Assessing Algorithmic Risk Assessments:
Challenges in Transparency and Oversight

Basil Hariri

Executive Summary

Algorithmic risk assessments are a relatively new and controversial tool used to predict negative criminal justice outcomes in many stages of the American incarceral system. These risk assessments are generally developed by corporations and used by county and state jurisdictions when making case-related decisions such as pretrial detention status, sentence length, and guilt. There is no discernable difference in accuracy between algorithmic risk assessments and the analogous actuarial tools currently in use; however, algorithmic risk assessments are highly opaque to the public and those who use them with little to no official oversight. The current lack of transparency, oversight, and any notable improvements over their more common actuarial counterparts raises concerns regarding the adoption of algorithmic risk assessments.

Risk Assessments in the Criminal Justice System

Risk assessment instruments (RAIs) are used in almost every stage of the American criminal justice system. They distill several factors of an individual case into a risk score that can be used "to assess an individual's risk of reoffending (or noncompliance with justice requirements)" (*What is Risk Assessment*, n.d.). They are often used by corrections officers to assign appropriate programming to inmates in incarceration; by probation and parole officers to determine levels of supervision post-release; and by judges in determining pretrial detention, sentencing length, and even guilt (*What is Risk Assessment*, n.d.; *AI in the Criminal Justice System*, n.d.). Judges and incarceral practitioners use risk assessments to standardize their treatment of similar cases.

A Short History of Risk Assessments

In the early 1900s, criminal justice RAIs began as informal clinical or correctional staff interviews with and examinations of defendants to determine their likelihood of reoffending or judicial noncompliance. These assessments were inconsistent and heavily dependent on the staff involved, their training, and their biases (*History of Risk Assessment,* n.d.).

In need of a standardized protocol for assessing risk, the Illinois parole system implemented the first actuarial risk assessment tool in the United States criminal justice system in 1933 (Burgess, 1936). Instead of relying on professional opinions, the Illinois tool examined factors of individual cases that correlated with violations of parole. It based defendants' risk on "previous criminal record, previous work record, whether married or single, conduct in the institution and so forth" (Burgess, 1936). Actuarial RAIs (like the Illinois parole tool) predict defendant risk by fitting the facts of a case to statistical models built on historical data.

Over the past century, risk assessment tools have been shown to reduce pretrial detention rates, improve efficiency of heuristics, and improve outcomes in recidivism and noncompliance with justice requirements (About the Public Safety Assessment, n.d.). Early successes in these criteria likely influenced the adoption of risk assessments. Today, over 60% of Americans live in a jurisdiction that uses some form of risk assessment for pretrial detention, the majority of which are actuarial (*How Many Jurisdictions,* n.d.).

Algorithmic Risk Assessments

Algorithmic RAIs are a recent type of risk assessment that generate a risk score by training a machine learning algorithm on historical criminal justice data. To create an algorithmic RAI, software developers create a relatively naive algorithm, decide what criminal justice data to train it with, then feed that historical data into the program. As it ingests historical data (including recidivism and failure-to-appear outcomes), the algorithm begins to make associations between variables in the data and the outcomes of interest. The model's ability to make predictions about new input data becomes more powerful as it is given more historical data (similar to classical statistical methods). Once an algorithmic RAI has been trained on historical cases, the facts of a current case are input and the model generates a risk score for practitioners to use.

Ultimately, algorithmic risk assessments fill the same role as actuarial assessments, but they replace standard statistics methods (e.g. multivariate regressions) with methods based on recent advancements in computer science and data analysis.

Current Use in Criminal Justice

The largest and most widely used algorithmic RAI in the United States today is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm, which is developed and owned by the private corporation Equivant[1] (*State v. Loomis, 2017*). The COMPAS system is used in at least 11 counties in the states of New York, Wisconsin, and California; with at least 4.3 million people currently living in jurisdictions that use it (*How Many Jurisdictions*, n.d.; Kirkpatrick, 2017). The COMPAS system includes a pretrial release scale, a general recidivism scale, and a violent recidivism scale which can be used at different points in the criminal justice process (Northpointe, 2015).

Other notable assessments include CommandCentral Predictive, HunchLab, PredPol, and Patternizr each of which offers a variety of products and services (Chohlas-Wood, 2020; Kirkpatrick, 2017). These risk assessment tools are largely marketed to and used by police departments in a growing trend known as "predictive policing" (Lau, 2020). While differing in implementation and specific offerings, these systems generally use "a combination of historical crime data and other data (time of year, day of the week, proximity to bars, lighting, weather, etc.) to generate a forecast" of where, when, and by whom future crimes will be committed (Cheetham, 2019).

To date, algorithmic RAIs show no improvements in performance over actuarial assessments currently in use. Research indicating high levels of success from algorithmic RAIs tend to emphasize general success rates or compare the results to using no risk assessment at all (Kirkpatrick, 2017). When compared to actuarial RAIs, the COMPAS algorithm does not predict outcomes better than simple actuarial tools currently in use: "[D]espite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear classifier with only two features" (Dressel & Farid, 2018).[2]

---

[1] Formerly known as Northpointe.

[2] Performance data on other algorithmic risk assessments is largely unavailable or difficult to compare to other RAIs

Trade-Offs in Accuracy and Equity

Given that Black bodies are more likely to be punished at every stage of the American incarceral system, risk assessments based on that system will necessarily suggest that Black people are at higher risk of recidivism, failure to appear, and other negative criminal justice outcomes (Spohn, 2011). Creating a perfectly accurate assessment based on historically racist data will inherently provide racist predictions. This basis on racist data may allow an otherwise "race-neutral" tool to reify historically racist policy and outcomes.

Neutralizing the influence of race on risk scoring is made difficult by the strong correlation between race and poverty, homelessness, likelihood of being stopped by police, and most other predictors of arrest (Creamer, 2020; Pierson et al., 2020; Wiltz, 2019). Alternatively, incorporating race data and later reducing the scores of communities traditionally targeted by police may sacrifice some degree of public safety. Assessments based on racist historical data eventually must make tradeoffs between public safety and equity. After consideration of the tradeoffs between public safety and fairness in algorithmic risk assessments, Corbett-Davies and colleagues (2017) conclude that:

> Maximizing public safety requires detaining all individuals deemed sufficiently likely to commit a violent crime, regardless of race. However, to satisfy common metrics of fairness, one must set multiple, race-specific thresholds. There is thus an inherent tension between minimizing expected violent crime and satisfying common notions of fairness. This tension is real: by analyzing data from Broward County, we find that optimizing for public safety yields stark racial disparities; conversely, satisfying past fairness definitions means releasing more high-risk defendants, adversely affecting public safety.

Corbett-Davies et al. recognize that public safety and racial equity are fundamentally at odds in current risk assessment methodology because historical data reflects racist policing and incarceral practices. Intentionally or not, any risk assessment based on racist data ultimately makes trade offs between these two priorities, and algorithmic risk assessments are no exception.

Litigation and Skepticism: Loomis v. Wisconsin (2016)

Loomis v. Wisconsin, a 2016 Wisconsin Supreme Court case decision, started a nationwide debate about the use of algorithmic RAIs in criminal sentencing. The case focused on the fact that the COMPAS algorithm's code is a trade secret and is not shared with anyone not employed by Equivant (including defendants, defense attorneys, prosecutors, and judges). COMPAS users input the facts of the case and receive a risk score with no additional information or context.

Before the COMPAS system's assessment was introduced to his case, Loomis made a plea deal with prosecutors for one year in county jail with probation (Washington, 2019). At his trial, the judge referenced the algorithm's risk score when sentencing him to six years in prison with five years of extended supervision. Loomis appealed this ruling on the grounds that his right to due process was denied because he was not allowed to examine the algorithm used in his sentencing.

Ultimately, the Wisconsin Supreme Court upheld the initial sentence because "COMPAS uses only publicly available data and data provided by the defendant" (*State v. Loomis*, 2017). Even though he could not examine the code of the COMPAS system, the Court ruled that his due process rights had not been violated because Loomis could examine the input data that it used to generate his risk score.

In their decision, the Court noted some skepticism towards algorithmic RAIs and prescribed guidelines for their use going forward. However, some legal scholars expect that these guidelines will be ineffective because they "[fail] to specify the vigor of the criticisms of COMPAS, [disregard] the lack of information available to judges, and [overlook] the external and internal pressures to use such assessments" (*State v. Loomis*, 2017).

Lack of Transparency

Loomis' case centered around issues with transparency. These issues threaten the ability of defendants to understand why a risk assessment assigned them a particular risk score. Transparency in government is a virtue widely accepted and promoted by American politicians of all stripes, including both the Obama and Trump administrations (Exec. Order No. 13892, 2019; Obama, 2009). In the context of courts specifically, Voermans (2007) notes that "openness of proceedings, especially public hearings, rulings and verdicts, traditionally constitutes an important part of the right to a fair trial."

In response to the Loomis decision, Park (2019) argues that "trade secret law should not serve to bar defendants from raising and investigating valid due process questions…the law and policy governing trade secrets must be reformed to account for the individual rights and social interests at stake." In other words, Park posits that algorithmic RAI trade secrets threaten due process by reducing transparency in the court system.

Currently, algorithmic risk assessments suffer from both explicit and implicit transparency issues described below.

Explicit Transparency

In *Loomis v. Wisconsin*, the Wisconsin Supreme Court affirmed Equivant's right not to share its code with court practitioners and the defendant in order to protect its trade secrets (*State v. Loomis, 2017*). Because Equivant does not have to disclose its code, part of the mechanism by which some Americans are denied pretrial release or sentenced to jail today is completely opaque to both the defendant and the judge (or jury) making those decisions. This dynamic is true of any algorithmic risk assessment that does not fully disclose its assessment's code and the historical data on which it was trained.

Improvements to explicit transparency can be made by requiring (by law or policy) that practitioners only use algorithmic RAIs with code and training data made fully accessible to all involved parties.[3]

Implicit Transparency

While a powerful technology, machine learning algorithms are limited in that neither they nor their developers can fully explain the reasoning behind a generated prediction. Machine learning "predictive models can be such complicated functions of the variables that no human can understand how the variables are jointly related to each other to reach a final prediction" (Rudin

[3] No major algorithmic risk assessments currently do so. PredPol has published a portion of its model, but does not disclose its implementation or the machine learning involved (Predictive Policing Technology, n.d.).

& Radin, 2019). Any risk score generated by an algorithmic RAI is the product of both the algorithm's design and the historical data it was trained on, making it extremely difficult to separate the influence of the algorithm from the influence of the historical data. Because of this difficulty, explaining the reasoning behind the decisions made by current algorithmic risk assessments is not feasible.

Achieving internal transparency would require that the developers of these algorithmic RAIs create them in such a way that generating risk scores can be audited. This requirement conflicts with modern machine learning algorithm design and would likely require reworking current algorithmic RAIs (Jordan & Mitchell, 2015).

Lack of Oversight

Oversight of an institution or process hopes to ensure it complies "with applicable policies, laws, regulations, and ethical standards" (*What is Oversight,* n.d.). Berk et al. (2018) contextualize the need for oversight of algorithmic risk assessments aptly:

> [I]n the end, it will fall to stakeholders—not criminologists, not statisticians, and not computer scientists—to determine the trade-offs. How many unanticipated crimes are worth some specified improvement in conditional use accuracy equality? How large an increase in the false negative rate is worth some specified improvement in conditional use accuracy equality? **These are matters of values and law, and ultimately, the political process. They are not matters of science.**

They describe the tension between accuracy and equity in risk assessments, and mark their belief that resolving that tension is a matter of values, law, and politics rather than science.

Currently, values, law, and politics are not explicitly required of the development of algorithmic risk assessments, and developers are not beholden to practitioners or the public. As noted above, only those who build these assessments know how they predict risk. Without judicial or legal oversight, the decisions of unelected and unknown individuals influence tools used to determine pretrial release, verdict, sentencing length, and other impactful outcomes for individuals in the criminal justice system.

Successfully addressing complex social, historical, and political issues like racial equity in any risk assessment will require substantial thought and discussion. However, for current algorithmic risk assessments, that thought and discussion is done behind closed doors.

Conclusion: Unclear Need for Algorithmic Risk Assessments

Today, algorithmic risk assessments provide no advantage over actuarial assessments while introducing several concerns. Current providers of algorithmic RAIs reduce transparency by preventing judicial practitioners and participants from examining their code (explicit transparency) and by failing to distinguish between the influence of the algorithm and the data on which it was trained (implicit transparency). Additionally, jurisdictions using algorithmic risk assessments give unelected engineers and corporate entities influence over their decision making without requiring judicial or legal oversight. Although these issues remain unresolved, millions of Americans currently live in jurisdictions using algorithmic risk assessments.

Frequently Asked Questions

What is a risk assessment?

Risk assessments are used in many parts of the criminal justice system. Depending on the context, they can be used to predict the risk of an individual committing another crime, failing to appear in court, breaking parole requirements, or some other undesirable criminal justice outcome.

What are the different types of risk assessments?

Clinical risk assessments are risk assessments where a professional (e.g., therapist) makes an assessment of risk based on interviews with and observations of the defendant. Clinical risk assessments are used less often today than in the past.

Actuarial risk assessments are risk assessments based on statistical analysis of historical data. They look at many past cases to predict the expected risk involved in current cases. Actuarial risk assessments are widely used and the most popular form of risk assessment in the criminal justice system.

Algorithmic risk assessments are risk assessments based on the algorithmic analysis of historical data. These differ from actuarial risk assessments in that the analysis is conducted by a computer program that was given the historical data, built a model based on that data, and makes predictions using that model. Algorithmic risk assessment is a relatively new form of risk assessment that has grown in popularity in the criminal justice system.

Who uses risk assessments?

Risk assessments are used by judges and other criminal justice practitioners to determine pretrial detention, verdict, sentencing length, parole requirements, etc. Risk assessments are not used as the only piece of information in these decisions; most often they are considered as one of many factors.

What is the COMPAS system?

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system is the most popular algorithmic risk assessment being used in criminal justice today. It is used in several jurisdictions and was the focus of *Loomis v. Wisconsin.*

What are the primary benefits of algorithmic risk assessments?

Algorithmic risk assessments show promising predictive power, but they are currently no more accurate than commonly used actuarial risk assessments. This accuracy may increase as the technology advances. Better predictions will allow practitioners to place lighter restrictions on low and moderate-risk individuals.

What are the primary concerns about algorithmic risk assessments?

The two primary concerns with algorithmic risk assessments are a lack of transparency and a lack of oversight. The code used in algorithmic risk assessments is not required to be shared with those using it. Judges may use the score generated by algorithmic risk assessments to make decisions about a defendant without understanding why that score was assigned. Additionally, the code does not need to be audited by any public agency or elected official. This means that the score generation process is not accountable to the taxpayers affected by them or the government officials using them.

Are algorithmic RAIs better than current actuarial RAIs?

Algorithmic risk assessments currently provide no performance advantages over their actuarial counterparts. On the other hand, they present substantial concerns relating to a lack of transparency and oversight that actuarial risk assessments do not have.

References

About the Public Safety Assessment. (n.d.). *Advancing Pretrial Policy & Research (APPR)*. Retrieved March 9, 2022, from https://advancingpretrial.org/psa/research/

AI in the Criminal Justice System. (n.d.). *EPIC - Electronic Privacy Information Center*. Retrieved March 9, 2022, from https://epic.org/issues/ai/ai-in-the-criminal-justice-system/

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2018). Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research*, *50*(1), 3–44. https://doi.org/10.1177/0049124118782533

Burgess, E. W. (1936). Protecting the Public by Parole and by Parole Prediction. *Journal of Criminal Law and Criminology*, *27*(4), 491. https://doi.org/10.2307/1137495

Cheetham, R. (2019, January 23). Why We Sold HunchLab. *Azavea*. https://www.azavea.com/blog/2019/01/23/why-we-sold-hunchlab/

Chohlas-Wood, A. (2020, June 19). Understanding risk assessment instruments in criminal justice. *Brookings*. https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017, June 9). Algorithmic decision making and the cost of fairness. *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3097983.309809

Creamer, J. (2020, July 15). *Inequalities Persist Despite Decline in Poverty For All Major Race and Hispanic Origin Groups*. United States Census Bureau. https://www.census.gov/library/stories/2020/09/poverty-rates-for-blacks-and-hispanics-reached-historic-lows-in-2019.html

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1). https://doi.org/10.1126/sciadv.aao5580

D'Alessio, S. J., & Stolzenberg, L. (2003). Race and the Probability of Arrest. *Social Forces*, *81*(4), 1381–1397.

Exec. Order No. 13892, 3 C.F.R 55239 (October 15, 2019). https://www.federalregister.gov/documents/2019/10/15/2019-22624/promoting-the-rule-of-law-through-transparency-and-fairness-in-civil-administrative-enforcement-and

Griffard, M. (2019). A Bias-Free Predictive Policing Tool?: An Evaluation of the NYPD's Patternizr. Fordham Urban Law Journal, 47(1), 42. https://ir.lawnet.fordham.edu/ulj/vol47/iss1/2/

History of Risk Assessment. (n.d.). Bureau of Justice Assistance. Retrieved March 9, 2022, from https://bja.ojp.gov/program/psrac/basics/history-risk-assessment

How Many Jurisdictions Use Each Tool? (n.d.). Mapping Pretrial Injustice. Retrieved March 9, 2022, from https://pretrialrisk.com/national-landscape/how-many-jurisdictions-use-each-tool/

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kirkpatrick, K. (2017). *It's Not the Algorithm, It's the Data*. *60*(2), 21–23. https://doi.org/10.1145/3022181

Lau, T. (2020, April 1). *Predictive Policing Explained*. Brennan Center for Justice. https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained

Obama, B. (2009, January 21). *Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government*. The White House. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government

Park, A.L. (2019, February 19). *Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing*. UCLA Law Review. https://www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745. https://doi.org/10.1038/s41562-020-0858-1

Predictive Policing Technology. (n.d.). *PredPol*. Retrieved March 14, 2022, from https://www.predpol.com/technology/

Northpointe. (2015, March 19). *Practitioner's Guide to COMPAS Core*. https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf

Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Spohn, C. (2011). Race, Ethnicity, and Crime. *The Oxford Handbook of Crime and Criminal Justice*. https://doi.org/10.1093/oxfordhb/9780195395082.013.0011

State v. Loomis: Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing. (2017, March 10). Harvard Law Review. https://harvardlawreview.org/2017/03/state-v-loomis/

Voermans, W. (2007). Judicial transparency furthering public accountability for new judiciaries. *Utrecht Law Review*, *3*(1), 148–159. https://doi.org/10.18352/ulr.42

Washington, A. L. (2019). How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate. *The Colorado Technology Law Journal*, *17*(1), 37. https://ssrn.com/abstract=3357874

Wiltz, T. (2019, March 29). ‘A Pileup of Inequities’: Why People of Color Are Hit Hardest by Homelessness. *The Pew Charitable Trusts*. https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2019/03/29/a-pileup-of-inequities-why-people-of-color-are-hit-hardest-by-homelessness

What Is Oversight and How Does it Relate to Governance? (n.d.). Canadian Audit and Accountability Foundation. Retrieved March 14, 2022, from https://www.caaf-fcar.ca/en/oversight-concepts-and-context/what-is-oversight-and-how-does-it-relate-to-governance

What Is Risk Assessment. (n.d.). Bureau of Justice Assistance. Retrieved March 9, 2022, from https://bja.ojp.gov/program/psrac/basics/what-is-risk-assessment