

November 2022

Algorithmic Risk Assessment: A Progressive Policy in Pretrial Release

Melissa Hamilton

Follow this and additional works at: <https://digitalcommons.law.uidaho.edu/idaho-law-review>

Recommended Citation

Melissa Hamilton, *Algorithmic Risk Assessment: A Progressive Policy in Pretrial Release*, 57 IDAHO L. REV. (2022).

Available at: <https://digitalcommons.law.uidaho.edu/idaho-law-review/vol57/iss3/8>

This Article is brought to you for free and open access by Digital Commons @ UIdaho Law. It has been accepted for inclusion in Idaho Law Review by an authorized editor of Digital Commons @ UIdaho Law. For more information, please contact annablaine@uidaho.edu.

ALGORITHMIC RISK ASSESSMENT: A PROGRESSIVE POLICY IN PRETRIAL RELEASE

MELISSA HAMILTON*

TABLE OF CONTENTS

I. INTRODUCTION	615
II. BACKGROUND TO RISK ASSESSMENT.....	616
III. THE PROSPECTS FOR ALGORITHMS.....	619
A. Reduce Detention Rates	619
B. Reduce Reliance Upon Bail Systems	621
C. Mitigate Human Biases	622
D. Improve the Justifiability of Detention Decisions	624
E. Conserve Criminal Justice Resources	624
IV. CHALLENGES FACING ALGORITHMIC RISK	625
A. Accuracy.....	626
B. Demographic Equity.....	629
C. Levels of Risk	632
V. CONCLUSIONS	633

I. INTRODUCTION

The United States is unfortunately renowned for being the world’s leader in mass incarceration, sustaining a detention rate exceeding every other country.¹The overall proportion of prisoners is remarkable. One out of every five prisoners globally resides in America’s prisons and jails.² The situation has become untenable in part because of the sheer financial burden of operating the domestic prison industrial complex.³ Consequently, bipartisan political and public support have promoted policy reforms to slash incarceration numbers.⁴

Notably, algorithmic risk tools have become a weapon of choice to combat mass incarceration.⁵ Selected studies find support for the role of science-informed risk tools to help alleviate excessive rates of imprisonment by convincing decision-makers that substantially more subjects may be safely supervised in their

* Professor of Law & Criminal Justice at the University of Surrey School of Law; PhD in criminology and criminal justice, The University of Texas at Austin; J.D., The University of Texas School of Law. The statistical research presented herein was supported by funding from the Koch Foundation.

1. ROY WALMSLEY, INST. FOR CRIM. & JUST. POL’Y RSCH., WORLD PRISON POPULATION LIST (12th ed. 2020), https://www.prisonstudies.org/sites/default/files/resources/downloads/wppl_12.pdf.

2. *Id.*

3. Nicole Lewis & Beatrix Lockwood, *The Hidden Cost of Incarceration*, MARSHALL PROJECT (Dec. 17, 2019, 5:00 AM), <https://www.themarshallproject.org/2019/12/17/the-hidden-cost-of-incarceration> (the costs to American taxpayers to operate America’s prison systems are \$80 billion a year).

4. Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CAL. L. REV. 439, 447 (2020).

5. Malcolm M. Feeley, *How to Think about Criminal Court Reform*, 98 B.U. L. REV. 673, 690 (2018).

communities.⁶ The new algorithmic risk wave has been actively targeted to back-end decisions, such as sentencing and parole.⁷ More recently, the pretrial context has become a “flashpoint” in efforts to reduce mass incarceration because of the significant numbers of individuals who are detained pending their trials.⁸

The issues are more salient considering the risks of COVID-19 in confined spaces amid crowded conditions.⁹ “So now, whatever public-safety risk is posed by releasees must be weighed against the public-safety risk that detaining them under such conditions could pose by spreading the disease, in addition to the risk to prisoners themselves.”¹⁰

This essay reviews the advancement of risk assessment tools in criminal justice generally and then discusses their role more specifically in pretrial settings. Several potential advantages of risk tools as a progressive reform for pretrial are explored, followed by relevant questions that deserve attention. For example, while algorithmic risk may (or may not) achieve acceptable rates of accuracy in estimating future risk, there is the potential that risk predictions may still be inequitable based on minority or gender status. To highlight these salient issues, selected results are employed from the author’s previous empirical studies on the performance of two algorithmic risk tools in several pretrial jurisdictions.

II. BACKGROUND TO RISK ASSESSMENT

Risk assessment involves predicting an individual’s potential for a negative criminal justice outcome, such as flight from justice or a new arrest.¹¹ Appraising the potential for future failures plays a role in various outcomes during an accused’s journey through the criminal justice system, from decisions to arrest through parole outcomes.¹² A primary purpose of widespread risk predictions is to allow officials to triage their populations and thus to manage them accordingly.¹³ Considering the high-stakes environment of criminal justice, proper risk practices should strike a reasonable balance among interests in public safety, protecting individual rights, and the efficient use of limited resources.¹⁴

6. Kristin Bechtel et al., *A Meta-Analytic Review of Pretrial Research: Risk Assessment, Bond Type, and Interventions*, 42 AM. J. CRIM. JUST. 443, 447 (2017).

7. Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 67 (2017).

8. Nicholas Scurich & Daniel A. Krauss, *Public’s Views of Risk Assessment Algorithms and Pretrial Decision Making*, 26 PSYCHOL. PUB. POL’Y & L. 1, 2 (2020).

9. Lauryn P. Gouldin, *Reforming Pretrial Decision-Making*, 55 WAKE FOREST L. REV. 857, 880-81 (2020).

10. J.J. Prescott, Benjamin Pyle, & Sonja B. Starr, *Understanding Violent Crime Recidivism*, 95 NOTRE DAME L. REV. 1643, 1647 (2020).

11. Garrett & Monahan, *supra* note 4, at 450.

12. Jordan M. Hyatt et al., *Reform in Motion: The Promise and Perils of Incorporating Risk Assessments and Cost-Benefit Analysis into Pennsylvania Sentencing*, 49 DUQ. L. REV. 707, 724–25 (2011).

13. John Monahan, *Preface: Recidivism Risk Assessment in the 21st Century*, in HANDBOOK OF RECIDIVISM RISK/NEEDS ASSESSMENT TOOLS xxiii, xxiii (Jay P. Singh et al. eds., 2018).

14. Ellen A. Donnelly & John M. MacDonald, *The Downstream Effects of Bail and Pretrial Detention on Racial Disparities in Incarceration*, 108 J. CRIM. L. & CRIMINOLOGY 775, 788–89 (2019).

Historically, such predictions have been based on intuition or the personal experience of the particular official responsible for making the relevant decision.¹⁵ A more modern trend is to rely upon scientific evidence that explains offending patterns.¹⁶

The “evidence-based practices movement” is the now popular term to describe the turn to behavioral sciences data to improve risk-based classifications. Scientific studies targeting recidivism outcomes are benefiting from the compilation of large datasets (i.e., big data) of discharged offenders. Researchers track the offenders post-release, observe recidivism rates, and then statistically test which factors correlate with recidivism. Risk assessment tool developers use computer modeling to combine factors of sufficiently high correlation and weight them accordingly using increasingly complex algorithms.¹⁷

An algorithm is simply an equation in which inputs are received and processed to produce a quantitatively-derived output.¹⁸ Risk tools often convert such quantitative outputs into categorical buckets—e.g., low, medium, and high risk—to permit partitioning the correctional population according to levels of predicted likelihood of failure.¹⁹

Risk algorithms have transformed criminal justice into a “data-driven decision-making” machine.²⁰ This form of algorithmic governance²¹ is a novel strategy to battle mass incarceration.²² Officials are using risk tools to help them divert low-risk defendants or reserve prison beds for those at highest risk.²³ In other words,

15. Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 556 (2015).

16. Yoav Mehozay & Eran Fisher, *The Epistemology of Algorithmic Risk Assessment and the Path Towards a Non-Penology Penology*, 21 PUNISHMENT & SOC’Y 523, 531 (2019).

17. Melissa Hamilton, *Debating Algorithmic Fairness*, 52 U.C. DAVIS L. REV. ONLINE 261, 266 (2019). “Evidence-based practice . . . is any correctional practice or intervention whose effectiveness at achieving its stated goal is supported by ‘empirically sound’ research of some kind.” Klingele, *supra* note 15, at 556.

18. Reuben Binns, *Algorithmic Accountability and Public Reason*, 31 PHIL. & TECH. 543, 545 (2018). See also OSONDE A. OSOBA ET AL., RAND, ALGORITHMIC EQUITY: A FRAMEWORK FOR SOCIAL APPLICATIONS ix (2019), https://www.rand.org/content/dam/rand/pubs/research_reports/RR2700/RR2708/RAND_RR2708.pdf (characterizing algorithms as “advanced decision-making or decision support artifacts that exhibit ‘intelligent behavior’”).

19. Alice Ristroph, *Criminal Law as Public Ordering*, 70 U. TORONTO L.J. 64, 81 (2020).

20. DANIELE KEHL ET AL., ALGORITHMS IN THE CRIMINAL JUSTICE SYSTEM: ASSESSING THE USE OF RISK ASSESSMENTS IN SENTENCING 3 (2017), <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>.

21. Kelly Hannah-Moffat, *Algorithmic Risk Governance: Big Data Analytics, Race and Information Activism in Criminal Justice Debates*, 23 THEORETICAL CRIMINOLOGY 453, 454 (2018).

22. Malcolm M. Feeley, *How to Think about Criminal Court Reform*, 98 B.U. L. REV. 673, 690 (2018).

23. Sarah L. Desmarais et al., *Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings*, 13 PSYCHOL. SERVS. 206 (2016).

algorithm-based risk assessment tools are popular in informing decisions concerning whether to incarcerate or release.²⁴

Proponents thereby view algorithmic risk assessment as an objective and reasonable way to reduce mass incarceration without sacrificing public safety.²⁵ Across criminal justice systems, the “offender risk assessment enterprise has been moving at warp speed since the turn of the millennium and shows no signs of slowing”²⁶ and is widely considered to represent industry best practice in risk evaluation.²⁷

The algorithmic turn specifically assists in pretrial systems to respond to the significant contribution that pretrial detention numbers in local jails make to excessive incarceration rates.²⁸ The Prison Policy Initiative has crunched the numbers: “Pretrial detention is responsible for all of the net jail growth in the last twenty years.”²⁹ Today, twenty percent of the nation’s combined prison and jail population is comprised of individuals yet to be convicted.³⁰ At any time, about 500,000 inmates are awaiting their trial dates.³¹ In local jails, pretrial inmates outnumber those convicted of crimes and serving their sentences: three-quarters of inmates in America’s jails have not been convicted of any crime.³² A related consequence is that one in five local jails is operating at or over capacity.³³

Consequently, pretrial’s role in mass incarceration set the stage for adoption algorithmic tools as a main reform.³⁴ The *Science Bench Book for Judges* from the National Judicial College conceptualizes risk assessment algorithms as the “most

24. Carolyn McKay, *Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making*, 32 CURRENT ISSUES CRIM. JUST. 22–39 (2020); Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 494 (2018).

25. Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 305 (2018). See generally Brandon L. Garrett, *Evidence-Informed Criminal Justice*, 86 GEO. WASH. L. REV. 1490 (2019) (providing a background on criminal justice advancements and analyzing, as well as advocating, empirically based approaches to adopting policy).

26. J. Stephen Wormith, *Automated Offender Risk Assessment*, 16 CRIMINOLOGY & PUB. POL’Y 281 (2017).

27. Bechtel, *supra* note 6, at 446; Melissa Hamilton, *Risk-Needs Assessment: Constitutional and Ethical Challenges*, 52 AM. CRIM. L. REV. 231, 232 (2015) [hereinafter Hamilton 2], <http://epubs.surrey.ac.uk/id/eprint/842342>.

28. COLIN DOYLE ET AL., HARV. LAW SCH., BAIL REFORM: A GUIDE FOR STATE AND LOCAL POLICYMAKERS 7 (2019), <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=9a804d1d-f9be-e0f0-b7cd-cf487ec70339&forceDialog=0>.

29. Wendy Sawyer & Peter Wagner, *Mass Incarceration: The Whole Pie 2020*, PRISON POL’Y INITIATIVE (Mar. 24, 2020), <https://www.prisonpolicy.org/reports/pie2020.html>.

30. WILL DOBBIE & CRYSTAL YANG, HAMILTON PROJECT, PROPOSALS FOR IMPROVING THE U.S. PRETRIAL SYSTEM 4 (2019), https://www.brookings.edu/wp-content/uploads/2019/03/DobbieYang_PP_20190314.pdf.

31. ZHEN ZENG, U.S. DEP’T OF JUSTICE, JAIL INMATES IN 2018 1 (Mar. 2020), <https://www.bjs.gov/content/pub/pdf/ji18.pdf>.

32. Sawyer & Wagner, *supra* note 29.

33. ZENG, *supra* note 31, at 8.

34. Mitali Nagrecha et al., *Court Culture and Criminal Law Reform*, 69 DUKE L.J. 84 (2020), <https://dlj.law.duke.edu/2020/04/courtculture/>.

ubiquitous form of predictive technology used in the criminal pretrial context,” and constitutes a critical weapon in the judiciary’s “arsenal of justice.”³⁵

III. THE PROSPECTS FOR ALGORITHMS

Reformers advocate the deployment of algorithmic risk tools to serve several goals in pretrial justice reforms which are, directly and indirectly, related to reasonably curtailing pretrial jail populations. The objectives, albeit with some overlap amongst them, include curtailing rates of detention, reducing reliance upon bail, mitigating human biases in decision-making, improving the justifiability of detention decisions, and conserving criminal justice resources.

A. Reduce Detention Rates

In a pretrial setting, judges are typically the final arbiters of whether jailed individuals are eligible for release.³⁶ Detaining those not convicted might appear improper,³⁷ but the practice is lawful when due process is followed.³⁸ The United States Supreme Court in the case of *United States v. Salerno* (1987) confirmed the constitutionality of ordering pretrial detention reliant upon a judge’s own guesstimate of the defendant’s future dangerousness.³⁹ The Court accepted that an appropriate balance justified this conclusion: “In our society liberty is the norm, and detention prior to trial or without trial is the carefully limited exception.”⁴⁰

Judges at pretrial thereby face conflicting goals: “Pretrial decision-making involves a fundamental tension between the court’s desire to protect citizens from dangerous criminals, ensure that accused individuals are judged before the law, and minimize the amount of pretrial punishment meted out to legally innocent defendants.”⁴¹ Courts should also be cognizant that pretrial detention has

35. Veronica Alicea-Galvan, *Pre-Trial Criminal*, in SCIENCE BENCH BOOK FOR JUDGES 201, 205 (Peggy Hora et al. eds., 2020), <https://www.judges.org/wp-content/uploads/Chapter6-SBB.pdf>.

36. Brandon P. Martinez et al., *Time, Money, and Punishment: Institutional Racial-Ethnic Inequalities in Pretrial Detention and Case Outcomes*, 66 CRIME & DELINQ. 837, 839 (2020).

37. “Detaining a person before he is found guilty of a crime is a particularly draconian decision for the state to make. Pretrial detention is one of the most severe things that can happen to a person: the detainee immediately loses his freedom, and can also lose his family, health, home, job, and community ties.” OPEN SOC’Y JUST. INITIATIVE, PRESUMPTION OF GUILT: THE GLOBAL OVERUSE OF PRETRIAL DETENTION 7 (2014), <https://www.justiceinitiative.org/uploads/de4c18f8-ccc1-4eba-9374-e5c850a07efd/presumption-guilt-09032014.pdf>.

38. Hafsa S. Mansoor, *Guilty Until Proven Guilty: Effective Bail Reform as a Human Rights Imperative*, DEPAUL L. REV. (forthcoming 2020), https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID3566273_code3824600.pdf?abstractid=3566273&mirid=1.

39. *United States v. Salerno*, 481 U.S. 739, 751 (1987).

40. *Id.* at 755.

41. THOMAS BLOMBERG ET AL., VALIDATION OF THE COMPAS RISK ASSESSMENT CLASSIFICATION INSTRUMENT 4 (2010), <http://criminology.fsu.edu/wp-content/uploads/Validation-of-the-COMPAS-Risk-Assessment-Classification-Instrument.pdf>.

significant downstream consequences, including increasing the likelihood of conviction, the severity of sentence, the number of future contacts with criminal justice agencies, barriers to stable employment, and limits on access to social safety nets.⁴² A relevant formula suggests an algorithmic tool will allow judges to “(1) Maximize public safety and (2) Maximize court appearance while (3) Maximizing release from custody.”⁴³

The terminology has evolved since *Salerno* from the vagueness of “future dangerousness” to a more refined perspective of “risk assessment.”⁴⁴ Relatedly, such predictions have over the last three decades moved from the unstructured decision-making by individual judges to the standardization and science offered by algorithmic tools, at least in those jurisdictions who have adopted such technologies.⁴⁵

While an ethical rule of thumb might be to release those who have not yet been convicted of any crime, judges who tend to be risk averse order pretrial detention as a means to ensure the defendants’ appearance for trial and/or to protect the community from potential offending if released.⁴⁶ Erring on the side of caution thereby partly explains the significant number of pretrial detainees.⁴⁷

Reformers contend that reliance upon scientifically-informed algorithms will reduce the use of detention.⁴⁸ While algorithmic risk scores are not meant to entirely displace human judgements, these technological, decision-support datapoints may allow judges to isolate those who actually pose a high risk of failure; consequently, judges may be more confident to discharge the rest.⁴⁹ The evidence-based platform might further encourage bolstering release rates by insulating judges from allegations that they wrongfully released an individual who endangered the public.⁵⁰ Thus, if a released offender commits some bad act, the scientific foundation of the tools offers an objective explanation.⁵¹

42. See generally Paul Heaton et al., *The Downstream Consequences of Misdemeanor Pretrial Detention*, 69 STAN. L. REV. 711 (2017).

43. MICHAEL R. JONES, U.S. DEP’T OF JUST., UNSECURED BONDS: THE AS EFFECTIVE AND MOST EFFICIENT PRETRIAL RELEASE OPTION 4 (2013), <https://www.ncjrs.gov/App/AbstractDB/AbstractDBDetails.aspx?id=269164>.

44. KIRK HEILBRUN ET AL., EVALUATION OF VIOLENCE RISK IN ADULTS 74, 76 (2009), https://www.researchgate.net/publication/299925474_Evaluation_of_Violence_Risk_in_Adults.

45. Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537 (2015).

46. Brandon Buskey, *Wrestling with Risk: The Questions Beyond Money Bail*, 98 N.C. L. REV. 379, 384 (2020).

47. Laurel Eckhouse et al., *Layers of Bias: A Unified Approach for Understanding Problems with Risk Assessment*, 46 CRIM. JUST. & BEHAV. 185, 186 (2019).

48. John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1729 (2018).

49. Sandra G. Mayson, *Dangerous Defendants*, 127 YALE L.J. 490, 515 (2018).

50. Michael O’Hear, *Actuarial Risk Assessment at Sentencing, Potential Consequences for Mass Incarceration and Legitimacy*, 38 BEHAV. SCI. & L. 193, 199 (2020).

51. Harry Surden, *Ethics of AI in Law: Basic Questions*, in THE OXFORD HANDBOOK OF ETHICS OF AI (Marcus D. Dubber et al. eds., forthcoming 2020).

B. Reduce Reliance Upon Bail Systems

In many jurisdictions, even if a judge considers an individual at little risk of failure, the judge may order that release be contingent upon posting some sort of security to mitigate any risk.⁵² A bond is typically intended as a means of encouraging the individual's return for court dates without new arrest charges.⁵³ The amount of bond is not always tied to an estimate of the individual's ability to pay it or to the likelihood of compliance with the law.⁵⁴ Some systems employ bail schedules that use the pending criminal charge as a proxy for the risk of flight or rearrest; the more severe the charge, the greater the amount of security required.⁵⁵

Reliance upon bail systems contributes significantly to jail population counts, and thus exacerbates mass incarceration.⁵⁶ Bail systems that do not link the bond to correspond with the individual's ability to pay inevitably drive up their pretrial jail population numbers.⁵⁷ Many defendants otherwise eligible for release remain behind bars simply because they cannot afford the bond.⁵⁸ These results contradict the foundation of a risk assessment system as individuals who continue to be detained based on financial disability can present lower risk profiles than those who are released after posting bail.⁵⁹

Other contributions to incarceration counts are explained by studies that show that pretrial detention is associated with increases in conviction rates, the likelihood of a sentence involving a prison term, and sentence length.⁶⁰ Moreover, incarceration is itself criminogenic in nature in that those who are detained—including in a preconviction context—are more likely to offend in the future.⁶¹ In

52. Alexa Van Brunt & Locke E. Bowman, *Toward a Just Model of Pretrial Release: A History of Bail Reform and a Prescription for What's Next*, 108 J. CRIM. L. & CRIMINOLOGY 701, 728 n.132 (2018).

53. *Id.*

54. COLIN DOYLE ET AL., BAIL REFORM: A GUIDE FOR STATE AND LOCAL POLICYMAKERS 7 (2019), <http://cjpp.law.harvard.edu>.

55. VERA INST. OF JUST., INCARCERATION'S FRONT DOOR: THE MISUSE OF JAILS IN AMERICA 32 (2015), https://www.vera.org/downloads/publications/incarcerations-front-door-report_02.pdf.

56. Shima Baradaran Baughman, *Dividing Bail Reform*, 105 IOWA L. REV. 947, 949 (2020).

57. *Id.* at 949–50.

58. Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CAL. L. REV. 439, 442 (2020); SARAH PICARD ET AL., CENTER FOR COURT INNOVATION, BEYOND THE ALGORITHM: PRETRIAL REFORM, RISK ASSESSMENT, AND RACIAL FAIRNESS 3 (2019), <https://www.courtinnovation.org/sites/default/files/media/document/2019/beyondthealgorithm.pdf>.

59. EMILY TIRY ET AL., URBAN INST., ROAD MAP TO PRETRIAL REFORMS 2 (2016), https://www.urban.org/research/publication/road-map-pretrial-reforms/view/full_report; Arthur Rizer & Caleb Watney, *Artificial Intelligence Can Make our Jail System More Efficient, Equitable, and Just*, 23 TEX. REV. L. & POL. 181, 187 (2018).

60. Ellen A. Donnelly & John M. MacDonald, *The Downstream Effects of Bail and Pretrial Detention on Racial Disparities in Incarceration*, 108 J. CRIM. L. & CRIMINOLOGY 775, 791 (2018); Megan Stevenson & Sandra G. Mayson, *Pretrial Detention and Bail*, in REFORMING CRIMINAL JUSTICE: A REPORT BY THE ACADEMY FOR JUSTICE 21, 22 n.5 (Erik Luna ed., 2017) (citing studies).

61. Léon Digard & Elizabeth Swavola, VERA INST. OF JUST., JUSTICE DENIED: THE HARMFUL AND LASTING EFFECTS OF PRETRIAL DETENTION 6 (2019) (citing studies).

contrast, for many lower risk pretrial detainees, early release materially curtails their risk of failure, thereby allowing them to avoid the jail's revolving door, which otherwise further contributes to mass incarceration.⁶²

Thus, bail reformers, with the support of the American Bar Association and the National Association of Pretrial Services Agencies, advocate risk assessment tools in pretrial settings to reduce reliance upon money bail as a means of procuring defendants' compliance, with the result of boosting release rates.⁶³

C. Mitigate Human Biases

Historically, judgements about an individual's likelihood to fail have typically been based on the gut instinct or the personal experience of the official responsible for making the relevant management decision.⁶⁴ Critics challenge those sorts of predictions as suffering from implicit and explicit human biases.⁶⁵ Judges in a pretrial context may rely upon oversimplified heuristics because they typically have few interactions with the defendants, limited access to relevant information, and insufficient time for full deliberation.⁶⁶ Pretrial detention hearings are brief and often without the assistance of defense counsel who may otherwise provide the judge with more contextualizing information about the defendant's prospects for success that could counter these biases and heuristics.⁶⁷ Remarkably, these types of cognitive shortcuts are more likely to dictate detention. Judges' uncorrected biases tend toward the direction of overestimating the potential of pretrial defendants to flee or endanger public safety.⁶⁸

62. John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1769 (2018).

63. *E.g.*, TODD SHAUGHNESSY, J. ET AL., REPORT TO THE UTAH JUDICIAL COUNCIL ON PRETRIAL RELEASE AND SUPERVISION PRACTICES 16 (2015).

64. See Alfred Blumstein, *Some Perspectives on Quantitative Criminology Pre-JQC: And Then Some*, 26 J. QUANTITATIVE CRIMINOLOGY 549, 554 (2010).

65. Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. DAVIS L. REV. 1067, 1070 (2018); Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 523 (2019); see also Chelsea Barabas, *Beyond Bias: Re-Imagining the Terms of "Ethical AI" in Criminal Law*, 12 GEO. J. L. & MOD. CRITICAL RACE PERSP. 84, 95 (2020) ("[I]mplicit bias is understood as a phenomenon which 'is somehow apart from us yet can infect our decision-making . . . as opposed to something that is variously, but systematically, cultivated and maintained.'" (citing Anna Lauren Hoffman, *Where Fairness Fails: On Data, Algorithms, and the Limits of Antidiscrimination Discourse*, 22 INFO., COMM. & SOC'Y 900 (2019))).

66. Hafsa S. Mansoor, *Guilty Until Proven Guilty: Effective Bail Reform as a Human Rights Imperative*, DEPAUL L. REV. at 18, 21 (forthcoming 2020). Nonetheless, the speedy nature of pretrial decision-making is not solely a negative attribute. Quick decisions may lead to earlier releases, thereby reducing the negative consequences of pretrial detention. Paul Heaton, *The Expansive Reach of Pretrial Detention*, 98 N.C. L. REV. 369, 375 (2020).

67. Shima Baradaran Baughman, *Dividing Bail Reform*, 105 IOWA L. REV. 947, 962 (2020); Megan Stevenson & Sandra G. Mayson, *Pretrial Detention and bail*, in REFORMING CRIMINAL JUSTICE: A REPORT BY THE ACADEMY FOR JUSTICE 21, 25 (Erik Luna et al. ed., 2017) ("In practice, bail hearings are a messy affair.").

68. Lauryn P. Gouldin, *Defining Flight Risk*, 85 U. CHI. L. REV. 677, 681 (2018).

A racial element exists here as well. Pretrial detention disproportionately ensnares minority citizens.⁶⁹ One reason is the potential for race-based biases when humans drive pretrial outcomes.⁷⁰ America has a long history promoting stereotypes connecting darkness of skin with criminality and violence.⁷¹ Despite their professional training, judges use heuristics that can draw upon racialized stereotypes of criminality.⁷² Empirical works specifically find evidence of racial bias in pretrial decisions.⁷³ Another reason for higher minority detention rates is that minorities face more limited access to the financial resources necessary to successfully secure release.⁷⁴

Algorithms may ameliorate such race-based biases produced by subjective human intuitions.⁷⁵ Algorithmic risk instruments are seen as more objective information-processing devices.⁷⁶ There is also the hope for significant accuracy rates from algorithmic prediction⁷⁷ in that algorithmic risk forecasts are more likely to be correct than human oracles.⁷⁸ Behavioral psychological studies rather consistently find that judgements based on science are more accurate than those reliant upon individual intuition.⁷⁹

69. Wendy Sawyer, *How Race Impacts Who Is Detained Pretrial*, PRISON POL'Y INITIATIVE (Oct. 9, 2019), https://www.prisonpolicy.org/blog/2019/10/09/pretrial_race/ (citing studies).

70. Emily Berman, *A Government of Laws and Not of Machines*, 98 B.U. L. REV. 1277, 1327 (2018).

71. Lydette S. Assefa, *Assessing Dangerousness Amidst Racial Stereotypes: An Analysis of the Role of Racial Bias in Bond Decisions and Ideas for Reform*, 108 J. CRIM. L. & CRIMINOLOGY 653, 657–58 (2019).

72. Brandon P. Martinez et al., *Time, Money, and Punishment: Institutional Racial-Ethnic Inequalities in Pretrial Detention and Case Outcomes*, 66 CRIME & DELINQ. 837, 841 (2020).

73. Megan Stevenson & Sandra G. Mayson, *Pretrial Detention and Bail*, in REFORMING CRIMINAL JUSTICE: A REPORT BY THE ACADEMY FOR JUSTICE 21, 30 (Erik Luna ed., 2017) (citing studies). See generally Meghan Sacks et al., *Sentenced to Pretrial Detention: A Study of Bail Decisions and Outcomes*, 40 AM. J. CRIM. JUST. 661 (2015) (additional studies reporting racial disparities in pretrial); Tina L. Freiburger et al., *The Impact of Race on the Pretrial Decision*, 35 AM. J. CRIM. JUST. 76 (2010).

74. Muhammad B. Sardar, *Give Me Liberty or Give Me . . . Alternatives? Ending Cash Bail and its Impact on Pretrial Incarceration*, 84 BROOK. L. REV. 1421, 1423 (2019); PRETRIAL JUST. INST., PRETRIAL RISK ASSESSMENT CAN PRODUCE RACE-NEUTRAL RESULTS 2 (2017) (citing studies).

75. Jeffrey C. Singer et al., *A Convergent Approach to Sex Offender Risk Assessment*, in THE WILEY-BLACKWELL HANDBOOK OF LEGAL AND ETHICAL ASPECTS OF SEX OFFENDER TREATMENT AND MANAGEMENT 341, 343 (Karen Harrison & Bernadette Rainey eds., 2013).

76. Jessica Gabel Cino, *Deploying the Secret Police: The Use of Algorithms in the Criminal Justice System*, 34 GA. ST. U. L. REV. 1073, 1078 (2018); Faye S. Taxman, *The Partially Clothed Emperor: Evidence-Based Practices*, 34 J. CONTEMP. CRIM. JUST. 97, 99 (2018).

77. WILL DOBBIE & CRYSTAL YANG, THE HAMILTON PROJECT, PROPOSALS FOR IMPROVING THE U.S. PRETRIAL SYSTEM 7 (2019); SARAH L. DESMARAIS & EVAN M. LOWDER, PRETRIAL RISK ASSESSMENT TOOLS: A PRIMER FOR JUDGES, PROSECUTORS, AND DEFENSE ATTORNEYS 4 (2019).

78. Arthur Rizer & Caleb Watney, *Artificial Intelligence Can Make our Jail System More Efficient, Equitable, and Just*, 23 TEX. REV. L. & POL. 181, 183 (2019); Seth J. Prins & Adam Reich, *Can We Avoid Reductionism in Risk Reduction?*, 22 THEORETICAL CRIMINOLOGY 258, 259 (2018).

79. Karen Bogenschneider & Bret N. Bogenschneider, *Empirical Evidence from State Legislators: How, When, and Who Uses Research*, 26 PSYCHOL. PUB. POL'Y & L. 413, 413 (2020).

D. Improve the Justifiability of Detention Decisions

Algorithmic risk tools are often lauded for reducing the arbitrariness of pretrial release decisions.⁸⁰ Human judgements may vary for reasons that algorithms cannot, such as by the person being overworked, hungry, or preoccupied.⁸¹ Though the algorithms are not intended to entirely usurp human involvement in criminal justice outcomes,⁸² the standardization inherent in algorithmic outcomes offers several benefits. Algorithmic predictions can help structure and thus enhance the human aspect of decision-making.⁸³ The automation afforded constrains discretion⁸⁴ and thus potentially reduces inconsistencies in detention outcomes.⁸⁵ As well, the objectivity afforded by algorithmic prediction may make pretrial release decisions appear more equitable⁸⁶ and render release decisions more legally justifiable.⁸⁷

E. Conserve Criminal Justice Resources

The deployment of a risk tool in pretrial decision-making may save the government significant monies, considering the requirement of vast resources to service the state of mass incarceration. Pretrial incarceration is expensive.⁸⁸ Detaining those pending trial costs over \$13 billion annually in the United States.⁸⁹

80. Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242, 267 (2019). See also PARTNERSHIP ON AI, REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM 7 (2019), <https://www.partnershiponai.org/wp-content/uploads/2019/04/Report-on-Algorithmic-Risk-Assessment-Tools.pdf> (indicating criminal justice risk tools "provide rigor and reproducibility for life-critical decisions").

81. R. Barry Ruback et al., *Communicating Risk Information at Criminal Sentencing in Pennsylvania: An Experimental Analysis*, 80 FED. PROB. 47 (2016).

82. Reuben Binns, *Algorithmic Decision-Making: A Guide for Lawyers*, 25 JUD. REV. 2, 6 (2020).

83. Faye S. Taxman, *The Partially Clothed Emperor: Evidence-Based Practices*, 34 J. CONTEMP. CRIM. JUST. 97, 99 (2018).

84. Faye S. Taxman & Amy Dezember, *The Value and Importance of Risk and Need Assessment (RNA) in Corrections & Sentencing: An Overview of the Handbook*, in HANDBOOK ON RISK AND NEED ASSESSMENT: THEORY AND PRACTICE 22, 23 (Faye S. Taxman ed., 2017).

85. Cecelia Klingele, *Making Sense of Risk*, 38 BEHAV. SCI. & L. 218, 220 (2020).

86. SARAH L. DESMARAIS & EVAN M. LOWDER, PRETRIAL RISK ASSESSMENT TOOLS: A PRIMER FOR JUDGES, PROSECUTORS, AND DEFENSE ATTORNEYS 3 (2019).

87. Michael O'Hear, *Actuarial Risk Assessment at Sentencing: Potential Consequences for Mass Incarceration and Legitimacy*, 38 BEHAV. SCI. & L. 193, 199 (2020); Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans*, in IZA INST. OF LABOR ECON. DISCUSSION PAPER No. 12853, at 61 (Dec. 2019), <http://ftp.iza.org/dp12853.pdf>.

88. Brandon L. Garrett, *Evidence-Informed Criminal Justice*, 86 GEO. WASH. L. REV. 1490, 1506 (2018).

89. Brian P. Schaefer & Tom Hughes, *Examining Judicial Pretrial Release Decisions: The Influence of Risk Assessments and Race*, 20 CRIMINOLOGY, CRIM. JUST., L. & SOC'Y 47, 48 (2019); see also COLIN DOYLE ET AL., BAIL REFORM: A GUIDE FOR STATE AND LOCAL POLICYMAKERS 8 (2019), <https://ccjls.scholasticahq.com/article/9908-examining-judicial-pretrial-release-decisions-the-influence-of-risk-assessments-and-race> (citing \$14 billion a year on pretrial detention).

If risk tools lead to increasing the rate of pretrial release, states might be better able to curtail unnecessary costs of incarceration,⁹⁰ without unreasonably endangering the public.⁹¹

Realistically, bipartisan support for pretrial reforms is not necessarily indicative of a moral judgement denouncing the legitimacy of incarcerating large numbers of people presumed innocent. Instead, it is a manifestation of a “virtue of frugality” in that minimizing imprisonment allows for more conservative resource allocations.⁹²

Other forms of savings from automated risk tools are possible. The churn of defendants through jails is significant. In 2018 alone, local jails counted almost 11 million bookings.⁹³ Technology-enhanced risk tools thus may make pretrial detention decision-making more economically efficient by curtailing the judicial processes and time spent on the heavy volume of cases.⁹⁴ A further anticipated benefit is that risk tool adoption could shield jurisdictions from the stresses and expenses from civil rights litigation brought to challenge practices relying heavily on money bail as the key to freedom.⁹⁵

IV. CHALLENGES FACING ALGORITHMIC RISK

The potential advantages to algorithmic risk in alleviating mass incarceration in pretrial jurisdictions are encouraging. Still, naïve utilization of algorithmic risk may undermine efforts. Several cautions are warranted for stakeholders to consider. Findings from the author’s prior empirical studies on two popular algorithmic risk tools used in pretrial settings are incorporated in this Section as relevant to provide context for the issues raised. The tools are COMPAS (the acronym for Correctional Offender Management Profiling for Alternative Sanctions) and the PSA (the acronym for the Public Safety Assessment). COMPAS is a proprietary set of risk tools that separately predicts for the risks of general recidivism and of violent recidivism.⁹⁶ The PSA is designed exclusively for pretrial,

90. Michael Baglivio et al., *The Usefulness of a General Risk Assessment, the Static Risk Assessment (SRA), in Predicting Pretrial Failure: Examining Predictive Ability Across Gender and Race*, 3 JUST. EVALUATION J. 1, 3 (2020), <https://doi.org/10.1080/24751979.2019.1668242>.

91. Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 304, 305 (2018).

92. Cecelia Klingele, *The Promises and Perils of Evidence-Based Corrections*, 91 NOTRE DAME L. REV. 537, 565 (2015).

93. OFF. OF JUST. PROGS., U.S. DEP’T OF JUST., NCJ253044, JAIL INMATES IN 2018 1 (Mar. 2020), <https://www.bjs.gov/content/pub/pdf/ji18.pdf>; WENDY SAWYER & PETER WAGNER, PRISON POL’Y INITIATIVE, MASS INCARCERATION: THE WHOLE PIE 2020 (2020), <https://www.prisonpolicy.org/reports/pie2020.html>.

94. Ric Simmons, *Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System*, 52 U.C. DAVIS L. REV. 1067, 1074 (2018).

95. COLIN DOYLE ET AL., BAIL REFORM: A GUIDE FOR STATE AND LOCAL POLICYMAKERS 8 (2019), <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=9a804d1d-f9be-e0f0-b7cd-cf487ec70339&forceDialog=0>.

96. Anne L. Washington, *How to Argue with an Algorithm: Lessons from the COMPAS-Propublica Debate*, 17 COLO. TECH. L. J. 131, 133 (2018).

with scales evaluating both failure to appear and general recidivism.⁹⁷ Both COMPAS and the PSA are marketed as tools that may be implemented in jurisdictions across the country and are popular in American criminal jurisdictions.⁹⁸

A. Accuracy

In order to combat mass incarceration, officials must get risk assessment “right” to raise its legitimacy in facilitating releases of defendants at lesser risk of failure to appear.⁹⁹ Overclassifying individuals as high-risk results in “gratuitous surveillance and overtreatment,”¹⁰⁰ while underpredicting endangers public safety.

Importantly, the guise of science should not relieve algorithmic tools from empirical, legal, and ethical scrutiny. *The Science Bench Book for Judges* warns of the myth of scientific objectivity as scientists, too, can be emotionally and professionally attached to their work, causing their research conclusions to become unconsciously biased or otherwise skewed.¹⁰¹ Hence, the fact that risk tool developers are satisfied with their science-based products does not oblige criminal justice officials to find these tools are sufficiently ready for real-world use, considering the significant negative consequences that potentially result. “While researchers may accept a moderate or less performance for a tool, for the practice of justice this should not be accepted or tolerated since it diminishes the value of the . . . tool.”¹⁰²

In the forensic sciences field, the concept of tool accuracy is often discussed in conjunction with the term *validation*. Validation involves an empirical review, demonstrating that the tool performs its intended functions at an acceptable level.¹⁰³ A tool with results establishing strong performance numbers encourages confidence by stakeholders and users.¹⁰⁴ In contrast, an ineffective or highly

97. PRETRIAL JUST. INST., SCAN OF PRETRIAL PRACTICES 27 (2019), <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=b2bd6339-8201-60f4-c262-a6317a409b82&forceDialog=0>; Kristian Lum et al., *The Impact of Overbooking on a Pre-Trial Risk Assessment Tool* (Jan. 23, 2020), <https://arxiv.org/pdf/2001.08793>.

98. Tim Brennan & William Dieterich, *Correctional Offender Management Profiles for Alternative Sanctions*, in HANDBOOK OF RISK/NEEDS ASSESSMENT TOOLS 3 (Jay P. Singh et al. eds., 2018) (regarding COMPAS); Arnold Found., *Developing a National Model for Pretrial Risk Assessment*, ARNOLD FOUND. 3 (2013), https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/LJAF-research-summary_PSA-Court_4_1.pdf (regarding PSA).

99. Brandon L. Garrett & John Monahan, *Judging Risk*, 108 CAL. L. REV. 439, 446 (2020).

100. Whitney Threadcraft-Walker et. al., *Gender, Race/Ethnicity and Prediction: Risk in Behavioral Assessment*, 54 J. CRIM. JUST. 12, 13 (2018).

101. Brian Abbott, *What is Science?*, in SCIENCE BENCH BOOK FOR JUDGES 5, 10 (Peggy Hora et. al. eds., 2019), <https://www.judges.org/wp-content/uploads/Chapter2-SBB.pdf>.

102. Faye S. Taxman & Amy Dezember, *The Value and Importance of Risk and Need Assessment (RNA) in Corrections & Sentencing: An Overview of the Handbook*, in HANDBOOK ON RISK AND NEED ASSESSMENT: THEORY AND PRACTICE 22, 37 (Faye S. Taxman ed., 2017).

103. Natalie Ram, *Innovating Criminal Justice*, 112 NW. U. L. REV. 659, 688 (2018).

104. Zachary Hamilton et al., *Customizing Criminal Justice Assessments*, in HANDBOOK ON RISK AND NEED ASSESSMENT: THEORY AND PRACTICE 536, 567 (Faye S. Taxman ed., 2017).

inaccurate tool offers little utility, with the resulting waste of resources potentially undermining any benefits the tool may otherwise offer.¹⁰⁵

Several reasons exist to suspect that tools may not perform as accurately as stakeholders might expect or desire. One is that many agencies take an off-the-shelf approach by adopting an available risk instrument and implementing it, and often without empirically testing the tool's ability to function adequately within their own populations.¹⁰⁶ Such a strategy is scientifically unreliable as a tool is not necessarily generalizable to other jurisdictions.¹⁰⁷ Tools are developed on distinct datasets and thus reflect the risk-relevant attributes of the sample(s) tested, which carries a limiting function: "Analytic methods can only work with the data given. Unmeasured heterogeneity across offenders, neighborhoods, and jurisdictions could possibly affect predictive accuracy."¹⁰⁸ Risk factors, even the more objective ones, are not necessarily universal.¹⁰⁹ Criminal acts and their correlates can vary depending on personal characteristics and experiences, times, geographies, sites, environments, and circumstances.¹¹⁰ In sum, judges and legal practitioners, even if they do not become experts in the science of risk assessment, should be mindful of "what it is the oracle is predicting" and "of how often this oracle is right and wrong."¹¹¹

Regrettably, the deployment of algorithmic risk tools is generally ahead of the empirical evidence to substantiate their abilities.¹¹² In other words, relatively few validation studies exist (at least in the public realm) to allow for a sufficient knowledge base to confirm that algorithmic tools offer a considerable degree of accuracy.¹¹³ The conclusions of authors of a recent meta-analysis review of risk tools across criminal justice, funded by the Council of State Governments Justice Center, are salient in these respects. The review was unable to verify that available

105. Richard Berk, *Accuracy and Fairness for Juvenile Justice Risk Assessments*, 16 J. EMPIR. LEG. STUD. 175 (2019).

106. Robert Werth, *Risk and Punishment: The Recent History and Uncertain Future of Actuarial, Algorithmic, and "Evidence-Based" Penal Techniques*, 13 SOC. COMPASS 1, 9 (2019).

107. James T. McCafferty, *The Importance of Counties: Examining the Predictive Validity of a State Juvenile Risk Assessment Instrument*, 55 J. OFFENDER REHABILITATION 377, 378 (2016).

108. James Hess & Susan Turner, *Accuracy of Risk Assessment in Corrections Population Management: Where's the Value Added?*, in HANDBOOK ON RISK AND NEED ASSESSMENT: THEORY AND PRACTICE 157, 162 (Faye S. Taxman ed., 2017).

109. Arthur Rizer & Caleb Watney, *Artificial Intelligence Can Make our Jail System More Efficient, Equitable, and Just*, 23 TEX. REV. L. & POL. 181, 192 (2019).

110. Keith Soothill, *Sex Offender Recidivism*, 39 CRIME & JUST. 145, 176 (2010).

111. Rhys Hester, *Risk Assessment Savvy: The Imperative of Appreciating Accuracy and Outcome*, 38 BEHAV. SCI. & L. 246, 248 (2020).

112. Seth J. Prins & Adam Reich, *Can We Avoid Reductionism in Risk Reduction?*, 22 THEORETICAL CRIMINOLOGY 258, 259 (2018).

113. Brandon L. Garrett & Megan Stevenson, *Open Risk Assessment*, 38 BEHAV. SCI. & L. 279, 280 (2020); John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, 93 WASH. L. REV. 1725, 1795 (2018); William D. Burrell, *Risk and Needs Assessment in Probation and Parole: The Persistent Gap Between Promise and Practice*, in HANDBOOK ON RISK AND NEED ASSESSMENT: THEORY AND PRACTICE 54, 80 (Faye S. Taxman ed., 2017).

tools were highly accurate, and further found that performances from any single tool varied in different settings and for types of failures predicted.¹¹⁴ The review concluded that for most of the tools, predictive validity statistics could only be ascertained from one or two studies, and that such studies tended to be carried out by the tool's own developers.¹¹⁵

Consistent therewith, a review of the literature on the two popular risk tools that were the focus of prior empirical studies (COMPAS and the PSA) revealed that the available validation studies were typically performed by employees, consultants, or research funding recipients of the tools' owners.¹¹⁶ Inherent conflicts of interest suggest diligence by external researchers to independently audit risk tool performance to offer critical insights into how well risk assessment tools actually perform in the field.¹¹⁷ My prior empirical research on risk assessment performance in field settings was meant to respond to such calls.

For example, investigations of the COMPAS general and violent recidivism scales (using one large sample) and the PSA failure to appear and general recidivism scales (three large samples) showed some promise whereby each was able to distinguish low, medium, and high-risk offenders in that the failure rates of released individuals increased in a linear fashion. That is, for the full sample sets, the failure rates increased as the ordinal prediction moved from low to medium to high risk.¹¹⁸ However, this result was not consistent across sub-samples as will be addressed farther below with respect to minority differences.¹¹⁹ In any event, the ability to rank groups of individuals as being more or less likely to fail is an exercise in comparing on a relative basis and not a measure of absolute accuracy.

Alternative metrics attest to a tool's ability to estimate actual failure rates. Several examples from the prior studies may be illustrative. One of the projects on the COMPAS violence risk scale found a true positive rate of 24 percent and a true negative rate exceeding 90 percent.¹²⁰ In other words, 24 percent of those predicted at high risk of violence actually were arrested for a new violent crime, while over ninety percent of those predicted at low or medium risk of violence were not arrested for a new violent crime.¹²¹ As another example, the PSA across samples

114. Sarah L. Desmarais et al., *Performance of Recidivism Assessment Instruments in U.S. Correctional Settings*, in HANDBOOK OF RISK/NEEDS ASSESSMENT TOOLS 3, 21–22 (Jay P. Singh et al. eds., 2018).

115. *Id.* at 20.

116. Melissa Hamilton, *Evaluating Algorithmic Risk Assessment*, 24 NEW CRIM. L. REV. 156, 181 (2021) [hereinafter Hamilton 3].

117. Alicia Solow-Niederman et. al., *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L. J. 705, 742 (2019); PARTNERSHIP ON AI, REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM 31 (2019), <https://www.partnershiponai.org/wp-content/uploads/2019/04/Report-on-Algorithmic-Risk-Assessment-Tools.pdf>.

118. Hamilton 3, *supra* note 116, at 197 tbl. 2; Melissa Hamilton, *Investigating Algorithmic Risk and Race*, 5 UCLA CRIM. JUST. L. REV. (forthcoming 2021) [hereinafter Hamilton 4]; Melissa Hamilton, *The Biased Algorithm: Disparate Impact for Hispanics*, 56 AM. CRIM. L. REV. 1553, 1565 (2019) [hereinafter Hamilton 5]; Melissa Hamilton, *The Classification and Predictive Validity of a Violence Risk Scale* (unpublished manuscript) (on file with author) [hereinafter Hamilton 6].

119. *See infra* note 141 and accompanying text.

120. Hamilton 6, *supra* note 118.

121. Hamilton 6, *supra* note 118.

performed poorly in predicting those who would fail to appear if released (i.e., a true positive rate of approximately 28 percent), but was much better at predicting those who would appear for their court dates (i.e., a true negative rate of 84–88 percent, depending on the sample).¹²² True positive rates such as these below 30 percent indicate significant error rates for high-risk predictions specifically. Still, these results with weak true positive rates do not mean that these tools are meaningless but suggest some caution. Judges using these scales might give much stronger weight to a prediction of success than of failure.

B. Demographic Equity

An important hope of advocates in adopting algorithmic risk is to improve the equity of pretrial release decisions where experience finds that certain minority groups suffer disproportionately in the state of mass incarceration.¹²³ Risk tool developers tend to characterize their tools as racially neutral: “Vendors promote [algorithmic] models to the public and to the agencies that use them as the answer to human bias, arguing that computers cannot harbor personal animus or individual prejudice based on race.”¹²⁴

Nonetheless, certain stakeholders have expressed concerns as to whether algorithmic tools are as fair as expected, particularly concerning the potential for adverse consequences to protected groups.¹²⁵ A coalition of over 100 groups self-described as comprising “civil and human rights, community, and data justice leaders from across the United States” (e.g., ACLU, NAACP, Electronic Frontier Foundation) signed “A Shared Statement of Civil Rights Concerns” expressing unease with fairness of risk assessment tools used in criminal justice.¹²⁶ The joint statement expressly calls for greater transparency and for third party audits of the underlying algorithms to check for race-based issues.¹²⁷

One potential negative attribute would exist if a tool systematically produced higher risk scores for minorities. Thus, if the tool predicts a greater proportion of one group to fail if released than another, the result could signify disparate impact on the higher scoring group.¹²⁸ Notably, disparate impact does not require

122. Hamilton 3, *supra* note 116, at 200 tbl. 12.

123. Laurel Eckhouse et al., *Layers of Bias: A Unified Approach for Understanding Problems with Risk Assessment*, 46 CRIM. JUST. & BEHAV. 185, 186 (2019).

124. *Id.*

125. Sandra Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218 (2019).

126. African American Ministers in Action et al., *The Use of Pretrial “Risk Assessment” Instruments: A Shared Statement of Civil Rights Concerns* 2 (2018), <http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf>.

127. *Id.* at 7.

128. Jennifer Skeem & Christopher Lowenkamp, *Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism*, 38 BEHAV. SCI. & L. 259, 262 (2020); Thomas Cohen & Christopher Lowenkamp, *Revalidation of the Federal PTRAs: Testing the PTRAs for Predictive Biases*, 46 CRIM. JUST. & BEHAV. 234, 235 (2019).

discriminatory motivation or intent¹²⁹ and, thus, applies to a facially neutral policy or practice,¹³⁰ such as a science-derived risk tool. “[A] transparent, facially neutral algorithm can still produce discriminatory results.”¹³¹ A reason to suspect that minorities will be predicted as riskier is that these tools commonly weigh criminal history heavily in their algorithms. Because statistics show that minorities in America tend to be burdened with more serious criminal records, tools that include criminal history predictors will tend to output higher risk scores for minorities.¹³² The few studies to date on the performance of algorithmic risk with minorities shows some evidence that minorities are more likely to be ranked at higher risk, though this result is not consistent across studies and for all tools.¹³³

Illustrative results from COMPAS and the PSA find some evidence for these concerns. A prior study of the COMPAS general recidivism tool in a large Florida county found that it classified as high risk 27 percent of black defendants compared to 11 percent of white defendants.¹³⁴ A research project on the PSA general recidivism scale likewise produced variations in that, at two separate sites, black defendants were far less likely to be classified in the low risk bin, about one-third of black defendants were assigned to the low risk bin compared to almost half of white defendants.¹³⁵ These results show evidence of disparate impact in real-world settings, confirming a cause for concern.

Another negative attribute might exist if an algorithmic tool is unable to produce similar accuracy rates for the different subsets of a population. For example, a risk assessment scheme that presumes risk tools are somehow universal, generic, and culturally neutral may well be flawed and thus not perform equally across minority groups.

The over or under estimation of risk that can ensue from this process is entirely plausible given (a) the potential omission of meaningful risk items specific to minority populations, (b) the inclusion of risk factors that are more relevant to White offenders, and (c) variation in the cross-cultural manifestation and expression of existing risk items.¹³⁶

A risk tool could yield unequal results for cultural minority groups if it fails to incorporate or otherwise consider their unique “behavioral practices and expectations, health beliefs, social/environmental experiences, phenomenology, illness narratives, deviant conduct, and worldview.”¹³⁷

129. Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 121–22 (2017); Mark MacCarthy, *Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms*, 48 CUMB. L. REV. 67, 80 (2017).

130. Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 554 (2018).

131. Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1024 (2017).

132. Skeem & Lowenkamp, *supra* note 128, at 267.

133. Threadcraft-Walker et al., *supra* note 100 (citing studies).

134. Hamilton 1, *supra* note 17, at 290 n.143.

135. Hamilton 4, *supra* note 118.

136. Stephane M. Shepherd & Roberto Lewis-Fernandez, *Forensic Risk Assessment and Cultural Diversity: Contemporary Challenges and Future Directions*, 22 PSYCHOL. PUB. POL'Y & L. 427, 429 (2016).

137. *Id.*

Drawing on the prior empirical studies may provide relevant context. Studies of COMPAS in the Florida jurisdiction revealed differences in predictive ability based on race and ethnicity.¹³⁸ COMPAS had a higher accuracy rate in predicting general recidivism for black defendants, but a lower accuracy rate for predicting non-recidivism, than white defendants.¹³⁹ In a separate analysis dividing the sample differently into two groupings of (a) Hispanics and (b) non-Hispanics (the latter combining whites, blacks and other), findings showed that the low, medium, and high risk bins held meaning for increasing rates of violent recidivism for non-Hispanic defendants, but not for Hispanic defendants.¹⁴⁰ For Hispanic individuals, the violent reoffending rate for medium risk was actually below that of the low risk bin.¹⁴¹ In other words, a single scheme of categorization was inappropriate: for the COMPAS violent risk scale a three risk bin strategy (i.e., low, medium, high) had meaning for non-Hispanic defendants, but for Hispanic defendants a two risk bin scheme (e.g., low versus high) appeared more appropriate. Another finding was that COMPAS frequently overpredicted the likelihood of general and violent recidivism of Hispanic defendants.¹⁴²

The prior study of the PSA similarly revealed differences in performance based on group status, but not consistently. Comparing overall accuracy rates in two separate samples for black versus white defendants on the new criminal offense scale, the PSA's ability to predict were similar for blacks and whites in one sample, but not in the other.¹⁴³ These results for the PSA indicate that racial equivalence may exist for the same tool in certain criminal justice sites, but results may vary by race to a statistically significant degree at other sites.

Algorithmic tools may be biased regarding gender as well. Most risk assessment tools contain what are purported to be "gender neutral" factors such as criminal history, age, mental health, and substance abuse.¹⁴⁴ Some observers have objected that purportedly neutral factors may not be so evenhanded in that many of these factors result in higher correlations to recidivism for men, meaning that the factors were far better predictors for males.¹⁴⁵ The prior study of COMPAS supported these suggestions in finding the tool systematically and significantly overpredicted both general recidivism risk and violent recidivism risk for women.¹⁴⁶

In the end, the presence of racial, ethnic, or gender differentials do not automatically signify that these more systematic risk assessment tools should be

138. See *infra* notes 139–142 and accompanying text.

139. Hamilton 1, *supra* note 17, at 286.

140. Hamilton 5, *supra* note 118, at 1565–66.

141. Hamilton 5, *supra* note 118, at 1565. For the general recidivism scale, while the low, medium, and high bins indicated linearly greater rates for non-Hispanics, the rates at the medium and high bins for Hispanics were similar. *Id.*

142. Hamilton 5, *supra* note 118, at 1566.

143. Hamilton 4, *supra* note 118.

144. Patricia Van Voorhis et al., *Women's Risk Factors and Their Contributions to Existing Risk/Needs Assessment*, 37 CRIM. JUST. & BEHAV. 261, 262 (2010).

145. VALERIE R. BELL, GENDER-RESPONSIVE RISK ASSESSMENT IN CORRECTIONS 108 (2014).

146. Melissa Hamilton, *The Sexist Algorithm*, 37 BEHAV. SCI. & L. 145, 152 (2019) [hereinafter Hamilton 7].

entirely banned. Stakeholders might need to consider the reasonableness of tradeoffs between demographic injustice in the algorithms and the benefits from safely reducing rates of detention. Notably, increasing release rates across a jurisdiction may in other ways benefit protected groups.¹⁴⁷

C. Levels of Risk

Risk tool outcomes often use risk bins to classify the population into ordinal rankings. Descriptive labels are commonly applied, such as low, medium, and high risk.¹⁴⁸ Still, such categorical labeling is not consistent across tools as their numbers (such as between three and six) and meanings (in terms of the degree of likelihood of failure associated therewith) vary. This sort of simplistic system provides an easy way to partition the population as more or less risky. However, it is important to note that these are merely relative categorizations of the particular population evaluated.

In terms of absolute estimates, no consensus exists as to what categorical bins mean as the percentile probabilities of failure.¹⁴⁹ Clinicians have no commonly agreed definition of risk categories,¹⁵⁰ statisticians have no accepted metric, and there are no normative or legal distinctions for such labels.¹⁵¹

Individual risk tools may provide detailed context for each of its own categories (e.g., provide a definition and corresponding recidivism rate estimate), but these depictions will be unique to the specific tool. The recidivism rate associated with the “high risk” group for one tool could be 70 percent, while in another tool “high risk” might indicate an observed failure rate of 15 percent. Hence, assumptions that even a “high risk” label signifies a serious risk of failure, much less a more likely than not presumption of failure, are erroneous. In pretrial, then, judges should be cognizant of the probabilities that are relevant in their jurisdictions with the specific tool deployed.

Referring to the prior research again for illustration purposes, the study of the COMPAS general recidivism tool indicated a significant failure rate for those classified as “high risk,” as over 70 percent were rearrested.¹⁵² However, there were significant ethnic differences when comparing Hispanic versus non-Hispanic defendants: with COMPAS high risk classifications, 57 percent of Hispanic defendants were rearrested compared to 75 percent of non-Hispanics

147. Skeem & Lowenkamp, *supra* note 128, at 260.

148. Ashley B. Batastini et al., *Communicating Violence Risk During Testimony: Do Different Formats Lead to Different Perceptions Among Jurors?*, 25 PSYCHOL. PUB. POL'Y & L. 92, 93 (2019).

149. Daniel A. Krauss et al., *Risk Assessment Communication Difficulties: An Empirical Examination of the Effects of Categorical Versus Probabilistic Risk Communication in Sexually Violent Predator Decisions*, 36 BEHAV. SCI. & L. 532, 534 (2018).

150. Daniel J. Neller & Richard I. Frederick, *Classification Accuracy of Actuarial Risk Assessment Instruments*, 31 BEHAV. SCI. & L. 141, 142 (2013).

151. J.C. Oleson et al., *Training to See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessments Among Federal Probation Officers*, 75 FED. PROB. 52, 55 (2011).

152. Melissa Hamilton, *Risk Assessment Tools in the Criminal Legal System: Theory and Practice* 33 tbl. 3 (2020), <https://www.nacdl.org/getattachment/a92d7c30-32d4-4b49-9c57-6c14ed0b9894/riskassessmentreportnovember182020.pdf> [hereinafter Hamilton 8].

defendants.¹⁵³ In studying the COMPAS violent recidivism scale regarding gender differences, a high risk classification of violent offending yielded a 52 percent failure rate for men, whereas the failure rate for women was half of that at 25 percent, suggesting a gendered disparity in absolute risk despite the same risk label assigned.¹⁵⁴ In both cases, the stigmatizing label of high risk in the same samples meant a dramatically different likelihood of reoffending based on Hispanic ethnicity and gender.

The studies into the PSA failure to appear tool likewise found that categorical labels had no consistent meaning in the field. While the high-risk bin achieved relatively equivalent failure rates across sites (about 28 percent), the low and medium risk bin failure rates each differed to a statistically significant degree.¹⁵⁵ In other words, the predictions of low and medium risk with this PSA scale were associated with different proportions of failure depending on the jurisdiction.

The foregoing results, in addition to showing evidence that categorical risk bins are not necessarily associated with equivalent rates of failure, also highlight another issue. Notice the results specifically for the COMPAS violent tool in scoring women with the PSA failure to appear tool (for both genders combined) in that they yielded relatively modest percentages in its high-risk bin.¹⁵⁶ A COMPAS prediction of high risk of violence was associated with a 25 percent rate of violent arrest for women, while a PSA prediction of high risk of failure to appear was associated with a 28 percent failure rate.¹⁵⁷ These findings exemplify the caution about any presumption that an algorithmic “high risk” outcome suggests the individual is highly likely to fail if released in percentage terms. A high-risk labeling may not mean in real terms that the individual is more likely than not to fail. This sentiment has policy relevance. The goal of reducing mass incarceration could be thwarted if a “high-risk” prediction itself justified detention because the vast majority of individuals so labeled, at least in these real-world examples, would likely have succeeded if released.

Overall, the foregoing examples for COMPAS and the PSA highlight the need for more clarity in the algorithmic risk field about how risk categorizations are used and the particular failure rates that are associated with them. These meanings vary by tool, setting, type of failure event, and for sub-groups.

V. CONCLUSIONS

The general comments from the institutional creator of the PSA—used as a case example herein—demonstrate the goals that proponents seek in advocating the implementation of algorithmic risk tools to facilitate release decisions.

From the beginning, we believed that an easy-to-use, data-driven risk assessment could greatly assist judges in determining whether to

153. Hamilton 5, *supra* note 118, at 1565.

154. Hamilton 7, *supra* note 146, at 150.

155. Hamilton 4, *supra* note 116.

156. Hamilton 8, *supra* note 154.

157. Hamilton 8, *supra* note 152; Hamilton 5, *supra* note 118.

release or detain defendants who appear before them. And this could be transformative. In particular, we believed that from switching from a system based solely on instinct and experience to one in which judges have access to scientific, objective risk assessment tools could further our central goals of increasing public safety, reducing crime, and making the most effective, fair, and efficient use of public resources.¹⁵⁸

In the end, overly optimistic—or overly pessimistic—views on the role of algorithmic risk tools to help reduce mass incarceration through pretrial jail detentions are not confirmed by the state of the scientific studies. Trade-offs are necessary when attempting such a significant reform in criminal justice that impacts many lives and institutions. The existing evidence suggests that continued efforts to improve algorithmic tools are justified. Stakeholders are encouraged though to also ramp up efforts to fund independent researchers to inform on how well the tools actually perform and to suggest ways to reduce any flaws discovered. Tackling mass incarceration through algorithmic risk requires a multidisciplinary engagement involving the law and empirical sciences.

158. *Developing a National Model for Pretrial Risk Assessment*, ARNOLD FOUND., 2 (2003), <https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=794208cc-3e92-23fe-40a7-6f02885b01a0&forceDialog=0>.