outcomes").

of risk assessment come from sources that range from detail-light non-academic reports put out by the agencies who designed the risk tool to nothing more than a single slide in a Power Point presentation. Somehow, criminal justice risk assessment has gained the near-universal reputation of being an evidence-based practice despite the fact that there is virtually no research showing that it has been effective.

There is ample research by social scientists suggesting that risk assessment tools *should* have beneficial effects. Risk assessment tools have been shown to be predictive of future arrest, and there is research suggesting (although not definitively) that they are better at predicting future arrest than judges are. This is the evidence that has earned risk assessment the "evidence-based" moniker, and the sheen of scientific credibility that this moniker entails likely contributed to the exponential growth in its use. But transforming a practice that "should" be beneficial to one that *actually does* provide benefit is not always straightforward. The same human foibles that champions of risk assessment point to when arguing for the adoption of risk assessment tools also complicate risk assessment as a policy. For instance, risk assessment tools may not be used as designed: they may be ignored or used "off-label" to accomplish something other than what was intended. Judges may not understand exactly what the risk score is measuring, or what level of statistical risk is associated with each risk category. The tool may be good at predicting misconduct, but the interventions taken to ameliorate risk may actually exacerbate it. The pressures of re-election or re-appointment may impact how and when the risk tool is used. And so forth.

This Article attempts to shift the conversation on risk assessment away from the hypothetical and towards the practical. It argues that transforming risk assessment into effective policy requires thinking carefully about context and the details of implementation. The impacts of a risk assessment tool depend on much more than the quality of the tool itself, they depend on numerous design choices: what level of judicial discretion to allow, what criminal justice interventions are recommended for each risk group, how to communicate statistical risk to the decision-makers, what accountability measures are in place, etc. Getting these design choices right may take time and revision; determining what constitutes "right" takes discussion amongst stakeholders.

This Article also presents some of the first rigorous empirical evidence on the impacts of risk assessment in practice. In particular, it focuses on the role of risk assessment in the rapidly proliferating bail reform movement. In the last few years, dozens of jurisdictions have adopted risk assessment tools to help determine pretrial custody. This study evaluates risk assessment in a state that has been heralded as a leader in pretrial practices and whose use of pretrial

risk assessment is often cited as an example for other jurisdictions: Kentucky.[12] Kentucky has had some sort of pretrial risk assessment tool available to judges since 1976, however its use was optional and many judges disregarded it. In 2011 Kentucky passed a law (House Bill 463, or HB 463)[13] that made use of the pretrial risk assessment tool mandatory and declared a presumptive default of immediate release (without monetary bail) for all low and moderate risk defendants.[14][15] This template – the use of risk assessment as a method of moving away from monetary bail – is the basis of the bail reform movement that has been rapidly gaining momentum across the country.[16]

　　Using detailed data on more than a million criminal cases, this Article analyzes the use of pretrial risk assessment in Kentucky, with particular attention on the impacts of the 2011 law that made risk assessment use

---

[12] Kentucky and Washington D.C. are likely the two most commonly cited examples of exemplary pretrial practices. *See, e.g.,* Brief of Amici Curaie Current and Former District and State's Attorneys, State's Attorneys General, United States Attorneys, Assistant United States Attorneys, and Department of Justice Officials, in Support of Plaintiffs-Appellees, O'Donnell v. Harris County, Texas, 227 F. Supp. 3d 706 (S.D. Tex. 2016) (citing both Kentucky and Washington D.C. as an example of good pretrial practices). *See also* Arthur W. Pepin, Conference of State Court Administrators, *Evidence Based Pretrial Release,* 8-9 (2012-2013), https://www.pretrial.org/download/policy-statements/Evidence%20Based%20Pre-Trial%20Release%20-%20COSCA%202012.pdf (citing Kentucky as an example of "successful implementation of evidence-based pretrial assessments"); Shaila Dewan, *Judges Replacing Conjecture with Formula for Bail*, N.Y. TIMES, (June 28, 2015), https://www.nytimes.com/2015/06/27/us/turning-the-granting-of-bail-into-a-science.html?_r=0 ("Kentucky has used a risk-assessment tool for decades, and is a leader among states when it comes to court appearance rates and low recidivism."); Criminal Justice Policy Program, Harvard Law School, *Moving Beyond Money: A Primer on Bail Reform*, 19 (October 2016), http://cjpp.law.harvard.edu/publications/primer-bail-reform (highlighting Kentucky's use of pretrial risk assessment as an example that other jurisdictions are following).

[13] Public Safety and Offender Accountability Act, H.B. 463, 2011 Gen. Assemb., Reg. Sess. (Ky. 2011) [hereinafter HB 463].

[14] Ky. Rev. Stat. Ann. § 431.066(2) (codifying H.B. 463) (instructing judges to consider the risk assessment when considering release and bail); Ky. Rev. Stat. Ann. § 431.066(3) (instructing release on unsecured bond or own recognizance for low risk defendants); Ky. Rev. Stat. Ann. § 431.066(4) (instructing release on unsecured bond or own recognizance for moderate risk defendants with possible supervision, monitoring or other conditions of release); Ky. Rev. Stat. Ann. § 27A.096(1,2,3) (instructing judges to follow guidelines set by the Supreme Court on pretrial release or supervision for moderate and high risk defendants); Supreme Court of Kentucky, 2011-12, *Order Approving Judicial Guidelines for Pretrial Release And Monitored Conditional Release,* http://courts.ky.gov/courts/supreme/Rules_Procedures/201112.pdf (generally affirming the centrality of the risk assessment tool in the release decision although granting judges the latitude to deviate from it; instructing pretrial services to develop a risk reduction plan including various conditions of release for judges to consider for high risk defendants).

[15] 90% of defendants were ranked as low or moderate risk.

[16] *See infra* Part I.c (discussing the current bail reform movement).

mandatory. The primary conclusion is that risk assessment in Kentucky came "not with a bang but a whimper."[17] Despite being crafted with the explicit goal of lowering incarceration rates,[18] HB 463 led to only a trivial increase in pretrial release. Furthermore, the increase in the release rate was matched by an uptick in failures-to-appear (FTAs) and pretrial crime; a disappointing counter to hopes that all three margins could be improved simultaneously. The low increase in releases is partly because judges took advantage of the discretion allowed to them by law and ignored the recommendations in the majority of cases. But this is not the whole story. In fact, HB 463 led to a marked change in bail setting practices. There was a 63% increase in the rate at which judges gave low risk defendants non-financial release, and a more moderate increase in non-financial release for moderate risk defendants. High risk defendants were released at lower rates. Thus, while there was a change in the type of defendants released, as well as the conditions of release, the net effects on the overall release rate were small. Furthermore, they were not permanent: the sharp change in practices and outcomes that occurred right after the law was implemented eroded over time as judges returned to their previous bail-setting practices.[19] Within a couple of years, the release rate was lower than it was before the bill.

As for racial disparities, the story is less straightforward. Facially, HB 463 benefited white defendants more than blacks. However, this is not because the risk assessment was more racially biased than judicial discretion. Rather, it is due to regional differences in how judges responded to HB 463. Judges from predominantly white rural counties liberalized their bail setting practices more than judges from more racially mixed urban areas, but *within* the same county, white and black defendants saw similar increases in release. Once county effects were taken into account, racial disparities remain constant throughout the time period of the analysis.

In 2013 Kentucky adopted a new risk assessment tool called the PSA. This tool was developed by the Laura & John Arnold Foundation using a nationally representative dataset of more than 1.5 million observations. It has received considerable national attention and has become one of the most widely used pretrial risk assessment tools.[20] The switch from Kentucky's local risk

---

[17] T.S. Eliot, "The Hollow Men", 1925.

[18] HB 463 was drafted with the goals of reducing incarceration rates while maintaining public safety. *See* Sen. Tom Jensen & Rep. John Tilley, *HB 463 – Statement From Sponsors*, CRIMINAL LAW REFORM: THE FIRST YEAR OF HB 463, 1 (June 6, 2012) http://c.ymcdn.com/sites/kybar.site-ym.com/resource/resmgr/2012_Convention_Files/ac2012_2.pdf.

[19] The fact that judges drifted back to their previous bail setting habits means that a randomized control trial that evaluated only short term effects would overstate its impact.

[20] *See* Laura & John Arnold Foundation, *Public Safety Assessment,* http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-

assessment tool to the PSA did not result in any noticeable improvement in outcomes. There was a small increase in the use of non-financial bond, and essentially no effect on releases, failures-to-appear, pretrial crime, or racial disparities in detention.

It should be noted that these results are directly counter to two widely-cited previous reports on risk assessment in Kentucky that found that HB 463 and the adoption of the PSA led to a *decrease* in pretrial crime.[21] These reports suffered from an error in methodology that created the erroneous impression that risk assessment had more beneficial effects than it actually did. Such errors underline the importance of third-party academic research.

The Kentucky example is not expected to be fully dispositive of the impacts of risk assessments in all places and times. An evaluation of risk assessment in practice is always a joint evaluation of the tool and the way it was used, and this can vary depending on context and details of implementation. However, Kentucky's method of implementation during the time period of this analysis was fairly typical of how pretrial risk tools are used in most jurisdictions. Judges were provided with the labels "low", "moderate", and "high" instead of the actual statistical risk, judges were allowed full discretion in the final bail decision, and there was no public forum of transparency and accountability whereby people could evaluate the frequency which which judges deviated from the presumptive default of non-monetary release. If these details of implementation were otherwise, risk assessment may have had a different impact.

As a case study, however, Kentucky offers important lessons for the bail reform movement, as well as for jurisdictions that have implemented or are considering implementing risk assessment in other criminal justice contexts.[22] First, Kentucky's experience should temper hopes that risk assessment will provide some sort of magic bullet that will lead to a large increase in the number of people released pretrial with no concomitant costs in terms of the crime or appearance rate. Risk assessment may offer improvements over the status quo, but reform requires more than simply adopting a risk assessment and calling it done.[23] Second, the Kentucky findings should ease (but not eliminate) concerns that risk assessment tools will exacerbate racial disparities. While

---

assessment/ (stating that the PSA is currently being used in 29 jurisdictions including 3 entire states).

[21] *See infra* notes 91, 92 & accompanying text.

[22] Dozens of jurisdictions have recently adopted risk assessment tools, and others are actively considering it. The PSA was piloted in Kentucky in 2013 and is now in use in 29 jurisdictions including 3 entire states (*see* PUBLIC SAFETY ASSESSMENT, *supra* note 20). Risk assessment is being used in sentencing in at least twenty states (*see* Starr, *supra* note 6, at 1).

[23] *See supra* Part II.b.

pretrial risk assessment did not affect racial disparities in Kentucky once regional trends were accounted for, scholars should continue to evaluate this question in other jurisdictions. Third, Kentucky demonstrates the challenges of trying to change criminal justice decision-making while retaining judicial discretion. Kentucky's statutes suggest a strong presumption of pretrial release, which accords with the stated goals of the bill's sponsors.[24] If judges followed the recommendations associated with the risk assessment, 90% of defendants would be granted immediate non-financial release. In practice, only 29% are released on non-monetary bond at the first bail-setting. If judges are not convinced or coerced to follow statutory guidelines, a risk assessment tool will not be an effective method of liberalizing release.

Finally, this Article calls for a change in how evidence-based criminal justice is practiced and conceived. A practice should not be considered evidence-based because it is associated with big data sets and sophisticated techniques, it should be considered evidence-based because its impacts have been carefully researched and understood. Rapid proliferation of a method with no knowledge of its effects is risky. Further, it precludes meaningful dialogue between the many well-intentioned individuals who want our criminal justice system to improve but have differing expectations about what the new tool will bring.

While Kentucky's experience showed that the benefits of pretrial risk assessment might not be as easy to achieve as researchers had hoped, this Article should *not* be read as proof that risk assessment tools are useless. Risk assessment can be used in a variety of ways and towards a variety of goals. In fact, Kentucky recently made substantial changes to how pretrial risk assessment is used in their state. Aware that pretrial detention rates had been rising, Kentucky recently limited judicial discretion by granting automatic non-financial release for low and moderate risk defendants charged with non-serious crimes. (This change was too recent to be included in this analysis.) In its commitment to trying new methods, evaluating what works, adjusting to improve, and evaluating again, Kentucky provides an example of evidence-based practice at its best.

This Article proceeds as follows. Part I provides a brief overview of evidence-based criminal justice, risk assessments, and the current bail reform movement. Part II discusses the empirical literature on risk assessment: the papers that claim that risk assessment tools are better at predicting future crime than judges, the slim set of research on the impacts of risk assessment in practice, and recent studies of racial bias in risk assessments. Part III presents an empirical evaluation of pretrial risk assessment in Kentucky. In particular, it

---

[24] Jensen & Tilley, *supra* note 18, at 2 ("The reforms in House Bill 463 are expected to bring a gross savings of $422 million over ten years by reducing the state's burgeoning prison population.").

uses graphical time-trend analysis to show how HB 463 and the adoption of the PSA affected bail practices, release rates, pretrial misconduct, and racial disparities.   Part IV discusses various lessons that can be drawn from Kentucky's experience with risk assessment.

I. A BRIEF OVERVIEW OF EVIDENCE-BASED CRIMINAL JUSTICE, RISK ASSESSMENT, AND BAIL REFORM

A. *Evidence-Based Criminal Justice*

The "evidence-based" moniker is used in a variety of subjects and refers to the idea that practices should be rigorously evaluated for their efficacy. The phrase was first used in the early 1990's in the medical literature:[25] "evidence-based medicine" became the key term to describe a movement towards medical practices that had been proven effective in clinical trial as opposed to those that were only supported only by anecdote or opinion.[26]   The phrase "evidence-based" was first applied to criminal justice in the late 1990's,[27] but a shift towards evaluating criminal justice programs for their efficacy had begun long before that. In 1974 Robert Martinson published a synthesis of research in corrections that was broadly interpreted as showing that "nothing works", i.e. that programs designed to rehabilitate offenders do not actually lower crime.[28] This study led to a shift away from the rehabilitative model of corrections that had previously dominated and is often marked by scholars as the beginning of the "New Penology":[29] a paradigm in criminal justice where the goal is to manage risk, not rehabilitate those who commit crime. Criminologists, however, did not abandon hopes that certain criminal justice programs were effective. The rapid expansion of computer power in the 1980's and 90's was paralleled by a rapid expansion of criminal justice research, and scholars began to identify a selection of policies that appeared to be effective at reducing crime.[30]   The idea that "nothing works" slowly lost ground in favor of the idea that some methods *do* work.   The evidence-based criminal justice movement

---

[25] David Sackett et al., Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine, 268 JAMA 2420 (1992).

[26] A 1976 report from the U.S. Office of Technology Assessment stated that "only 10 to 20% of all procedures used in present medical practice have been proven by clinical trial." (Stan Orchowsky, Justice Research and Statistics Program, *An Introduction to Evidence-Based Practices,* (April 2014) at 2-3).

[27] Lawrence W. Sherman, Police Foundation, *Evidence Based Policing* (1998).

[28] Robert Martinson, *What Works?—Questions and Answers About Prison Reform*, 35 PUB. INT. 22, 25 (1974).

[29] Malcolm M. Feeley & Jonathan Simon, *The New Penology: Notes on the Emerging Strategy of Corrections and its Implications*, 30 CRIMINOLOGY 449 (1992).

[30] See Orchowsky, *supra* note 26, at 3-4; Klingele, *supra* note 1, at 544-55.

has the goal of identifying and expanding the use of practices that social science research has demonstrated to be effective.[31] Partly as a result of efforts by organizations such as the National Institute of Corrections and the Justice Reinvestment Initiative, the ideas associated with evidence-based criminal justice gained in popularity throughout the 2000s and are now core to law and policy around the country.[32]

The Office of Justice Programs provides a useful definition of key terms. They consider practices to be evidence-based "when their effectiveness has been demonstrated by causal evidence, generally obtained through high-quality outcome evaluations."[33] A program is "effective" if it achieves its intended outcomes, which often include reducing criminal behavior or misconduct.[34] "Outcome evaluations" refer to social science research that attempts to infer the causal impact of a particular program or policy by comparing outcomes for a group of people who were affected by that policy to outcomes for a control group of people who were not affected by that policy. (For example, this Article compares pretrial release rates for the group of defendants who were booked right before HB 463 was introduced, to pretrial release rates for defendants who were booked right after HB 463 was implemented.) The extent to which research can be interpreted as evidence of a causal relationship between a policy and an outcome depends on the extent to which other explanations for the correlation can be ruled out.[35]

Determining whether a particular policy is evidence-based depends on the quality, quantity and consistency of the social science research demonstrating its impact. One popular system of determining the extent to which a practice is evidence-based is shown visually as a pyramid, where practices can be classified, from top to bottom, as gold, silver, bronze, iron or dirt.[36] Gold standard programs are supported by high-quality experimental evidence, significant impacts on outcomes, and multiple site replications.[37] Lower ranked programs are supported by lower-quality evidence (e.g. research where alternative mechanisms could explain the results or where the sample is small) and the lowest ranked programs are either lacking in conclusive research or have been proven to be ineffective. According to this ranking, risk assessment would not be considered evidence-based, as there is no high quality evidence showing that is has been effective.

---

[31] IMPLEMENTING EVIDENCE BASED PRACTICES, *supra* note 4.

[32] For an excellent overview of the rise of evidence-based criminal justice see Klingele, *supra* note 1, at 551-567.

[33] CrimeSolutions.gov, https://www.crimesolutions.gov/GlossaryDetails.aspx?ID=15.

[34] CrimeSolutions.gov, https://www.crimesolutions.gov/GlossaryDetails.aspx?ID=13.

[35] CrimeSolutions.gov, https://www.crimesolutions.gov/GlossaryDetails.aspx?ID=7.

[36] See IMPLEMENTING EVIDENCE BASED PRACTICES, *supra* note 31, at 17.

[37] Id.

The term "evidence-based" is sometimes used in a looser way, as simply integrating the best available research into decision making and practice. When Kentucky passed the law mandating the use of pretrial risk assessment they did so on the best evidence that was available at the time. They were pioneers in pretrial risk assessment, and new techniques will never be fully proven when they are adopted.

The ideas and practices associated with evidence-based criminal justice has made significant headway into law and policy, at both the state, local, and federal level.[38] This can include both general instructions to use evidence-based principles,[39] specific instructions for the fraction of state expenditures that must be spent on evidence-based practices,[40] and orders to adopt specific evidence-based practices.[41]

B. *Criminal Justice Risk Assessment*

Evaluating the risk of future criminal activity has long been part of practice in criminal justice. The term "risk assessment", however, usually refers to the use of formal, actuarial, and algorithmic methods of predicting the likelihood of future crime or misconduct. (In practice, however, they predict what is visible: arrest, conviction, reincarceration, probation revocation, etc.) Actuarial risk assessment tools have been in use in criminal justice since the 1920s,[42] but their use has been rapidly accelerating over the last ten years.[43] They are used for purposes of determining bail or the conditions of release, in setting the sentence length or exempting offenders from minimum sentences, in determining the level of supervision for probationers, in evaluating a request for

---

[38] The Justice Reinvestment Initiative has helped spread evidence-based practices into 27 states (*see* Bureau of Justice Statistics, Justice Reinvestment Initiative: JRI Sites, https://www.bja.gov/programs/justicereinvestment/jri_sites.html (last visited August 5, 2017)); The director of the Administrative Office of the United States Courts attributes a decline in recidivism to their use of evidence-based practices (Annual Report 2016, Director's Message, http://www.uscourts.gov/statistics-reports/annual-report-2016 (last downloaded August 5, 2016).

[39] *See e.g.,* AZ ST Code of Jud. Admin., § 6-201.01 (instructing probation to develop evidence-based policies and procedures); Idaho Code § 20-219-5 (instructing the state board of corrections to use evidence-based practices in supervising probationers and parolees)

[40] *See e.g.,* Ky. Rev. Stat. Ann. § 27A.097(5) (stating that by July 1, 2016 75% of state moneys extended on supervision and intervention programs shall be for programs that are in accordance with evidence-based practices).

[41] *See e.g.,* 42 PA. CONS. STAT. § 2154.7 (adopting a risk assessment in sentencing); Washington SB-6204 Sec. 11.1&2 (adopting swift and certain sanctions in community supervision).

[42] Howard G. Borden, *Factors for Predicting Parole Success,* 19 JOURNAL OF THE AMERICAN INSTITUTE OF CRIMINAL LAW AND CRIMINOLOGY 328, (1928).

[43]

parole, and in choosing the appropriate rehabilitative program or restriction on liberty for juvenile offenders.[44]

Most risk assessment tools currently in use are fairly simple "checklist-style" tools.[45] These tools take a set of inputs, usually between six and twenty, and assign a certain number of points to each input. The points assigned to each input are determined through statistical analyses that evaluate how well each input predicts the outcome that the tool is designed to predict. The inputs to a risk assessment algorithm almost always include criminal history or criminal-justice-related misconduct, some also include socio-economic factors such as education level, marital status, or home neighborhood.[46] Age and gender are sometimes included, but race is not. The risk score is then calculated by summing the points assigned to each input. Usually, the risk score is then aggregated to a small group of risk classifications: people with the lowest scores are labeled low risk, those with medium scores are labeled moderate risk and those with the highest scores are labeled high risk. Determining what level of risk warrants the high-risk label is a normative choice.

In addition to the checklist style risk assessments described above, there are also more complicated methods of evaluating risk that are developed through a method called machine learning. Machine learned risk assessment tools are designed by a computer itself, with a little guidance from the person that develops them. The researcher tells the computer which inputs to use, which outcomes to predict, and which learning method to use. The computer does the rest. Machine learned risk assessments tend to be black-box mechanisms: it's hard to understand why they yield the predictions that they yield. This is because the relationship between the inputs and the risk score is non-linear and varied. For example, a machine learned risk instrument might show that the impact age has on the likelihood of future arrest is different for people who are facing drug charges than for those who are facing domestic violence charges. Machine learned predictions are usually more accurate than the simpler checklist style tools. However, they are still uncommon in criminal justice. The black-box nature of the tool makes them non-transparent, which raises legal and ethical issues:[47] it's difficult to challenge a high risk

---

[44] *See supra* note 5 & accompanying text.

[45] *See* Mayson, *infra* note 159, at 13-18.

[46] *See* Mayson, *infra* note 159, at 13-18 (explaining the construction of common pretrial risk assessment instruments); Melissa Hamilton, *Adventures in Risk: Predicting Violent and Sexual Recidivism in Sentencing Law,* ARIZONA STATE L. J. 1, 14-18 (explaining the construction of risk assessment instruments predicting violent or sexual crime).

[47] *See* Kelly Hannah-Moffatt, *Actuarial Sentencing: An Unsettled Proposition,* 30 JUSTICE QUARTERLY 270, (2013) (discussing transparency concerns with risk assessment); Melissa Hamilton, *Risk Needs Assessment: Constitutional and Ethical Challenges,* 52 AMERICAN CRIM. L. REV. 231, 267-271 (2015) (discussing legal issues related to transparency in risk assessment).

classification if one does not know the reasons behind the classification. Furthermore, they require a higher level of technical training to build and implement.

Often, the risk classifications come with explicit recommendations for action. A basic (although not universal) principle is that both rehabilitative interventions and restrictions on liberty increase as the risk level increases. The choice of what type of risk to predict (i.e. what outcome in what time window), which algorithm to use to predict that risk, how to divide the group into different classification levels, and which criminal justice actions (e.g. bail amounts, sentence lengths, etc.) are appropriate for each risk level are all choices that depend, at least partially, on the normative and legal landscape.

Risk assessment tools are one of the most prominent and widely adopted methods associated with the evidence-based criminal justice movement. The National Institute of Corrections, an organization that has been deeply involved in the advancement of evidence-based criminal justice, places risk assessment tools at number one in a list of evidence-based ways to reduce recidivism.[48] Risk assessment tools are so closely tied to the evidence-based movement that the terminology is sometimes interchangeable: the use of risk assessment in sentencing is often referred to as simply "evidence-based sentencing".[49]

C. *Risk Assessment in Bail Reform*

The method of determining which defendants are released, released on conditions, or detained pretrial has been one of the most rapidly changing areas of criminal justice over the last couple of years.[50] The idea behind the current bail reform movement is to make this decision on the basis of the risk of flight or future crime, not the ability to pay bail. Critics of the monetary bail system argue that conditioning release on money results in racial and wealth-based disparities in detention, a waste of taxpayer money, and harm to public safety.[51] Defendants who pose a low risk of crime or flight, they argue, should not be detained due to an inability to pay monetary bail. Conversely, wealthy defendants who pose a high risk of serious crime should not be released simply because they can afford bail. Many, including this author, have argued that pretrial detention or electronic monitoring should be reserved for those who

---

[48] *Supra* note 4.

[49] See Sonja Starr, Evidence-Based Sentencing and the Scientific Rationalization of Discrimination, 66 STANFORD L. REV. 803, 805 (2014).

[50] For an overview of the history of bail see Tim Schacke, Fundamentals of Bail: A Resource Guide for Pretrial Practitioners and a Framework for American Pretrial Reform, (Aug. 2014) http://www.clebp.org/fundamentalsofbail.html.

[51] *See e.g.* MOVING BEYOND MONEY, *supra* note 12, at 6-8.

pose a high risk of violent crime or flight.[52] According to Chief Justice Rehnquist, "In our society liberty is the norm, and detention prior to trial, or without trial, is the carefully limited exception."[53]

Dozens, if not hundreds, of jurisdictions are pursuing or have recently implemented wholesale changes to their bail practices.[54] By a recent count, bail reform efforts are active in all but a handful of states.[55] Support for reform can be found across party lines and across agencies: public defenders, district attorneys, judges, governors, sheriffs, and so forth. Bail reform is a rare area of bipartisan cooperation in the U.S. Senate: Kamala Harris (Dem.) and Rand Paul (Rep.) recently introduced a joint bill to use federal funding to encourage states to reform or replace the practice of money bail.[56] Change has come in the form of new legislation,[57] revisions to state constitutions,[58] new judiciary rules as decreed by state courts,[59] and as the result of civil rights litigation.[60] Class action lawsuits have been filed in jurisdictions across the country claiming that current bail practices violate due process protections and the Equal Rights Amendment.[61] These lawsuits have resulted in a number of consent decrees entailing reform to local pretrial processes, as well as a landmark federal ruling, *O'Donnell v. Harris County,* ordering pretrial release of misdemeanor defendants who can't afford bail.[62] This ruling is currently under appeal at the

---

[52] Megan T. Stevenson and Sandra Mayson, *Bail Reform: New Directions for Pretrial Detention and Release,* 21 (March 13, 2017). *In* ACADEMY FOR JUSTICE, A REPORT ON SCHOLARSHIP AND CRIMINAL JUSTICE REFORM (Erik Luna ed., 2017, Forthcoming).

[53] United States v. Salerno, 481 U.S. 739 (1987).

[54] Pretrial Justice Institute, *Where Pretrial Improvements are Happening,* July 2017 (providing an overview of recent changes to pretrial practices across the United States)

[55] *Id.* at 15-17.

[56] The Public Safety and Accountability Act of 2017.

[57] *See e.g.,* Mark Pazniokas & Keith M. Phaneuf, *Bail Reform Wins Final Passage in the Senate,* ctmirror.org, June 7, 2017, https://ctmirror.org/2017/06/07/bail-reform-wins-final-passage-in-senate/ (describing new bail reform law in Connecticut).

[58] *See infra,* note 67 & accompanying text.

[59] *See e.g.* Michael Dresser, *Maryland Court of Appeals: Defendants Can't be Held in Jail Because they Can't Afford Bail,* BaltimoreSun.com, Feb. 8, 2017, http://www.baltimoresun.com/news/maryland/bs-md-bail-rule-20170207-story.html (discussing a new rule by Maryland's highest court that is designed to reduce the use of monetary bail).

[60] Litigation designed to reform bail practices is either active or recently resolved in Texas, Louisiana, Illinois, Massachusetts, Georgia, Tennesssee, Alabama, California, Kansas, Mississippi and Missouri. *See* Civil Rights Corps, *Ending Wealth Based Pretrial Detention,* CivilRightsCorps.org, http://www.civilrightscorps.org/ending-wealth-based-pretrial-detention (last visited Aug, 8, 2017);

Equal Justice Under All, *Ending the American Money Bail System,* EqualJusticeUnderAll.org, http://equaljusticeunderlaw.org/wp/current-cases/ending-the-american-money-bail-system/ (last visited Aug. 8, 2017).

[61] Id.

[62] O'Donnell v. Harris County, Texas, 227 F. Supp. 3d 706 (S.D. Tex. 2016).

Fifth Circuit; if it stands it is likely to transform bail systems around the country and decrease jail populations considerably.

In shifting the emphasis towards risk as opposed to the ability to pay bail, the recent bail reform movement has been intimately linked with the adoption of actuarial risk assessment instruments.[63] While exact details differ across jurisdictions, the new model generally involves reducing or eliminating the use of monetary bail, and adopting a risk assessment tool to help the judges make decisions about pretrial custody).[64] This has resulted in a rapid proliferation of the use of pretrial risk assessments. The pretrial risk assessment tool developed by the Arnold Foundation has been adopted by dozens of jurisdictions and three entire states in the last few years.[65] The Harris-Paul bail reform bill encourages states to replace money bail with pretrial risk assessment.[66] States such as New Jersey and New Mexico have revised their constitution to allow for direct orders of detention on the basis of risk as determined, at least in part, by actuarial risk assessment.[67] Across the country, as a result of changes enacted by the executive branch, legislature, and the judiciary, jurisdictions are adopting pretrial risk assessment.[68]

The current wave of bail reform is still in flux. The extent and the exact nature of the changes depend partially on battles that are being waged in city halls, courthouses, and the court of public opinion around the country. Risk assessments are controversial, and not all agree that they should play a central role in bail reform.[69] Currently, however, risk assessments are a dominant theme in a rapidly accelerating reform movement.

---

[63] *See* Mayson, *infra* note 159, at 12-20 (an overview of the recent bail reform movement).

[64] *Supra* note 54 (listing adoption of risk assessment as part of bail reform in many jurisdictions).

[65] *Supra* note 20.

[66] *Harris, Paul Introduce Bill to Encourage States to Reform or Replace Unjust, Costly Money Bail System,* (July 20, 2017), https://www.harris.senate.gov/content/harris-paul-introduce-bill-encourage-states-reform-or-replace-unjust-costly-money-bail.

[67] *See* Judge Glenn A. Grant, *infra* note 112, at 3 (discussing role of risk assessment in N.J.'s reforms); Mathew Coyte, *New Pretrial Risk Assessment More Fair*, Albuquerque Journal, June 14, 2017, https://www.abqjournal.com/1017512/new-pretrial-risk-assessment-more-fair.html (discussing recent N.M.'s recent constitutional changes and the implementation of risk assessment Albuquerque's most populous county).

[68] *Supra* note 54, at 10-14.

[69] *See e.g.,* John Raphling, Human Rights Watch, *Human Rights Watch Advises Against Using Profile Based Risk Assessment in Bail Reform*, (July 17 , 2017), https://www.hrw.org/news/2017/07/17/human-rights-watch-advises-against-using-profile-based-risk-assessment-bail-reform (arguing against the use of risk assessment tools in bail reform).

II. Risk Assessment: The Slim Evidence on this "Evidence-Based" Practice

This section provides an overview of the empirical literature that has been used to substantiate beliefs about the impacts of risk assessment.[70] It discusses the literature arguing that risk assessment tools are better at predicting future crime than judges, the slim empirical evidence on the impacts of risk assessment in practice, and the empirical studies that serve to raise or mitigate concerns about racial bias in risk assessment.

A. *Algorithmic Prediction vs. Human Intuition*

The most common argument in support of risk assessment is that formal, actuarial, and algorithmic methods of prediction perform better than the intuitive methods used by judges or other experts.[71][72] Thus, by making smarter decisions about who to release, we could decrease detention rates while keeping crime and non-appearance rates constant, or vice versa. Most of the research that on this question comes from the 1950s-1980s psychology literature. A broadly cited meta-analysis of such papers claims that "on average, mechanical [i.e. actuarial and algorithmic] prediction techniques were about 10% more accurate than clinical [i.e. human] predictions."[73] This meta-analysis cites 10 papers that compare algorithmic to human predictions in the criminal justice context.[74] All of these papers date to before 1988 and most use small samples and questionable analytic techniques.[75] In addition to the older literature,

---

[70] This section does not include studies showing that risk assessments are effective at predicting criminal activity unless the study compares the predictive power of risk assessments to the informal predictions of judges or other criminal justice practitioners.

[71] *See, e.g.,* Sarah Picard-Fritsche et al., Center for Court Innovation, *Demystifying Risk Assessment: Key Principles and Controversies*, 8 (2017) ("On balance, actuarial—or data-driven— risk models have tended to outperform the judgments of individual practitioners, including clinical professionals, in accurately assessing risk. Thus the rationale behind expanding the use of formal risk assessment tools is that they offer the potential for helping justice agencies make more informed decisions.").

[72] Sam Wiseman provides an interesting argument for adopting risk assessment as a way to counter a principal-agent problem: by shielding judges from personal responsibility their actions may closer reflect society's interests. Samuel R. Wiseman, *Fixing Bail,* 84 George Washington L. Rev. 417, 438-454 (2016).

[73] William M. Grove et al. *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 12 Psychological Assessment 19, 19 ("On average, mechanical predictions were about 10% more accurate than clinical predictions").

[74] *Id.* at 22-24.

[75] Thomas R. Litwack *Actuarial Versus Clinical Assessments of Dangerousness*, 7 Psychology 409, (2001) (generally critiquing analytic techniques in Grove et. al.); Starr, supra note 49, at 850-855 (arguing that there is not yet any persuasive evidence that actuarial instruments outperform judge's predictions).

several more recent papers argue that statistical tools are better at predicting future offending than judges or magistrates.[76] These papers use much larger data sets and more advanced methodologies than the previous ones. However, both the older papers and the new follow a similar pattern and are susceptible to many of the same critiques.

The idea that actuarial tools outperform human intuition in predicting crime has become broadly accepted.[77] This may be true, and indeed many papers suggest that it is. However, this is not something that is easy to prove. Ideally, research comparing the two methods of prediction would be explicitly set up as a horse race between the two approaches. Both humans and algorithms would be informed of the goals of the study and would make predictions accordingly. A winner would be declared based on the accuracy of their predictions. Unfortunately, prior studies comparing different methods of predicting crime do not follow such an approach.

Instead, prior studies have had to make do with second best situations: real world settings where the predictions of the humans may not be directly observed, where humans may have different goals than the algorithm, and where future crime is influenced by, and thus a direct function of, the predictions. A recently released paper called *Human Decisions and Machine Predictions* is one of the more carefully executed instances of the literature, and is a good demonstration of its strengths and limitations.[78] This paper uses detailed data on pretrial defendants in New York City to estimate the risk of failing to appear in court or committing another crime. The authors use machine learning techniques – complex computer-based methods of predicting risk – to build a risk prediction for each defendant.[79] The inputs to the predictions include criminal history, offense and age as predictors for misconduct (rearrest or nonappearance in court) among the group of defendants who were released pretrial.[80] The authors then conduct a policy simulation in which they estimate what the crime rate would have been if, instead of

---

[76] *See* Baradaran & McIntyre, *supra* note 7, at 553 ("Even with this increase in releases, because we are better targeting which defendants to release, pretrial violent-crime rates would decrease"); Kleinberg et al., *supra* note 8, at 28, ("While there is some variation across judges, the algorithm dominates each judge.").

[77] *See, e.g.*, Steven L. Chanenson & Jordan M. Hyatt *The Use of Risk Assessment at Sentencing: Implications for Research and* Policy, VILLANOVA UNIVERSITY CHARLES WIDGER SCHOOL OF LAW PUBLIC LAW & LEGAL THEORY WORKING PAPER NO, 2017-1040, 10 (2017) ("there is a significant literature that suggests that, with regard to accuracy, statistical methods generally outperform subjective clinical judgments"); Eric S. Janus & Robert A. Prentky, *Forensic Use of Actuarial Risk Assessment With Sex Offenders: Accuracy, Admissibility and Accountability*, 40 AM. CRIM. L. REV. 2 (2003) ("Our thesis is straightforward: actuarial methods have proven equal or superior to clinical judgements").

[78] *See* Kleinburg et al., *supra* note 7

[79] *Id.* at 16

[80] Id.

following the status quo procedure, the decision on whether to release or detain a defendant had been made solely by the machine learned algorithm. They estimate that if the detention decision was made by their tool that crime could be dramatically lowered while the detention rate remains constant, or that the detention rate could be dramatically lowered while the crime rate remains constant.[81]

One of the main challenges to determining whether the algorithm outperforms the judges is that the authors do not directly observe the judges' predictions. They attempt to infer the predictions by looking at which defendants were detained pretrial.[82] There is undoubtedly a connection between the predictions of the judge and the detention status of defendants, but this connection is noisy and mediated by several other factors. Most notably, judges are not, by and large, directly determining detention. They are setting bail, and the defendant will be released if he or she posts bail. The bail amount is not supposed to keep a defendant detained (although it can be used that way), it is supposed to provide incentive for a released defendant to return to court. Thus, the judges must predict several unknowns simultaneously: the risk of crime or failures-to-appear (FTA), the likelihood the defendant will post a given amount of bail, and the impact that bail will have on the defendant's pretrial misconduct. A defendant who was detained pretrial is not necessarily someone who the judge considered higher risk than one who was released.

Second, judges are likely to have more complicated preferences than the algorithm. For one, they may be taking into account factors other than risk. For example, a judge might find it inappropriate to detain someone on very minor charges, even if they pose a high statistical risk of future offending. Further, there are multiple types of risk that judges consider – FTA, violent crime, drug crime, etc. – and judges are likely to vary in the extent to which they are concerned with each. Aggregating multiple judges together means that even if each were performing optimally according to his or her own preferences, a risk algorithm could outperform the group of them on any single dimension.

Third, the ability to compare the risk prediction tool to the judges' intuition relies on the assumption that the crime risk of a detained defendant is equivalent to the crime risk of a released defendant with similar visible characteristics (ie criminal record, etc.).[83] This is almost certainly not true. For more serious crimes, where release rates are low, the released defendants will not be at all representative.

---

[81] *Id.* at 6.
[82] *Id.* at 37.
[83] *Id.* at 6, 19-24

The authors are not unaware of these confounds and make some clever attempts to address these issues in their paper.[84] They make some headway in providing evidence that the confounds cannot explain away their results. However, these are not minor confounds, they are fundamental challenges to the research design. Ultimately it is unclear how much more accurate the risk prediction algorithm is, if at all.

There are studies in other contexts that are explicitly set up as a horse race between the two types of prediction. These studies do not suffer from the confounds described above and provide more convincing evidence of the effectiveness of actuarial tools. In one study, a large group of law professors, law school deans and appellate attorneys were asked to predict outcomes in Supreme Court rulings related to their field.[85] Predictions were also generated using an algorithm that had been developed before the beginning of the study. Both sets of predictions were publicly posted before each ruling, and the accuracy was compared after each ruling. While the legal experts were correct 59% of the time, the model was correct 75% of the time.[86]

Despite the weakness of the literature on human vs actuarial prediction of crime, it remains reasonable to believe that a well-built actuarial tool can out-predict a judge on future offending. This is likely to be particularly true in the rapid, assembly-line style proceedings that characterize many bail hearings. However, the margin of improvement is not clear.

B. *Research on the Impacts of Risk Assessment*

Even if risk assessment tools are significantly better than judges at predicting future offending, that doesn't automatically mean that its adoption will lead to large benefits. The impact that risk assessment has in practice will depend on the way it is used. A variety of contextual and policy details will be influential: the amount of judicial discretion allowed, the judge's incentive structure, the fraction of defendants in each risk classification, the specific action recommendations associated with each risk classification, the court culture, etc. For one, judges may broadly ignore the tool. In a recent survey of judicial attitudes towards risk assessment at sentencing, less than 10% thought that the actuarial tools would predict better than judges.[87] A survey of more than 2000 probation and parole officers found that even among the most compliant officers, "practitioners routinely exercise substantial discretion to

---

[84] *Id.* at 18-24, 29-37.
[85] Theodore W. Ruger et al., The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking, 104 COLUMBIA L. REV. 1150, (2004).
[86] *Id.* at 1150.
[87] *See* Chanensen & Hyatt, *supra* note 77, at 10.

choose interventions that are more restrictive or intensive that the tool recommends."[88] Furthermore, judges may also be selective in when they believe the risk assessments are accurate: discounting its predictive validity for certain races, for example. For instance, a judge may think a black defendant is higher risk than their risk score indicates because of racist stereotypes. Alternatively, she may think a black defendant is lower risk than is indicated by the risk score because she is aware that racially disparate policing practices contributed to his previous arrests. While the question of racial disparity is often posed as a choice between biased instruments or biased judges, in practice, the important question is "how do the two interact?"

The adoption of a risk assessment tool is often paired with explicit or implicit recommendations for how judges are supposed to treat defendants in each risk classification. Simply dividing the group into different classification levels contains information about what fraction of defendants should be thought of as "low", "moderate" or "high" risk, and where the cutoffs for each group begin. In many cases, the action recommendations that correspond with each risk group are explicit. In the pretrial context, it is common to recommend non-financial release for low risk defendants, conditional release (with supervision or low monetary bond) for moderate risk defendants, and supervision or detention for high risk defendants.[89] Even if the policy is not explicit about what actions should be taken with each group, the action is usually understood. "High" risk defendants should be treated as high risk defendants are supposed to be treated in that particular jurisdiction; usually this translates into greater restrictions on liberty. An evaluation of the impacts of risk assessment in practice will thus be a joint evaluation of the tool, the manner in which it is used by the judge, and the recommendations that accompany it.

Such evaluations are very scarce. This is true in all areas where criminal justice risk assessments are used, and it is particularly true in the pretrial context. The little evidence that is available about the impacts of pretrial risk assessment come from detail-light non-academic reports usually put out by the organization who designed or implemented the tool. Two of the most commonly cited reports use data from Kentucky.[90] One is a report put out

---

[88]Joel Miller & Carrie Maloney, *Practitioner Compliance with Risk/Needs Assessment Tools* 40 CRIMINAL JUSTICE AND BEHAVIOR 716, 728 (2013).

[89] See, e.g., The Sentencing Project, Reducing Racial Disparity in the Criminal Justice System: A Manual for Practitioners and Policymakers, (2008), at 31.

[90] A third study also analyzes HB 463, but relies heavily on data from Kentucky Pretrial Services, *Report on Impact of House Bill 463: Outcomes, Challenges and Recommendations*, (2012). This study argues that judicial discretion has undermined the effectiveness of HB 463. (Robert Veldman, *Pretrial Detention in Kentucky: An Analysis of the Impact of House Bill 463 During the First Two Years of its Implementation,* 102 KENTUCKY LAW JOURNAL 777, 778 (2013)).

by Kentucky Pretrial Services evaluating House HB 463.[91] Another is a report put out by the Laura & John Arnold Foundation evaluating the adoption of their risk assessment tool, the PSA.[92]

Both reports are brief and show simply that the average detention and pretrial rearrest rates are lower in the period after the risk assessment change than they were in the period before.[93] (The Kentucky Pretrial Services report also claim that FTA rates were lowered.)[94] While these findings are often cited as evidence that risk assessment can jointly decrease both detention rates and crime, the articles provide little evidence that the changes cited come from the risk assessment. For one, it is possible that there was a steady decline in both detention rates and rearrest that started long before the period of analysis and had nothing to do with risk assessment. However, there is a more fundamental reason why the statistics presented in these reports cannot be interpreted as an evaluation of risk assessment. Both reports were published very soon after the change they are analyzing, and before all the cases in the sample were resolved.[95] The analysts do not correct for the fact that defendants whose cases were not yet resolved had, on average, less time in which to be rearrested than defendants whose original arrest occurred earlier.[96] This artificially deflates the rearrest rate for defendants whose arrest occurred after the adoption of the risk assessment. In statistical terms, this is called "truncation bias", and it erroneously made it appear like risk assessment led to lower instances of

---

[91] REPORT ON IMPACT OF HOUSE BILL 463, *supra* note 90.

[92] Laura & John Arnold Foundation, *Results from the First Six Months of the Public Safety Assessment – Court in Kentucky,* 1 (2014), (This study has recently been removed from the Arnold Foundation's website but is on file with the author. Representatives of the Arnold Foundation explained that it was removed due to concerns about the quality of the data used in the report.).

[93] REPORT ON IMPACT OF HOUSE BILL 463, *supra* note 90, at 6; RESULTS FROM THE FIRST SIX MONTHS, *supra* note 92, at 1.

[94] REPORT ON IMPACT OF HOUSE BILL 463, *supra* note 90, at 6.

[95] The publication date of REPORT ON IMPACT OF HOUSE BILL 463, *supra* note 90, is June, 2012. The post-HB-463 period is the year from June 2011-June 2012. Thus, the post-HB-463 period includes many cases which have not yet been adjudicated. The publication date of FIRST SIX MONTHS, *supra* note 93, is July 2014. The post-PSA period extends from July-December 2013 while the comparison group includes cases from July 2009 to July 2013. Only 83% of the released defendants whose cases originated in the six month post-PSA period were resolved by the date of the Arnold publication, compared with 97% of the pre-PSA cases. The report contains a footnote acknowledging that pretrial rearrest rates may rise since some cases remain open. In fact, the fraction of released defendants who were arrested pretrial rose an additional 2.5 percent since the publication of the report, demonstrating that the PSA did not lead to a reduction in pretrial crime.

[96] External pressures forced a rapid release of this report despite significant concerns about the data that were expressed by Kentucky's head of Pretrial Services. (Based on a telephone conversation with Tara Boh Blair, head of Kentucky Pretrial Services, dated May 15, 2017).

misconduct.[97][98] In Part IV of this study I provide evidence that neither the 2011 law nor the adoption of the PSA led to a lower rate of pretrial rearrest or FTA.[99]

It is not uncommon to find statistics from other jurisdictions cited as evidence that pretrial risk assessment led to a decrease in detention rates, FTAs and crime, but the research supporting these claims are as tenuous – or more so – than the studies cited above. The Arnold Foundation released a report stating that the use of their risk assessment tool in Lucas County, Ohio, led to an doubling in the number of defendants granted non-financial release and a decrease in pretrial rearrest and failures-to-appear, but this one-page press release contains little detail besides that.[100] Mecklenberg, North Carolina is also supposed to have seen a dramatic drop in their jail population after adopting the Arnold Foundation's risk assessment, but the evidence that is cited to support this consists entirely of slides taken from two Power Point presentations.[101] A report put out by Mecklenberg County Jails makes no mention of risk assessment but does indeed show that the jail population declined after 2010 (the year pretrial risk assessment was adopted).[102] However, the jail population had been steadily decreasing prior to 2010 as well.[103] This report provides scant reason to believe that risk assessment was responsible for the post-2010 decline, as the slope is not visibly different from the long term trend.[104]

Multnomah County, Oregon is cited as a successful example of risk assessments leading not only to lower detention rates among juveniles, but also

---

[97] The same error is responsible for the erroneous conclusion in REPORT ON IMPACT OF HOUSE BILL 463, *supra* note 90, at 6, that rates of non-appearance were lower after HB 463 than before.

[98] The Arnold Foundation report acknowledges in a footnote that the post-PSA rearrest and non-appearance rates may rise since some cases remained open. However, in conversations with them, they expressed the opinion that the differences in results are mostly due to differences in the how the data was pulled and processed.

[99] The Arnold Foundation also released a report stating that non-financial release is up, and crime and FTAs are down in Toledo, Ohio after implementing the PSA. There are not enough details in the report to assess these claims. *See* Laura & John Arnold Foundation Press Release *New data: Pretrial risk assessment tool works to reduce crime, increase court appearances*, August 8, 2016.

[100] Laura and John Arnold Foundation, *New Data: Pretrial Risk Assessment Tool Works to Reduce Crime, Increase Court Appearances,* (Aug. 8, 2016) http://www.arnoldfoundation.org/new-data-pretrial-risk-assessment-tool-works-reduce-crime-increase-court-appearances/.

[101] Both MOVING BEYOND MONEY, *supra* note 12, at 21, and REDUCING RACIAL AND ETHNIC DISPARITIES IN JAILS, *infra* note 105, at 28, state that risk assessment led to a drop in incarceration in Mecklenberg, but Power Point slides are the only references cited in these papers that directly support this claim.

[102] Mecklenberg County Criminal Justice Services, *Jail Population Trend Report*, (July-Sept. 2014).

[103] *Id.*

[104] The Arnold Foundation has recently funded several third-party evaluations of their risk assessment tool; the results of these studies will hopefully be available before long.

lower racial disparities in detention.[105] This study does indeed show evidence that racial disparities dropped after the implementation of risk assessment, but the authors of the study do not attribute the change to risk assessment per se. Risk assessment was but one of a multitude of changes that were adopted at the same time, all with the express goal of reducing racial disparities, and it is unclear which one made the difference.[106]

A study in Virginia evaluated whether training pretrial officers in a series of actions ("praxis") to take on the basis of the risk assessment evaluation affected release recommendations of the officers, release and supervision decisions by the judge, and pretrial misconduct from defendants.[107] The study, put out by the company that developed and implemented the risk tool and the praxis, states that the training led to increased release recommendations, increased release, and lower misconduct.[108] The research methodology compared agencies that were randomly assigned to receive training against agencies that were randomly assigned to be a control group.[109] Random assignment to treatment and control groups is often a very rigorous method of evaluating the impacts of a policy. However, there were a limited number of agencies, raising concerns that there were pre-existing differences in practices and outcomes between the agencies assigned to each group.[110] The study does not provide the details necessary to evaluate the claim that the differences in outcomes are caused solely by praxis training.[111]

New Jersey has recently implemented dramatic reform to their pretrial system. They shifted from a fairly typical system in which judges set money bail and defendants are detained only if they cannot afford to pay to a system in

---

[105] Justice Policy Institute, *Reducing Disproportionate Minority Confinement: The Multnomah County, Oregon Success Story and its Implications* (January 2002).

[106] REDUCING DISPROPORTIONATE MINORITY CONFINEMENT, *supra* note 105, at 15-16 ("It is difficult to assess what any one detention reform strategy (alternatives to incarceration, objective risk assessments, expedited case processing, sanctions grid for VOPs) or explicit DMC reduction strategy (diversity training, additional public defender resources, staff diversification, data collection and research, new coalitions with other agencies and groups, diversification of the delivery system) made the difference in Multnomah.")

[107] Mona J.E. Danner et al., Luminosity, Risk Based Pretrial Release Recommendations and Supervision Guidelines, 1-2 (August 2015), http://luminosity-solutions.com/site/wp-content/uploads/2014/02/Risk-Based-Pretrial-Guidelines-August-2015.pdf

[108] Id.

[109] *Id.* at 4

[110] *Id.* at 4.

[111] There were 29 agencies randomized; presumably these varied in size and were associated with different regions and regional practices. A typical RCT where randomization was conducted over a small number of groups would show evidence about the extent to which outcomes differed across treatment and control groups before the experimental intervention occurred. If the outcomes differed across treatment and control before the intervention, then post-intervention differences in outcomes cannot be attributed to the intervention.

which money bail is virtually never used and defendants can be directly detained on the basis of crime or flight risk as measured by the PSA.[112] This required an amendment to the constitution, the creation of a pretrial services organization, bi-partisan support, and cooperation from a broad range of agencies.[113] New Jersey has seen a dramatic decline in the rates of pretrial detention since bail reform was implemented,[114] and money bail has been used only rarely.[115] (The impact on pretrial crime and FTAs is still unknown.) Under New Jersey's revised laws, judges retained considerable discretion to order pretrial detention or set money bail, so it is interesting to note how much practices have changed. Judges are appointed by the governor in New Jersey, and Governor Chris Christie has been a big supporter of New Jersey's bail reform. This may have influenced the potency of the new laws.

Cook County, Illinois, home to Chicago, also recently adopted the PSA to aid in pretrial decision making. However, the impacts in Cook County may have been less than in New Jersey. An article by the Chicago Sun Times reports that Cook County judges deviate from the bail recommendations of the risk assessment 85% of the time.[116] For reference, judges in Cook County are either elected, or appointed by the elected judges.

Outside of the pretrial context, Virginia is sometimes cited as an example of successful use of risk assessment in sentencing. Despite claims that the introduction of risk assessment led to the diversion of "25 percent of nonviolent offenders from prison without raising the crime rate",[117] the studies

---

[112] Judge Glenn A. Grant, Acting Administrative Director of the Courts, *Remarks Before the Senate Budget and Appropriations Committee*, at 3 (Thursday, May 4, 2017), http://www.judiciary.state.nj.us/pressrel/2017/SenateBudgetCommitteeRemarks_May_4_2 017.pdf.

[113] *Id.* at 2 (discussing the broad coalition behind reform) and 4 (mentioning the newly created pretrial services unit); Matt Arco, *Christie signs bail reform measure, lauds lawmakers for bipartisanship,* NJ.COM (August 11, 2014), (discussing bi-partisan support for the constitutional amendment), http://www.nj.com/politics/index.ssf/2014/08/christie_signs_bail_reform_measure_lauds_l awmakers_for_bipartisanship.html.

[114] Joe Hernandez, *N.J.'s Jail Population Dropped 10 Percent in Two Months After It Scrapped Cash Bail*, NEWSWORKS (Mar. 14, 2017), http://www.newsworks.org/index.php/local/new-jersey/102173-njs-jail-population-dropped-10-percent-in-two-months-after-it-scrapped-cash-bail

[115] Based on communications with Roseanne Scotti, Drug Policy Alliance.

[116] Frank Main, *Cook County Judges Not Following Bail Recommendations: Study,* chicago.suntimes.com, (July 3, 2016), http://chicago.suntimes.com/chicago-politics/cook-county-judges-not-following-bail-recommendations-study-find/.

[117] As cited in Skeem & Lowenkamp, *infra* note 139, at 681, and DEMYSTIFYING RISK ASSESSMENT, *supra* note 71, at 12 ("A key case study that bears this out is the state of Virginia, where the use of a validated risk tool in multiple jurisdictions allowed for the diversion of 25 percent of nonviolent, prison-bound offenders over a three-year period without increasing crime.").

that are cited to support these claims say nothing of the sort.[118] Virginia's intention was to divert 25% of nonviolent offenders with the use of the risk assessment,[119] but neither study provides any evidence that the introduction of the risk assessment led to any change in practice. In fact, neither study analyzes data from the period before the risk assessment at all.[120]

There is a limited literature that does directly address the impacts of risk assessment in practice. A study by Richard Berk showed that parole board members in Pennsylvania did not change their release decisions very much when risk assessments were available.[121] Berk finds tentative evidence that the risk assessment tool lowered recidivism rates, but cautions against firm conclusions due to weakness in the research design.[122] A different paper by Berk and coauthors evaluates a large randomized control trial that evaluated the efficacy of assigning prisoners to different security level prisons using a machine learning tool risk assessment tool. Inmates in the treatment group were assigned to prisons of differing security levels by the machine learning risk assessment tool, and the control group inmates were assigned using the existing scoring system. There was no decrease in inmate misconduct for defendants who were assigned to facilities using the new tool,[123] but the misconduct was shifted towards higher security facilities, suggesting that the tool was effective in predicting misconduct.[124] A third study by Berk and colleagues uses a randomized control trial to evaluate whether probationers and parolees who had been labeled as low risk by a risk assessment tool could be placed under low supervision.[125] This study is better thought of as an evaluation of supervision levels than an evaluation of risk assessment, since it does not compare decision making with risk assessment against the status quo decision-making method.[126]

---

[118] The studies cite to Brian J. Ostrom et al., Report funded by NIJ, *Offender Risk Assessment in Virginia*, 15 (2003) and Mathew Kleiman et al., *Using Risk Assessment to Inform Sentencing Decisions for Nonviolent Offenders in Virginia*, 53 CRIME AND DELINQUENCY 106, 112 (2007).

[119] Ostrom et al., *supra* note 118.

[120] Kleiman et al., *supra* note 118, at 112 ("The analysis is based on 555 nonviolent offenders who were recommended by the sentencing guidelines for incarceration but were ultimately diverted."); Ostrom, *supra* note 118, at 31 (2003) (describing the sample as those who had a risk assessment score and were eligible for diversion).

[121] *See* Berk, *supra* note 123, at 12, Table 2.

[122] *Id.* at 23.

[123] Richard Berk et al., *A Randomized Experiment Testing Inmate Classification* Systems, 2 CRIMINOLOGY & PUB. POL'Y 215, 232 (2003).

[124] *Id.* at 233.

[125] Richard Berk et al., When Second Best is Good Enough: A Comparison between a True Experiment and a Regression Discontinuity Quasi-Experiment, J. EXP. CRIMINOLOGY 1, 4 (2010).

[126] *Id.*

In sum, there is a sore lack of research on the impacts of risk assessment in practice. There is next to no evidence that the adoption of risk assessment has led to dramatic improvements in either incarceration rates or crime without adversely affecting the other margin. The research on whether it should *theoretically* (due to improvements in predictive accuracy) is far from definitive. Nonetheless, it is a broadly held belief that the adoption of risk assessment tools will lead to clear improvements in the efficiency of criminal justice.[127]

C. *Racial Disparities and Risk Assessment*

Recently, there has been increased debate about whether risk assessment tools will worsen racial disparities in criminal justice. Risk assessment proponents argue that the objective rankings of a risk tool will be less biased than the subjective evaluations of potentially-racist judges.[128] Critics counter that the risk tools themselves may be racially biased.[129] Some of the confusion lies in a lack of clear language about what constitutes "racial bias" in risk assessment. Sandra Mayson provides an overview of the different, and sometimes conflicting, measures of racial equality in risk assessment[130], but for the purposes of this Article, I consider a tool to be racially biased if it systematically assigns higher risk scores to defendants from a particular race than their true risk warrants. (While "true risk" is hard to define or measure on the individual level, on the group level it refers to the average incidence of the predicted outcome.) This is how the term "bias" is used in statistics[131] and is similar to the common language usage of the term.[132]

---

[127] *See, e.g.*, Skeem & Lowenkamp, *supra* note 139, at 680 ("One way to unwind mass incarceration without compromising public safety is to use risk assessment instruments in sentencing and corrections."); DEMYSTIFYING RISK ASSESSMENT, supra note 71, at 22 ("there is a growing professional consensus that the careful and ethical implementation of risk assessment tools can facilitate improved criminal justice outcomes."); Anne Milgram et al. *Pretrial Risk Assessment: Improving Public Safety and Fairness in Pretrial Decision Making*, 27 FEDERAL SENTENCING REPORTER 216, 219 (2015) ("judges' use of [risk assessment] instruments has helped them make pretrial release decisions that have reduced pretrial crime, kept dangerous offenders off our streets, and reduced the number of low-risk defendants detained before trial").

[128] See, e.g., Robert D. Hoge, *Standardized Instruments for Assessing Risk and Need in Youthful Offenders,* 29 CRIMINAL JUSTICE AND BEHAVIOR 380, 387 (2002).

[129] *See, e.g.,* Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FEDERAL SENTENCING REPORTER 237, 1 (2015).

[130] Sandra Mayson, *Bias in, Bias out,* (unpublished manuscript, May 2017) (on file with author).

[131] "Systematic error or bias refers to deviations that are not due to chance alone. The simplest example occurs with a measuring device that is improperly calibrated so that it consistently overestimates (or underestimates) the measurements by X units." *Lesson 4:*

Using this definition, there are a number of reasons why risk assessment tools could be biased against blacks.[133] The most common argument is that inputs to risk assessment – prior convictions, prior incarceration sentences, education, employment, etc. – are themselves the result of racially disparate practices. While two defendants may pose a similar crime risk, the defendant living in a heavily-policed minority neighborhood is likely to have a lengthier criminal record and thus a higher risk score than one who lives in a less heavily policed neighborhood. Similarly, a risk algorithm that is trained to predict an outcome that is the result of racially disparate law enforcement or prosecution practices also incorporates bias into the algorithm. While these sources of potential bias almost certainly affect the risk assessment, they are hard to correct for, and few even try. Actual rates of offending are unknown and the gap between behavior and criminal record can only be guessed at.

Another place where bias can enter the risk tool is in the design of the instrument. In the simpler, "checklist" style instruments the designers choose both the inputs and the weights on the inputs. If a designer puts more weight on inputs that correlate with race than their crime-predictiveness warrants, the tool will be biased. In machine-learned risk assessment tools this type of bias is less of a concern. The weights that the algorithm places on the different inputs will generally reflect only the extent to which these inputs are predictive of what it is trained to predict.

Empirical research on racial bias in risk assessment is both thin and recent. In 2016, ProPublica released a study that claimed to have found evidence that a proprietary risk assessment tool called COMPAS, used to help make decisions about pretrial release in Broward County, Florida, was biased against black defendants.[134] To support this claim, they show that "black defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts."[135] In technical terms, this is a disparity in "false-positive rates", or the fraction of non-recidivating defendants who were ranked as high risk. Many researchers countered this argument with the point that disparate false positive rates will be

---

Bias and Random Error, Stat 509 Design and Analysis of Clinical Trials, https://onlinecourses.science.psu.edu/stat509/node/26 (last visited July 5, 2017).

[132] The online Oxford English Dictionary defines bias as "Inclination or prejudice for or against one person or group, especially in a way considered to be unfair." OXFORDDICTIONARIES.COM https://en.oxforddictionaries.com/definition/bias (last visited July 5, 2017).

[133] *See generally* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact,* 104 Cal. L. Rev. 671, 677-692 (2016) (discussing ways in which data mining can discriminate).

134 Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016) ("There's software used across the country to predict future criminals. And it's biased against blacks.").

[135] *Ib*.

present every time there are disparate rates of offending – and thus disparate average risk levels – across groups.[136] The intuition behind this is simple. A false positive rate is a ratio: the denominator is the total number of people who don't recidivate and the numerator is the number of people who don't recidivate but are classified as "high risk". As the risk level of a group increases, there will be fewer people who don't recidivate and more people who are labeled high risk. Since there are fewer non-recidivists the denominator will decrease. Since there are more people labeled "high risk" – some fraction of which do not recidivate – the numerator increases. Thus, in simple mathematical terms, as the risk level goes up, the false positive rate will go up too. In other words, differing levels of offending will lead to disparate false positive rates *even if we knew the true risk of each group and even if the tool is completely unbiased.*[137] In fact, when there are disparate base rates of offending, one would have to program a risk tool to *be* biased (so that one group systematically gets a lower or higher risk classification than their true risk level warrants) in order to eradicate disparate false positive rates.[138]

While ProPublica framed this as being about actuarial risk assessment, it's actually relevant to the entire project of using risk to make decisions in criminal justice. The key points apply equally regardless of how the risk evaluation was achieved: through proprietary black box risk assessment tools, transparent checklist instruments, and to judges' intuitive assessment of risk. If it's concerning that black defendants who don't recidivate are more likely to be labeled high risk than white defendants who don't recidivate (and there are plenty of reasons why this should be concerning!) then this calls into question the entire regime of using risk as a basis of restricting liberties, not simply actuarial risk assessment instruments.

Disparate false positive rates are not a measure of racial bias under the definition used in this Article. Most other researchers do not measure racial

---

[136] *See, e.g.,* William Dieterich et al., Northepointe Inc., *Compas Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, (July 8, 2016), https://university.pretrial.org/viewdocument/compas-risk-scales-demonstrating-a.*;* Jennifer L. Doleac & Megan Stevenson, *Are Criminal Risk Assessment Scores Racist?,* BROOKINGS INSTITUTE (August 22, 2016), https://www.brookings.edu/blog/up-front/2016/08/22/are-criminal-risk-assessment-scores-racist/.

[137] While disparate impact is not inherently unfair, it can be unfair when the costs are born by a marginalized group and the benefits accrue to the dominant group. The unfairness arises not from the disparate impact per se but from the presumption that the benefits to the dominant group were given disproportionate weight in the policy decision, while the costs to the marginalized group were discounted.

[138] Jon Kleinberg et al. *Inherent Tradeoffs in the Fair Determination of Risk Scores*, November 17, 2016 (showing that, except for in very specialized circumstances, achieving equal false positive rates would require a risk tool where the same risk classification would correspond with different levels of actual risk across the two groups), https://arxiv.org/pdf/1609.05807.pdf.

bias using disparate false positive rates either.[139] Instead, they measure bias using "predictive parity": similar recidivism rates among white and black defendants with the same risk score. If a risk score is racially biased using the definition provided above, then the likelihood of committing crime would be lower for black defendants than it would be for white defendants with the same risk score. The company that owns COMPAS responded to ProPublica's article by showing that there is no evidence that the tool is biased using predictive parity tests.[140] Similar results have been found for the Post Conviction Risk Assessment (PCRA) in a detailed study of racial disparities and risk assessment.[141] The PSA likewise shows predictive parity for white and black defendants,[142] as does the Virginia Pretrial Risk Assessment Instrument (VPRAI).[143] Most studies that test for racial bias in the risk assessment using predictive parity find no evidence of bias.

However, there are reasons to question the predictive parity approach. First, it's impossible to test for predictive parity in rates of reoffending; one can only test for predictive parity in something visible, like arrest or conviction. Given the differences in how different neighborhoods are policed – as well as the many other opportunities for racial bias or racial disparity to affect the likelihood of arrest or conviction – a group of black defendants who are rearrested at the same rate as a group of white defendants may have committed fewer crimes. The race gap between the rate of offense and rate of rearrest is thought to be lower for violent crimes than for less serious crimes.[144] The Skeem and Lowenkamp study on racial bias in the PCRA focuses primarily on predictive parity for violent crime rearrest for this express reason.[145]

A second concern with using predictive parity as a measure is that the *rate* of reoffending does not directly measure the *risk* of reoffending at the time the risk was evaluated. Actual reoffending is a joint combination of a person's propensity to commit crime and the opportunities and incentives that she

---

[139] *See* Jennifer Skeem & Christopher Lowenkamp, *Risk, Race, & Recidivism*, 52 CRIMINOLOGY 680, 685 (2016) ("There is substantial agreement on the empirical criteria that indicate when a test is biased… the paramount indicator of test bias is *predictive bias"*).

[140] Dieterich et al., *supra* note 136, at 2-3.

[141] Skeem & Lowenkamp *supra* note 139, at 690.

[142] RESULTS FROM FIRST SIX MONTHS, *supra* note 93, at 4.

[143] *See* Mona J.E. Donner et al., Luminosity, *Race and Gender Neutral Pretrial Risk Assessment*, at 11, http://luminosity-solutions.com/site/wp-content/uploads/2014/02/Race-and-Gender-Neutral-Pretrial-Risk-Assessment-November-2016.pdf.

[144] *See, e.g.,* Skeem & Lowenkamp*, supra* note 139, at 2-3.

[145] Skeem & Lowenkamp, *supra* note 139, at 690 (describing why they choose risk of violent crime as the main focus of analysis).

faces.[146] Risk evaluations, both formal and informal, influence these opportunities and incentives. In particular, they influence the likelihood that a defendant will be incarcerated, supervised, provided treatment, and so forth. If judges treat black defendants differently than white defendants this would bias measures of predictive parity. For example, if judges are more likely to assign white defendants to take an effective drug treatment program, their rate of reoffending will be lower than black defendants who were less likely to be assigned to the effective program.[147]

Determining whether or not a risk tool is racially biased is probably redundant. As Princeton computer scientist Aylin Caliskan says: "Machines are trained on human data. And humans are biased."[148] The important question is whether the use of actuarial risk assessment tools results in more disparate outcomes than the status quo, or other viable alternatives. Outside of the research presented in this study, the empirical research on this is next to non-existent.[149]

III. THE IMPACTS OF PRETRIAL RISK ASSESSMENT IN KENTUCKY

This section provides some of the first rigorous empirical evidence on the impacts of risk assessment in practice. It provides important information about risk assessment's effects in a state that has been held up as a leader in pretrial reform, as well as insight about a risk assessment tool that has been widely adopted in other jurisdictions. It also serves as an empirical case study through which to explore, as is done in Part IV, the myriad ways that the impacts of risk assessment in practice may be different, and more complicated, than previously thought. Such differences underline the importance of constantly evaluating new methods: a habit that is too rarely present in criminal justice despite the lip service paid to evidence-based practices.

A. *Overview of Pretrial Risk Assessment in Kentucky*

Kentucky has often been held up as a leader in pretrial practices. It is noted as an early adopter of pretrial risk assessment tools and is often cited as

---

[146] Shawn Bushway & Jeffrey Smith, *Sentencing Using Statistical Treatment Rules: What We Don't Know Can Hurt Us*, 23 JOURNAL OF QUANTITATIVE CRIMINOLOGY 377, 378 (2007).
[147] Doleac & Stevenson, *supra* note 136.
[148] Brian Resnick, *How Artificial Intelligence Learns to be Racist*, VOX.COM (April 17, 2017), https://www.vox.com/science-and-health/2017/4/17/15322378/how-artificial-intelligence-learns-how-to-be-racist.
[149] THE MULTNOMAH COUNTY SUCCESS STORY, *supra* note 106, is the only study that the author is aware of that could even tentatively be taken as evidence on how risk assessment in practice affects racial disparities relative to the status quo.

an example that other jurisdictions are following, or should follow.[150] It is also known as one of only four states that have outlawed the commercial bail industry.[151] In that regard Kentucky is closer to the rest of the world; outside of the United States there are very few other countries whose bail system is dominated by the commercial provision of bail.[152] Furthermore, Kentucky's pretrial services agency has earned both national and local respect for their adoption of evidence-based practices,[153] their low FTA and pretrial crime rates,[154] and their rigorous data collection.[155]

From inception in 1976, Kentucky's pretrial services agency used a formal risk assessment tool to aid in the pretrial release decision.[156] The risk assessment originally used was a simple six question instrument developed by the Vera Institute in 1961.[157][158] This instrument, which focused on strength of family and community ties to identify defendants who were a good candidate for release, was in use for thirty years. In 2006, Kentucky adopted a new risk assessment instrument, modernized to reflect new knowledge about the best predictors of risk, and similar in many ways to other pretrial risk tools currently in use.[159] It was a checklist-style instrument that put heavy weight on criminal history and prior FTAs. It also included several non-criminal-justice inputs, such as whether the defendant has stable employment, housing, and a reference who would be willing to attend court or co-sign the bond. It was validated (i.e. shown to be predictive of pretrial rearrest and FTA) in a 2010 study conducted

---

[150] *See supra* note 12 (demonstrating that Kentucky's use of pretrial risk assessments is an example for other jurisdictions).

[151] Justice Policy Institute, For Better or for Profit, How the Bail Bond Industry Stands in the Way of Effective Pretrial Justice, at 40 (Sept. 2012).

[152] F.E. Devine, Commercial Bail Bonding 12, (1991).

[153] The head of Kentucky Pretrial Services received Kentucky's 2012 Public Advocate Award for her work promoting pretrial justice. Jamie Neal, *State Pretrial Services leader receives 2012 Public Advocate Award*, KY.GOV (Oct. 2, 2012) http://courts.ky.gov/pages/newsroom.aspx?viewMode=PressRelease&pressReleaseGUID=%7BD15D83C0-10F7-4E92-9CB1-A775C14F2DCC%7D.

[154] *See* Table 1.

[155] *See* Part IV.c for a discussion of why Kentucky Pretrial Services is an example of evidence-based criminal justice at its best.

[156] Kentucky Pretrial Services, *Pretrial Reform in Kentucky,* Administrative Office of the Courts, at 10 (Jan. 2013), https://www.pretrial.org/download/infostop/Pretrial%20Reform%20in%20Kentucky%20Implementation%20Guide%202013.pdf.

[157] Id.

[158] Charles E. Ares et al., *The Manhattan Bail Project: An Interim Report on the Use of Pre-Trial Parole*, 38 N.Y.U. L.REV. (1963)

[159] For a summary of common pretrial risk assessment tools and their inputs see Sandra G. Mayson, *Dangerous Defendants* (forthcoming in the YALE LAW JOURNAL) 16 (2017)

by the JFA Institute.[160] Among defendants who were released pretrial, a low risk classification corresponded to a 10.4% failure (rearrest or FTA) rate, a moderate risk classification corresponded to a 20.9% failure rate, and a high risk classification corresponded to a 27.8% failure rate.[161]

Between 2000 and 2010, Kentucky's incarcerated population – both jail and prison – grew by 45%, more than three times the U.S. average.[162] Correctional costs were placing considerable pressure on the state budget. In 2010, Kentucky convened a bi-partisan, inter-agency task force to research methods of reducing the jail and prison population without harming public safety.[163] The task force released a report in January of 2011 with a number of specific recommendations for reform. Many of the task force's recommendations were adopted in a bill called House Bill 463 (HB 463).[164] This bill was introduced in February of 2011, passed almost unanimously that March, and was made law as of June 8, 2011.[165]

As far as pretrial issues are concerned, the "most significant advancement [of HB 463] is the mandatory use of a 'research-based, validated risk assessment tool' to measure a defendant's risk of flight or of posing a risk to the public."[166] Before HB 463, use of the pretrial risk assessment tool was optional. Judges who were not interested in the tool were not required to look at it, and many did not use it at all.[167] HB 463 made consideration of the risk assessment a mandatory part of determining bond.[168] It instructed judges to grant non-monetary release (release on recognizance or unsecured bond) to low and moderate risk defendants.[169] Moderate and high risk defendants were to be

---

[160] James Austin et al., *Kentucky Pretrial Risk Assessment Instrument Validation*, The JFA Institute, at 1 (Oct. 29, 2010).

[161] The 2010 validation study recommended some minor modifications to the risk assessment; these were enacted in March of 2011. Ninety five percent of defendants have the same risk classification in both instruments, thus the impacts of this modification are expected to be minimal and are not explored in this Article.

[162] Id.

[163] Kentucky Department of Corrections, *House Bill 463 Implementation Report*, at 3 (Dec. 1, 2012); Legislative Research Commission, *Report of the Task Force on the Penal Code and Controlled Substances Act*, at 1-2 (Jan. 19, 2011).

[164] Sen. Tom Jensen (R - London) and Rep. John Tilley (D - Hopkinsville), *HB 463 – Statement From Sponsors*, in CRIMINAL LAW REFORM: THE FIRST YEAR OF HB 463 (June 6, 2012) at 1.

[165] *See* IMPLEMENTATION REPORT, *supra* note 163, at 3.

[166] Damon Preston, Deputy Public Advocate of Kentucky, *Criminal Law Reform: The First Year of HB 463*, Kentucky Bar Association 2012 Annual Convention, 5 (June 6, 2012)

[167] As reported by Tara Boh Blair, Chief Operations Office, Kentucky Pretrial Services, in a telephone conversation on May 15, 2017.

[168] KY. REV. STAT. ANN. § 431.066(2) ("In making [the pretrial release and bail determination], the court shall consider the pretrial risk assessment").

[169] Ky. Rev. Stat. Ann. § 431.066(3,4).

considered for GPS monitoring or supervision.[170] Defendants were granted a $100-per-day credit towards the bail amount for each day they spend in jail,[171] the bail amount was capped at the maximum fine for crimes that were punishable by fine only,[172] and non-monetary release was recommended for defendants charged with crimes where the presumptive punishment is probation.[173] However, nowhere in HB 463 was judicial discretion limited. In a Kentucky Supreme Court Order that clarified how judges should respond to HB 463, this was made abundantly clear. "Nothing in these guidelines should be construed to limit the court's discretion as to whether or not to grant pretrial release to a defendant."[174]

In addition to changes to the pretrial system, HB 463 made a number of changes to other parts of the criminal proceeding. Risk assessment tools were made mandatory in the pre-sentencing report, for determining supervision levels among defendants on probation or parole, and in determining parole suitability.[175] Other changes include mandatory re-entry supervision, a speeding up of the parole process, and reduced penalties for drug possession and minor drug trafficking.[176] (Due to the variety of changes that occurred around the same time, this Article places minimal focus on long-term trends in conviction, sentencing or recidivism. Instead, the focus of the empirical analysis will be on sharp, discrete changes in pretrial outcomes for defendants who were booked right before and right after HB 463, and right before and after the adoption of the PSA.)

In July of 2013, Kentucky adopted a new risk assessment tool: the Public Safety Assessment Court (PSA). This tool was developed by the Arnold Foundation using a large dataset on pretrial releases in more than 300 jurisdictions.[177] The PSA evaluates risk along three dimensions: risk of FTA, risk of new arrest, or risk of new arrest for a violent crime.[178][179] The inputs for

---

[170] HB 463 stated that moderate risk defendants should be released on supervision, and that the Kentucky Supreme Court should create guidelines for high risk defendants. The Supreme Court's guidelines stated that the court should consider a risk reduction plan and supervision strategy as developed by the pretrial officer. *Infra* note 174 at 6.

[171] Ky. Rev. Stat. Ann. § 431.066(5)(a).

[172] Ky. Rev. Stat. Ann. § 431.525(2).

[173] Ky. Rev. Stat. Ann. § 218A.135(a).

[174] Supreme Court of Kentucky, 2011-12, *Order Approving Judicial Guidelines For Pretrial Release and Monitored Release*, at § 11, available at http://courts.ky.gov/courts/supreme/Rules_Procedures/201112.pdf.

[175] *See* IMPLEMENTATION REPORT, *supra* note 163, at 4.

[176] Public Safety and Offender Accountability Act (HB 463): Justice Reinvestment Summary, http://www.ncsl.org/documents/nalfo/JusticeReinvestmentMikeMullins.pdf.

[177] Laura & John Arnold Foundation, *Developing a National Model for Pretrial Risk Assessment*, at 3 (November 2013).

[178] *Id.* at 4

the PSA are similar to those used in Kentucky's previous risk tool,[180] although the weights are different and the non-criminal-justice items were eliminated.

Kentucky was the first jurisdiction in which the PSA was piloted.[181] Throughout 2013 and 2014 the Arnold Foundation continued to do research on their tool and made several modifications.[182] In July of 2014 Kentucky switched to a modified version of the PSA: one that is currently in use in jurisdictions around the country.[183] Age at arrest was added as an input to the new criminal activity score, and the weighting was adjusted slightly.[184]

The risk assessment is conducted by the pretrial services officer right after the defendant is arrested and booked into jail. Using information gathered from the interview as well as the defendant's criminal records, the pretrial officer will calculate the defendant's risk score and classification and present it to the judge during the bail hearing.[185] In Kentucky, the bail hearing is supposed to occur within 24 hours of booking. In many Kentucky counties, this occurs via a phone call between the pretrial officer and the judge. The pretrial officer informs the judge of the details of the alleged offense as well as the risk level of the defendant. The judge decides a bail amount, supervision status, and any other conditions of release. If the defendant does not post bail within 24 hours, the pretrial officer notifies the court that the defendant has not made bail. At this point, the judge can choose to change the bond. If the judge does not alter the bond, or if the defendant still does not post, the defendant usually must wait for the first appearance to have the bond reconsidered.

B. *Description of the Data*

The data used in this study was provided by Kentucky's Administrative Office of the Courts and covers all defendants who were arrested and booked into jail in Kentucky between July 1, 2009 and July 1, 2016. The data was

---

[179] Lauryn Golding provides compelling arguments for why it is important to predict flight risk and danger separately. Lauryn Golding, *Disentangling Flight Risk from Dangerousness,* 2016 B.Y.U L. Rev. 837, (2016).

[180] The inputs include: pending charge, prior misdemeanor conviction, prior felony conviction, prior FTAs, prior violent conviction, prior incarceration, violent current offense, violent current offense for someone under 21. (Laura & John Arnold Foundation, *Developing a National Model for Pretrial Risk Assessment*, 4 (November 2013))

[181] *Supra* note 177, at 3.

[182] Based on documentation provided by Kentucky Pretrial Services (on file with author).

[183] Id.

[184] Id.

[185] Virtual Tour of Kentucky Pretrial Services http://courts.ky.gov/courtprograms/pretrialservices/Pages/virtualtour.aspx (last visited Aug 10, 2017) (providing detailed information about pretrial procedure in Kentucky).

extracted in May 2017 from records maintained by Kentucky Pretrial Services and Kentucky courts. The original data set contains more than 1.5 million criminal cases. The analysis presented in this Article includes only cases that originate with an arrest for a new criminal offense. Cases where the original arrest was for a probation or parole violation, a failure-to-appear, or a violation of conditions of pretrial release are omitted, leaving 1,030,732 criminal cases.[186]

Table 1 presents a selection of statistics describing the sample used for analysis. The first column refers to misdemeanor cases (65% of the sample) and the middle column refers to felony cases. For reference, the rightmost column provides statistics from a national sample of felony defendants in large urban counties.[187] This represents the most expansive data set publicly available to describe court processes nationally. The Kentucky sample is mostly male, with an average age of 33. The sample is disproportionately black as compared to Kentucky's population: only 8% of Kentucky's population is black, versus 17% of defendants.[188] Compared to the national sample, however, Kentucky is disproportionately white; 77% of felony defendants in Kentucky are white compared to 30% nationally. Although the fraction of felony defendants facing violent charges is much lower in Kentucky, the fraction who remain detained until the case is disposed is similar to the national average. The fraction of felony defendants who are released within a day is slightly lower in Kentucky and the fraction that is granted non-financial release is considerably lower: less than half of the national average. Misdemeanants have a slightly higher release rate than felony defendants, but still almost a quarter are detained until disposition and a third spend more than one day in jail.

Bail amounts for defendants who are required to pay bond are lower in Kentucky as well, possibly because there are no bail bondsmen to loan money for bond. The fraction of released defendants with an FTA or pretrial rearrest are both lower in Kentucky than the national average.

Table 1 - Comparing Kentucky to a national sample

Characteristic	Kentucky misdemeanor	Kentucky felony	National sample felony
Male	70%	73%	83%

---

[186] While bail decisions for violation/FTA cases are also interesting, there are good reasons why they may be different than bail decisions for an original arrest; including both groups would complicate the interpretation of results. All of the main results, however, are still found when analyzing the full sample of cases.

[187] Bureau of Justice Statistics, Felony Defendants in Large Urban Counties,2009 – Statistical Tables, (December 2013).
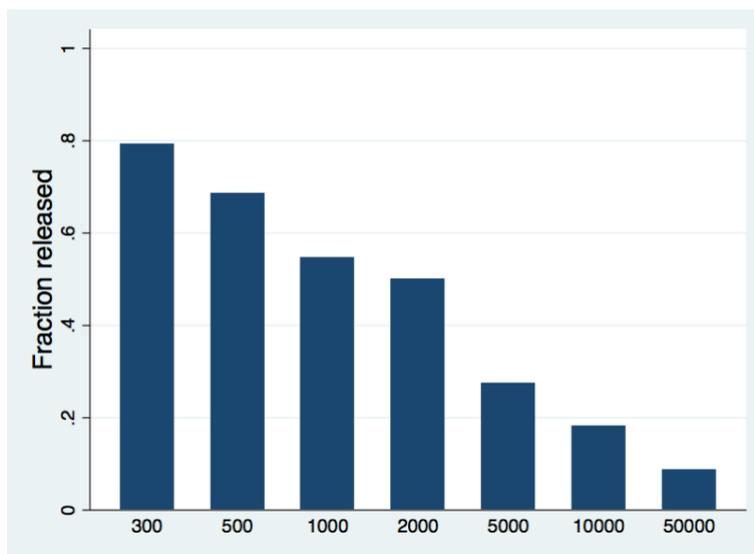
[188] Population of Kentucky, 2010 Census, http://censusviewer.com/state/KY (last visited July 7, 2017).

Age	34	33	32
Black	16%	19%	45%
White	80%	77%	30%
Hispanic	3.6%	1.7%	24%
Has violent felony charge	NA	10%	25%
Prior felony conviction	33%	46%	43%
Release prior to disposition	77%	62%	62%
Release within a day	66%	29%	31%
Non-financial release	37%	16%	40%
Median bail for those not given non-financial release	$1000	$5000	$10000
Median bail for detainees	$1000	$10000	$25000
Fraction of releasees with at least one FTA	14%	10%	17%
Fraction of releasees who were rearrested pretrial	10%	13%	16%
Fraction of defendants with at least one FTA	10%	6%	NA
Fraction of defendants who were rearrested pretrial	8%	8%	NA

As in other jurisdictions, a large fraction of defendants who are required to pay monetary bond to secure their release fail to post within three days of the bail hearing. Figure 1 shows the fraction of defendants with a given amount of bail who are released within three days. Around 20% of defendants with bail set at $500, and half of those with bail set at $2000, remain in jail for more than three days beyond their booking date.[189]

---

[189] The provision in HB 463 that granted $100 per day in bail credit was routinely ignored. According to the author's own calculations, about a third of all judges never allowed this for any defendants.

Figure 1 - Fraction released at various levels of cash bail

Note: Each bar indicates the fraction of defendants who are released within three days among those who had monetary bail set at the amount shown.

C. *HB 463 Led to an Increased Use of Risk Assessment (And an Overview of Empirical Methods)*

This section has two goals: to demonstrate that the 2011 law making the use of risk assessment mandatory actually resulted in an increased use of risk assessment, and to explain the empirical methodology used throughout the remainder of Part III. The two goals are combined because it can be useful to discuss methods with aid of an example, as opposed to discussing them abstractly.

The empirical methods used in this paper consist mostly of graphical time-trend analysis: a visual representation of trends and changes to pretrial outcomes.[190] The focus of the analysis is on sharp changes that occur right around HB 463 and the adoption of the PSA. A sharp, discrete change to pretrial practices or outcomes whose timing coincides exactly with the implementation of a new law or a new risk tool can likely be attributed to

---

[190] There are more than one million observations in this data set – with such a large sample most changes that can be seen visually would also be statistically significant. While the formal statistical tests are not presented in this text, they have been conducted.

that law or tool. The causes of longer term trends are harder to identify, and thus are not a primary focus of this article.
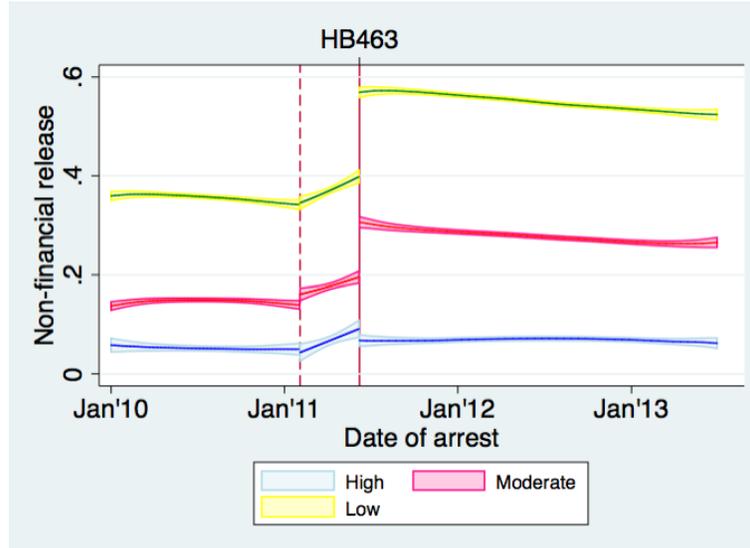
Demonstrating that judges increased their use of the risk assessment instrument when HB 463 made it mandatory requires showing that bail practices changed in accordance with the recommendations associated with each risk classification. In other words, it requires showing that judges became more lenient with defendants who were classified as lower risk, and more strict with defendants who were classified as high risk.

Figure 2 shows a time trend in the fraction of defendants in each of the three risk groups who are granted non-financial release at the first bail hearing.[191] The horizontal axis indicates the booking date and the vertical axis is the fraction of defendants who were granted a non-financial release at the first bail hearing. The dotted vertical line indicates the date that HB 463 was introduced as legislation and the solid vertical line indicates the date it was implemented. The colored lines are estimates of the time trend in non-financial release for defendants in each of the three risk classification groups. The time trends are estimated using local linear smoothing with a bandwidth of 120 days.[192] The local linear smoothing is employed because on any given day the actual number of defendants who are granted non-financial release can be higher or lower than expected due to idiosyncratic factors. This idiosyncratic fluctuation, often referred to as "noise", is visually distracting, and so time trend graphs will almost always use some method of smoothing to make the trend easier to see.

---

[191] The unit of analysis in this, and in the remainder of this Article, is a case. For conciseness, however, the time trends are described as referring to defendants, not cases. Using more precise language, Figure 2 shows a time trend in the fraction of cases in which defendants received non-financial release.

[192] Details about this technique can be found in many places. *See, e.g.,* the Stata manual for the command "lpoly" at 4-8, http://www.stata.com/manuals13/rlpoly.pdf.

Figure 2 - Non-financial release

Note: The top, middle, and bottom line indicate the fraction of low, moderate and high risk defendants who are granted non-financial release. The dashed vertical line is the date that HB 463 was introduced as legislation; the solid line indicates the date it was implemented.

Each point on the yellow line represents the fraction of low risk defendants who are expected to get a non-financial release on a particular date, and so forth for the other risk groups. The shading around each line represents the 95% confidence interval – a measure of uncertainty – for the time trend. There are "cuts" in the smoothing of the time trend at times when one might expect sharp changes to the trend: in Figure 2 there is a cut at the date when legislation was introduced and another at the date it was implemented. The cuts function by limiting the data that is used to build the trend line to only one side of the cut point. In other words, the trend line for defendants who were booked right before a cut date will be estimated using only data from defendants booked before the cut date, and the same for defendants booked right after a cut date.

(In some circumstances, the choice of smoothing technique and cut location can give misleading visual impressions. The Appendix provides an alternative method of graphing time trends – binned scatter plots – which demonstrates each of the key empirical claims made in this paper without the use of smoothing or cuts.)
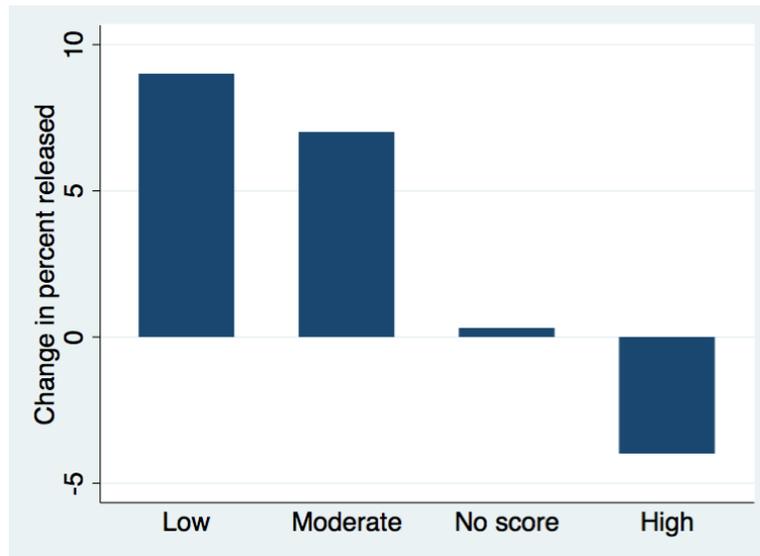
Figure 2 shows a dramatic increase in the fraction of low risk defendants who were granted non-financial release around the time of HB 463. Before the bill was introduced only about 35% of low risk defendants were granted non-financial release, but after its implementation that number had risen to 57%: a 22 percentage point increase, or a 63 percent increase relative to the earlier mean. The pink line shows a 16 percentage point increase in non-financial releases for moderate risk defendants and the blue

line shows that the fraction of high risk defendants receiving non-financial release remained essentially the same. This figure shows that while judges are deviating from the presumption of non-financial release for all low and moderate risk defendants more often than not, HB 463 did result in a marked change in practices, which corresponded closely with the classifications of the risk assessment.

Overall, HB 463 led to a sizeable decrease in bail for defendants who were ranked as low risk, a more moderate decrease in bail for defendants ranked as moderate risk, and an increase in bail for defendants ranked high risk, as shown in Table 2. Release rates changed accordingly. Figure 3 shows the changes in the fraction of defendants who are released within three days of the bail hearing before and after HB 463.[193] HB 463 led to a 9 percentage point (p.p.) increase in releases for low risk defendants, a 7 p.p. increase in releases for moderate risk defendants, and a 4 p.p. decrease in releases for high risk defendants. Interestingly, there was no change in the release rate for defendants who did not receive a risk score due to difficulties in verifying key inputs. This further supports the claim that the change in bail setting practices after HB 463 is due to the information provided by the risk assessment.

---

[193] Defendants who were detained until the case was disposed, but for whom disposition occurred within three days of the bail hearing, are counted as released within three days.

Figure 3 - The Impact that HB 463 had on the Release Rate of Defendants with Various Risk Classifications

Note: This figure shows the change in release rates between the two months before HB 463 was introduced and the two months after it was implemented. The change in release rates is shown for defendants who were rated low, moderate, or high risk, as well as for defendants who did not receive a risk score. A positive change means that defendants were more likely to be released after HB 463 than they were before.

Table 2 - Impacts of HB 463 for low, moderate and high risk defendants

Outcome measure	Group	Before HB 463	After HB 463	Difference
Non-financial bond	Low risk	35%	57%	+22 p.p.
	Moderate risk	15%	31%	+16 p.p.
	High risk	5%	7%	+2 p.p.
Low cash bail ($1000 or less)	Low risk	23%	14%	-9 p.p.
	Moderate risk	29%	22%	-7 p.p.
	High risk	32%	25%	-7 p.p.
Moderate-high cash bail (greater than $2500)	Low risk	24%	18%	-6 p.p.
	Moderate risk	35%	31%	-4 p.p.
	High risk	45%	48%	+3 p.p.
Release within 3 days of booking	Low risk	73%	81%	+9 p.p.
	Moderate risk	50%	57%	+7 p.p.
	High risk	34%	30%	-4 p.p.

D. *Risk Assessment's Impact on Bond Setting and Release*

While the previous section focused on differing impacts for defendants with different risk classifications, this section shows the overall effect. In particular, this section analyzes the impact that HB 463 and the adoption of the PSA had on bond setting and release across all defendants.

Figure 4 shows a time trend in the fraction of all defendants granted non-financial release at the first bail hearing. From left to right, the vertical lines indicate the date when HB 463 was introduced, the date that HB 463 was implemented, the date that the PSA was adopted, and date when the PSA was revised to the version that is now broadly used around the country.

Figure 4 - Non-financial release rates over time for all defendants

Note: This figure shows the fraction of defendants who are granted non-financial release over time. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 4 shows a sharp jump up in the fraction of defendants who are granted non-financial release coinciding exactly with HB 463. The increase begins as soon as the bill was introduced (it passed almost unanimously)[194] and accelerates at the time of implementation. In total, there is a 13 percentage point jump in non-financial releases from January to June of 2011. Almost immediately, however, the rate of non-financial releases begins to fall. It declined steadily until July 2013, when the PSA is adopted. There is a smaller spike upwards after the adoption of the PSA, then the non-financial release rate declines again after that, with virtually no change as the PSA is revised. By January of 2016, more than half of the increase in non-financial releases that came as a result of HB 463 had disappeared.

Figure 5 shows the fraction of defendants who are given a low cash bond (requiring a monetary payment of $1000 or less) at the first bail hearing. Interestingly, we see almost the exact inverse of the pattern we saw in Figure 4. HB 463 results in a sharp *drop* in the fraction of defendants receiving low cash bail, an *increase* over time as practices move back towards their previous state, a small *jump down* in low cash bail around the adoption of the PSA, and an *increase* after that. This suggests that judges responded to the risk assessment changes analyzed in this Article by substituting non-financial release for low-cash bail. As time went on, however, they returned to their previous bail setting practices.

---

[194] *See* REPORT ON IMPACT OF HOUSE BILL 463, *supra* note 90, at 4.

Figure 5 - Low cash bail over time

Note: This figure shows the fraction of defendants who are given bail of $1000 or less. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 6 shows a time trend in the fraction of defendants who are released within three days of booking.[195] For visual simplicity, and because there are very little changes that occur around that time, there is no cut in the time trend estimation at the time the PSA is revised. As can be seen in Figure 6, neither HB 463 nor the PSA has a big effect on the release rate. HB 463 led to only a 4 percentage point increase in the fraction of defendants who are released within three days of booking, and the adoption of the PSA led to a barely perceptible 1 percentage point increase in releases. It appears that most of the defendants who were granted a non-financial release as a result of these changes would have gotten out on a low cash bond regardless. Moreover, the small increase in releases was short-lived: by 2015, the release rate was lower than it had been before HB 463.

---

[195] About 5% of defendants have a holder, which decreases the release rate somewhat.

Figure 6 - Release within 3 days

Note: This figure shows the fraction of defendants who are released within 3 days of booking. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

E. *Risk Assessment's Impacts on Pretrial Misconduct*

The small increase in releases as a result of HB 463 was accompanied by an increase in the likelihood that defendants would fail to appear in court. Figure 7 shows a sharp jump up in the failure-to-appear rate (defined as the fraction of all defendants who fail to appear for at least one court date) from before the legislation was introduced to after the new law was implemented. The size of the increase – about 3 percentage points – was not large in and of itself, but it is large relative to the base level: about a 40 percent increase over the mean. The introduction of the PSA did not lead to a decline in failures-to-appear. If anything, the FTA rate is slightly higher after the PSA was adopted than before. This does not necessarily reflect on the PSA, however, as there is no sharp change in FTAs that coincides with the date that the PSA was adopted. The drift upward in FTAs during that time period could have been caused by some other factor.
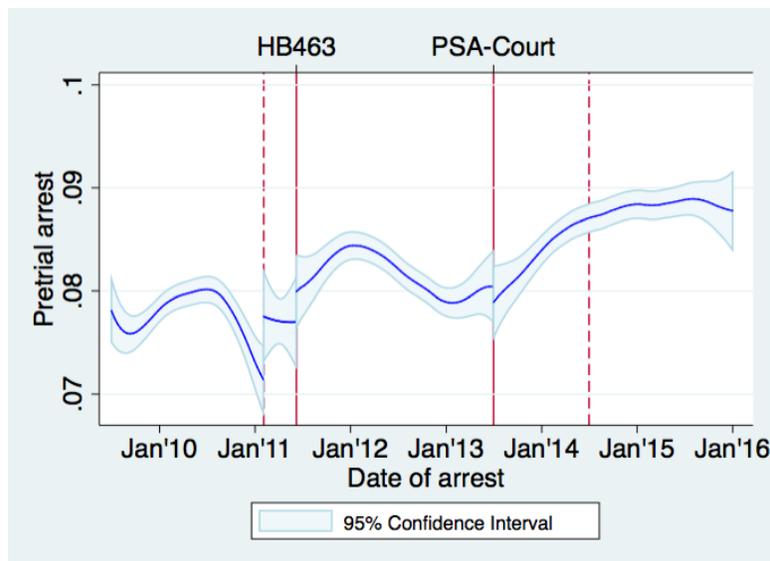
Figure 7 - FTAs over time

Note: This figure shows the fraction of defendants who fail to appear in court at least once. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 8 shows a time trend in the fraction of all defendants who were arrested for a new offense during the pretrial period.[196] The graph shows an increase in rearrests around the time of HB 463. The increase is less of a stark and indisputable break in trend than was seen for FTAs in the previous figure. Inferring that HB 463 led to an increase in rearrests requires inferring that the drop in rearrests right before the introduction of the legislation was indicative of a meaningful change in trend that would have continued in the absence of the law. Alternatively, one could argue that the drop down in rearrests towards the end of 2010 was just an idiosyncratic fluctuation in the rearrest rate, and the rise after the legislation was introduced was simply more idiosyncratic fluctuation. Alternative analysis, shown in the appendix, suggests that the former interpretation is more likely. Regardless, it is clear that the increased use of risk assessments as a result of the 2011 law did not result in a decline in the pretrial rearrest rate.

---

[196] The pretrial rearrest rate captures only arrests that are for new crimes, not arrests for violation of court orders or FTAs.

There is no sharp change in the pretrial rearrest rate around either the adoption or modification of the PSA. The pretrial rearrest rate is slightly higher after the adoption of the PSA, but this appears to be part of a general upward drift in the pretrial rearrest rate and thus not likely to be due to the change in risk assessment tools.[197]
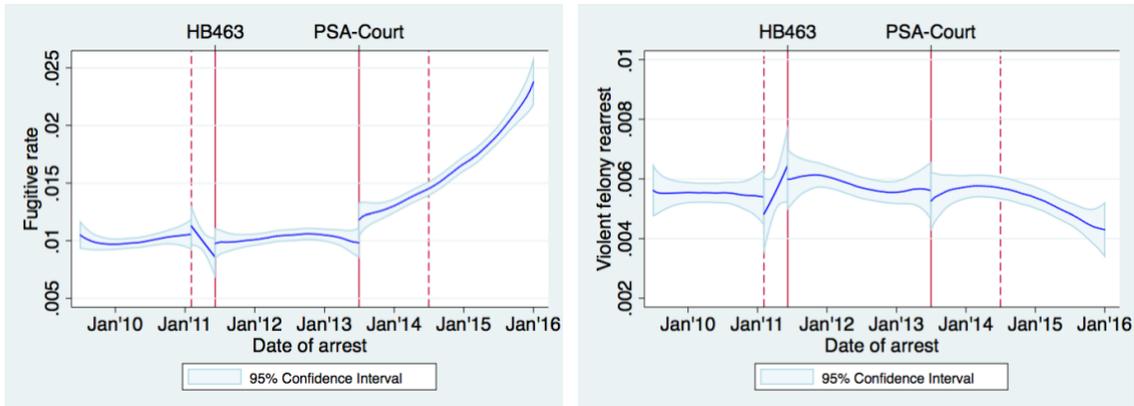
Figure 8 - Pretrial rearrests

Note: This figure shows the fraction of defendants who are arrested for a new offense during the pretrial period. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 9 shows two other important pretrial outcomes: the fraction of defendants who remain a fugitive at the time the data was compiled (May 2017) and the fraction of defendants who are rearrested for a violent felony pretrial. There are no visually discernible changes in the fugitive rate or the violent felony rearrest rate occurring as a result of either HB 463 or the

---

[197] The Arnold Foundation report, *supra* note 92, that claimed that the PSA led to lower rates of pretrial rearrest used a slightly different sample (all cases, not just cases that began with an arrest for a new offense) and a different measure (the fraction of released defendants with a pretrial arrest, not the fraction of all defendants with a pretrial rearrest). This is not the cause of the disparity between their results and those shown in this Article. Using their methods, this author was able to replicate their findings and show that the post-PSA pretrial rearrest rate rose from 8.5% at the time that their report was published to 11% once all cases had resolved. *See supra* note 95 & accompanying text (providing more discussion about the differences in results).

adoption of the PSA. Furthermore, these rates are very low. Less than one percent of all defendants are rearrested for a violent felony (murder, non-negligible manslaughter, forcible rape, robbery or aggravated assault) during the pretrial period. The fugitive rate is also quite low, ranging from 1 to 2.5% of all defendants.[198] This indicates that failures-to-appear and flight are far from synonymous: the large majority of defendants with an FTA eventually either reappear or are apprehended.

Figure 9 - Fugitive rate and rate of violent felony rearrest

Note: The figure on the left shows the fraction of defendants who remain a fugitive as of May 2017 and the figure on the right shows the fraction of defendants who are arrested for a violent felony during the pretrial period. From left to right, the vertical lines in each chart indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

The changes that were shown graphically in this section are summarized in Table 3. The left-most numerical column shows outcomes for all defendants booked during the two months before HB 463 legislation was introduced: December 2010 and January 2011. The next column shows outcomes for all defendants booked in the two months after HB 463 was implemented: July and August of 2011. The final two columns show the two month averages before and after the adoption of the PSA: May and June of 2013 and July and August of 2013.

Table 3 - Impacts of HB 463 and PSA for all defendants

Outcome	Before HB 463	After HB 463	Before PSA 1	After PSA 1

---

[198] This rate is higher for defendants who were booked relatively recently, possibly because there has been less time elapsed during which to locate the missing defendant.
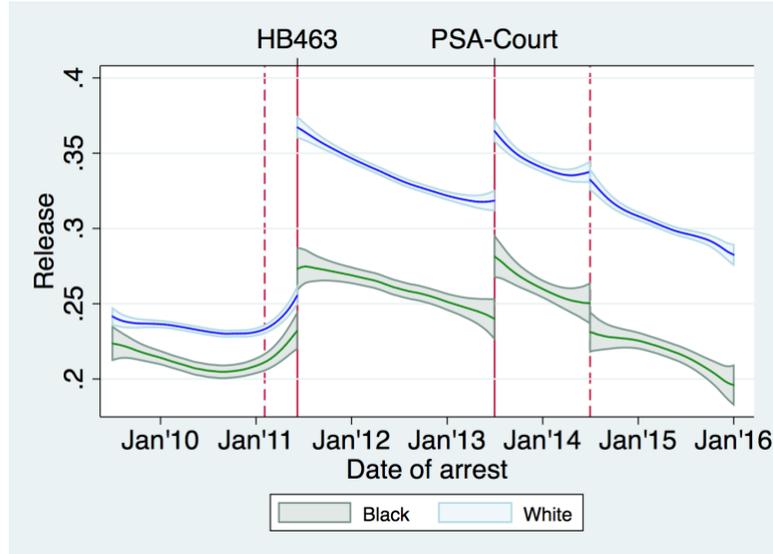
Non-financial release within 3 days	21%	34%	31%	35%
Low cash bail ($1000 or less)	32%	25%	26%	23%
Release within 3 days	63%	67%	64%	65%
FTA	7.6%	9.6%	9%	9.4%
Pretrial rearrest	7.3%	8%	8%	8%
Fugitive rate	0.99%	0.95%	0.98%	1.18%
Violent rearrest pretrial	0.52%	0.59%	0.56%	0.52%

The Appendix provides several figures to demonstrate that the results presented in Parts III.d and III.e are robust to alternative specifications. In particular, the Appendix shows that the key results are not caused by changes in the types of defendants who are arrested, are robust to alternative methods of measuring pretrial rearrest and FTA, do not depend on specific choices regarding smoothing and cut points, and are prevalent among a group of defendants who are least likely to be affected by other non-risk-assessment related aspects of HB 463.

F. *Racial and Regional Disparities in Bond and Release*

This subsection evaluates whether risk assessment affected racial disparities in the likelihood that a defendant is granted non-financial release or is otherwise released within three days of booking. Figure 10 shows time trends in the fraction of white defendants who are granted non-financial release (shown in blue) and the fraction of black defendants who are granted non-financial release (shown in green). Relative to black defendants, white defendants are more likely to be granted non-financial release throughout the entire time period of the sample. There was an increase in non-financial release for both groups as a result of HB 463, however the increase was larger for white defendants than it was for blacks. The racial gap jumped from about 2 percentage points to 10 percentage points after HB 463 was implemented and remained relatively constant through January of 2016.
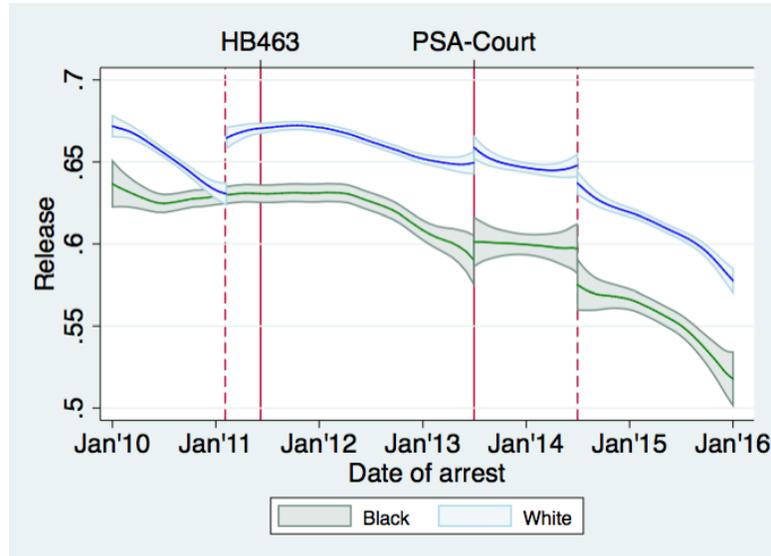
Figure 10 - Racial disparities in non-financial release

Note: The blue line shows the fraction of white defendants who are granted non-financial release and the green line shows the fraction of black defendants who are granted non-financial release. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 11 shows time trends for both races in the likelihood of being released within three days of booking. We see a similar but more attenuated pattern; the race gap increased after HB 463 and then remained relatively constant at about 5 percentage points. In fact, despite the increase in the likelihood of being granted non-financial release, HB 463 did not lead to a visually discernible increase in the likelihood of being released within 3 days for black defendants.
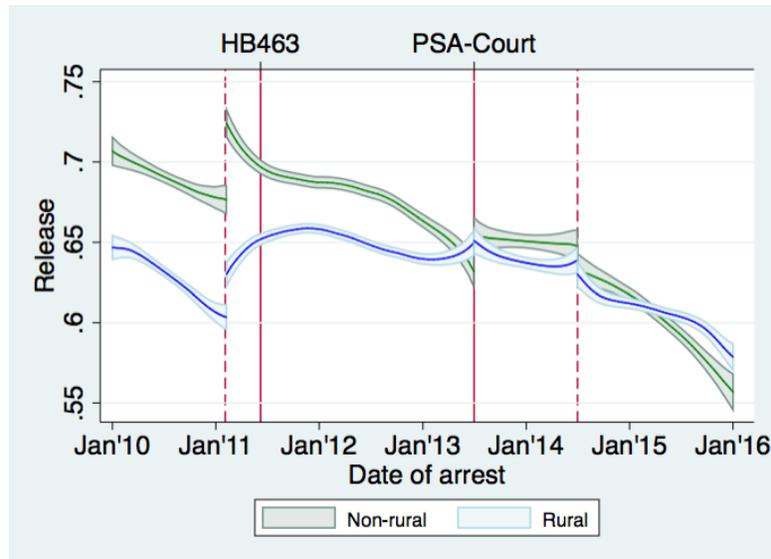
Figure 11 - Racial disparities in release within 3 days of booking

Note: The blue line shows the fraction of white defendants who are released within three days and the green line shows the fraction of black defendants who are released within three days. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 12 shows that there are regional disparities in the likelihood of being detained pretrial. The white line shows a time trend in the fraction of rural defendants who are released within three days of booking and the green line shows the same for non-rural defendants. Before HB 463, rural defendants were about 8 percentage points more likely to be detained pretrial than those living in cities or suburban areas. However, this gap shrunk and then reversed itself over time. The gap shrunk partly because rural regions responded more to HB 463 than non-rural regions. It also shrunk because the release rate dropped precipitously for non-rural regions over the six years of analysis: from a high of about 70% in January 2010 to a low of 55% in January 2016.

Figure 12 - Rural/non-rural disparities in release

Note: The blue line shows the fraction of rural defendants who are released within three days and the green line shows the fraction of non-rural defendants who are released within three days. From left to right, the vertical lines indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.
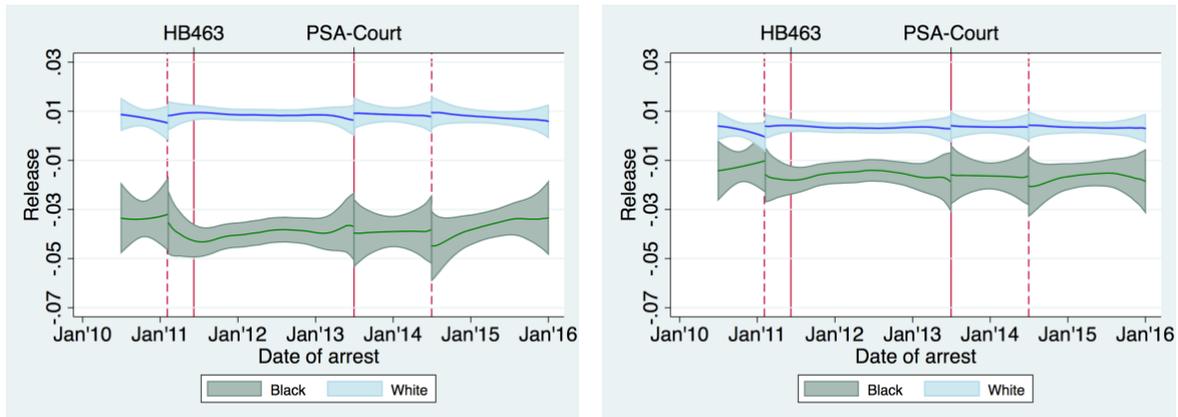
The differing trends in rural and non-rural regions complicate the analysis of racial disparities, since rural regions have a high percentage of white defendants (85%) while non-rural regions are more mixed (around 68% white and 30% black). Thus the fact that white defendants appear to have been advantaged by HB 463 more than black defendants could simply be because they live in regions where the judges changed their bond setting habits more as a result of the law.

The graph to the left of Figure 13 shows racial disparities over time once county effects and regional time trends have been accounted for. This was accomplished by estimating the average release rate for all races in each county by month by year.[199] Figure 13 plots the difference between the actual and predicted release rate for white and black defendants. As can be

---

[199] Formally, this is constructed by regressing a dummy for being released pretrial on county fixed effects as well as circuit-by-month-by-year fixed effects and then collecting the residuals from that regression. The left graph in Figure 13 is a local linear time trend of those residuals for black and white defendants.

seen, once county effects and varying time trends at the circuit level have been accounted for, the racial gap in the likelihood of being released is pretty constant over time at about 5 percentage points. While this research design is not well suited for detecting small changes, there is no visible evidence to suggest that risk assessment affected racial disparities once differing regional trends were accounted for.

Figure 13 - Racial disparities in pretrial release after accounting for county effects (left) and county as well as charge (right)

Note: In each figure, the blue line shows a time trend in releases for white defendants and the green line shows a time trend in releases for black defendants. The left figure shows the difference in release rates once county effects and time trends have been accounted for. The right figure shows the difference in release rates once county effects, time trends, offense, age, gender and recent criminal history have been accounted for. From left to right, the vertical lines in each chart indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

The graph on the right in Figure 13 shows that about half of the racial gap in release rates disappears once gender, age, detailed information about the charge, and recent criminal history is accounted for.[200] This graph shows the difference between the actual release rates and the predicted release rates, using predictions which take into account not only county by month by year effects, but also age, gender, the top charge, the total number of charges, the level of the charges, and whether or not the defendant has a

---

[200] The graph on the right of Figure 13 shows residuals from a regression of a release dummy on county fixed effects, circuit-by-month-by-year fixed effects, the exact charge for the 42 most common top charges, the total number of charges, whether the defendant had at least one class A, B, C or D felony, whether the defendant had at least one class A or B misdemeanor, the age at arrest, gender and whether the defendant had a prior arrest, a prior FTA or a pending charge within the year before booking.

pending case, prior case, or FTA within the year before the booking date.[201] Even after accounting for these variables, black defendants are still about 2-3 percentage points more likely to be detained than white defendants. There are a number of potential explanations for this gap. For one, the data does not include the full criminal history. It's possible that black defendants have more prior arrests/FTAs and thus had higher bail. Racial bias could also lead the judge to set higher bail, although this is less likely in Kentucky since judges are often unaware of the race of the defendant when setting bail. Third, due to correlations between race and income, black defendants may be less able to afford a given amount of bail than white defendants are.

In sum, Part III provided evidence that judges did use the risk assessment more when it was made mandatory in HB 463. HB 463 resulted in a 22 percentage point increase in the likelihood of non-financial release for low risk defendants and a 16 percentage point increase in the likelihood of non-financial release for moderate risk defendants. However some of those who were released on non-financial bond as a result of HB 463 would have otherwise been released on low cash bond. Thus the net effects on the release rate were attenuated. HB 463 led to a 9 percentage point increase in total releases (both non-financial and on money bond) for low risk defendants, a 7 percentage point increase in releases for moderate risk defendants and a 4 percentage point decrease in releases for high risk defendants. In total, this resulted in a 4 percentage point increase in the release rate for all defendants, which eroded over time as judges returned to their previous bail setting habits. FTAs increased by 3 percentage points after HB 463 was implemented, and pretrial rearrest increased by about 1 percentage point. The adoption of the PSA had negligible effects on the overall release rate, FTA rate, or pretrial rearrest rate. Neither HB 463 nor the PSA had any effect on racial disparities once regional differences were accounted for.

IV. LEARNING FROM KENTUCKY'S EXPERIENCE WITH RISK ASSESSMENT

This section discusses various implications of the empirical results presented in Part III. It begins by exploring potential reasons why the large gains that many had assumed would accompany the adoption of the risk assessment tool were not realized in Kentucky. It discusses ways that Kentucky's experience with pretrial risk assessment should and should not

---

[201] The criminal history is limited to a year before the booking date since the data begins in July of 2009. Thus, estimating more than a year of criminal history data would not be possible for defendants who are booked towards the beginning of the data set.

affect expectations about the impacts of risk assessment in other jurisdictions. Finally, it calls for new direction in the evidence-based criminal justice movement: a deeper integration of evaluation into the process of adopting new methods.

A. *Why No Efficiency Gains?*

After HB 463, judges incorporated the risk assessment into their bail practices significantly more than they had previously. Defendants ranked as low risk were more 9 percentage points *more* likely to be released and defendants ranked high risk were 4 percentage points *less* likely to be released. If the risk classifications of the risk assessment instrument were much more accurate than the judge's intuitive assessment of risk one might expect a gain in efficiency: simultaneously decreasing detention rates, FTAs and pretrial crime – or at least decreasing one without increasing the others. This did not occur. Why not?

First, risk assessment tools may not have provided as large a gain in predictive power as was expected. As discussed in Part II.a of this Article, the research arguing that actuarial tools out-perform human intuition in predicting crime is far from definitive. While there are good reasons to believe that risk assessment tools provide new and useful information, the margin of gain is unclear.

Another possibility is that judicial discretion was used not to correct the risk assessment when it erred, but to override the risk assessment when it was correct. Human decision-making has been shown to be subject to a variety of foibles: false heuristics, over-weighting of small probabilities, over-confidence, risk aversion, etc.[202] While these types of human error are part of the reason to expect that actuarial prediction tools can predict better than human intuition, they may also be reasons why actuarial prediction tools are not that useful in practice. The policy-relevant question is not "Is the actuarial tool better at predicting misconduct than the judge" but rather "Does the judge make better decisions when given access to actuarial predictions?". A recent survey indicates that only a small minority of judges think that a risk assessment tool does a better job at predicting future crime than themselves.[203] Given this skepticism, it is unclear under what circumstances judges make different decisions as a result of the tool than they would have otherwise. If the prediction tool fails to influence

---

[202] *See, e.g.,* Amos Tversky & Daniel Kahneman, *Judgement Under Uncertainty: Heuristics and Bias,* 185 SCIENCE 1124, (1974) (generally discussing a number of different cognitive heuristics and biases that affect the ability to assess probabilities).
[203] *See supra* note 87 & accompanying text.

decisions in circumstances where the predictive gains are the greatest, the usefulness of the tool will be curtailed.

It's also possible that the use of actuarial risk tools *did* lead to a substantial increase in the predictive capacity of judges, but that this information did not translate into improved outcomes. Identifying the appropriate interventions for different risk levels is an important challenge to the successful use of risk assessment. Part II.b of this Article discussed a paper by Richard Berk and coauthors, in which an experiment in using actuarial risk assessment to assign prisoners to different security classifications did not lead to lower rates of offending while in prison. It did, however, appear to be effective at sorting prisoners based on offending level: while the total offending rates were the same, the rates were higher in high security prisons and lower in low security prisons. If the use of risk assessment did not lead to lower total offenses, it may simply have been because placement in high security prisons was not effective at preventing offending.

One of the most dramatic changes in bail setting practice as a result of HB 463 is an increase in non-financial release as opposed to release on low monetary bond. While this likely resulted in a decrease in the number of defendants detained pretrial due to an inability to pay bail, it may have reduced the incentives for released defendants to show up in court. Alternative methods of increasing appearance rates, such as court notifications, were rare: less than 5% of released defendants were assigned by the court to receive phone call reminders of their next appearance.[204] (Kentucky has since dramatically expanded their use of court reminders.) If the recommendations associated with being classified as low risk included robust support to help defendants overcome barriers to appearance (difficulties with transportation, getting time off work, arranging child care, etc.), the use of the tool may have been more effective.

B. *Lessons for Other Jurisdictions*

Jurisdictions around the country differ widely in their criminal procedure, culture and demographics. The experience other jurisdictions have with risk assessment will not, in general, be an exact mirror of Kentucky's. Nonetheless, certain lessons can be drawn from Kentucky's experience that should influence what to expect from pretrial risk assessment in other areas.

First, Kentucky's experience with risk assessment should temper hopes that the adoption of risk assessment will lead to a dramatic decrease

---

[204] Kentucky has since begun sending automatic text message court reminders to all defendants.

in incarceration with no concomitant costs in terms of crime or failures to appear. That's not to say that risk assessment brought no benefit; just because Kentucky was not able to simultaneously improve along all three margins (detention, crime and FTAs) that doesn't mean that the tool wasn't useful. It simply means that realizing large gains in practice are not as easy as realizing them in a hypothetical policy simulation. While it's certainly possible that other jurisdictions will experience a larger efficiency gain than Kentucky, there is no strong a priori reason to expect this to be the case. The risk tools used in Kentucky are similar or identical to other pretrial risk assessment tools currently in use.[205] They were shown to be predictive of future offending and non-appearance.[206] The recommendations associated with the tool – non-financial release for low risk defendants, release onto supervision for moderate risk defendants, and supervision or detention for high risk defendants – are fairly typical of the recommendations used in other jurisdictions. Kentucky's practice of allowing judges the discretion to deviate from these recommendations if they find a crime or flight risk is also typical of pretrial policy.

Kentucky does differ, however, in that it was an early adopter of risk instruments. This meant that the margins of change analyzed in HB 463 were not the difference between having and not having a risk instrument, but the difference between having a risk instrument that was not heavily used, and being required to consider it as part of the release decision. Furthermore, the fact that Kentucky was an early adopter means that the change being analyzed happened before risk assessment tools had gained the popularity that they currently have. This cultural shift may affect judges' openness to these tools. For both of these reasons, the margin of change before and after HB 463 is lower than it might be in other jurisdictions.

While the Kentucky experience should *temper* hopes that pretrial risk assessment will result in a dramatic decline in detention rates with no increase in FTAs or pretrial crime, it does not mean those hopes should be abandoned. As discussed in the previous section, the usefulness of risk assessment in practice depends on a number of factors that are, as of yet, poorly understood. Future studies may show that risk assessment has been more successful in other contexts, and may provide insight on how to replicate and expand that success. One potential avenue is the use of machine-learned prediction tools, as opposed to the checklist style tools currently in use. Machine learned tools provide more accurate – some

---

[205] *See supra* note & accompanying text.
[206] FIRST SIX MONTHS, *supra* note 93, at 3-4; JFA, *supra* note 160 & accompanying text.

argue much more accurate – predictions than simpler tools.[207] These tools come at a cost however – they use a "black box" technique, meaning that it is next to impossible to understand why a person got the score that they did.[208]

As for racial disparities, it is unclear whether the Kentucky experience with risk assessment will be replicated in other jurisdictions. Kentucky is a largely rural, predominantly white state. Racial dynamics in Kentucky are not expected to be representative of racial dynamics in dense urban areas, in the heavily Latino southwest, or in the black rural south. That doesn't, however, mean that Kentucky's experience provides no useful knowledge. In some regards, Kentucky provides a particularly stringent test of racial bias in risk assessment. Bail hearings in Kentucky usually happen over the phone between the judge and the pretrial officer.[209] Thus, the judge is less likely to be aware of the race of the defendant, which should minimize the incidence of explicit racial bias. Demonstrating that risk assessment does not increase racial disparities relative to the status quo when the status quo is not likely to be heavily biased is a stronger finding than showing that it does not increase racial disparities relative to potentially racist judges. Thus Kentucky's experience with risk assessment should somewhat assuage concerns about expanded racial disparities, but further research is needed.

Jurisdictions adopt risk assessment for a variety of reasons. In addition to hopes of increased efficiency, jurisdictions may look to risk assessment as a way to centralize and standardize pretrial decision-making. This is likely to be particularly appealing to bail reform advocates who seek to lower pretrial detention rates. In fact, lowering the jail population was one of the goals of HB 463.[210] Kentucky demonstrates some of the challenges with this technocratic approach to bail reform. While judges certainly changed behaviors as a result of HB 463, they deviated from the recommendations of the risk assessment more often than not. If the hope is to use risk assessment to coax pretrial practices in a certain direction, careful thought should be given to how to achieve this goal. Likely this involves either establishing clear guidelines for when deviation from

---

[207] Richard Berk & Jordan Hyatt, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions,* 27 FEDERAL SENTENCING REPORTER 222, 227 (2015) ("Machine learning offers superior forecasting accuracy and more, at least compared to traditional methods of forecasting and unstructured clinical judgment.").

[208] *But see* Joshua Kroll et al., *Accountable Algorithms*, 165 UNIVERSITY OF PENNSYLVANIA L. REV. 633, (2017) (discussing reasons why transparency may not be a good method of ensuring accountability).

[209] Virtual Tour of Kentucky Pretrial Services, http://courts.ky.gov/courtprograms/pretrialservices/Pages/virtualtour.aspx.

[210] *See supra* note 18 & note 24.

recommendations is or is not allowed, making deviation costly for the judge in some way (e.g. requiring a detailed written explanation of the reasons for deviation), or nurturing a culture change among judges. These strategies may differ in jurisdictions where judges are elected, like Kentucky,[211] and in jurisdictions where judges are appointed, like New Jersey.

The limits of enacting criminal justice reform via statute alone are not limited to risk assessment. In Kentucky, the tenuous connection between statute and practice permeates the pretrial process. For example, Kentucky has a statute stating that defendants can earn a $100 credit towards the payment of bail for each day detained pretrial.[212] Yet a vaguely worded loophole (except if "found by the court to present a flight risk or to be a danger to others") can result in blanket override of the statute if a judge so chooses.[213] In fact, one third of the judges never allow jail time credit for any defendant. Even clearly written law from the Kentucky constitution is routinely ignored. The constitution states that defendants have a right to bail except in capital cases.[214] Yet of the 24,000 defendants who were denied bond during the time period of the analysis, 90% of them were charged with only a misdemeanor or level D felony. Anecdotally, these were mostly defendants who demonstrated a persistent pattern of failing to appear in court,[215] but a reasonable explanation does not negate the violation of constitutional rights.

C. *Towards a New Direction in Evidence-Based Criminal Justice*

Data, science, and technology have been rapidly changing all aspects of modern life, from how we work, to how we learn, to how we spend time with our friends and family. Tech-industry enthusiasts describe this process as "creative disruption": a dramatic change in how people accomplish certain tasks with the advent of a new, more effective method. Many people would agree that the criminal justice system is itself in need of some creative disruption. Billions of dollars are spent each year on policing, prosecuting, incarcerating and monitoring our communities, yet few are satisfied with the results. Crime rates remain high in many neighborhoods, racial disparities abound, and the system is commonly viewed as opaque, ad-hoc, unfair, outdated and ineffective.

---

[211] Judicial elections occurred in Kentucky in November 2010 and November 2014.
[212] Ky. Rev. Stat. Ann. § 431.066(5)(a).
[213] Ky. Rev. Stat. Ann. § 431.066(5)(b)(2).
[214] KY Const. § 16 ("All prisoners shall be bailable by sufficient securities, unless for capital offenses when the proof is evident or the presumption great").
[215] As per a telephone conversation with Tara Boh Blair dated May 15, 2017.

The ideas and practices associated with evidence-based criminal justice have likely advanced in no small part from a hope that data, science and technology will bring improvements to a system that many believe is in need of reform. However, enthusiasm for the potential of new technologies may have led us to put the cart before the horse: widescale adoption of risk assessment before knowing anything about what its impacts were in practice. Risk assessment wears the clothes of an evidence-based practice – they are developed with the use of large data sets and sophistical techniques, and endorsed by social scientists running policy simulations – but risk assessment tools should not be considered evidence-based until they have shown to be effective.

This Article advocates a new direction in evidence-based criminal justice: one in which an iterative process of evaluation and adaption is central. This does not constitute wholesale change from the current ideals of evidence-based criminal justice, but it shifts the focus towards integrating evaluation into the everyday operations of criminal justice. When a new technique is adopted, outcomes should be monitored to see if the desired effects were achieved. If they were not, adjustments can be made accordingly. In this paradigm, a method would be neither championed nor pilloried until its impacts in practice are clearly understood. This paradigm is characterized by informed curiosity: a willingness to try new techniques, but also a willingness to learn and adjust if the new techniques did not work as hoped.

In many ways, Kentucky Pretrial Services embodies that ideal. Over the years they have shown a continued willingness not only to try new methods, but also to evaluate how those methods have affected key outcomes, and change practices if need be. This capacity did not materialize out of thin air. For one, it requires a data infrastructure that took many years of hard work to develop and implement. Once developed, however, their data systems allowed them to monitor changes and trends in bail, release, and pretrial misconduct. For example, they have been aware that the release rate has been dropping precipitously, particularly in urban areas. In cooperation with Kentucky Pretrial Services, Kentucky's highest court has recently declared a major revision in the way pretrial risk assessment is used in their state. As of January 1, 2017, all defendants who are rated low and moderate risk and who are charged with low level crimes (non-violent and non-sexual misdemeanors as well as certain Class D felonies) are granted immediate non-financial release. No bail hearing is required, thus no judicial discretion is involved in the decision. If the goal is to liberalize release for low level defendants, Kentucky's new method of using risk assessments may prove more effective than how they were used previously. Hopefully future studies will chart the impacts of this change, and help advance our knowledge about the different ways risk assessments can be used in practice.

CONCLUSION

This Article began with a quote stating that we are beyond the point that risk assessment can be thought of as a trend, and into a "risk assessment era".[216] That one of the foremost examples of evidence-based criminal justice has advanced as far as it has with so little evidence on its impacts is a little unnerving. While evidence-based criminal justice is commonly cited as an ideal, we are still far from embodying it in practice.

This Article evaluated the impacts of pretrial risk assessment in a state that has been widely heralded as a leader in pretrial reform. It showed that pretrial risk assessment in Kentucky led to neither the dramatic efficiency gains predicted by risk assessment's champions, nor the increase in racial disparities predicted by its critics. While discussion and research about the *expected* outcomes of a change in policy will always be important, real world implementation can differ from what theory predicts in a number of ways.

Empirical research evaluating risk assessment will expand, and we will learn more about the impacts of risk assessment in different contexts. Kentucky's experience should temper expectations but not eliminate hopes; risk assessment tools may prove to be a highly beneficial input to criminal justice, but understanding how and under what conditions is likely to take time and careful research.

---

[216] Starr, *supra* note 6.

This Appendix provides several figures to demonstrate that the results presented in Part III are robust to alternative specifications. The four graphs in Figure 14 show, clockwise from top left, a variant of the time trend in non-financial release, release within three days, pretrial rearrest, and FTAs. Instead of showing the actual fraction of defendants for whom each outcome was present, the figures show residuals from a regression of the outcome on detailed variables describing the offense, basic demographics, and recent criminal history. These residuals are the difference between the actual outcome and the predicted outcome (where the predictions are based on the descriptors listed above). This process helps remove the effect of any change in defendants over time. For example, the fact that the release rate is declining over time could have been explained by the defendants arrested towards the end of the sample have committed more serious crimes than those who were arrested towards the beginning of the sample. If the charges that defendants are facing grow more serious, it would not be surprising that the release rate fell.

The trends shown in Figure 14 look quite similar to the time trends shown in Part 0 and Part III.e of this paper. The trends are centered at zero, since the vertical axis is measuring the difference between the predicted rates and the actual rates. However the patterns are qualitatively quite similar, as are the magnitudes of change. Thus the evidence presented in Parts 0 and III.e are likely explained by differences in pretrial practices as opposed to a change in the type of people who are arrested.

Like Figure 14, Figure 15 shows time trends in non-financial release, release within three days, FTA and pretrial rearrest. However, these trends are not built using linear smoothing like the figures in the main body of the text do. Figure 15 consists of scatter plots, where each dot represents the average outcome for all defendants booked within a two month span. As such, the figures are visually somewhat noisier. Nonetheless, the patterns remain the same. This eases concerns that any specific choices about the method of linear smoothing or the cuts in the time trend that were used in the graphs shown in the main body of the text created misleading visual impressions.
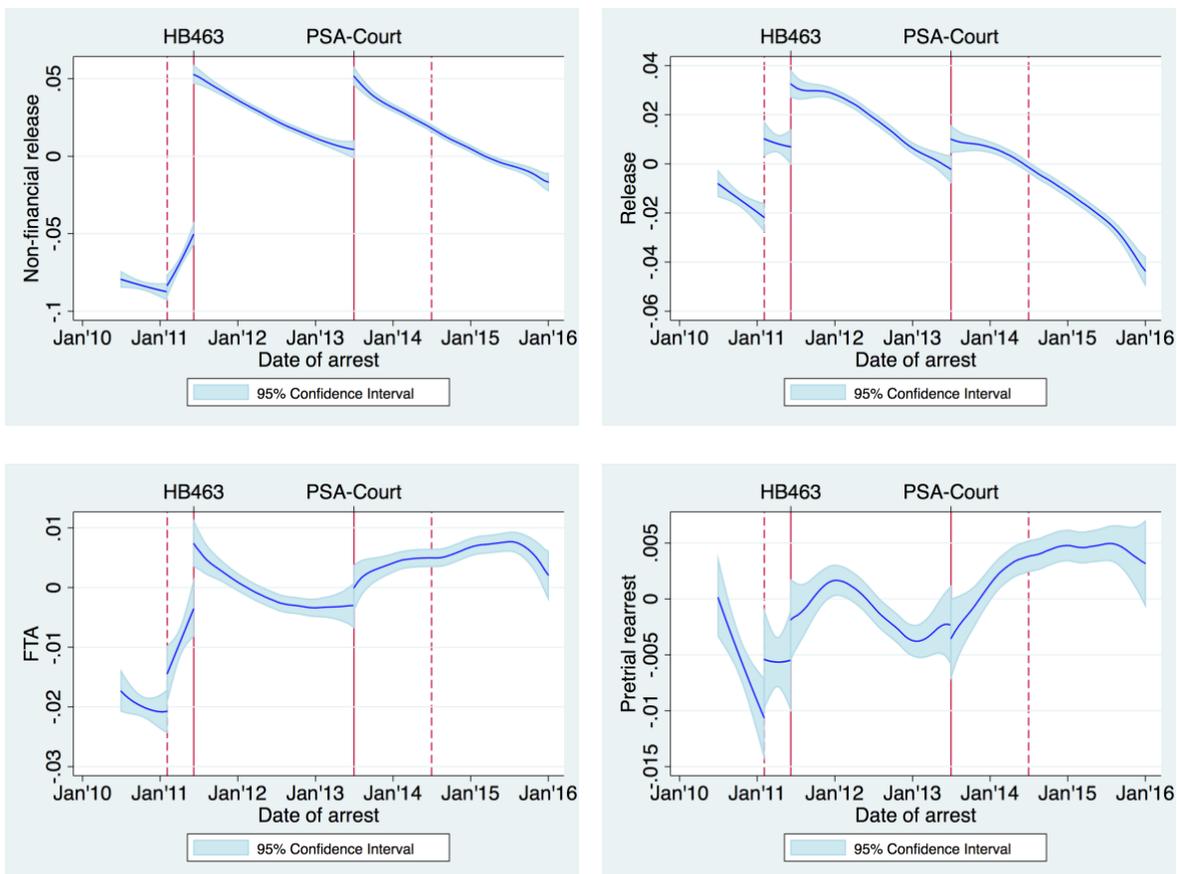
Figure 16 shows time trends in the same four outcome measures that were shown in the previous figures, but the sample is limited to felony defendants who are not facing any drug charges. Since the other pretrial-related changes that were enacted as part of HB 463 are expected to mostly affect drug offenders and misdemeanants, this specification helps ensure that the patterns we are seeing are truly a result of risk assessment. Once again, the results are qualitatively very similar: the same sharp changes are seen around the time of HB 463 and very little change around the adoption of the PSA.

Finally, Figure 17 provides alternative specifications for evaluating time trends in FTAs and pretrial rearrest. The pretrial misconduct measures used in Part III were inputted by the pretrial officers who were charged with monitoring defendants and recording misconduct. If the pretrial officers were better about inputting data after HB 463 this could result in a false impression that pretrial misconduct had increased, when in fact it was simply better recording practices. Thus I construct alternative measures of FTA and pretrial rearrest that do not depend on the reporting habits of the pretrial officers. The measures used in the top part of Figure 17 were constructed from the data: a defendant was considered to have an FTA (or a pretrial rearrest) if the data shows that the same person was arrested for non-appearance (or a new offense) after the original booking date. These measures will not be exactly the same as those inputted by the pretrial officers; for example, the pretrial officers see all FTAs while the data shows only FTAs that resulted in an arrest. Nonetheless, this provides a robustness check – an alternative method that shows similar results, particularly as they pertain to the crucial time periods before and after HB 463 and before and after the adoption of the PSA. Just like in figures shown in the main body of the text, Figure 17 show an increase in FTAs and pretrial rearrest that occurs immediately after HB 463, and no change after the adoption of the PSA.

While Figure 7 and Figure 8 in the main text show the fraction of *all* defendants with an FTA or pretrial rearrest, the bottom two graphs in Figure 17 show the fraction of *released* defendants who had an FTA or pretrial rearrest.[1] This allows us to evaluate the extent to which the increase in misconduct occurred solely because there were more people released. The bottom left graph in in Figure 17 shows that even looking solely at released defendants, the fraction with a FTA increases after HB 463. Thus the changed conditions of release (non-financial release vs. release on low cash bond) or a change in the type of people who were released is likely responsible for the increase in FTAs once risk assessment became mandatory. However, the bottom right graph in Figure 17 shows that the fraction of released defendants who have a pretrial rearrest, however, does not exhibit much of an increase after HB 463. Thus the increase shown in Figure 8 is likely a result of an increase in the number of people released.
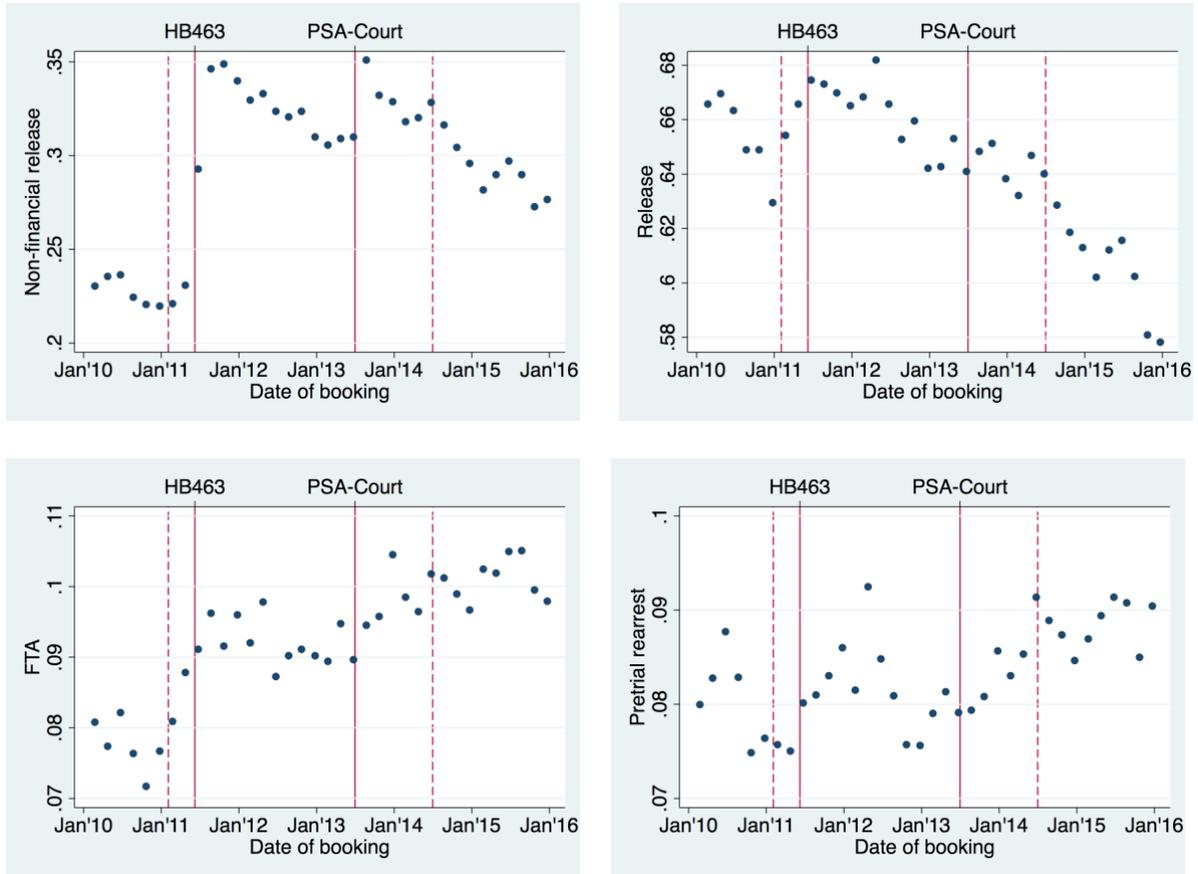
---

[1] The measures used here are the ones that were inputted by the pretrial officers, however the graphs look very similar if using the ones constructed by the data.

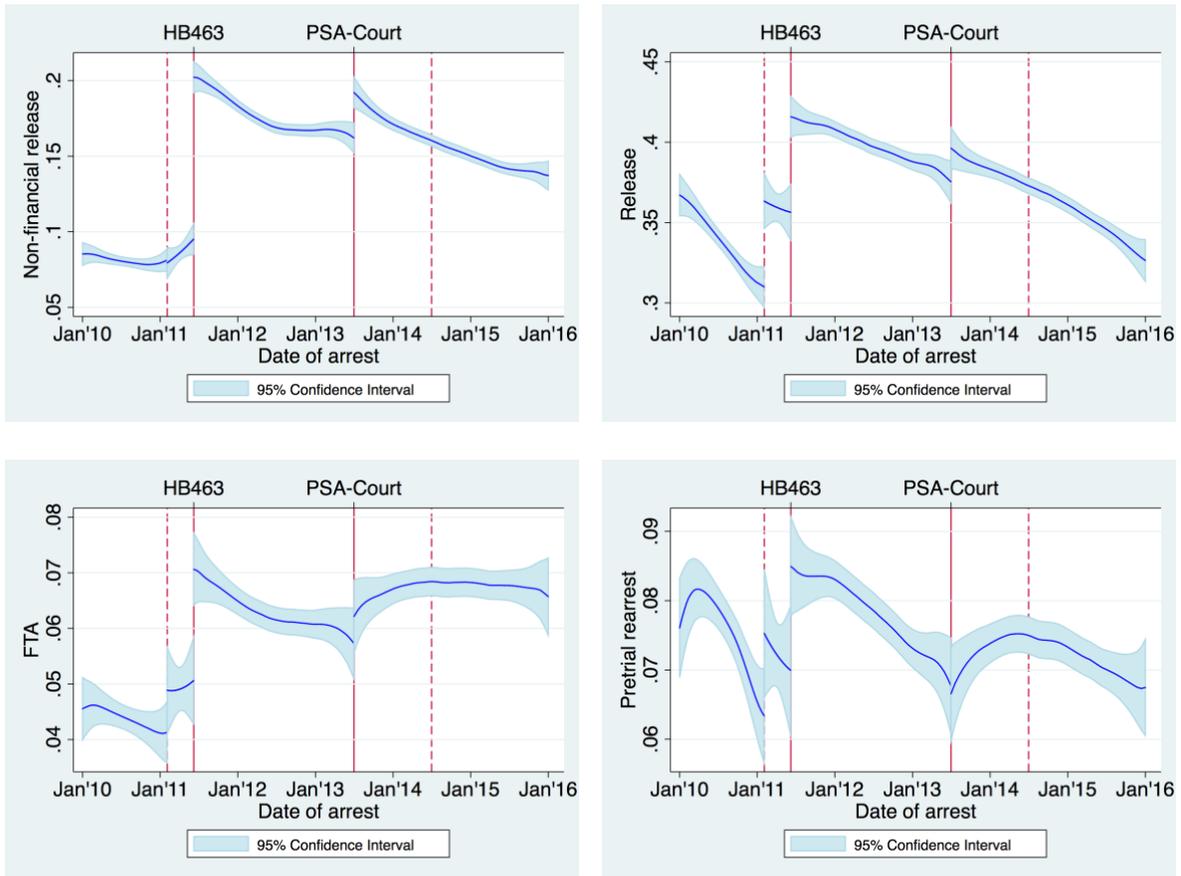# Figure 14 - Adjusting for offense, demographics and criminal history

Note: Clockwise from top left, the figures show time trends in the fraction of defendants granted non-financial release, the fraction of defendants released within three days, the fraction who are arrested for a new offense during the pretrial period and the fraction of defendants fail to appear to at least one court date. From left to right, the vertical lines in each chart indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified. The horizontal axis is the booking date and the vertical axes are residuals from regressions where the predictor variables consist of the exact charge (for the 42 most common top charges), the total number of charges, whether the defendant had at least one class A, B, C or D felony, whether the defendant had at least one class A or B misdemeanor, the age at arrest, gender and whether the defendant had a prior arrest, a prior FTA or a pending charge within the year before booking. The time trends begin in July of 2010 so that all defendants have at least one year of criminal history data.

Figure 15 - Two month averages

Note: Clockwise from top left, the figures show two month averages in the fraction of defendants granted non-financial release, the fraction of defendants released within three days, the fraction who are arrested for a new offense during the pretrial period and the fraction of defendants fail to appear to at least one court date. From left to right, the vertical lines in each chart indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.
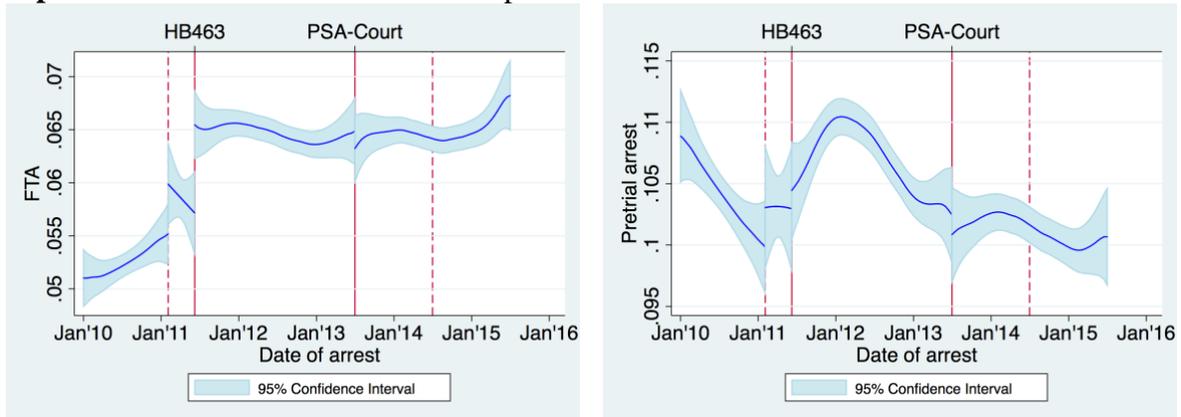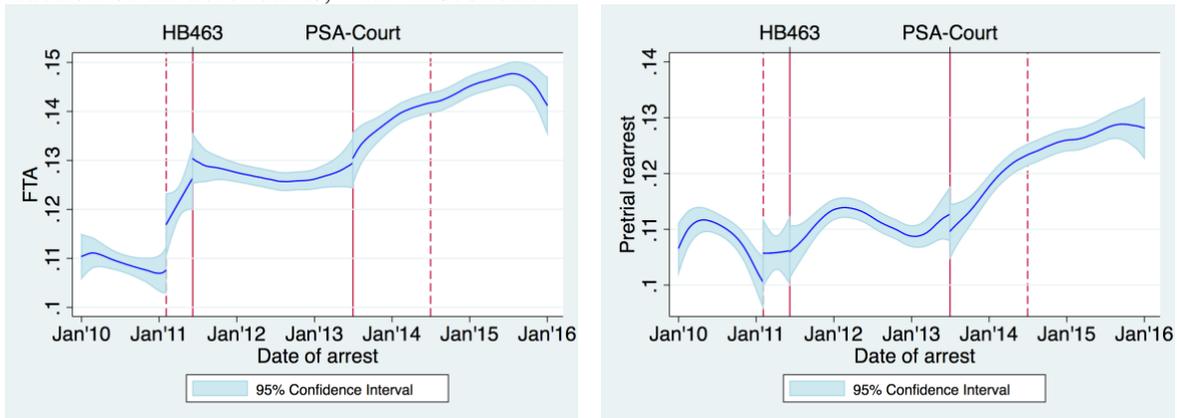
Figure 16 - Non-drug felonies

Note: Clockwise from top left, the figures show time trends in the fraction of defendants charged with a non-drug-related felony who were granted non-financial release, released within three days, arrested for a new offense during the pretrial period and who fail to appear for at least one court date. From left to right, the vertical lines in each chart indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.

Figure 17 - Alternative specifications

Top: Alternative measures of FTA and pretrial rearrest

Bottom: FTA rate and pretrial rearrest rate defined as fraction of released defendants, not fraction of all defendants, with misconduct

Note: The top left shows a time trend failures-to-appear and the top right shows a time trend in pretrial rearrest. While the FTA and pretrial rearrest measures used in the main body of the text were as reported by the pretrial officers, these measures were constructed from the data. A defendant was considered to have an FTA (pretrial rearrest) if the data shows that they were rearrested for an FTA (new offense) between the time of the original arrest and the time of disposition. The bottom two figures show the same FTA and pretrial measure used in the main body of the text, but the time trend is the fraction of released defendants with one of these outcomes, not the fraction of all defendants. From left to right, the vertical lines in each chart indicate the date HB 463 was introduced as legislation, the date it was implemented as law, the date the PSA was adopted, and the date it was modified.