

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing

Citation

Kehl, Danielle, Priscilla Guo, and Samuel Kessler. 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing

Danielle Kehl, Priscilla Guo, and Samuel Kessler
Responsive Communities

I. Introduction 2

II. The History of Risk Assessment in the Criminal Justice System 3

A. The Past as Prelude: The Selective Incapacitation Movement of the 1980s 3
B. Rehabilitation: A Shift Toward Individual Sentencing and Its Discriminatory Effects 6
C. Retributivism and the Rise of Evidence-Based Sentencing 7
D. The Evolution of Risk Assessment Tools 8
E. Enter the Algorithms: Risk Assessment Software 9
F. Risk-Assessment Validity and Adoption 11

III. Algorithms and Criminal Sentencing 13

A. The Move from Parole and Pre-Trial to Sentencing Risk Assessments 13
B. The Sentencing Process 14
C. Evidence-Based Sentencing and the Embrace of Risk Scores 15

IV. Legal Issues Raised By Risk Assessments in Sentencing 18

A. COMPAS Considered in Wisconsin: The Loomis Case 18
B. Constitutional Issues Implicated By Risk Assessment Algorithms 21
C. Related Sentencing Issues: Managing Risk in the Criminal Justice System 26

V. Challenges Presented by the Use of Risk Assessment Algorithms in Sentencing 28
A. Opacity 28
B. Bias and Lack of Reliability 28
C. Diverging Concepts of Fairness 30

VI. Recommendations for the Use of Risk Assessment Algorithms 32

A. Transparency 32
B. Accountability and Oversight 33
C. Robust and Holistic Approach to Fairness 34

VII. Further Areas for Research 35

VIII. Conclusion 36

Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing

Danielle Kehl, Priscilla Guo, and Samuel Kessler

I. Introduction

In the summer of 2016, some unusual headlines began appearing in news outlets across the United States. "Secret Algorithms That Predict Future Criminals Get a Thumbs Up From the Wisconsin Supreme Court," read one.[1] Another declared: "There's software used across the country to predict future criminals. And it's biased against blacks."[2] These news stories (and others like them) drew attention to a previously obscure but fast-growing area in the field of criminal justice: the use of risk assessment software, powered by sophisticated and sometimes proprietary algorithms, to predict whether individual criminals are likely candidates for recidivism. In recent years, these programs have spread like wildfire throughout the American judicial system. They are now being used in a broad capacity, in areas ranging from pre-trial risk assessment to sentencing and probation hearings.

This paper focuses on the latest—and perhaps most concerning—use of these risk assessment tools: their incorporation into the criminal sentencing process, a development which raises fundamental legal and ethical questions about fairness, accountability, and transparency. The goal is to provide an overview of these issues and offer a set of key considerations and questions for further research that can help local policymakers who are currently implementing or considering implementing similar systems. We start by putting this trend in context: the history of actuarial risk in the American legal system and the evolution of algorithmic risk assessments as the latest incarnation of a much broader trend. We go on to discuss how these tools are used in sentencing specifically and how that differs from other contexts like pre-trial risk assessment. We then delve into the legal and policy questions raised by the use of risk assessment software in sentencing decisions, including the potential for constitutional challenges under the Due Process and Equal Protection clauses of the Fourteenth Amendment. Finally, we summarize the challenges that these systems create for law and policymakers in the United States, and outline a series of possible best practices to ensure that these systems are deployed in a manner that promotes fairness, transparency, and accountability in the criminal justice system.

1 Ethan Chiel, *Secret Algorithms That Predict Future Criminals Get a Thumbs Up From Wisconsin Supreme Court*, Fusion (July 27, 2016), http://fusion.net/story/330672/algorithms-recidivism-loomis-wisconsin-court/.

2 Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Suggested Citation: Kehl, Danielle, Guo, Priscilla, Kessler, Samuel. *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. (July 2017). Responsive Communities. Available at: https://cyber.harvard.edu/publications/2017/07/Algorithms.

II. The History of Risk Assessment in the Criminal Justice System

The past decade has witnessed an explosion in the use of algorithms in the public sphere in the United States. The rapid and unprecedented rise of predictive algorithms has been fueled by a number of factors, including the vast amounts of data generated by ubiquitous use of the internet and smart devices and a growing emphasis on data-driven decision-making in both our private lives and public policy.[3] Unsurprisingly, this emphasis on the use of data in government has permeated many stages of the criminal justice system as well, from predictive policing to risk assessment in the corrections system.[4] But while data-driven approaches may explain the recent *expansion* in the use of risk assessment tools, the algorithmic revolution was not responsible for their conception. Risk assessment tools—and the principles underlying their development—have actually been a part of the criminal justice system for decades. In this section, we discuss the predecessors to modern risk assessment software and how their use has evolved and shifted in response to various competing theories of criminal punishment. In particular, we highlight the parallels between the modern emphasis on these tools in sentencing and a controversial (and ultimately unsuccessful) movement in the 1980s known as "selective incapacitation."

A. The Past as Prelude: The Selective Incapacitation Movement of the 1980s

The modern debate about risk assessment algorithms in sentencing bears a striking similarity to a 1980s movement that the *New York Times* described as a "quiet revolution" in the criminal justice system: the selective incapacitation movement.[5] Selective incapacitation theory was based on the premise that the justice system should seek to identify, or "select," a subset of individuals who are particularly prone to violence or recidivism—colloquially known as "career criminals"[6]—and incapacitate them by keeping them in prison for longer periods of time.[7] Removing these criminals from the general population, in theory, would lead to an overall reduction in the crime rate.[8] Although it was ultimately short-lived, the theory of selective incapacitation and the controversy surrounding its practical and ethical implications offer some critical insights into today's debate about risk assessment instruments that similarly purport to identify individuals who are at high risk for both general and violent recidivism and inform judges of those characteristics during the sentencing process. The selective incapacitation debate also suggests that policymakers should proceed cautiously and deliberately when embracing the use of modern risk assessment software, balancing their interest in reducing future crimes against concerns about accuracy and individual fairness.

Crime prediction has been a feature of the United States criminal justice system since the early 1920s.[9] Beginning in the late 1960s and early 1970s, crime prediction research focused

3 *See, e.g.*, *Digital Decisions*, Ctr. for Dem. & Tech. (2016), https://cdt.org/issue/privacy-data/digital-decisions/.

4 Mara Hvistendahl, *Can 'Predictive Policing' Prevent Crime Before It Happens?*, Science (Sept. 28, 2016), http://www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens.

5 Tamar Lewin, *Making Punishment Fit Future Crimes*, N.Y. Times (Nov. 14, 1982), http://www.nytimes.com/1982/11/14/weekinreview/making-punishment-fit-future-crimes.html.

6 *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, 96 Harv. L. Rev., 511, 511 (1982).

7 Thomas Mathiesen, *Selective Incapacitation Revisited*, 22 L. and Human Behavior, 455, 455 (1998), www.jstor.org/stable/1394595.

8 Selective incapacitation is different from collective incapacitation, which is used to punish all persons convicted of similar offenses in the same way. The strategy is used on broad categories of criminals, such as those who committed major felonies or those who have the same number of petty crimes. *Id.* at 455-56.

9 *Id.* at 458.

primarily on identifying an element of "dangerousness" in offenders—namely, the capacity to commit violent crimes.[10] However, predicting dangerousness turned out to be quite complex, and early attempts resulted in a striking number of false positives. For example, some studies in the late 1960s and early 1970s mistakenly identified between 54 and 99 percent of participating individuals as "dangerous."[11]

Nonetheless, despite the difficulty of predicting dangerousness, proponents of the selective incapacitation movement proposed to punish certain individuals more severely based entirely on a predicted future rate of offending.[12] This concept of punishing criminals not for what they had done in the past but for what they could do in the future represented a radical shift in theories of sentencing in criminal justice. The ethical considerations underlying the selective incapacitation strategy embodied a conflict between utilitarianism and the idea that criminals should get their "just deserts." Under a utilitarian approach, selective incapacitation could be justified if it would reduce crime overall and ultimately protect the most number of people from danger. By contrast, the idea of just deserts focuses entirely on punishing criminals for past conduct and emphasizes that "it is unfair to punish for choices expected which have not yet been made — that is, for expected crimes that *might never be committed*."[13]

The utilitarian goals were the focus of the selective incapacitation movement: decreasing the crime rate by imprisoning the most dangerous felons[14] and reducing mass incarceration.[15] But the theory relied on two major assumptions: (1) Career criminals are responsible for the bulk of serious crimes in America and (2) these career criminals can be identified through characteristics like their personal and criminal history.[16] The first assumption was proved through a variety of studies, most notably the 1972 study entitled "Delinquency in a Birth Cohort."[17] After careful analysis of the criminal records of 10,000 males in Philadelphia, the study found that 51.9 percent of the total offenses were committed by just 18 percent of the group, otherwise known as the chronic offenders.[18] The second assumption, however, never substantiated the correlations that it drew between personal and criminal histories and the potential for recidivism.

In 1982, a report from the RAND Corporation expounded on the potential benefits of selective incapacitation theory and stimulated discussion in the academic and criminal justice community on the validity of the theory. The report's authors, Peter Greenwood and Allan Abrahamse, surveyed 2100 male inmates in California, Texas, and Michigan prisons and jails over a six-year period.[19] They gathered information directly from prisoners through interviews about their crimes and data compiled into self-reports, and then included information from their official crime records in the report. As evidence that a small percentage of the prison population was particularly prone to criminal activity, the researchers noted that "[a]mong active burglars, 50 percent committed fewer than 6 per year, while 10 percent committed more than 230 per year."[20] Furthermore, they found strong correlations between recidivism and factors such as "juvenile convictions, heroin or barbiturate use, unemployment and prior imprisonment."[21]

Yet there were significant limitations to the RAND report. First, the assumptions were based on robbery and burglary crimes only. Moreover,

10 Mathieson, *Selective Incapacitation: Reducing Crime Through Predictions of Recidivism*, *supra* note 7, at 515.

11 JOHN MONAHAN, PREDICTING VIOLENT BEHAVIOR: AN ASSESSMENT OF CLINICAL TECHNIQUE 244, 246-50 (1981).

12 Jacqueline Cohen, *Incapacitation as a Strategy for Crime Control: Possibilities and Pitfalls*, 5 CRIME AND JUSTICE, 1, 12 (1983), www.jstor.org/stable/1147469.

13 Mathiesen, *Selective Incapacitation Revisited*, *supra* note 7, at 460 (emphasis added).

14 Stephen D. Gottfredson & Don Gottfredson, *Selective Incapacitation?*, 478 THE ANNALS OF THE AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE 135, 142 (Mar. 1985).

15 Lewin, *Making Punishment Fit Future Crimes*, *supra* note 5.

16 Cohen, *Incapacitation as a Strategy for Crime Control*, *supra* note 12, at 8-9.

17 MARVIN E. WOLFGANG, ROBERT M. FIGILIO & THORSTEN SELLIN, DELINQUENCY IN A BIRTH COHORT 327 (1972).

18 *Id.*

19 Peter W. Greenwood with Allan Abrahamse, *Selective Incapacitation*, RAND CORPORATION (Aug. 1982), https://www.rand.org/content/dam/rand/pubs/reports/2007/R2815.pdf (hereinafter "RAND Report").

20 *Id.* at xiii.

21 Lewin, *Making Punishment Fit Future Crimes*, *supra* note 5.

the report was a retrospective analysis of past crimes committed. There was no actual test of predicted future behavior. Greenwood and Abrahamse *did* develop a predictive scale for identifying risk in offenders, labeling those most likely to reoffend as "high-rate" and the rest as "medium-rate" or "low-rate." Their predictive scale was fairly accurate in predicting which criminals would be low-rate offenders, with a rate of 76 percent correctness.[22] However, the scale was extremely inaccurate for high-rate offenders. According to researcher Jacqueline Cohen, only 45 percent of the criminals categorized as high-rate offenders were correctly identified, resulting in a false-positive rate of 55 percent of survey respondents.[23] In other words more than half of supposed high-rate offenders were incorrectly labeled.

The RAND report faced other criticisms as well. There were no validation tests conducted on the report. Moreover, the data was highly speculative since the researchers obtained much of the personal and criminal history through interviews with the offenders themselves. In 1986, researcher Christy Visher reanalyzed the report and concluded that "reduction of crime would decline further" if the model was completed with more official criminal records.[24]

More importantly, predicting recidivism was susceptible to the risk of false negatives and false positives that could undermine the entire purpose of the theory. In false negative cases, individuals were mistakenly predicted as unlikely to recommit but subsequently did. In these cases, predictive failure allowed individuals back into a society where they could commit additional crimes. False positives, on the other hand, represented an error that threatened individual liberty. Individuals would be mistakenly identified to be recidivists and imprisoned for crimes that they had no intention of actually committing. Notably, the RAND report acknowledged that the problem of false positives raised concerns about selective incapacitation because it undermined the foundational presumption of innocence until proven guilty. The study acknowledged that "[a]s long as our ability to discriminate between high and low-rate offenders is imprecise, there will be legitimate concern about those who are improperly classified… Furthermore, there will be differences of opinion as to the legitimacy of using some of the factors that are correlated with rates of offending (e.g. juvenile record, drug use, employment) for sentencing purposes."[25]

Proponents of the theory, however, argued that selective incapacitation was still an improvement over relying solely on human judgment in criminal sentencing, citing the need for guidelines and "orderly assessment schemes."[26] In other words, they argued that these predictive instruments were much more accurate than our intuitive methods.[27] Predictive instruments could help judges identify who was truly risk for recidivism, thereby limiting the imposition of long prison sentences that human judges tend to dole out somewhat arbitrarily.[28]

Ultimately, selective incapacitation never became a mainstream concept, largely due to concerns about predictive accuracy and individual fairness.[29] Yet its principles have lingered on in the criminal justice system. Many states today have "repeat-offender laws, prosecutorial units devoted to career criminals and sentencing policies that consider prior offenses, job stability and other personal data."[30] And despite the criticisms of its methods, the RAND report inspired the precursors to various modern-day prediction tools, including the INSLAW instrument (developed for U.S. federal prosecutors to carry out risk assessment of offenders), the Salient Factor Score (developed as a risk assessment scale for U.S. Parole Commission), and the Canadian

22 Gottfredson & Gottfredson, *Selective Incapacitation?*, *supra* note 14, at 140.

23 Cohen, *Incapacitation as a Strategy for Crime Control*, *supra* note 12, at 48-49.

24 Christy Visher, *The Rand Inmate Survey: A Reanalysis*, *in* A. Blumstein et al., Criminal Careers and "Career Criminals" 205-226 (1986).

25 RAND Report, *supra* note 19, at 22-23.

26 C.D. Webster, *Comment on Thomas Mathiesen's Selective Incapacitation Revisited*, 22 L. and Human Behavior, 473, 473 (1998), available at www.jstor.org/stable/1394596.

27 R.A. Wright, In Defense of Prisons (1994).

28 Mathiesen, *Selective Incapacitation Revisited*, *supra* note 7, at 466.

29 Kathleen Auerhahn, *Selective Incapacitation and the Problem of Prediction*, 37 Criminology 703, 703 (1999).

30 Lewin, *Making Punishment Fit the Future Crime*, *supra* note 5.

Dangerous Behavior Rating Scale for Metropolitan Toronto Forensic Service.[31]

Although selective incapacitation is primarily seen as a historical footnote today, the movement sheds light on today's discussion of risk assessment algorithms in sentencing, which is plagued by many of the same concerns about accuracy and fairness toward the individual defendant. Much like proponents of selective incapacitation in the 1980s, advocates for the widespread use of risk assessments today appear to be doing so out of a genuine desire to reduce mass incarceration without increasing the crime rate and to use data and technical analysis to improve upon untethered human judgment.[32] But doing so successfully and fairly may be a far more difficult task than it seems, particularly in the sentencing context, where risk assessment could ultimately turn into a resurrection of the ideas behind selective incapacitation theory.

B. Rehabilitation: A Shift Toward Individual Sentencing and Its Discriminatory Effects

Beyond the selective incapacitation context, today's risk-assessment algorithms are the product of broader philosophical debate in the United States regarding the objectives of our criminal justice system. In the late nineteenth century, the American criminal justice system began to shift away from capital and corporal punishment and towards rehabilitation. This rehabilitative focus dominated criminal justice discussions until the 1970s and it emphasized assigning punishments based on an individual's characteristics rather than just the crimes that they committed. In other words, instead of simply punishing people in proportion to the severity of their crimes, individuals were given unique sentences and treatment with the ultimate goal of rehabilitation, in order to prepare them for safe reentry into society. With rehabilitation as the central goal, strict guidelines and sentences were not considered appropriate. Thus, in order to ensure individual treatment, judges were granted extraordinary discretion in regards to sentencing decisions.[33]

Yet greater sentencing discretion may have had negative effects. In particular, it quickly became clear that minorities were being treated disproportionately compared to their white peers in sentencing.[34] In 1977, Senator Edward Kennedy explained the disparate impact of contemporary sentencing practices on minorities:

> During the past few years a quiet but constructive debate has ensued over the issue of comprehensive criminal sentencing reform. The debate has involved judges, lawyers, corrections officials, law enforcement officers, members of the academic community and others. It has focused primarily on two interrelated problems–the total absence of any prescribed guidelines to aid judges during the sentencing process and the wide disparity in the sentences actually imposed in criminal cases.... The result has been chaotic–all too often two defendants with similar backgrounds, convicted of the same crime, receive widely disparate sentences.[35]

Although increased judicial discretion was intended to serve a rehabilitative end, the disparate impact that it has had on minorities suggested the approach also had a discrimina-

31 Mathiesen, *Selective Incapacitation Revisited*, *supra* note 7, at 460.

32 *See, e.g.*, CSG Justice Center Staff, *Risk and Needs Assessment and Race in the Criminal Justice System*, The Council of State Governments (May 31, 2016), https://csgjusticecenter.org/reentry/posts/risk-and-needs-assessment-and-race-in-the-criminal-justice-system/ (noting that "validated risk and needs assessment is necessary to more accurately determine the risk of recidivism and criminogenic needs of people involved in the criminal justice system—and to inform how the system responds to reduce that risk and address those needs—than by relying on subjective, individual judgment.").

33 Douglas A. Berman, *Re-balancing Fitness, Fairness, and Finality for Sentences*, 4 Wake Forest J. Law & Policy 151, 157-8 (2014).

34 Joshua B. Fischman & Max M. Schanzenbach, *Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. of Empirical Legal Stud. 729, 729-64 (2012) (noting that United States Sentencing Guidelines were introduced to mitigate the disparate impact of judicial discretion on judges, but might have had a counterproductive effect).

35 Pierce O'Donnell et al., Toward a Just and Effective Sentencing System: Agenda for Legislative Reform (1977).

tory effect. Many more policymakers eventually joined Senator Kennedy in questioning whether better guidelines might be necessary to assist in sentencing decisions in order to mitigate the system's disproportionate sentencing practices.[36]

C. Retributivism and the Rise of Evidence-Based Sentencing

With these concerns in mind, the sentencing reform movement of the 1970s and 1980s shifted back towards the retributive notion that criminal sentences should be based primarily on the crime committed rather than on the criminal himself.[37] The primary result of this shift was the establishment of clearer sentencing practices and increased use of sentencing guidelines. As a part of this reform movement, Congress passed the Sentencing Reform Act (SRA) in 1984.[38] The SRA was predicated on the idea that sentencing practices had become unfair and uncertain under the prevailing rehabilitative model, and it formalized federal sentencing through the establishment of the U.S. Sentencing Commission.[39] The SRA also prescribed a clear sentencing structure under the federal sentencing guidelines.[40]

While the extraordinary discretion granted to judges under the rehabilitative approach may have produced discriminatory effects, strict retributivism was criticized for ushering in the era of mass incarceration, which arguably had its own discriminatory impact.[41] Policymakers soon began to grapple with the problems created by America's ever-expanding prison population and the harsh realities of these new sentencing requirements.

In recent years, there has been a move towards evidence-based practices (EBP), which strive to improve sentencing decisions by incorporating scientific and quantitative methods. Evidence-based practices take an actuarial approach to assessing and treating risk, using the scientific method to predict future behavior.[42] Although the EBP movement has received some criticism for having had little effect on the mass incarceration problem (or potentially making it worse),[43] EBP is intended to improve sentencing outcomes by using empirical assessment to inform sentencing decisions.[44]

In the context of the criminal justice system, evidence-based practices utilize data to assess the risk of re-offense, or recidivism. The goal of these methods is to reduce recidivism rates by focusing on particular offender character-

36 *The Sentencing Reform Act of 1984: Principle Features Affecting Guideline Construction*, U.S. Sentencing Commission (n.d.), http://www.ussc.gov/research/research-and-publications/simplification-draft-paper-2.

37 Letter from Jonathan Wroblewski, Director, Office of Policy and Legislation, U.S. Department of Justice's Criminal Division, to the Honorable Patti B. Saris, Chair, U.S. Sentencing Commission (Jul. 29, 2014), https://www.justice.gov/sites/default/files/criminal/legacy/2014/08/01/2014annual-letter-final-072814.pdf (hereinafter "DOJ Letter").

38 Sentencing Reform Act, P.L. §98-473 (1984).

39 *An Overview of the United States Sentencing Commission*, United States Sentencing Commission (Jan. 5, 2011), http://www.ussc.gov/sites/default/files/pdf/about/overview/USSC_Overview.pdf.

40 Nathan James, *Risk and Needs Assessment in the Criminal Justice System*, Congressional Research Services 1, 13 (Oct. 13, 2015), https://fas.org/sgp/crs/misc/R44087.pdf. It should be noted that although the SRA established the U.S. Sentencing Commission to create clearer sentencing guidelines, these guidelines only apply on a federal level. Most criminal prosecution still happens in state courts, where these federal guidelines have some persuasive influence but are not legally binding.

41 *See, e.g.*, *The Moral Failures of America's Prison-Industrial Complex*, The Economist (Jul. 8, 2015), http://www.economist.com/blogs/democracyinamerica/2015/07/criminal-justice-and-mass-incarceration.

42 Actuarial assessment is "a formal method . . . [that provides] a probability, or expected value, of some outcome. It uses empirical research to relate numerical predictor variables to numerical outcomes. The sine qua non of actuarial assessment involves using an objective, mechanistic, reproducible combination of predictive factors, selected and validated through empirical research, against known outcomes that have also been quantified." Kirk Heilbrun, *Risk Assessment in Evidence-based Sentencing: Context and Promising Uses*, 1 Chapman J. Crim. Just. 1, 134 (2009).

43 Cecelia Kingele, *The Promises and Perils of Evidence-Based Corrections*, University of Wisconsin Law School, Legal Studies Research Paper Series No. 1368 (Nov. 30, 2015), http://www.wisconsinappeals.net/wp-content/uploads/2015/12/Klingele-article.pdf.

44 Richard E. Redding, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, 1 Chapman J. Crim. Just. 1, 3-4 (2009), http://works.bepress.com/richard_redding/11.

istics and criminogenic needs—factors which are believed to increase a person's propensity to commit crimes in the future.[45] Criminals are generally grouped by their risk, and assigned a high, medium, or low risk score. Consistent with rehabilitative approaches, this risk score is supposed to help determine the treatment and interventions an offender will receive in prison.[46] Factors that increase and decrease the likelihood of recidivism are both considered, and sentencing as well as treatment are assigned with these factors in mind.[47]

Some experts have praised evidence-based practices for their potential to find a constructive middle ground between the extreme results produced by placing a stronger emphasis on either rehabilitation or retributivism alone. As Chapman University's Dr. Richard E. Redding explains:

> [T]he evidence-based approach will likely result in sentencing decisions that more comprehensively consider relevant utilitarian and retributive considerations. '[R]etribution-oriented judges may concern themselves with the story of crime, and perhaps proceed to construct a narrative about the offender's criminal history, but they are unlikely to construct a story of the offender's life as a rehabilitation oriented judge would be likely to do.' Risk and needs assessments force judges to focus on both stories–the offense and offense history as well as the risk and protective factors relevant to rehabilitation, all in a more precise and accurate way."[48]

Far from a complete departure from rehabilitation and retributivism, the evidence-based risk/needs assessment model, which we describe in the next section, embraces the principles of rehabilitation while attempting to preserve some of the standardization provided by retributive approaches. As mentioned above, these evidence-based practices also place a renewed emphasis on recidivism because of its central role in decisions about how to treat offenders, particularly when trying to balance public safety against a desire to reduce mass incarceration and prison overcrowding.[49] This shift has led to the development of risk assessment tools that are aimed at predicting an individual's likelihood of recidivism.

D. The Evolution of Risk Assessment Tools

There have been roughly four different generations of risk assessment tools over the course of the past century.[50] The focus on rehabilitation from the first half of the twentieth century can be seen in the first generation, where risk assessment was conducted on a case-by-case basis by correctional staff and clinical professionals working in prisons.[51] These actors would generally rely on their own professional judgment when making decisions for individuals about sentencing, supervision, and treatment. But over time, the way in which risk is measured has evolved considerably.

---

45 *Id.*

46 Melissa Hamilton, *Risk-Needs Assessment: Constitutional and Ethical Challenges*, U. OF HOUSTON LAW CENTER NO. 2014-W-2 (Jan. 26, 2015), https://ssrn.com/abstract=2506397.

47 Redding, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, *supra* note 44, at 3-4.

48 *Id.* at 9.

49 According to the National Institute of Justice, recidivism is important due its interplay with incapacitation, specific deterrence, and rehabilitation: "*Incapacitation* refers to the effect of a sanction to stop people from committing crime by removing the offender from the community. *Specific deterrence* is the terminology used to denote whether a sanction stops people from committing further crime, once the sanction has been imposed or completed. *Rehabilitation* refers to the extent to which a program is implicated in the reduction of crime by "repairing" the individual in some way by addressing his or her needs or deficits." *Why Recidivism is a Core Criminal Justice Concern*, NATIONAL INSTITUTE OF JUSTICE (Oct. 3, 2008), https://www.nij.gov/topics/corrections/recidivism/pages/core-concern.aspx.

50 Susan Turner et al., *Development of the California Static Risk Assessment (CSRA): Recidivism Risk Prediction in the California Department of Corrections and Rehabilitation*, CENTER FOR EVIDENCE-BASED CORRECTIONS, UNIVERSITY OF CALIFORNIA-IRVINE (2013), http://ucicorrections.seweb.uci.edu/files/2013/12/Development-of-the-CSRA-Recidivism-Risk-Prediction-in-the-CDCR.pdf; James Bonta & D.A. Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, PUBLIC SAFETY CANADA (2006-07), https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-nd-rspnsvty/rsk-nd-rspnsvty-eng.pdf.

51 Bonta & Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, *supra* note 50, at 3.

The largest evolution of risk assessment came with the aforementioned shift towards evidence-based practices and the development of sophisticated tools to measure risk. Evidence-based risk/needs assessment instruments consider the interplay between *static* and *dynamic* risk factors. Dynamic risk factors are any factors that contribute to recidivism risk that can change over time. For rehabilitative tools, these factors—which include current age, employment status, and whether a person is in treatment for substance/alcohol abuse—are treated through targeted interventions that are intended to decrease the likelihood of recidivism.[52] These dynamic factors are also referred to as "criminogenic needs" since they can be addressed via treatment.[53] For example, an offender with alcohol problems might be placed in programming aimed at treating his addiction, which could ultimately decrease his likelihood of reoffending. On the other hand, static risk factors—which include criminal history, age at first arrest, and gender—are also correlated with risk, but they are not targeted for treatment since they cannot be changed. Static factors are, however, often used alongside dynamic factors to evaluate risk of recidivism.[54]

The second generation of risk assessment tools, which emerged in the 1970s, primarily embraced static factors for measuring risk.[55] Many second-generation tools abandon dynamic risk-factors altogether,[56] and the immutable nature of static factors makes it difficult (if not impossible) for these tools to account for positive changes or progress.[57] Since the offender cannot alter static factors, tools which rely upon them might have a discriminatory effect—judging people for factors over which they have no control. The third generation of risk assessments attempted to solve for the shortcomings of static risk factors by considering static and dynamic factors in tandem with one another.[58] This generation, of which risk/needs assessment is a part, is especially useful to rehabilitative models where changing offender characteristics matter. Finally, the fourth generation of risk assessment tools builds off of the third generation but it embraces a more "systematic and comprehensive" approach to measuring recidivism and treating offenders based on their specific risk factors and characteristics.[59]

E. Enter the Algorithms: Risk Assessment Software

Today's fourth-generation risk-assessment tools are far more technically sophisticated and widely available than the rudimentary tools that had been used in the United States to inform parole decisions since the 1920s.[60] A number of modern risk-assessment tools take advantage of machine learning algorithms, which generate risk models based on vast quantities of data. As these algorithms are used over time, their models often dynamically adjust to new data. Risk assessment tools and software–many of which incorporate machine learning–are now being used in a variety of contexts, including prison rehabilitation programs, pretrial risk assessment, and sentencing. In this subsection, we describe the primary tools and models used in these three areas.

i. Rehabilitation-Specific Risk Assessment Tools

The foundation of most rehabilitative risk/needs assessment (RNA) tools is the risk-needs-responsivity (RNR) model, which rests on the aforementioned concept of responding to recidivism

52 James, *Risk and Needs Assessment in the Criminal Justice System*, *supra* note 40, at 3. For example, if an offender has a history of alcohol or drug abuse (a dynamic factor), they may receive some kind of addiction treatment.

53 D.J. Simourd, *Use of Dynamic Risk/Need Assessment Instruments Among Long-Term Incarcerated Offenders*, 31 Crim. Just. and Behav., 306, 306-323 (2004).

54 James, *Risk and Needs Assessment in the Criminal Justice System*, *supra* note 40, at 3.

55 Bonta & Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, *supra* note 50, at 3.

56 Turner et al., *Development of the California Static Risk Assessment*, *supra* note 50, at 5

57 Bonta & Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, *supra* note 50, at 4.

58 *Id.*

59 *Id.*

60 Richard A. Berk & Justin Bleich, *Statistical Procedures for Forecasting Criminal Behavior*, 12 Criminology & Pub. Policy 1, 2 (2013). *See also* Howard G. Borden, *Factors For Predicting Parole Success*, 19 J. Crim. Law & Criminology 328, 328-36 (1928), http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=2101&context=jclc.

risk and criminogenic needs through the most appropriate treatment.[61] The RNR model, which rose to prominence in the third and fourth generations of risk assessment, is based on three principles:

1. The risk principle, which asserts that risk is predictable, and high-risk offenders should receive different and more intensive treatment than low-risk offenders.[62]
2. The needs principle, which suggests rehabilitative treatment and sentencing decisions should respond to criminogenic needs which contribute to criminal behavior.[63]
3. The responsivity principle, which describes how treatment should be tailored to the specific offender.[64]

Many RNA instruments are used in prison rehabilitation programs, and these tools use the RNR model to rehabilitate as well as incapacitate offenders. Canada was a trailblazer in this area of using evidence-based methods for rehabilitation,[65] but California and other states in the U.S. have followed suit by implementing RNA and rehabilitation into treatment and sentencing. Rehabilitative tools like those developed for use in Canada and California target dynamic risk factors for treatment, and use static risk factors to measure risk.[66]

ii. Pretrial Detention and Release

Another use of risk-assessment tools is for pre-trial detention and release decisions, which generally places more focus on static risk factors. One such pre-trial tool, the Public Safety Assessment (PSA), is used in 29 American jurisdictions including three entire states: Arizona, Kentucky, and New Jersey. The PSA, which was developed by the Laura and John Arnold Foundation, was built using data from 1.5 million crimes spanning 300 U.S. jurisdictions, and it measures risk using a very narrow set of static risk factors relating primarily to the defendant's age and criminal history. The PSA does not seek to identify rehabilitative treatments for offenders, but rather was built to help make decisions about whether an individual should be detained or released before going to trial.[67] The instrument makes a risk determination based on the aforementioned static risk factors, and this risk classification is used to determine whether a person is low-risk, and can therefore safely be released, or is high-risk, and should be detained.[68]

iii. Sentencing

Although there has been considerable focus on using risk assessment algorithms in rehabilitation and pretrial decision-making, they have recently drawn attention for their use in sentencing—the primary focus of this paper.[69] In 1994, Virginia was the first state to implement a risk assessment instrument for use in sentencing. The instrument, which was created by the Virginia Criminal Sentencing Commission, was designed to identify low-risk felons in order to assign them a more suitable type of punishment.[70] These alternative punishments include diversion from prison to jail, diversion from jail to community service or home-arrest, and fines.[71] Virginia remains unique in its approach to developing risk assessment tools. While a handful of states

61 Bonta & Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, *supra* note 50, at 1.

62 James, *Risk and Needs Assessment in the Criminal Justice System*, *supra* note 40, at 3.

63 *Id.*

64 *Id.* at 6-7.

65 Bonta & Andrews, *Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation*, *supra* note 50, at 3.

66 James, *Risk and Needs Assessment in the Criminal Justice System*, *supra* note 40, at 1, 13.

67 *See infra* Part III.A.

68 *Public Safety Assessment*, Laura and John Arnold Foundation, http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-assessment/ (accessed Dec. 15, 2016).

69 *See, e.g.*, Julia Angwin, *Make Algorithms Accountable*, N.Y. Times (Aug. 1, 2016), http://www.nytimes.com/2016/08/01/opinion/make-algorithms-accountable.html?_r=0; Sari Horwitz, *Eric Holder: Basing Sentences on Data Analysis Could Prove Unfair to Minorities*, Washington Post (Aug. 1, 2014), https://www.washingtonpost.com/world/national-security/us-attorney-general-eric-holder-urges-against-data-analysis-in-criminal-sentencing/2014/08/01/92d0f-7ba-1990-11e4-85b6-c1451e622637_story.html?utm_term=.18af89d61814.

70 Brian Ostrom, *Offender Risk Assessment in Virginia: A Three-stage Evaluation: Process of Sentencing Reform, Empirical Study of Diversion and Recidivism, Benefit-cost Analysis*, National Center for State Courts: Virginia Criminal Sentencing Commission (2002), http://www.vcsc.virginia.gov/risk_off_rpt.pdf.

71 *Id.* at 19.

like Virginia and Pennsylvania use risk-assessment tools that have been developed by (or in partnership with) the state government, many more states and jurisdictions have implemented or adapted one of several existing commercial systems.[72]

One of the first and most popular commercial risk-assessment tools to be used in sentencing is called the Level of Service Inventory – Revised (LSI-R). LSI-R, which was developed by the Canadian company Multi-Health Systems, pulls information from a survey containing a wide set of static and dynamic factors. These factors, which range from criminal history to personality patterns, are used to determine a person's risk for recidivism as well as the best sentencing options. The tool was initially developed for use in rehabilitation, but it subsequently has been adapted for use in sentencing. LSI-R and adapted versions of it are used to assist sentencing in a number of states and jurisdictions, including Washington[73] and California.[74]

Another popular tool, COMPAS, was created by the company Northpointe. COMPAS assesses variables under five main areas: criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion. It uses a combination of static and dynamic factors in order to assess recidivism risk, and it can be programmed for a variety of use cases.[75]Although COMPAS can be employed for purposes beyond sentencing, a number of states, including Wisconsin, Florida, and Michigan, use COMPAS to assist judges with sentencing decisions.[76] Because COMPAS is proprietary software, it is not subject to federal oversight and there is almost no transparency about its inner workings, including how it weighs certain variables. COMPAS has created a considerable amount of controversy for this very reason.

F. Risk-Assessment Validity and Adoption

Accuracy is of paramount concern when it comes to using risk assessment instruments, especially in the sentencing context. A 2006 study in the *Journal of Criminal Justice* that examined the importance of implementation integrity for LSR-I noted that while it is important that high-risk offenders receive more severe sentences, it is equally important that low-risk offenders receive less severe sentences.[77] Risk-assessment algorithms are useful for identifying these high and low-risk offenders, but it is important that they are identified accurately since inaccuracies would not only be unjust, but could actually make individuals likely to recidivate.[78]

Research has generally confirmed that risk assessment instruments can predict who is at risk to recidivate with at least some degree of accuracy.[79] Furthermore, a number of academics like James Bonta argue that actuarial assessment, which is at work in risk-assessment algorithms, is preferable to clinical assessment.[80] Bonta notes that studies have generally credited greater accuracy and predictive validity to the objectivity

72 *Algorithms in the Criminal Justice System*, Electronic Privacy Information Center (n.d.), https://epic.org/algorithmic-transparency/crim-justice/ (accessed Dec. 15, 2016).

73 *Sex Offender Sentencing in Washington State: Predicting Recidivism Based on the LSI-R*, Washington State Institute for Public Policy (2006), http://www.wsipp.wa.gov/ReportFile/935/Wsipp_Predicting-Recidivism-Based-on-the-LSI-R_Predicting-Recidivism-Based-on-the-LSI-R.pdf.

74 Turner et al., *Development of the California Static Risk Assessment (CSRA)*, *supra* note 50.

75 *Practitioner's Guide to COMPAS Core*, Northpointe (Mar. 19, 2015), https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf.

76 *Algorithms in the Criminal Justice System*, *supra* note 72.

77 Anthony W. Flores et al., *Predicting Outcome with the Level of Service Inventory-Revised: The Importance of Implementation Integrity*, 34 J. Crim. Just. 523, 523-29 (2006), http://www.sciencedirect.com.ezp-prod1.hul.harvard.edu/science/article/pii/S0047235206000833 (noting that "the incorporation of the risk principle of offender classification dictates that higher risk individuals warrant the majority of correctional attention, including the most intensive levels of both rehabilitative service and supervision. Conversely, and arguably as important, is the need to leave lower risk individuals free from intense levels of intervention to avoid interference with the protective factors that are likely present in their environment and within themselves.").

78 *Id.*

79 James, *Risk and Needs Assessment in the Criminal Justice System*, *supra* note 41, at 3.

80 James Bonta, *Offender Assessment: General Issues and Considerations*, Compendium on Effective Correctional Programming (2000), http://www.csc-scc.gc.ca/005/008/compendium/2000/chap_4-eng.shtml.

of actuarial tools compared to the theoretical nature professional clinical judgment. Nevertheless, no instrument is completely accurate, and it has even been suggested that there might be some "natural limit" to the accuracy of risk-assessment algorithms.[81] Yet risk assessment has received widespread support and is generally considered to be a valid method for predicting risk.

81 John Monahan and Jennifer L. Skeem, *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning*, 26 Fed. Sent'g Rep. 162 (Feb. 2014).

III. Algorithms and Criminal Sentencing

While the previous section described the history of risk assessment in the criminal justice system broadly, this section focuses specifically on the use of modern risk assessment tools in sentencing decisions. We discuss the inherent challenges of adapting these tools from the parole and pre-trial context to sentencing, and then explain the mechanics of how the scores are currently incorporated into the sentencing process.

A. The Move from Parole and Pre-Trial to Sentencing Risk Assessments

The fact that these algorithms have been successfully used in other parts of the criminal justice system may help explain why lawmakers and judges have been relatively quick to embrace them in the sentencing context. But these risk assessment tools may be better suited and easier to assess in other contexts, such as during pre-trial release, when a judge is evaluating whether a criminal defendant should be held in jail prior to her scheduled appearances in court. The goals of a pre-trial risk evaluation are relatively well defined. A judge is trying to predict whether the defendant will appear in court when she is supposed to, and whether she is likely to commit any crimes in the meantime. If a defendant poses a significant flight risk or a danger to the public, the judge will likely recommend against release, whereas a defendant that appears low risk in both categories is likely to be set free before trial. It is not surprising, therefore, that it has become increasingly common to augment judicial decision-making with risk assessment software like Public Safety Assessment in order to help reduce the number of individuals behind bars before trial without increasing risk to the public.[82]

Sentencing, by contrast, involves a much broader range of considerations. A sentencing decision involves first deciding *how* to punish someone and then, if a judge chooses incarceration, *how long* a sentence should be. Determining the severity and length of punishment often draws upon a number of different theories of punishment, including individual retribution, rehabilitation, deterrence, and incapacitation.[83] Judges often base their decisions on multiple theories, despite their varied goals.[84]

As discussed above, there is a clear relationship between recidivism and the goals of rehabilitation and incapacitation: individuals who are unlikely to reoffend are typically considered good candidates for rehabilitation and less severe forms of punishment, whereas a high risk of recidivism may support an argument for long-term or permanent incapacitation to protect society against the defendant's future dangerousness.[85] But the links between recidivism and the punishment goals of deterrence and retribution are more tenuous.[86] To the extent that a longer

82 Matthew Conlen, Reuben Fischer-Baum & Andy Rossback, *Should Prison Sentences Be Based on Crimes that Haven't Been Committed Yet?*, FiveThirtyEight Politics (Aug. 4, 2015), http://fivethirtyeight.com/features/prison-reform-risk-assessment/ (noting that "[t]here is little question that well-designed risk assessment tools "work," in that they predict behavior better than unaided expert opinion."). *See, e.g.*, J.C. Oleson et al., *Training To See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessment Among Federal Probation Officers*, 75 Fed. Prob. 52 (Sept. 2011) (finding that federal probation officers "made more consistent and accurate assessments of offender risk when using [a risk assessment tool] than when using unstructured clinical judgment" or relying on professional experience). Several professors that we interviewed for this paper also indicated that forthcoming studies find similar results in the pre-trial risk assessment context, where decisions aided by algorithmic risk assessment tools are more accurate at predicting which offenders are likely to commit crimes if released than decisions made solely relying on judge's unguided human judgment.

83 *See* Model Penal Code § 1.02(2), which notes that the general purposes of sentencing include, among others, "prevent[ing] the commission of offenses," "promot[ing] the correction and rehabilitation of offenders," and "differentiat[ing] among offenders with a view to a just individualization in their treatment."

84 In one study, eighteen judges were asked to report information their decisions on 1000 adult offenders, and the results suggested they rarely attributed their decision to any one goal. Gottfredson, *Selective Incapacitation?*, *supra* note 14.

85 Bernard Harcourt, Against Prediction 31-34 (2005).

86 *See, e.g.*, Paul Gendreau et al, *The Effects of Prison Sentences on Recidivism*, Department of the Solicitor General of Canada (1999), https://www.prisonpolicy.org/scans/gendreau.pdf (concluding that "[p]risons should not be used with the expectation of reducing criminal behavior" and "[t]he primary justification of prison should be to incapacitate offenders (particularly

prison sentence deters the individual who is receiving the sentence from committing future crimes, it arguably has an impact on recidivism. But we tend to think of deterrence in terms of society more broadly, and how the decision to punish an individual for a crime will impact others who might be inclined to commit the same crime. This broader conception of deterrence bears little relation to an individual's risk of committing future crimes. Finally, retribution, although focused entirely on the individual criminal, is a backward-looking assessment of his blameworthiness. A criminal's future dangerous has little relevance to ensuring that he gets his "just desserts" for the crime he previously committed. Thus, in sentencing decisions, although recidivism may be a relevant factor, it is hardly the only consideration—and may not even be a central or determinative one.[87]

Moreover, regardless of a judge's primary theory of punishment, it is less clear how he or she should use a risk assessment score to inform a sentencing decision as opposed to the pre-trial release context. Before trial, a judge faces a decision that is essentially binary: should the prisoner stay in jail for the duration of the pre-trial period or not? But at sentencing, a judge also has to decide how long the punishment should be. There is little positive evidence supporting the notion that a longer criminal sentence has a significant impact on an individual's recidivism.[88] And so it does not necessarily follow that a longer sentence will decrease the likelihood that a criminal will commit crimes again in the future.[89] A judge therefore faces a more complicated question about how to use the information provided in the risk assessment score, and his answer may be highly dependent on his own primary theory of punishment. Or, he may simply take a risk-averse approach and impose more stringent sentences on criminals who are labeled high risk in order to avoid potential blame for a high-risk criminal who received a less severe sentence and ultimately did reoffend.[90]

These differences do not necessarily suggest that these tools should only be used in the pre-trial risk assessment context, but rather that expanding to other, more complicated areas like sentencing requires a great deal of thought. In particular, sentencing authorities need to consider which goals of punishment they are trying to achieve and how algorithmic tools could help maximize for those goals, if possible.[91] Part of this process may also involve thoughtful deliberation about how to quantify effects like deterrence and retribution, which are harder to mathematically measure than recidivism but may be valuable ends. Moreover, it highlights the need for research to inform our understanding of how factors like the type and length of the sentence impact future outcomes.

B. The Sentencing Process

Despite the complexity of using these instruments in sentencing, as noted above, states are increasingly recommending or mandating their use. In this subsection we provide some context about how sentencing works generally and how these risk scores are specifically being incorporated into that process today.

---

those of a chronic, high-risk nature) for reasonable periods and to exact retribution.").

87 For further discussion of these concepts, see Harcourt, Against Prediction, *supra* note 85, at 31-34, 188-89.

88 Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 Stan. L. Rev. 803, 855-56 (2014) (noting that "[t]he instruments tell us, at best, who has the highest risk of recidivism…. not… whose risk of recidivism will be reduced the most by incarceration" and that EBS "predictions are not conditional on the sentence.").

89 We should note that there is potential for a false positive here: a criminal who is identified as high risk for recidivism and subsequently given a longer sentence may therefore appear as a recidivism "success" because the fact that he is in prison for longer deprives him of the opportunity to commit future crimes. However, that does not mean that the punishment *itself* lowers his risk of recidivism, but rather that his incapacitation makes it difficult or impossible to commit crimes. For further discussion, see Jennifer L. Doleac & Megan Stevenson, *Are Risk Assessment Scores Racist?*, Brookings Inst. (Aug. 22, 2016), https://www.brookings.edu/blog/up-front/2016/08/22/are-criminal-risk-assessment-scores-racist/.

90 Judges are likely to overcorrect and err on the side of a higher rate of false positives rather than bear the personal and societal risk of a recidivist committing a crime.

91 Interview with Jim Greiner, Professor, Harvard Law School, and Chris Griffin, Research Director, Harvard Law School's Access to Justice Lab (Nov. 7, 2016). Greiner and Griffin noted that in order for a risk assessment tool to "work," it has to know how success is defined and maximize toward that goal.

A criminal sentencing typically unfolds as follows: after a defendant has been convicted, the judge or sentencing authority requests a pre-sentence investigation report (PSI) with pertinent information about the defendant's life and background. This report is usually prepared by an officer of the court with a background in social work—not a lawyer—and may include information about a defendant's criminal record, details from interviews with the defendant's family, friends, and former employers, and other personal and biographical details. From a legal standpoint, there are few restrictions on what this pre-sentence investigation report may contain. Although strict rules govern what evidence can be introduced during the guilt phase of a trial, at sentencing a judge is free to consider a wide range of additional evidence without running afoul of a defendant's right to due process.[92] The rationale for the distinction is that sentencing is not just about the narrow issue of guilt, but is also informed by a defendant's life and characteristics. In our system, not every offense in a particular legal category calls for an identical punishment absent consideration of the past life, habits, and prior criminal record of a particular offender.[93]

Once the pre-sentence investigation report has been compiled, it is provided to the judge for review. Although the information in the PSI is generally made available to the defendant or his counsel as well, certain information or parts of the report the report may be considered confidential and kept from the defendant. The justification for this selective redaction is that the individuals speaking with the social worker compiling the report may wish to do so in confidence, especially if they fear reprisal from the defendant—and without a guarantee that the defendant will not be able to see that information, they might be hesitant or altogether unwilling to talk, thereby reducing the amount of information upon which a judge can base her sentencing decision.[94] Once the judge receives the PSI and any additional evidence presented at a sentencing hearing, she is free to use that information however she sees fit in making a final determination.

C. Evidence-Based Sentencing and the Embrace of Risk Scores

In recent years, as legal experts and legislatures have embraced the idea of evidence-based sentencing (EBS), they have aggressively encouraged judges to consider broader studies and risk assessments at sentencing. For example, the latest proposed revision of the sentencing sections of the Model Penal Code (MPC) explicitly endorses the use of risk assessment instruments in the shift to EBS.[95] The Conference of Chief Justices, the Conference of State Court Administrators, and the National Center for State Courts have also begun working together on a project to develop evidence-based sentencing practices.[96]

Like proponents of selective incapacitation in the 1980s, EBS advocates' goals goals are largely framed in progressive terms: to reduce incarceration and save money by identifying low-risk offenders who can be punished without going to jail.[97] Yet, like many of the state statutes discussed below, the current draft language in the MPC is relatively broad in its endorsement of

92 Williams v. New York, 337 U.S. 241, 251 (1949). Although *Williams* has been overruled in the death penalty context, *see* Gardner v. Florida, 430 U.S. 349 (1977) (imposing heightened evidentiary requirements for the punishment phase of a capital trial), the holding remains intact for other criminal cases.

93 *Williams*, 337 U.S. at 251-52.

94 *See Gardner*, 430 U.S. at 358-59 (noting that "an assurance of confidentiality to potential sources of information is essential to enable investigators to obtain relevant but sensitive disclosures from persons unwilling to comment publicly about a defendant's background or character. The availability of such information… provides the person who prepares the report with greater detail on which to base a sentencing recommendation and, in turn, provides the judge with a better basis for his sentencing decision."). In *Gardner*, the Supreme Court ruled that this confidentiality is unconstitutional in the capital sentencing context, but did not impose any such requirement on ordinary criminal trials.

95 According to the latest available draft, § 6B.09 of the revised MPC will endorse the use of "actuarial instruments or processes, supported by current and ongoing recidivism research, that will estimate the relative risks that individual offenders pose to public safety," including their formal incorporation into sentencing guidelines "[w]hen these instruments or processes prove sufficiently reliable."

96 Redding, *Evidence-Based Sentencing: The Science of Sentencing Policy and Practice*, *supra* note 44, at 7-8.

97 Starr, *Evidence-Based Sentencing*, *supra* note 88, at 816.

risk assessment tools. Only the advisory notes indicate any caution or need for "adequate protections" in order to ensure these tools are used fairly or highlight the importance of validity studies and other research to ensure accuracy.[98] This embrace of EBS is also at odds with other voices in the criminal justice system, including the Department of Justice, which has taken a more skeptical approach toward the use of algorithms in sentencing.[99] In 2014, the Department of Justice noted that "experience and analysis of current risk assessment tools demonstrate that utilizing such tools for determining prison sentences to be served will have a disparate and adverse impact on offenders from poor communities already struggling with many social ills."[100]

The statutory language that currently authorizes—and in some cases requires—the use of these tools varies widely across jurisdictions.[101] At least five states now require the use of risk assessments in criminal sentencing, but in different ways. Arizona, for example, specifically requires that the presentence reports in all probation-eligible cases "contain case information related to criminogenic risk and needs as documented by the standardized risk assessment and other file and collateral information."[102] Similarly, Oklahoma requires the use of an assessment and evaluation instrument designed to predict risk of recidivism to determine eligibility for any community punishment.[103] The Kentucky statute requires that pre-sentence investigation report must include a defendant's risk and needs assessment, and that sentencing judges must "consider" the results and "likely impact of a potential sentence on the reduction of the defendant's potential future criminal behavior."[104] The Ohio legislature took the approach of mandating that the Ohio Department of Rehabilitation and Correction "select a single validated risk assessment tool for adult offenders" that will be used for a variety of purposes that include sentencing.[105] Accordingly, the state uses the Ohio Risk Assessment System (ORAS), which it developed in partnership with the University of Cincinnati.[106] Similarly, Pennsylvania required that its sentencing commission adopt a risk assessment instrument to help determine appropriate sentences,[107] which resulted in a lengthy process undertaken by the state Sentencing Commission to develop its own custom tool and a series of guidelines for its use.[108] The extensive processes undertaken in Ohio and Pennsylvania in consultation with a wide range of experts and academics provide a stark contrast to those states which embraced these tools with just a few lines in a statute and have largely left it to individual judges to sort out.

A number of other states merely permit the use of risk assessments in criminal sentencing, acknowledging their potential to guide judicial decision-making and reduce mass incarceration. In Idaho, for example, *if* a court orders a presentence investigation, the report for all offenders sentenced to prison time and for certain offenders receiving probation must include information about current recidivism rates, differentiated based on whether the offender risk level is low, moderate, or high.[109] Louisiana similarly allows courts to use a presentence investigation validated risk and needs assessment tool prior to sentencing an adult offender who is eligible for assessment.[110] In Indiana, the state supreme court has recommended that evidence-based offender assessment instruments be used at criminal sentencing.[111] The West Virginia Supreme Court has indicated in an unpublished decision that although the legislature requires probation officers to conduct standardized risk and needs assessments,[112] the court retains dis-

98 *Id.*
99 DOJ Letter, *supra* note 37, at 7.
100 *Id.*
101 For a broader overview of how various risk assessment algorithms are used state-by-state, see *Algorithms in the Criminal Justice System*, *supra* note 72.
102 Ariz. Code of Judicial Admin. § 6-201.01(J)(3).
103 Okla. Stat. tit. 22, § 988.18(B).
104 Ky. Rev. Stat. Ann. § 532.007(3)(a)-(b).
105 Oh. Rev. Code Ann. § 5120.114(A)(1)-(3).
106 *Ohio Risk Assessment System*, Ohio Department of Rehabilitation and Correction, http://www.drc.ohio.gov/oras.
107 42 Pa. Cons. Stat. § 2154.7(a)
108 *See supra* note 72 and accompanying text.
109 *See* Idaho Code § 19-2517.
110 Louisiana Stat. Ann. § 15:326(A).
111 Malenchik v. State, 928 N.E.2d 564, 575 (Ind. 2010) (holding that "trial courts are encouraged to employ evidence-based offender assessment instruments… as supplemental considerations in crafting a penal program tailored to each individual defendant.").
112 *See* W. Va. Code § 62-12-6(a)(2).

cretion to decide how to use these tools to inform sentencing decisions.[113]

113 State v. Rogers, No. 14–0373, 2015 WL 869323, at *4 (W.Va. Jan. 9, 2015).

IV. Legal Issues Raised By Risk Assessments in Sentencing

When critics discuss these risk assessment tools, one of the first questions that comes up often centers on the legality of their use. In this section, we explore the primary legal issues raised by the use of risk assessments in sentencing. We begin with an analysis of the leading case on this issue from the Supreme Court of Wisconsin. We then discuss the broader constitutional issues implicated by these algorithms and related sentencing questions.

A. COMPAS Considered in Wisconsin: The *Loomis* Case

In the summer of 2016, the Supreme Court of Wisconsin considered the legality of using risk-assessment software in criminal sentencing.[114] *State v. Loomis* is one of the first major cases in the United States to address concerns about whether a judge's consideration of a software-generated risk assessment score during sentencing constitutes a violation of due process or overt discrimination.[115] The decision generated mixed reactions from both academics and the public for its endorsement of the use of risk assessment scores in sentencing despite clear hesitation on the part of all three judges in the panel about the potential for bias and other troubling implications of the use of these algorithms.[116]

Eric Loomis, the defendant in the case, was arrested for operating the vehicle during a drive-by shooting and pled guilty to lesser charges of fleeing the police and driving a stolen car.[117] After he pled guilty, the court requested a pre-sentence investigation report, which included among other information a risk score calculated using COMPAS. Loomis was designated by the COMPAS algorithm as high risk for all three types of recidivism measured by the program: pre-trial recidivism, general recidivism, and violent recidivism. The fact that he was a registered sex offender likely contributed to that score,[118] although the proprietary nature of the software makes it difficult to pinpoint exactly why he was designated high risk. Nonetheless, the state argued that the court should consider all three high-risk scores when determining the appropriate sentence.[119] Loomis received a six-year prison sentence, and at the hearing Judge Scott Horne told him: "The risk assessment tools that have been utilized suggest that you're extremely high risk to reoffend."[120]

Loomis challenged his sentence, arguing that the judge's use of the risk assessment score violated his right to due process—that is, his constitutional right to a fair trial. Specifically, he argued that it violated due process for three reasons: (1) it violated his right to be sentenced based on accurate information because the proprietary nature of the COMPAS software prevented him from assessing the accuracy of the score; (2) it violated his right to an individualized sentence because it relied on information about the characteristics of a larger group to make an inference about his personal likelihood to commit future crimes; and (3) it improperly used "gendered assessments" in calculating the score.[121] Ultimately, the court rejected Loomis's claims and held that COMPAS could be used at sentencing, although it made several recommendations about limiting COMPAS's use in fu-

114 State v. Loomis, 881 N.W.2d 749 (Wisc. 2016).

115 Joe Palazzolo, *Wisconsin Supreme Court to Rule on Predictive Algorithms Used in Sentencing*, Wall St. J. (Jun. 5, 2016), http://www.wsj.com/articles/wisconsin-supreme-court-to-rule-on-predictive-algorithms-used-in-sentencing-1465119008.

116 *See, e.g.*, Chiel, *Secret Algorithms That Predict Future Criminals Get a Thumbs Up from the Wisconsin Supreme Court*, *supra* note 1; Interview with the Sonja Starr, Professor, University of Michigan Law School (Oct. 28, 2016); Interview with Jim Greiner, Professor, Harvard Law School, and Chris Griffin, Research Director, Harvard Law School's Access to Justice Lab (Nov. 7, 2016).

117 Megan Garber, *When Algorithms Take the Stand*, The Atlantic (Jun. 30, 2016), http://www.theatlantic.com/technology/archive/2016/06/when-algorithms-take-the-stand/489566/.

118 *See* Mitch Smith, *In Wisconsin, A Backlash Against Using Data to Foretell Defendants' Futures*, N.Y. Times (Jun. 22, 2016), http://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html.

119 *Loomis*, 881 N.W.2d at 755.

120 Palazzolo, *Wisconsin Supreme Court to Rule on Predictive Algorithms Used in Sentencing*, *supra* note 115.

121 *Loomis*, 881 N.W.2d at 757.

ture cases.[122]

In response to the accuracy argument, the court acknowledged that the proprietary nature of COMPAS prevented Loomis from seeing exactly how his score was calculated. However, since most of the information the algorithm used came from a questionnaire that he completed and public records, the court concluded that he had an opportunity to ensure that the information was accurate.[123] "[T]o the extent that Loomis's risk assessment is based upon his answers to questions and publicly available data about his criminal history, Loomis had the opportunity to verify that the questions and answers listed on the COMPAS report were accurate."[124]

The court responded to his argument about his right to an individualized sentence by distinguishing a hypothetical case where the risk assessment score was either the *only* factor or the *determinative* factor in a sentencing decision from the present case, whereas here the risk score was simply one piece of information among many that the judge considered in the sentencing decision.[125] The court suggested that a due process challenge might succeed if the risk assessment score was the determinative or role factor the judge considered, but rejected Loomis's argument that considering it at all constituted a due process violation.[126] The court emphasized: "COMPAS has the potential to provide sentencing courts with more complete information to address [the] enhanced need [for more complete information up front]."[127] In support of this assertion, the court cited *Malenchik v. State*, a 2010 Indiana Supreme Court decision that looked at similar risk assessment tools and found that they help judges "more effectively evaluate and weigh several express statutory sentencing considerations such as criminal history, the likelihood of affirmative response to probation or short term imprisonment, and the character and attitudes indicating that a defendant is unlikely to commit another crime."[128]

Finally, the court considered Loomis's challenge to the use of gender as a variable that can change a defendant's risk score. This issue was complicated by the fact that the COMPAS algorithm is proprietary, and the parties in the case disputed the mechanics of how COMPAS takes gender into account.[129] Loomis argued that the algorithm considered gender as a criminogenic factor, whereas the state argued that it is used solely for "statistical norming," that is, to compare each offender to a "norming" group of his or her own gender. Nonetheless, Loomis objected to any use of gender in calculating the scores; the state, in response, argued that gender needs to be considered in a risk assessment to achieve statistical accuracy because men and women have different rates of recidivism and different rehabilitation potential.[130] The court, rejecting Loomis's argument, found that "if the inclusion of gender promotes accuracy, it serves the interests of institutions and defendants, rather than a discriminatory purpose."[131]

Loomis further argued that even if the statistical generalizations based on gender were accurate, they were unconstitutional. In support of this claim, he cited *Craig v. Boren*, a 1976 case where the Supreme Court held that an Oklahoma law that treated men and women differently was unconstitutional even though it was based on empirical data that supported the gender-based difference in the law.[132] The Supreme Court reasoned in *Craig* that "the principles embodied in the Equal Protection Clause [of the Fourteenth Amendment] are not to be rendered inapplicable by statistically measured but loose-fitting generalities concerning the… tendencies of aggregate groups."[133] Loomis, however, failed to raise his claim as an Equal Protection violation, as the court had found in *Craig v. Boren*, instead arguing that the use of gender violated his right to due process. But the Wisconsin court found that Loomis had not met the burden of proving that the court actually relied on gender as a factor in imposing his sentence, especially since the judge did not mention it in explaining his rationale.[134]

122 *Id.*
123 *Id.* at 761-62.
124 *Id.*
125 *Id.* at 765.
126 *Id.*
127 *Id.*
128 Malenchik v. State, 928 N.E.2d 564, 574 (Ind. 2010).
129 *Loomis*, 881 N.W.2d at 765.
130 *Id.*
131 *Id.* at 767.
132 Craig v. Boren, 429 U.S. 190, 208-10 (1976).
133 *Id.* at 208-09.
134 *Loomis*, 881 N.W.2d at 767 (noting that the judge

Having rejected all three of Loomis's due process claims, the Wisconsin court approved the use of COMPAS in this particular case, but it did express some hesitation about its future use absent clear limitations. The court first outlined permissible uses for the software, noting that while COMPAS cannot be determinative, the risk scores can be considered a relevant factor in several circumstances, including: (1) "diverting low-risk prison-bound offenders to a non-prison alternative," (2) assessing the public safety risk an offender poses and whether he can be safely and effectively supervised in the community rather than in prison, and (3) to inform decisions about the terms and conditions of probation and supervision.[135]

The court went on to prescribe key limitations for its use. While the risk score can help a judge better understand a defendant's unique situation and relevant factors, the court held that it should not be used to determine the length or severity of the punishment, and certainly should not be counted as an official aggravating or mitigating factor in a sentencing decision.[136] The court acknowledged that COMPAS was not designed with all of the goals of punishment in mind, but rather a focus on recidivism alone. Its lack of relevance to other important sentencing aims like retribution (which is a backward-looking assessment of an individual's blameworthiness) and deterrence (a broader concept that goes beyond the individual) makes it a "poor fit" for determining the length of the sentence.[137] In order to ensure that these limitations are being followed, the court mandated that a judge must explain at sentencing "the factors in addition to a COMPAS risk assessment that independently support the sentence imposed."[138]

The court also addressed the information that should be included in any pre-sentence investigation report containing a COMPAS score. This "written advisement of its limitations" should explain that:

1. COMPAS is a proprietary tool, which has prevented the disclosure of specific information about the weights of the factors or how risk scores are calculated;
2. COMPAS scores are based on group data, and therefore identify groups with characteristics that make them high-risk offenders, not particular high-risk individuals;
3. Several studies have suggested the COMPAS algorithm may be biased in how it classifies minority offenders;
4. COMPAS compares defendants to a national sample, but has not completed a cross-validation study for a Wisconsin population, and tools like this must be constantly monitored and updated for accuracy as populations change; and
5. COMPAS was not originally developed for use at sentencing.[139]

The concurring opinions reiterated the note of caution about relying on the COMPAS score in a meaningful way. Chief Judge Patience Drake Roggensack wrote separately to clarify that while a sentencing judge may *consider* a COMPAS score, he may not *rely* on it in making his sentencing decision.[140] Judge Shirley Abrahamson also wrote separately to emphasize that in considering COMPAS or other tools in sentencing, a judge "must set forth on the record a meaningful process of reasoning addressing the relevance, strengths, and weaknesses of the risk assessment tool" as a means to address concerns about their use.[141] She also noted that the lack of understanding about COMPAS and how it works was a "significant problem" in this case.[142]

The *Loomis* case was a landmark decision, since it was the first time a U.S. court evaluated these algorithms head on. The post-decision headlines made sweeping declarations like "Secret algorithms that predict future criminals get a thumbs up from Wisconsin Supreme Court"[143]

specifically referenced "your history, your history on supervision, and the risk assessment tools that have been utilized" when explaining to Loomis why he was at a high risk to reoffend).

135 *Id.* at 767-78.

136 *Id.* at 768.

137 *Id.* at 769.

138 *Id.*

139 *Id.* at 769-70.

140 *Id.* at 772 (Roggensack, C.J., concurring).

141 *Id.* at 774-75 (Abrahamson, J., concurring).

142 *Id.* at 774.

143 Chiel, *Secret Algorithms that Predict Future Criminals Get a Thumbs Up From Wisconsin Supreme Court*, *supra* note 1.

and "(Un)fairness of Risk Scores in Criminal Sentencing."[144] Yet legal experts noted that the court's analysis was more sophisticated and did not simply rubber stamp the use of programs like COMPAS without any safeguards whatsoever.[145] Moreover, its implications are limited, not only because it is binding only in the state of Wisconsin, but also because Loomis chose not to bring all possible claims challenging its constitutionality. First, he opted not to contest any of the socioeconomic variables used in COMPAS. And, as the opinion noted, while Loomis argued that the use of gender as a variable was problematic, he did not frame it as an Equal Protection violation, altering the court's analysis of the issue.[146]

Some critics have also pointed out flaws in the court's analysis of the issues that were before it. University of Michigan Law Professor Sonja Starr argues that the court erred in its analysis of the gender issue, and that, under existing constitutional doctrine, saying that the inclusion of gender makes the instrument more accurate is simply not enough to justify including it.[147] Moreover, by expressing concerns about the potential for unfairness and discrimination in using COMPAS but still approving it in this case, the court may ultimately fail to meaningfully restrict the use of the instrument. It is unclear, for example, how a judge might use a risk score if he cannot change the length of the sentence based on that number.[148] Nor does the opinion acknowledge that in jurisdictions like Wisconsin, where judges are elected, it is difficult to imagine that a "high risk" label will not result in a longer sentence.[149] Although it is not a given that elected judges will impose harsher sentences, when campaigning they may find it extraordinarily difficult to defend a decision to give a light sentence to a "high risk" offender, especially if that individual actually does commit future crimes.

The court arguably addressed the surface issues by adding caveats and mandating certain disclosures accompany COMPAS scores in presentence investigation reports, but it was silent on the underlying question of why the scores are being included in the report at all if they should not affect that length of the sentence.[150] As the highest court in Wisconsin, the judges deciding this case certainly had the authority to take a stronger stand and tell lower courts in the state not to consider these scores at all, but they did not.

As one of the first cases to meaningfully address the most recent incarnation of risk assessment scores, *Loomis* is significant but by no means determinative. The opinion demonstrated not only the challenges that the court faced in understanding how programs like COMPAS work, but also the fact that there is little helpful precedent to guide judges' decision-making when it comes to assessing their legality and crafting meaningful restrictions. And it is clearly not the end of the discussion. In the spring of 2017, the U.S. Supreme Court asked the federal government to weigh in on the question of whether it should hear Loomis' petition for a writ of certiorari, an indication of some interest in the issue—although the high court has yet to decide if it will allow the appeal to go forward.[151]

B. Constitutional Issues Implicated By Risk Assessment Algorithms

Broadly speaking, risk-assessment systems raise two primary constitutional concerns: their impact on an individual's right to due process, and the potential that the inclusion of certain variables constitutes an equal protection viola-

144 Danielle Citron, *(Un)Fairness of Risk Scores in Criminal Sentencing*, FORBES (Jul. 13, 2016), http://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-risk-scores-in-criminal-sentencing/#386c67d24479.

145 Lauren Kirchner, *Wisconsin Court: Warning Labels Are Needed for Scores Rating Defendants' Risk of Future Crime*, PROPUBLICA (Jul. 14, 2016), https://www.propublica.org/article/wisconsin-court-warning-labels-needed-scores-rating-risk-future-crime.

146 Starr describes why this argument fits into the Equal Protection Clause of the Fourteenth Amendment in a 2014 *Stanford Law Review* article. *See infra* Part IV.B.

147 Interview with the Sonja Starr, Professor, University of Michigan Law School (Oct. 28, 2016).

148 *Id.*; Interview with Jim Greiner, Professor, Harvard Law School, and Chris Griffin, Research Director, Harvard Law School's Access to Justice Lab (Nov. 7, 2016).

149 Interview with the Sonja Starr, Professor, University of Michigan Law School (Oct. 28, 2016).

150 *State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*, 130 HARV. L. REV. 1530 (Mar. 10, 2017).

151 Adam Liptak, *Sent to Prison by a Software Program's Secret Algorithms*, N.Y. TIMES (May 1, 2017), https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html?_r=2.

tion. It is worth noting that many of these same issues came up during the selective incapacitation movement in the 1980s, which raised concerns about both individualized sentencing and fairness.[152] Although the Due Process and Equal Protection claims are related, we treat them separately here in order to make the argument clearer.

i. Due Process Challenges: The Right to Review and Verify Sentencing Information

As explained above, the information that a judge considers at sentencing is not constrained by traditional evidentiary rules. Judges traditionally have discretion to consider a wide range of factors about the defendant's personal history, prior criminal record, and other details as part of the decision-making process; in *Williams v. New York*, the Supreme Court explained why such information, typically provided through a pre-sentence investigation report, might be useful to a sentencing judge and why it is not unfair to the defendant to rely on such information even if it not admissible during the guilt phase of the trial.[153] At the same time, however, the Supreme Court has subsequently recognized in *Gardner v. Florida* that the sentencing process itself must satisfy the requirements of the Due Process clause of the Fourteenth Amendment.[154] Although *Gardner* is a capital case—and therefore is subject to certain heightened restrictions compared to ordinary criminal sentencing cases—it raises the question about whether a defendant has a meaningful opportunity to refute, supplement, or explain the information upon which his sentencing decision is based.[155]

Because the use of risk assessment algorithms is so new, there have not been many legal challenges under the Due Process clause. The *Loomis* case, discussed above, challenged the use of COMPAS as a violation of the defendant's due process rights. But the Wisconsin Supreme Court held that Loomis's challenge did not clear the constitutional hurdles. Importantly, the case relies on two prior state court decisions: *State v. Samsa*, which considered the court's reliance on COMPAS scores provided in pre-sentence investigation reports (but did not address due process considerations),[156] and *State v. Skaff*, a 1989 decision which held that the right to be sentenced based on accurate information includes the right to review and verify information contained in the pre-sentence investigation report.[157] The crux of the court's reasoning in the *Loomis* decision was the fact that the COMPAS score cannot be the *only* thing the sentence is based on, or even the determinative factor, thereby arguably ensuring that the judge will consider other information about the particular case and assign an individual sentence based on the totality of the circumstances.[158] The court also reasoned that the right to review and verify information in the PSI was satisfied because the defendant could review and correct the public records upon which COMPAS relies, and the rest of the information was provided by the defendant himself in a questionnaire.[159]

Loomis's failure to succeed in his due process claim does not, of course, foreclose this line of argument in the future. It remains to be seen whether the limitations described by the court are actually sufficient to protect a defendant's right to an individual sentence. It is particularly difficult to assess whether the risk score was a determinative factor in the judge's decision-making process, which the court suggests *would* rise to the level of a due process viola-

152 Mathieson, *Selective Incapacitation: Reducing Crimes Through Predictions of Recidivism*, *supra* note 7.

153 *See* Williams v. New York, 337 U.S. 241, 251 (1949).

154 Gardner v. Florida, 430 U.S. 349, 359 (1977) (noting that "[t]he defendant has a legitimate interest in the character of the procedure which leads to the imposition of sentence even if he may have no right to object to a particular result of the sentencing process.").

155 *Id.*

156 State v. Samsa, 359 Wis.2d 580, 590 (Ct. App. Wisc. 2014) (rejecting a challenge to a sentence on the basis that "COMPAS is merely one tool available to a court at the time of sentencing and a court is free to rely on portions of the assessment while rejecting other portions.").

157 State v. Skaff, 152 Wis.2d 48, 57-58 (Ct. App. Wisc. 1989).

158 It is likely significant that the judge told Loomis at the sentencing hearing that the COMPAS score was one of multiple factors that he weighed when ruling out probation and assigning a six-year prison term: "In terms of weighing the various factors, I'm ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that you're extremely high risk to re-offend." State v. Loomis, 881 N.W.2d 749, 755 (Wisc. 2016).

159 *Id.* at 765.

tion,[160] but without identifying a specific test that could be used to help make that determination. Nor does the decision address the fact that there is a plausible distinction between being able to review and rebut the individual pieces of information that are fed into the algorithm and being able to actual review how the score itself was calculated.

ii. Equal Protection: Are We Embracing Explicit Discrimination Under Technocratic Framing?

In *Griffin v. Illinois*, a seminal 1956 case about the rights of indigent defendants, Justice Hugo Black wrote that "providing equal justice to poor and rich, weak and powerful alike" as "the central aim of our entire judicial system—all people charged with crime must, so far as the law is concerned, stand on an equality before the bar of justice in every American court."[161] The concept of individualism is at the heart of the Supreme Court's Equal Protection jurisprudence, which flows from the clause in the Fourteenth Amendment of the U.S. Constitution providing that no state shall "deny to any person within its jurisdiction the equal protection of the laws."[162]

In a 2014 speech to the National Association of Criminal Defense Lawyers, former Attorney General Eric Holder expressed serious concerns about the use of risk assessment software and its potential to undermine this central tenet of the criminal justice system. He told the audience:

> Although these measures were crafted with the best of intentions, I am concerned that they may inadvertently undermine our efforts to ensure individualized and equal justice. By basing sentencing decisions on static factors and immutable characteristics – like the defendant's education level, socioeconomic background, or neighborhood – they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.[163]

Holder went on to urge extreme caution when sentencing criminals not based on the facts of the crimes committed and the defendant's criminal history, but also on factors outside his or her control "or on the possibility of a future crime that has not taken place."[164] A few days prior to Holder's speech, the Department of Justice's Criminal Division had sent a letter to the Chair of the U.S. Sentencing Commission expressing similar concerns about the use of predictive analysis in criminal sentencing.[165] The letter noted that these risk assessment instruments "raise constitutional questions because of the use of group based characteristics and suspect classifications in the analytics."[166] Although the Equal Protection Clause does not require that the government treat every person exactly the same, it does prohibit discrimination if it is based upon impermissible classifications.

In a 2014 *Stanford Law Review* article published the same year, Professor Starr lays out a detailed strategy for challenging the use of these risk-assessment instruments under the Equal Protection Clause. Her basic thesis is that using risk assessment scores in criminal sentencing represents "an explicit embrace of otherwise-condemned discrimination, sanitized by scientific language."[167] By including variables like age and gender as well as socioeconomic factors like employment and education, these systems are enabling judges to consider factors that have long been considered inappropriate to bring into criminal sentencing. While we would object to the idea of judges systematically imposing harsher sentences on defendants who are poor or uneducated or from a certain demographic group, we are essentially sanc-

160 *Id.* (noting that "[i]f a COMPAS risk assessment were the determinative factor considered at sentencing this would raise due process challenges regarding whether a defendant received an individualized sentence.").
161 Griffin v. Illinois, 351 U.S. 12, 16-17 (1956) (quotation marks omitted).
162 U.S. CONST. amend. XIV, § 1.
163 *Attorney General Eric Holder Speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference*, U.S. DEPT. OF JUSTICE (Aug. 1, 2014), https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th.
164 *Id.*
165 DOJ Letter, *supra* note 37, at 4-8.
166 *Id.* at 7.
167 Starr, *Evidence-Based Sentencing*, *supra* note 88, at 803, 806.

tioning the practice by encouraging the use of risk systems that—despite their "technocratic framing"—take these variables into account. Moreover, Starr argues, these systems are unconstitutional, because the Supreme Court has consistently held that otherwise-impermissible discrimination cannot be justified by statistical generalizations about groups, such as a particular race or gender—even if those generalizations are, on average, accurate.[168] Our criminal justice system is premised on the idea that people have a right to be treated as individuals under the law.

iii. Race as a Variable

Virtually everyone agrees that race would be a constitutionally impermissible factor to include, and thus it is not included as an explicit variable in of any of these systems.[169] Explicit race-based classifications are subjected to the highest level of scrutiny by the courts, and when strict scrutiny applies it is virtually always fatal to the law or regulation being challenged. Thus if race was explicitly included as an input in the COMPAS algorithm, its use in sentencing criminal defendants would almost certainly constitute an Equal Protection violation.

However, excluding race itself does not necessary mean that factors that correlate heavily to an individual's race—serving essentially as proxies for race—are excluded from these algorithms.[170] Nor are factors that have disparate impact based on the race of the individual, such as a question that asks a criminal defendant the number of times he or she has been stopped by the police.[171] As data scientist Cathy O'Neil writes her book, *Weapons of Math Destruction*:

> [I]t's easy to imagine how inmates from a privileged background would answer one way and those from tough inner-city streets another. Ask a criminal who grew up in comfortable suburbs about "the first time you were ever involved with the police," and he might not have a single incident to report other than the one that brought him to prison. Young black males, by contrast, are likely to have been stopped by police dozens of times, even when they've done nothing wrong… So if early "involvement" with the police signals recidivism, poor people and racial minorities look far riskier.[172]

Unfortunately, while O'Neil and other critics correctly point out that using factors which correlate with race may be troubling, existing constitutional doctrine does not suggest that their inclusion in a risk assessment instrument would constitute an Equal Protection violation. The current standard for evaluating whether a facially neutral law (or in this case, the use of a facially neutral factor, like the number of reported contacts with the police) that has a racially disparate impact violates the Equal Protection Clause comes from *Washington v. Davis*.[173] *Washington v. Davis* held that while disproportionate impact on the members of a particular racial group is not irrelevant, strict scrutiny is only triggered if the individuals challenging the law can show that it was also adopted with a racially discriminatory intent. If not, rational basis review applies, a highly deferential standard. In the case of risk assessment algorithms, a criminal defendant challenging his or her sentence would have to be able to prove that the variable that correlated heavily to race was included *for the purpose* of racial discrimination, which is an extraordinarily difficult burden to meet.[174] Only a handful of cases in the forty years since *Washington v. Davis* was decided have successfully proven racially discriminatory intent, and they

168 *See, e.g.*, Craig v. Boren, 429 U.S. 190, 210 (1976), discussed *supra* note 132 and accompanying text.

169 Luis Daniel, *The Dangers of Evidence-Based Sentencing*, GovLab Blog (Oct. 31, 2014), http://thegovlab.org/the-dangers-of-evidence-based-sentencing/ (noting that "[o]verwhelmingly, states do not include race in the risk assessments since there seems to be a general consensus that doing so would be unconstitutional.").

170 Michal Kosinski et al., *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 Proceedings of the National Academy of Sciences of the United States of America, 5802, 5802-05 (2012) (finding that easily accessible digital records such as Facebook "likes" can be used to automatically and accurately predict highly sensitive personal information, including ethnicity).

171 *Id.*

172 Cathy O'Neil, Weapons of Math Destruction 25-26 (2016).

173 Washington v. Davis, 426 U.S. 229 (1976).

174 *See* Personnel Administrator v. Feeney, 442 U.S. 256 (1979) (holding that in order to find discriminatory intent, a state legislature has to have acted "because of," not "in spite of," the effects of a statute in relatively disadvantaging members of a particular minority group).

have mostly occurred in the jury selection context. It is therefore unlikely that a constitutional challenge against factors that correlate heavily to race will succeed under current doctrine.

A recent Supreme Court case, however, offers a sliver of hope. In early 2017, the Supreme Court ruled in *Buck v. Davis*, a case which addresses a related topic: the constitutionality of a death sentence in a case where an expert witness testified at sentencing that a black defendant was more likely to be dangerous in the future (which is an aggravating factor in the Texas death sentencing scheme) because of his race.[175] Although the procedural history of the case is complex—and the question before the Supreme Court focused on whether Buck's counsel gave him ineffective assistance in not objecting to the testimony—the issue of whether the racial nature of the expert witness's testimony tainted the sentencing decision remains at the heart of the case. As many experts predicted,[176] the Court ruled in Buck's favor in February, allowing him to appeal his death sentence.[177] Chief Justice Roberts, writing for the majority, openly acknowledged that "[a]s an initial matter, this is a disturbing departure from a basic premise of our criminal justice system: Our law punishes people for what they do, not who they are. Dispensing punishment on the basis of an immutable characteristic flatly contravenes this guiding principle."[178] Although the case does not overrule any precedent that relates directly to claims of racially discriminatory impact, certain language in the opinion suggests that the Supreme Court might be uncomfortable with sentences that are clearly based on unchangeable characteristics like race, which some risk assessment tools arguably do.

iv. Gender as a Variable

In contrast to race, systems like COMPAS and LSI-R do take gender into account, despite the fact that gender classifications are subject to an intermediate level of scrutiny[179] that requires an "exceedingly persuasive justification" to hold up under the Equal Protection Clause.[180] In the few instances where the issue has been raised directly, courts have generally held that it is impermissible to base sentences on gender,[181] which does not bode well for risk-assessment systems that produce different results based on the gender of the defendant.[182] The issue is by no means settled, so it is entirely possible that in a future challenge the courts would find that the gender classification in risk assessment algorithms constitutes a constitutional violation—especially because, as mentioned above, the defendant in *State v. Loomis* failed to raise his gender discrimination claim under the Equal Protection Clause, and as such the court did not have to address it.

175 Amy Howe, *Argument Preview: Justices to Consider Role of Racial Bias in Death Penalty Case*, SCOTUS Blog (Sept. 28, 2016), http://www.scotusblog.com/2016/09/argument-preview-justices-to-consider-role-of-racial-bias-in-death-penalty-case/; Nina Totenberg, *Supreme Court To Hear Death Penalty Case Based On Racially Tainted Testimony*, NPR (Oct. 5, 2016), http://www.npr.org/2016/10/05/496630474/supreme-court-to-hear-death-penalty-case-based-on-racially-tainted-testimony.

176 *See, e.g.*, Amy Howe, *Argument Analysis: Justices Appear Inclined to Rule in Favor of Texas Death Row Inmate in Racial Bias Case*, SCOTUS Blog (Oct. 5, 2016), http://www.scotusblog.com/2016/10/argument-analysis-justices-appear-inclined-to-rule-in-favor-of-texas-death-row-inmate-in-racial-bias-case/.

177 Buck v. Davis, 580 U.S. ___ (2017).

178 *Id.* at 21.

179 Intermediate scrutiny falls in between the heavily burdensome strict scrutiny that applies to race-based classifications and almost always results in invalidation, and the highly deferential rational basis review, under which very few laws are declared unconstitutional. Under this standard of review, the burden falls on the government to prove that a classification is substantially related to the achievement of an important government purpose, Craig v. Boren, 429 U.S. 190 (1976), which the court later suggested required an "exceedingly persuasive justification," United States v. Virginia, 518 U.S. 515, 531 (1996).

180 United States v. Virginia, 518 U.S. at 531 (noting that "equal protection principles, as applied to gender classifications, mean state actors may not rely on "overbroad" generalizations to make "judgments about people that are likely to … perpetuate historical patterns of discrimination").

181 *See, e.g.*, United States v. Maples, 501 F.2d 985, 989 (4th Cir. 1974); Williams v. Currie, 103 F. Supp. 2d 858, 868 (M.D.N.C. 2000). *See also* Carissa Byrne Hessick, *Race and Gender as Explicit Sentencing Factors*, 14 J. Gender Race & Just. 127, 137 (2010).

182 It is worth noting that empirical research suggests that female defendants on average receive more lenient treatment than male defendants, but judges and prosecutors do not explicitly endorse such differential treatment. *See* Sonja B. Starr, *Estimating Gender Disparities in Federal Criminal Cases*, Univ. of Mich. L. Sch. & Econ. Research Paper Series, Paper No. 12-018, 3-4, 17 (2012), *available at* http://ssrn.com/abstract=2144002.

In *Craig v. Boren*, moreover, the U.S. Supreme Court rejected a defense of a gender classification that was grounded in statistical generalizations about women, even if those generalizations were empirically supported.[183] The challenged Oklahoma statute allowed women to buy certain types beer once they reached the age of 18 but prohibited men from buying it until they were over 21 because statistical evidence suggested that men between the ages of 18 and 21 were over ten times more likely than their female peers to drive drunk.[184] There is a plausible case, therefore, to argue that the incorporation of gender into risk assessment calculations is unconstitutional, even if the government argues that it has a substantial interest in the inclusion of gender because it improves the accuracy of the algorithm. This important distinction highlights an underlying tension between the legal and technical approaches to these issues. In the world of machine learning, accuracy is valued above all else, whereas our legal system tends to place a greater emphasis on the principle of fairness—even if it requires eschewing empirical results.[185]

v. Socioeconomic Status as a Variable

The use of socioeconomic variables might also qualify as an impermissible wealth classification, although the argument is not quite as clear-cut as those that apply to race or gender. A number of these risk assessment instruments incorporate data about a defendant's employment status, income, education, and job skills. Despite indicating in *Griffin v. Illinois* that the Supreme Court might subject wealth-related classifications to the strict scrutiny that applies to discrimination based on race and national origin,[186] the Supreme Court later held that poverty is not inherently suspect.[187] Even so, there is ample case law that recognizes that we should not place special burdens on indigent defendants, especially in the sentencing context.

In *Bearden v. Georgia*, the Supreme Court rejected the argument that a defendant's poverty could be considered a factor that increased his likelihood of recidivism and therefore justified additional incapacitation.[188] The court held that a sentence increase cannot not be based on "lumping [the defendant] together with other poor persons and thereby classifying him as dangerous. It would be little more than punishing a person for his poverty."[189] Although financial background is not considered completely *irrelevant* to sentencing—judges have traditionally been allowed to consider financial history and employment background at sentencing—there is a distinction when those factors are used to trigger "extra, unequal punishment" for poor defendants.[190] There is also general support for the idea that lower socioeconomic status should not be considered an aggravating factor justifying a higher sentence.[191]

Thus, while the argument that using factors related to socioeconomic status is unconstitutional is not frivolous, it is by no means a clear-cut one. Most of the relevant precedent involves situations where an individual cannot pay for something (such as bail or court fees) because of his poverty and is therefore subjected to greater punishment as a result—which is distinct from independently using socioeconomic status as a dynamic factor in a risk assessment. For the Supreme Court to invalidate a factor like employment status or income based on this line of cases would certainly be a novel—albeit not unprecedented—application.

C. Related Sentencing Issues: Managing Risk in the Criminal Justice System

While we wait for the constitutional challenges to unfold, however, there are other, more immediate legal and policy reasons to scrutinize these systems and the factors upon which they rely. In general, the rush by state legislatures

183 Craig v. Boren, 429 U.S. 190, 210 (1976).

184 *Id.*

185 We owe this point to Ben Green, a fellow at the Berkman Klein Center for Internet & Society and a PhD candidate in Applied Mathematics at Harvard University. We discuss this tension in greater depth in Part V.C.

186 Griffin v. Illinois, 351 U.S. 12 (1956).

187 *See, e.g.*, Maher v. Roe, 432 U.S. 464, 471 (1977) (noting that the Court has not held that "financial need alone identifies a suspect class for purposes of equal protection analysis").

188 Bearden v. Georgia, 461 U.S. 660, 661-62 (1983).

189 *Id.* at 671.

190 Starr, *Evidence-Based Sentencing*, *supra* note 88, at 831-32.

191 The Federal Sentencing Guidelines, for example, forbid consideration of socioeconomic status.

and the scholars revising the MPC to embrace the idea of evidence-based sentencing begs the question of whether managing risk should be so heavily emphasized among the multiple purposes of criminal sentencing.[192] Although it has long been clear that managing risk is a part of the sentencing consideration, the use of these algorithms almost certainly increases the prominence of risk assessments in the decision-making process. Yet judges might not be the best or most appropriate actors to try to manage these risks. Nor is there a significant body of evidence at this point that suggests we are actually good at predicting or managing risk—or that longer sentences, for example, might decrease the risk of recidivism.[193]

Moreover, these algorithms likely do not consider the fact that many of the factors that increase a risk score might also be considered mitigating evidence. A young, poor, or uneducated defendant might be at a higher risk for recidivism, but those same circumstances might also diminish his culpability and justify a more lenient sentence, rather than a harsher one. The Supreme Court confronted this very issue in *Penry v. Lynaugh*, a capital case where the Court called the defendant's intellectual disabilities a "two-edged sword."[194] Because the defendant's mental handicap prevented him from learning from his mistakes, it arguably increased his future dangerousness and could be considered an aggravating factor.[195] At the same time, it was also a mitigating factor because it reduced his blameworthiness for the crime he committed.[196] *Penry* highlights an inherent tension in the justice system between competing concerns for public safety and individual liberty—concerns that are equally implicated by the rise of risk assessment instruments used in sentencing.

192 DOJ Letter, *supra* note 37, at 8 (emphasizing that "[d]etermining imprisonment terms should be primarily about accountability for past criminal behavior. While any effective sentencing and corrections policy will take account of future behavior to some extent—incapacitating those more likely to recidivate and utilizing effective reentry efforts to reduce the likelihood of recidivism—we believe the length of imprisonment terms should mostly be about accounting for past conduct. As analytics evolve, we are concerned about the implications of sentencing policy moving away from this precept.").

193 *See* Gendreau et al, *The Effects of Prison Sentences on Recidivism*, *supra* note 86.

194 Penry v. Lynaugh, 492 U.S. 302, 324 (1989).

195 *Id.* at 323.

196 *Id.* at 324.

V. Challenges Presented by the Use of Risk Assessment Algorithms in Sentencing

Drawing on the analysis of the history of risk assessments and the legal and ethical concerns that they raise, this section attempts to summarize key concerns related to the use of these tools. In particular, we focus on three issues: opacity, bias and unreliability, and diverging concepts of fairness.

A. Opacity

In her concurring opinion in the *Loomis* case, Wisconsin Judge Shirley Abrahamson lamented that "this court's lack of understanding of COMPAS was a significant problem in the instant case. At oral argument, the court repeatedly questioned both the State's and defendant's counsel about how COMPAS works. Few answers were available."[197] Abrahamson's concurrence highlights one of the critical challenges identified by both legal and technical experts: the lack of transparency about how these tools work.[198] Although the details vary widely among the different systems, the broad concerns relate to: (1) the inputs themselves, (2) how those inputs are weighted by the algorithm, and (3) whether specific factors (or combinations of factors) may end up serving as proxies for problematic or impermissible variables like race and poverty. These challenges can be compounded by a lack of information about the underlying assumptions made by the computer scientists developing the algorithms or conflicting purposes when a tool is developed for one context, such as pre-trial risk assessment, and then adapted for another like sentencing.

The challenges presented by this opacity are two-fold. First, they make it difficult for researchers and outside experts to evaluate and audit the algorithms in order to test for accuracy and bias.[199] The lack of information about how inputs are weighted also makes it harder to bring legal challenges to the use of these tools, since criminal defendants cannot say for sure whether or how suspect factors like gender or racial proxies may have influenced the risk assessment score or the judge's ultimate sentencing decision.[200] In the *Loomis* case, for example, the court dismissed the gender claim because the sentencing judge did not mention it specifically when explaining his decision[201]—a distinction which seems to ignore the fact that a judge may never explicitly mention a factor like gender when it is quietly incorporated into an opaque risk score rather than considered openly in the pre-sentence investigation report or at a hearing.

It is also worth noting the distinction here between algorithms developed by for-profit companies like Northpointe and Multi-Health Systems and those created by or in conjunction with non-profits, researchers, and academics, like Public Safety Assessment and the state of Pennsylvania's risk assessment algorithm. While all of these tools may look like "black boxes" to outsiders and are susceptible to concerns about opacity, the proprietary tools developed by for profit companies present unique challenges. Those companies have both a greater interest in shrouding their products in secrecy in order to remain competitive and more legal tools at their disposal to keep their algorithms away from public scrutiny.[202] Academic researchers and governments, by contrast, tend to have more incentives to make the details of their algorithms publicly available and ensure that they are subject to appropriate scrutiny and oversight.

B. Bias and Lack of Reliability

In May 2016, ProPublica released an in-depth report about COMPAS suggesting that it was both racially biased and inaccurate.[203] According to

197 State v. Loomis, 881 N.W.2d 749, 774 (Wisc. 2016).

198 *See, e.g.*, Nicholas Diakopoulos, *We Need to Know the Algorithms the Government Uses to Make Important Decisions About Us*, The Conversation (May 23, 2016), https://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869.

199 George Joseph, *Justice By Algorithm*, CityLab (Dec. 8, 2016), http://www.citylab.com/crime/2016/12/justice-by-algorithm/505514/.

200 Interview with the Sonja Starr, Professor, University of Michigan Law School (Oct. 28, 2016).

201 *Loomis*, 881 N.W.2d at 767.

202 Companies like Northpointe can argue that the details of their algorithms constitute trade secrets that shield them from disclosure. O'Neil, Weapons of Math Destruction, *supra* note 172, at 29.

203 Angwin et al., *Machine Bias: There's Software Used*

ProPublica's analysis, the scores not only proved "remarkably unreliable" in forecasting violent crime, but they also contained significant racial disparities—even though the formula does not officially take race into account. COMPAS incorrectly labeled black defendants as more likely to commit crimes again than they actually were, while also frequently mislabeling white defendants as low risk.[204] The study was cited by the court in the *Loomis* case in the discussion of the controversy surrounding these tools, even though it did not ultimately factor into the court's analysis in the case.[205] Although the findings of the study have been disputed by Northpointe,[206] the research nonetheless highlights growing discomfort among members of the legal and academic communities that these tools, which have been embraced for ostensibly progressive reasons like reducing mass incarceration, may inadvertently reinforce or even exacerbate existing racial disparities.[207] As a group of computer science researchers wrote in the *Washington Post* in response to the debate between ProPublica and Northpointe: "Algorithms have the potential to dramatically improve the efficiency and equity of consequential decisions, but their use also prompts complex ethical and scientific questions.... We must continue to investigate and debate these issues as algorithms play an increasingly prominent role in the criminal justice system."[208]

The risk of bias may be compounded by algorithms that rely on other potentially biased data sets, such as those that are used for predictive policing.[209] The interaction between these algorithms is one of the central concerns expressed by O'Neil in *Weapons of Math Destruction*. O'Neil argues that that police essentially respond to two types of crimes: (1) crimes that are "reported," which usually refers to violent crimes (such as assault, homicide, and rape) and property crimes, and (2) crimes that are "found," such as when individuals are stopped and found to possess a small quantity of drugs or be engaged in otherwise illegal activity. Because of historic policing patterns—many of which are reinforced by new predictive tools—predominantly poor and minority neighborhoods tend to face a disproportionate amount of police activity with respect to "found" crimes.[210] Consequently, O'Neil argues, the data sets concerning "found" crimes are likely biased to suggest that poor and minority communities commit a higher proportion of these crimes than they actually do.[211] If that information is then incorporated into a recidivism risk calculation, it might falsely suggest that a poor or minority defendant is at a greater risk to commit future crimes and therefore assign that individual a higher risk score.

Of course, we should not pretend that inadvertent (and potentially overt) bias has not always played a role in judge's sentencing decisions.

Across the Country to Predict Future Criminals. And It's Biased Against Blacks., *supra* note 2.

204 *Id.* The study found that black defendants were almost twice as likely as white defendants to be labeled a higher risk but not actually reoffend, whereas white defendants were much more likely to be labeled lower risk but ultimately commit other crimes.

205 *Loomis*, 881 N.W.2d at 749 n. 2.

206 William Dieterich et al., *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, NORTHPOINTE (Jul. 8, 2016), http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf (explaining that "[b]ased on our examination of the work of Angwin et al. and on results of our analysis of their data, we strongly reject the conclusion that the COMPAS risk scales are racially biased against blacks.").

207 *See, e.g.*, Chris Griffin, *Fear and Loathing Over Risk Assessments, Part 2*, HARVARD LAW SCHOOL'S ACCESS TO JUSTICE LAB (Oct. 14, 2016), http://a2jlab.org/fear-and-loathing-over-risk-assessments-part-2/ (noting that the ProPublica piece "focuses on the troubling implications of racial imbalances in scores and predictive accuracy.").

208 Sam Corbett-Davies et al., *A Computer Program Used for Bail and Sentencing Decisions was Labeled as Biased Against Blacks. It's Actually Not that Clear.*, THE WASHINGTON POST (Oct. 17, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/

209 *See, e.g.*, Jack Smith IV, *Crime Prediction Tool PredPol Amplifies Racially Biased Policing, Study Shows*, MIC (Oct. 9, 2016), https://mic.com/articles/156286/crime-prediction-tool-pred-pol-only-amplifies-racially-biased-policing-study-shows#.Xp0PSJZA1.

210 Jacob Metcalf, *Ethics Review for Pernicious Feedback Loops: Reading Weapons of Math Destruction*, DATA & SOCIETY INST. (Nov. 7, 2016), https://points.datasociety.net/ethics-review-for-pernicious-feedback-loops-9a7ede4b610e#.3pfok1602.

211 O'NEIL, WEAPONS OF MATH DESTRUCTION, *supra* note 172, at 26-29; Interview with Cathy O'Neil, Author, *Weapons of Math Destruction* (Oct. 25, 2016).

A recent study found that judges in Florida, for example, sentence black defendants to 68 percent more time in prison for serious first-degree crimes even when they score the same as their white counterparts on the formula used to determine sentences.[212] But the fact that bias exists in the current system does not justify reinforcing—or even institutionalizing—bias by using risk assessment tools.[213]

Moreover, bias aside, many of these algorithms have not been evaluated for their accuracy in the specific contexts or geographic areas in which they are being deployed. According to the Electronic Privacy Information Center (EPIC), which has compiled an overview of state-by-state adoption of risk assessment algorithms, although some states have conducted validity studies that how well these algorithms perform with respect to their specific populations, many have yet to do so.[214] Indeed, the Wisconsin Supreme Court noted this summer that the state had not conducted a cross-validity study regarding COMPAS's accuracy, and recommended that the tools be constantly monitored and updated.[215] In states where validity studies *have* been conducted, it is similarly unclear whether any of those studies have or will be repeated regularly in order to ensure ongoing accuracy as the population changes.

C. Diverging Concepts of Fairness

To argue that risk assessment algorithms should be crafted fairly is uncontroversial, but the precise definition of "fairness" is hard to nail down. Whether an algorithm is *technically* fair is heavily reliant upon the precise objectives of that algorithm, and it raises a number of normative considerations. One might argue that an algorithm is technically fair as long as it makes accurate and consistent predictions. But in addition to the fact that the academic community has still not reached a consensus on an exact definition for fairness in the statistical context,[216] Dr. Jeremy Kun points out that an algorithm's training data may itself be flawed, indicating that the inputs themselves may not be "trustworthy."[217] Even if a perfectly accurate algorithm does exist, the fairness-as-accuracy definition might still come up short in the event that an algorithm leads to generalizations about particular groups. Consider an accurate algorithm that comes to the blanket conclusion that men tend to deserve higher risk scores than women. Whether or not the algorithm is accurate, would it be fair for individuals to be judged based on immutable characteristics such as gender? Such a circumstance is reminiscent of the one in the aforementioned case of *Craig v. Boren*,[218] which grappled with the tension between statistical generalizations and empirical validity. To avoid the possible unfairness that comes from pure "accuracy," one could argue instead that an algorithm is only fair if its outcomes favor no particular group. While this suggestion sounds reasonable on its face, it can lead to its own complex set of questions, such as whether this could inadvertently lead to reverse discrimination.

The legal concept of fairness is also nebulous, but in a different way. Legal fairness encompasses the idea that every individual is entitled to certain procedural rights designed to give them a "fair" shot in the justice system. In *McCleskey v. Kemp*, for example, the Supreme Court considered this issue when a death row prisoner argued that his sentence was unconstitutional because the process through which he was convicted was administered in a racially discriminatory manner.[219] The Court articulated its concept of fairness not necessarily in terms of reaching the correct outcome, but rather

212 Josh Salman et al., *Florida's Broken Sentencing System: Designed for Fairness*, HERALD TRIBUNE, http://projects.heraldtribune.com/bias/sentencing/. The *Herald Tribune* reviewed millions of records in the state of Florida and found that, across the board, "[w]hen defendants score the same points in the formula used to set criminal punishments — indicating they should receive equal sentences — blacks spend far longer behind bars" compared to white defendants.

213 *See, e.g.*, Starr, *Evidence-Based Sentencing*, *supra* note 88, at 806 (explaining that "[t]he technocratic framing of [evidence-based sentencing] should not obscure an inescapable truth: sentencing based on such instruments amounts to overt discrimination based on demographics and socioeconomic status.")

214 *Algorithms in the Criminal Justice System*, *supra* note 72.

215 State v. Loomis, 881 N.W.2d 749, 769-70 (Wisc. 2016).

216 Jeremy Kun, *One Definition of Algorithmic Fairness: Statistical Parity*, MATH ∩ PROGRAMMING (Oct. 19, 2015), https://jeremykun.com/2015/10/19/one-definition-of-algorithmic-fairness-statistical-parity/.

217 *See id.*

218 Craig v. Boren, 429 U.S. 190, 208-10 (1976).

219 McCleskey v. Kemp, 481 U.S. 279 (1987).

reaching a (hopefully) correct outcome through a process that gave the individual a fair opportunity and guaranteed his or her rights to due process. In the opinion, Justice Powell explained that that "our consistent rule has been that constitutional guarantees are met when 'the mode [for determining guilt or punishment] itself has been surrounded with safeguards to make it as fair as possible.'"[220] In other words, legal fairness tends to prioritize parity in the process by which an outcome is reached rather than the outcome itself.

220 *Id.* at 313 (quoting Singer v. United States, 380 U.S. 24, 35 (1965)).

VI. Recommendations for the Use of Risk Assessment Algorithms

Despite the concerns described above, we assume that risk assessment tools will continue to be used in the criminal justice system, including at sentencing, in light of both their widespread embrace in the United States and the potential benefits they offer if correctly implemented. Nonetheless, given the myriad challenges, we believe that policymakers should proceed cautiously and deliberately in implementing these systems. The goal of this section is to identify overall concepts to guide policymakers that ensure transparency, accountability, and fairness are given central. While these recommendations are not comprehensive, we believe that they represent a valuable starting point for conversations around the use of these tools.

A. Transparency

One of the central themes emphasized by both legal and technical experts is the need for greater transparency about how these algorithms were developed, the assumptions that were made in their design, how their factors are weighted, and how frequently they are assessed and updated. While transparency alone will not necessarily reduce the likelihood of bias, it remains valuable for a number of reasons. First and foremost, greater transparency can help facilitate audits by outside researchers.[221] It can also help increase the general understanding of these systems, how they work, and the tradeoffs involved in implementing them. More information about inputs and the weights of variables is also critical for any future constitutional challenges based on the use of impermissible or potentially impermissible factors.[222]

Law Professor Danielle Keats Citron and others have also developed and advocated for a concept known as technological due process—which aims to ensure that there is ample opportunity to challenge the decisions made by algorithms—that can be instructive in this context as well. The core values of the technological due process concept are transparency, accuracy, accountability, participation, and fairness.[223] Citron and Frank Pasquale call for increased federal regulatory oversight over scoring systems that collect data about individuals, generate scores from that data, distribute scores to decision makers, and use those scores in decision making.[224] They argue that individuals should have the "right to inspect, correct, and dispute inaccurate data and to know the sources (furnishers) of the data."[225] Furthermore, they believe that the algorithm that generates a score from said data needs to be public so that each process can be inspected. Finally, they emphasize that policymakers need to ensure that a score is fair, accurate, and replicable.[226]

The key mechanism behind technological due process is the requirement of audit trails that record correlations between rules and decisions made in algorithms. The audit trail would include a map of the facts and rules that were applied to each decision made in an algorithm.[227] Vendors should also make the source code for the algorithms available to the public, which will enable outsiders to test these algorithms, a standard practice among software developers.[228] Testing can detect patterns of problematic classifications based on race, nationality, sexual orientation, and gender.[229] By making the data public, academics will also be able to comment on the scoring systems and help ensure that they are infused with public values rather than

221 *See, e.g.*, Diakopoulos, *We Need to Know The Algorithms The Government Uses to Make Important Decisions About Us*, *supra* note 198.

222 *See* Starr, *Evidence-Based Sentencing*, *supra* note 88.

223 Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 90 Wash. U. L. Rev. 1, 20 (2014).

224 *Id.*

225 *Id.*

226 *Id.* at 22.

227 Danielle Keats Citron, *Technological Due Process*, 85 Wash. U. L. Rev. 1249, 1254 (2008).

228 Richard Berk, a statisticiation from the University of Pennsylvania who played a central role in the development of Pennsylvania's risk assessment program, also argues that all companies should be required to be disclose the complete contents of their algorithms. At the very least, Berk believes some government entity should be created or tasked with evaluating the full contents of risk-assessment algorithms, even if they are proprietary like COMPAS. Interview with the Richard Berk, Professor, University of Pennsylvania (Oct. 31, 2016).

229 Citron & Pasquale, *The Scored Society: Due Process for Automated Predictions*, *supra* note 223, at 25.

dictated solely by the whims of the programmer.[230] Although some opponents of disclosure have argued that it will threaten to innovation and or make it easier for participants to "game the system," these concerns can be mitigated by the fact that there is inadequate evidence of such behavior in other instances.[231] While full public disclosure would be ideal, policymakers can work with industry on a case-by-case basis to determine if more limited forms of disclosure would be more appropriate.[232]

Transparency should also inform a government's decisions about whether to use proprietary risk assessment software or work with academics or non-profits to develop tools specifically for a particular jurisdiction.[233] As noted earlier, proprietary tools like COMPAS are inherently subject to less scrutiny and oversight than their public counterparts might be. As a group of computer scientist researchers candidly explained, "Northpointe has refused to disclose the details of its proprietary algorithm, making it impossible to fully assess the extent to which it may be unfair, however inadvertently. That's understandable: Northpointe needs to protect its bottom line. But it raises questions about relying on for-profit companies to develop risk assessment tools."[234] The tension between the legitimate business interests of a private company that wants to protect and sell its product and the need for public accountability may not be easy to resolve.[235]

230 *Id.* at 26.

231 *Id.*

232 For example, one potential compromise is to provide limited public transparency but full disclosure to government agencies.

233 Professor Berk argues, for example, that the goals of a company with proprietary software may be fundamentally incompatible with these transparency requirements. Interview with the Richard Berk, Professor, University of Pennsylvania (Oct. 31, 2016).

234 Corbett-Davies et al., *A computer program used for bail and sentencing decisions was labeled as biased against blacks. It's actually not that clear.*, *supra* note 208.

235 Some experts, like Berk, suggest that the financial goals of private companies and the fairness requirements of the criminal justice system may ultimately be mutually exclusive.

B. Accountability and Oversight

While transparency is a foundational step, it is just the beginning. In order to promote maximum accountability, policymakers need to ensure that the systems they deploy have been designed for the purpose for which they are being used, that they are appropriate for the particular jurisdiction or geographic area, and that they are continually monitored and assessed for accuracy and reliability.[236] Any tools they adopt should be built with integrity, based on the best available science, and calibrated to minimize potential negative effects, such as the inclusion of problematic variables.

The need to conduct validity studies on a state-by-state—or potentially even more granular—level is clear.[237] A tool that has been tested on the national population or in other states may not be appropriate for a particular location. Local policymakers should possess that information *before* deciding to implement any risk assessment system, which requires validity studies and other research as a prerequisite to making any decisions. Moreover, testing and validity studies should not simply be completed once and then forgotten about. States should require regular repetition of validity studies and develop procedures to make appropriate alterations based on any changes in the population or new information that emerges about these tools. Policymakers should also talk to their peers in other jurisdictions to share best practices and look for opportunities for standardization among jurisdictions, so that an individual's protection against biased or unreliable algorithms is less dependent on what jurisdiction he or she happens to be in.

In addition to validity studies, facilitating outside research and auditing is also critical. Greater transparency will have little impact if outside researchers do not have access to the data and tools to evaluate and test the algorithms for bias. These tools should also be rigorously evaluated in comparison to existing mechanisms in the justice system to ensure that they actually repre-

236 Mark Ackerman, *Safety Checklists for Sociotechnical Design*, Data & Society Inst. (Oct. 26, 2016), https://points.datasociety.net/safety-checklists-for-sociotechnical-design-2cb9192e9e3b#.iipwibu0o.

237 *Algorithms in the Criminal Justice System*, *supra* note 72.

sent an improvement over the status quo.[238] The Access to Justice Lab at Harvard Law School, for example, recently started a project to evaluate the efficacy of risk assessment scores in pre-trial assessment, using randomized control trials in several jurisdictions around the United States.[239] Research projects like can offer critical insights into how these tools work, how judges actually use them, and how they might be deployed on a large scale in the best and most appropriate manner.[240] It may make sense to initiate similar efforts on a wider scale.

Finally, where risk assessment algorithms are concerned, maintaining oversight of implementation and ongoing use of these tools should not be a hands-off process.[241] Policymakers should be involved at all stages of the process, asking difficult questions and forcing their partners—whether they are for-profit companies, academic institutions, or non-profit organizations—to explain and justify any assumptions or decisions that they make in developing and using these tools. Especially in light of the serious constitutional concerns raised by scholars like Professor Starr, governments should not simply "outsource" risk assessments to private companies and assume that they will help guide judicial decision-making in a way that is both accurate and fair to the individual defendant. Rather, policymakers should maintain an active role in overseeing their use and ensuring that both the developers and those individuals employing them are aligned with the overall goals of the system and aware of any potential pitfalls.

C. Robust and Holistic Approach to Fairness

Based on the diverging concepts of technical and legal fairness described above, policymakers need to engage in a thorough dialogue about how to reconcile or prioritize these competing values. Absent such a conversation, it will be difficult to resolve disputes like the one between ProPublica and Northpointe about whether COMPAS is biased.[242] Legal scholars and technical experts need to engage with one another about the appropriate technical and legal measures that should be in place to guarantee that the algorithms do not inappropriately prioritize one type of fairness over another.

Critical issues also need to be addressed in the development phase of these algorithms, particularly with regard to the inputs and how they are used.[243] O'Neil, for example, argues that these risk assessment algorithms should eliminate as many unnecessary variables as possible, especially those that are potential proxies for race or rely on historically-biased data sets (such as the "found" crimes described in the previous section).[244] Indeed, some researchers have found that it is possible to duplicate the results of a system like COMPAS using far fewer variables—and far fewer problematic variables, at that.[245] This research suggests that some of the most problematic variables could be removed from these systems without sacrificing accuracy, although more studies are clearly required. O'Neil takes an even more controversial position: that troubling variables should be excluded *even if* their exclusion decreases the accuracy of the algorithm. "Are we going to sacrifice the accuracy of the model for fairness? Do we have to dumb down our algorithms?" she writes. "In some cases, yes. If we're going to be equal before the law, or be treated equally as voters, we cannot stand for systems that drop us into different castes and treat us differently."[246]

238 O'Neil, Weapons of Math Destruction, *supra* note 172, at 208.

239 *See Pre-Trial Release*, Access to Justice Lab at Harvard Law School, http://a2jlab.org/current-projects/signature-studies/pre-trial-release/.

240 O'Neil, Weapons of Math Destruction, *supra* note 172, at 209-10.

241 Michael Luca et al., *Algorithms Need Managers, Too*, Harvard Business Review (Jan./Feb. 2016), https://hbr.org/2016/01/algorithms-need-managers-too.

242 Corbett-Davies et al., *A computer program used for bail and sentencing decisions was labeled as biased against blacks. It's actually not that clear.*, *supra* note 208.

243 Ackerman, *Safety Checklists for Sociotechnical Design*, *supra* note 236.

244 O'Neil, Weapons of Math Destruction, *supra* note 172, at 210.

245 Sheldon X. Zhang et al., *An Analysis of Prisoner Reentry and Parole Risk Using COMPAS and Traditional Criminal History Measures*, 60 Crime & Delinquency 167, 187 (2014) (finding that a model assessing four static variables, gender, age, age of first arrest, and number of prior arrests, performed just as well as COMPAS in predicting prior arrests).

246 O'Neil, Weapons of Math Destruction, *supra* note 172, at 210.

Robust procedural safeguards will also help ensure that, once they enter the criminal justice system, these scores are used properly and that their inadvertent impact is minimized. The guidelines laid out by the *Loomis* case represent a decent first attempt to guide the use of these scores, but the holes in the court's analysis likely undermine their effectiveness. Much greater precision is required. In particular, it would be valuable to develop standards for the types of information provided to judges and sentencing authorities in the PSI regarding the risk assessment tools, how scores were calculated, and so on. These guidelines should include specific recommendations about how the information is actually presented to judges in the PSI.[247] These rules should also address a defendant's right to review and challenge this information, in light of precedent established in cases like *Gardner*.[248] Furthermore, policymakers need to think creatively about how to feasibly restrict judges from lengthening sentences based on the scores in the PSI, which is prohibited by the *Loomis* court but practically quite difficult given the amount of discretion that judges have in sentencing.

VII. Further Areas for Research

Beyond the recommendations described above, more scholarship is clearly needed to answer critical questions about the legality, fairness, and long-term impact of using risk-assessment algorithms in the sentencing context. There has been a substantial amount of research on the use of these risk-assessment algorithms in rehabilitation and pre-trial assessment,[249] but their use in sentencing is far newer and still warrants additional inquiry. This section identifies some key research questions about the technical and legal issues that are ripe for further inquiry.

The Science of Sentencing

- How does the length of a sentence impact prisoner behavior, particularly with regard to an individual's propensity for recidivism?
- Do risk assessment algorithms represent an improvement over unguided human judgment?

Technical Fairness and Accuracy

- Is there a particular accuracy threshold that should be required before a risk assessment tool can be used in sentencing? How should that threshold be established?
- Beyond statistical parity, how can we reconcile the concepts of technical and legal fairness for use in sentencing? How should fairness and accuracy be balanced against one another?
- Should fairness be defined differently in a sentencing context as compared to a rehabilitative or pre-trial context?
- What kind of data should jurisdictions be collecting or maintaining for use in sentencing algorithms, in order to ensure accuracy and fairness?
- Is there any advantage to using tools that emphasize static factors over dynamic factors (or vice versa)?

Legality and Transparency

- Is the incorporation of certain variables (e.g. race, gender, socioeconomic status) into these algorithms unconstitutional? In particular, does it violate the Due Process or Equal Protection clauses of the Fourteenth Amendment?
- Is the use of variables that correlate heavily with impermissible factors like race unconstitutional?
- How much information can private companies be required to disclose about their algorithms? How much information *should* they be required to disclose?
- What is the appropriate administrative agency or other government institution to whom the contents of these algorithms should be disclosed?
- Should the rules for decision-making incorporated into these algorithms be available

---

247 *Communicating Risk at Sentencing*, Pennsylvania Commission on Sentencing, Risk Need Assessment Project, Interim Report, 1, 8 (2014), http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-8-communicating-risk-at-sentencing/view (finding that the manner in which risk-assessment scores are presented to judges can have an effect on the degree to which these assessments are consider in sentencing decisions and that judges in Pennsylvania usually tend to prefer as much information as possible.)

248 *See supra* notes 154-55 and accompanying text.

249 *See supra* Part II.D.

for public comment and input?

- Should more explicit rules be developed to govern the form of risk assessment information provided to judges before sentencing? What should these rules look like?

Validity Testing

- What metrics should states and jurisdictions use when conducting validity tests?
- Can guidelines be established that are transferable across jurisdictions in order to supplement validity testing?
- What level of transparency is necessary in order for jurisdictions to conduct validity tests?

While this is by no means a complete list, these questions are intended serve as a useful jumping-off point for those who are in a position to conduct research on the use of risk-assessment algorithms in sentencing, or can provide funding for said research.

VIII. Conclusion

The growing use of risk assessment software in criminal sentencing is a cause for both optimism and skepticism. While these tools have the potential to improve sentencing accuracy in the criminal justice system and reduce the risk of human error and bias, they also have the potential to reinforce or exacerbate existing biases and to undermine certain basic tenets of fairness that are central to our justice system. In this report, we have tried to canvass a wide range of these legal and technical challenges in order to help policymakers make more informed decisions about whether and how to implement these systems in the future. Ultimately, we believe that the current trend toward greater use of these tools is likely to continue, and therefore we would urge policymakers to maintain a focus on fairness, accountability, and transparency when deploying these tools. There are important ethical and normative decisions that need to be made as these risk assessment tools are integrated into the existing system—and those decisions should not be made lightly or with insufficient information.