Bail or Jail? Judicial versus Algorithmic Decision Making in the Pretrial System

**Doaa Abu Elyounes**

Hardvard Law School

Berkmen Klein Center for Internet and Society

dabuelyounes@sjd.law.harvard.edu

**Abstract:** …

# Introduction

450,000 presumably innocent people are held in jail awaiting trial on any given day in the United States, a world leader in the total number of pretrial detainees.[[1]](#footnote-1) The pretrial phase begins when a judicial officer or grand jury determines that there is probable cause to support a criminal charge; and it ends when the charge is adjudicated or dismissed. Based on only two factors- (1) the risk that the defendant would fail to attend his/ her trial; and (2) the risk that the defendant will commit additional crime while awaiting trial; the judge decides whether the defendant will be released on his or her recognizance, released on conditions such as money bail, or community supervision; or if the defendant will be detained until the end of the trial. Preventive detention should be the exception and not the rule, and it is saved to the riskiest of defendants and the severe offences.[[2]](#footnote-2) Money bail is by far the most common condition of release. Nationwide, 9 out of 10 felony defendants who were detained pretrial had bail set and would have been released if they had posted it.[[3]](#footnote-3) Some jurisdictions use “bail schedules” a fixed amount that is attached to each offence in the criminal code, and in other jurisdictions the bail amount is subject to the discretion of the judge.[[4]](#footnote-4) Minority groups such as blacks, Latinos and native Americans are twice as likely to be detained in pretrial because they cannot afford the money bail set in their case. Usually these minorities face higher bail amount in comparison to white defendants with similar charges.[[5]](#footnote-5)

In recent years, criminal justice activists, judges, politicians and the general public, have been raising a strong voice against the outrages system that detains the poor, favor release for the rich; and increase racial gaps; by calling for a major bail reform.[[6]](#footnote-6) Bail reform often include two tails: one focuses on abolishing money bail practices and the second focuses on assessing the risk that defendants pose in a more empirical way and moving toward actuarial risk assessment tools. This paper will focus on the latter tail.

In order to differentiated between low, medium and high risk defendants; and to make sure that only those who pose a risk to public safety or might flea are detained, judges are using actuarial tools that meant to eliminate human biases and to rationalize the decision making process by summarizing all relevant information in a more efficient way than the human brain.[[7]](#footnote-7) The National Institute of Justice found that when an actuarial risk assessment tool is being used there is a significant reduction in jail population; and that more defendants were released in pretrial.[[8]](#footnote-8) Despite that, actuarial risk assessment tools are portrayed in a very negative way by the media and academics due to increased computational capabilities of the tools and the fear that they will take over such a sensitive decision about an individual in an opaque and irreversible way. Title such as “AI is sending people to jail-and getting it wrong”[[9]](#footnote-9) or “Courts are using AI to sentence criminals. That must stop now”[[10]](#footnote-10) are making their way to the front pages of the popular media without a proper explanation about the nature of these tools and their actual capabilities.

This paper aims to examine the hype and analyze whether deploying AI in the existing risk assessment tools will make the fears addressed in the media come true, or perhaps it will improve our criminal justice system. The goal of this paper is to shed some light on the way machine learning based algorithms operate; and what would be the implications of deploying them in the criminal justice system. The paper looks at the majority of risk assessment tools used across the country. It compares between 7 tools, used in more than 80 U.S. jurisdictions among them are 5 full states, big cities like New York and Chicago, mid size cities and small counties. 5 tools are based on traditional regression analysis and 2 have a certain component of machine learning. The paper concludes that, despite the hype, robots are very far from replacing judges, and in fact, there are more differences than similarities between tools based regression analysis and tools based machine learning. The adaptation of machine learning to the criminal justice is not easy. It requires policy makers and computer scientists to define together terms like fairness, transparency, accuracy and efficiency. The implementation require all actors in the field to balance the tradeoffs between the definitions and to determine complex moral, legal, and socio-technical questions. However, as of now, given the limited capabilities of machine learning, adding a component of machine learning to the existing tools, if done properly, is unlikely to harm due process or equal protection, on the contrary it has the potential to improve decision making in pretrial. Capitalizing on the strength of machine learning, the score given to each defendant can be more personalized and meaningful for all actors in the field. Judges will be able to understand better the risk that the defendant pose, and to impose the suitable conditions to mitigate this risk. The defendant too, will have a better understanding of the score and will be able to appeal it on a more meaningful grounds if needed

The paper will proceed as follows: first, the unique characteristics of the pretrial system in the U.S. and the legal framework that govern this regime will be discussed. Next, the paper describes the evolution of risk assessment in the criminal justice system, from the past to the future. This chapter conducts a comparison between regression analysis (the traditional statistical method used in the most tools as of today);[[11]](#footnote-11) and machine learning. The goal of this comparison is to provide the readers with no technical background some elementary knowledge about the differences between the techniques. In the next chapter, the article addresses the seven most commonly used risk assessment tools across the country. The first five tools are based on regression analysis and these are: a federal tool, a tool developed by a nonprofit and widely distributed; and three local tools developed by the states Virginia, Colorado and Ohio. The last two tools are more technologically advanced, these are the COMPAS tool developed by a for-profit company; and a tool that is not being used in practice yet, but it is based on machine learning and was developed in an academic environment. This will be the layout for the following chapter which is the core of this paper. Building on prior research from the field, this chapter will identify the most pressing policy issues that policy makers need to take into account when implementing a risk assessment tool; and it will assess the seven tools according to them. The goal of this assessment is to examine the performance of the seven tools in regard to each policy matter.

Due to inherent differences between the way regulation and policy are designed and the way mathematical algorithms operate, these tools can be more opaque, less transparent and less explainable. But these concerns could be addressed with proper safeguards and with full understanding of the technical component. The implementation will not be easy, but pretrial is a good place to start with because it is a short stage, that have simple and clear goals, it involves relatively straightforward legal questions and its outcomes are quicker and easier to measure.[[12]](#footnote-12) The only way to ensure that AI based risk assessment tools will benefit defendants as well as the system, is through a collaboration between experts from different disciplines. All actors in the field must recognize that adjusting the legal standards to the computing capabilities and vice versa is needed.

# Pretrial in the U.S.

Each year, almost 12 million defendants are admitted into local jail across the country, 60% of them are awaiting trial, and the majority of them will eventually be charged with low level, nonviolent offences.[[13]](#footnote-13) When a person is arrested, the court must determine whether the person will be unconditionally released pending trial, released subject to a condition or combination of conditions (such as money bond, community supervision, enrolling in a substance treatment program); or held in jail in a preventive detention. The type of conditions that each defendant is subject to depends on their own risk level. In most jurisdictions, the types of considerations that judges are allowed to take into account are failure to appear in court (whether the defendant will attend the actual trial); and public safety (whether the defendant will commit a crime while awaiting trial).[[14]](#footnote-14) Preventive detention should be a rare instance, in which a judge will determine that there is no condition or combination of conditions that can adequately address those risks. However, in reality, many defendants are being held in jail not because of the risk they pose to the community, but because they are simply poor. Precisely, 34% of pretrial detainees across the country are there because they cannot afford posting their bail.[[15]](#footnote-15) Traditionally bail determinations are made based on judges’ personal recognizance guided by the law or on fixed bail schedules, which provide pre-determined bail amounts based on current charges. In some cases, the defendant will be released only if they pay the full bail amount or certain percentage; and in other cases the obligation is to pay the amount only if the defendant will fail in pretrial.[[16]](#footnote-16) If the defendant cannot afford the money bail set in their case, a bondsman will post a surety bond on their behalf and charge the defendant a certain amount which is nonrefundable in exchange for taking on the risk that the defendant will fail to appear.[[17]](#footnote-17) As a result, the for-profit bond industry is a 2-billion dollar per year industry, which “exploit” minorities and poor communities.[[18]](#footnote-18) In recent years, bail reform movements urged policy makers to abolish the practice of money as a sole condition of release in return to freedom. As of now, New Jersey and California passed laws that eliminated money bail completely, and other states such as Massachusetts, Alaska, Illinois, Connecticut and Colorado put restrictions on the type of offences that judges can assign money bail or required the judges to consider the defendant’s ability to post the bail.[[19]](#footnote-19) other jurisdictions such ad Kentucky and Washington D.C. eliminated the commercial bail bond industry.[[20]](#footnote-20) Similar initiatives are happening also on the local level, by district attorneys who decided not to ask for bail in certain cases; and grassroots organizations that are fighting the phenomena in courts. It is important to mention that reducing the reliance on money bail is only one approach that jurisdictions can implement but there are other alternatives such as investing in pretrial services, finding cost-effective ways to increase court appearance; and implementing risk assessment tools.[[21]](#footnote-21) While this paper acknowledges that only a holistic approach that takes into account the different channels of reform will be truly effective, it focuses on the latter solution, investing in actuarial risk assessment tools.

## The Consequences of Awaiting Trial in Jail

Preventive detention entails many consequences for the defendant and society:

* Detention during pretrial is correlated with receiving longer sentence: a study using data from state courts, found that defendants who were detained for the entire pretrial period were at least three times more likely to be sentenced to jail or prison than defendants who were released at some point pending trial. And their sentences were significantly longer.[[22]](#footnote-22) A separate study found similar results in the federal system. The study analyzed 1,798 cases drawn from two federal districts (New Jersey and Pennsylvania Eastern) and after controlling for a number of variables found that being detained before trial is associated with increased sentence length.[[23]](#footnote-23) The studies described above demonstrate the strong correlation between pretrial detention and the length of the sentence, however the precise causation is hard to detect, because it is safe to assume that the reason why at least some of the defendants who were detained in pretrial is high risk, so it is logical that they will receive longer sentences. In order to cope with that, the studies control for variables such as the nature of the offence, criminal history and risk. It is safe to assume that at least part of the reason for longer sentence is the pretrial detention.[[24]](#footnote-24)
* Detention during pretrial is highly predictive of reoffending in the future: even for relatively short periods behind bars, defendants who were detained for more days were more likely to commit additional crimes in the pretrial period – and were also more likely to do so during the two years after their cases ended.[[25]](#footnote-25)
* Detention during pretrial is very costly: the estimated yearly cost of incarcerating pretrial detainees is 13.6 billion taxpayer dollars.[[26]](#footnote-26) The daily cost of detaining a pretrial defendant in a federal facility ranged from a low of $35.41 to a high of $163.35, with an adjusted average daily cost of detention while awaiting trial of $72.67. In stark contrast, the daily cost of releasing the defendant under the supervision of a federal probation officer is for an average of $8.21. In addition, many defendants are held far from the courts in which they appear; and that creates an additional transportation cost.[[27]](#footnote-27)
* Detention during pretrial limits the ability to prepare properly for the case: in jail, detainees have reduced access to their attorneys, which limits the defendant's ability to fully participate and contribute to the preparation of the defense case.[[28]](#footnote-28)
* Detention exerts great pressure on the defendant to plea bargain the disposition of the case: especially in low level crimes defendants incentivized to plead guilty even if they are innocent, because sometimes if they receive credit for the time spent in jail awaiting conviction, the remainder of their detention time can be short, or nonexistent, if they take a plea.[[29]](#footnote-29) For example, in Rikers Island (New York City’s main jail complex) only 165 felony cases proceeded to trial in 2011, in sharp contrast to the 3391 cases in which the defendants pled guilty.[[30]](#footnote-30)
* Detention creates negative perceptions of the detainee in the minds of the court/jury who convict and/or sentence the defendant.[[31]](#footnote-31)

Giving the severe consequences of preventive detention, it is important to ensure that defendant in their early stages of the trial, while they still enjoy the presumption of innocence under the law will be detained only in rare exceptions and not as a rule.[[32]](#footnote-32)

## The Legal Framework

Given the decentralization that characterize criminal law enforcement in the U.S., the legal framework governing pretrial is very broad, and changes from one jurisdiction to another.

### Constitutional Protection and Federal Laws

* The Presumption of Innocence dictates that a formal charge against a person is not evidence of guilt; in fact, a person is presumed innocent and the government has the burden of proving the person guilty beyond a reasonable doubt.[[33]](#footnote-33) Justice White wrote in 1895- “The principle that there is a presumption of innocence in favor of the accused is the undoubted law, axiomatic and elementary, and its enforcement lies at the foundation of the administration of our criminal law.”[[34]](#footnote-34) Therefore, any restriction imposed on the defendant in the pretrial stage is not for punishment, but rather for guarantying his/ her appearance during the trial where the presumption of innocence will be debated.
* The Due Process Clause, anchored in the Fifth and the Fourteen Amendments,[[35]](#footnote-35) it provides that the government shall not take a person's life, liberty, or property without due process of law. Due process is that which comports with the deepest notions of what is fair and right and just.”[[36]](#footnote-36) As it relates to restricting a pretrial defendant's liberty, due process requires, at a minimum, that the defendant receive the opportunity for a fair hearing before an impartial judicial officer, that the decision to restrict liberty be supported by evidence, and that the presumption of innocence be honored.[[37]](#footnote-37)
* Forbidding Excessive Bail- The eight amendment of the constitution mention bail explicitly and states “Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted.”[[38]](#footnote-38) Courts interpreted the requirement as setting an amount that reflects “adequate assurance “that the accused will attend the trial and comply with the sentence.[[39]](#footnote-39)

The Equal Protection Clause, anchored in the 14th amendment, embodies the principle that persons who are similarly situated ought to be treated alike. The right exemplifies the concept that individuals should be treated fairly in the exercise of fundamental rights and that there shall not be any distinction between groups based on impermissible criteria like the inability to pay money bail.[[40]](#footnote-40)

* The bail reform act: relevant only to defendants in the federal system. The Bail Reform Act enacted in 1966.[[41]](#footnote-41) The Act reinforced that the sole purpose of bail was to assure court appearance and that the law favors release pending trial. In addition, the Act established a presumption of release by the least restrictive conditions with an emphasis on non-monetary terms of bail.[[42]](#footnote-42) In the late 1970s, there was a shift toward protecting the community from the potential danger posed by defendants awaiting trial, and an amendment to the act granted the federal courts the authority to detain criminal defendants for preventive purposes.[[43]](#footnote-43) For certain Severe offenses, the Act state that the default is pretrial detention and it passes the burden of proof to the defendant to demonstrate otherwise.[[44]](#footnote-44) The Supreme Court upheld the constitutionality of some preventive pretrial detention in its 1987 decision United States v. Salerno.[[45]](#footnote-45) The Salerno case acknowledge the interest of the government in protecting the community, however it does not provide a blanket authorization for pretrial detention. It relies on the limitations set by congress in the statute and the fact that preventive detention is saved only for the riskiest of defendants and the serious offences.[[46]](#footnote-46)
* The Speedy Trial Act enacted in 1974,[[47]](#footnote-47) established pretrial services agencies in 10 judicial districts, in order to reduce crime by persons released to the community and to minimize unnecessary pretrial detention. The agencies were to interview each person charged with other than a minor crime, verify background information, and present a report to the judicial officer considering bail. The agencies also were to supervise persons released to their custody pending trial and connect them with community services.[[48]](#footnote-48) In 1982 pretrial services were expended to every federal judicial district with the enactment of the Pretrial Services Act.[[49]](#footnote-49)

### State Laws

Most state constitutions include provisions guaranteeing a right to bail. A typical right-to-bail provision states: “all persons shall be bailable by sufficient sureties, unless for capital offenses, where the proof is evident, or the presumption great.”[[50]](#footnote-50) In some states the courts interpret the word shall more broadly and except severe offences, defendants usually granted bail and will be detained only if they cannot depose it. In other states courts have more discretion in deciding who will be released and who will be detained, and the right to bail is not automatic. In 9 states the laws are stricter, and the only limitation is according to the 8th amendment, not imposing an excessive bail.[[51]](#footnote-51) As described above, some states also enacted specific provisions limiting money bail. In addition, since the popularity of risk assessment tools is growing, more and more jurisdictions are anchoring the requirement to adopt such tools in regulation.[[52]](#footnote-52)

# Risk Assessment Tools in Pretrial- From the Past to the Future

Traditionally, judges were relying on their own recognizance and experience to assess the risk that a certain defendant pose. But studies show that when judges rely on their intuitions, they do not use information reliably; they may assign weight to items that are in fact not predictive, or they may be overly influenced by causal attributions.[[53]](#footnote-53) Thus, judges started to use some version of risk assessment tools more than 50 years ago.[[54]](#footnote-54) The Vera Point Scale, considered to be the first actuarial pretrial risk assessment, was developed and adopted in Manhattan in 1961. Defendants were classified by the degree of risk they pose, and based on this classification, court officers developed the recommendation for release.[[55]](#footnote-55) Since then, the number and sophistication of these algorithms has vastly increased over the past decades, and today there are approximately 60 risk assessment tools used across the country.[[56]](#footnote-56) 24% of pretrial agencies use tools based on static factors, mainly criminal history, 12% of pretrial agencies rely on tools that include dynamic aspect and are based on interviews with the defendant and data about employment, education, family status etc. and 64% of pretrial agencies use tools that include a combination of the two.[[57]](#footnote-57)

Most scholars, criminal justice practitioners, and citizens see actuarial methods as a more efficient, rational, and wealth-maximizing tool to allocate limited law enforcement resources. We increasingly put our faith in actuarial instruments, thinking that an algorithm can do a much better job than a human-brain.[[58]](#footnote-58) However, beside the great potential to improve the system, there are significant risks that if not addressed could clash with the defendant’s constitutional rights. Due to their algorithmic nature, questions have been raised in regard to the validity of actuarial risk assessment tools, their opacity and lack of explainability; and there is a big conceptual gap between the camp of the supporters of algorithmic risk assessment tools and the camp of the opponents. This gap can be attributed to two things:

First, small details about the design and the implementation of risk assessment tools can inherently change the outcome. For example, while there are many empirical studies that prove that algorithms outperform human judgement in predicting recidivism risk in pretrial,[[59]](#footnote-59) some scholars pointed out that after implementing a risk assessment tool increased release rate occurred significantly more among white defendant, concluding that the algorithm might be biased toward black defendants.[[60]](#footnote-60) But in fact, there could be many explanation for such phenomena- there are differences in the way judges respond and interact with risk assessment tools, while in one jurisdiction judges can use the tool to “liberalize” their practices, in another jurisdiction judges can use the tool to reinforce their internal biases and deviate from the recommendation whenever it is convenient for them.[[61]](#footnote-61)

Second, there is a lack of understanding about the way traditional regression analysis algorithm operate and the way more complex algorithms that are based on AI and machine learning operate. Thus, the purpose of this chapter is to familiarize the readers with no technical knowledge with basic concepts of AI and machine learning that will help in putting the discussion about risk assessment tools in context.

## The Difference Between Regression Analysis and Machine Learning

The difference between regression analysis and machine learning is the volume of data and the human involvement in building a model. Machine Learning algorithms are capable of learning from trillions of observations; they make prediction and learn simultaneously. However, statistical modeling is generally applied for smaller data with less attributes.[[62]](#footnote-62) While the objective behind both machine learning and statistical regressions is learning from data, the approaches that each method is pursuing is different. Machine learning usually do not rely on rules, and statistical regressions formalize relationships between variables in the form of mathematical equations.[[63]](#footnote-63) “Machine learning (“ML”) refers to the capacity of a system to improve its performance at a task over time. Often this task involves recognizing patterns in datasets, although ML outputs can include everything from translating languages and diagnosing precancerous moles to grasping objects or helping to drive a car.”[[64]](#footnote-64) Machine learning is based on traditional statistical techniques that have been used for a while, including regression analysis. The increased capabilities attributed to machine learning are due to increased computing capabilities and a huge amount of data that was not available in the past.[[65]](#footnote-65)

Consider the following example, a newly appointed judge, who use to be a prosecutor before, need to decide under what conditions to release defendant until the end of the trial, or if to keep him in preventive detention. The defendant is a 28 years old, black male, a single father of two kids, work in a factory, accused of raping a 19 years old female that he met in a party, has two prior convictions one for sexual assault and the other for possessing drug for personal use, he has one failure to appear in court, he has a tenth grade education, a tattoo on his shoulder, history of alcoholism, and he receives food stamps and other social benefits. In the short pretrial hearing session, the defendant denies the accusation and any acquaintance with the female. The judge is confused, on one hand, she empathizes with his personal circumstances and the weak evidence that the prosecution provided so far, but on the other hand she is worried for the safety of young women in his neighborhood, especially giving his previous criminal record. In order to reach an informed decision, the judge turns to data about previous cases. Of course, an exact similar case is probably hard to find, so the judge needs to broaden the scope and look for defendants that are similarly situated and check if they fail to appear or end up committing a crime while awaiting trial. So, the judge for example would look for single fathers, between 20-30, who were accused in rape, with 2 priors, 1 failure to appear, did not graduate high school, have a tattoo and receiving social benefits. Let’s assume that there were 5 such cases in her jurisdiction, one defendant was detained until the end of the trial, 2 of them fail to appear; and 2 of them end up committing another crime. The sample is too small, and there are many variations that the judge can play with to increase the sample. For example, she can eliminate the tattoo criteria, she can look at males between 20-30, not necessarily single fathers. Maybe she should look up cases about males between 25-30, or maybe she would look for males with specifically one prior in sexual assault and one prior in drug usage. And what about the race factor, should the judge look up only black males? And how significant is the educational factor? The options are endless[[66]](#footnote-66) If in her jurisdiction a risk assessment tool is being used, most likely the best predictive factors were already defined, and the combination will generate a result along the scale of high, medium and low risk.

A machine-learning approach to this problem attempt to provide a more flexible solution, solution that takes into account the complex relationship between the variables and the outcome, and based on learning from the data itself.

## Types of Machine Learning Algorithms

Machine-learning is a broad field that compile many methods and approaches for solving one prediction problem. The question which algorithm to use depends on many factors such as the amount of the data and its quality, the specific task we want the algorithm to solve (for example, predicting a category, predicting a quantity or both) and the level of explainability of the result that we wish to have.[[67]](#footnote-67) It is not in the scope of this paper to cover all the possible algorithms and to dive into the technical details of each one. But to illustrate for the readers the capabilities of different machine-learning techniques, the difference between the three main branches of machine learning (supervised learning, unsupervised learning and reinforcement learning) will be explained.

### Supervised Learning

In supervised algorithms we know in advance what are the inputs (variables) and the outputs (outcome), and the algorithm apply learning technique to detect the correlation between the variable and the outcome. For example, in pretrial, the input would be the factors that we decide that the algorithm can consider (for example age, gender and number of priors), and the outcome could be release or jail.[[68]](#footnote-68) We would use a dataset to train the algorithm, if we have information about 1000 defendants that we know that half of them appeared in trial and half of them failed, we would feed the algorithm with information about 900 defendants (450 appeared and 450 failed) and we will let the algorithm to figure out on its own why each one of them appeared or failed, in other words, which combination between age, gender and priors lead to release or jail. Afterwards, we will use the remaining data about 100 defendants to validate the algorithm, we will not tell the algorithm if those defendants appeared or failed, we would let the algorithm guess the result and based on the information we have, we can determine the percentage in which the algorithm got it right and the error rate. If out of the 100 the model predicted correctly 89 cases, then the algorithm has 89% rate of accuracy.[[69]](#footnote-69) this type of algorithms called supervised algorithms because the learning process is similar to a teacher who supervise her students, the student learn and the teacher correct him or her when they make a mistake.[[70]](#footnote-70)

One popular method of supervised learning in the context of criminal justice is called Random Forest. One of the algorithms that will be discussed later on in the paper operates using similar technique. Random forest consists of two stages. The first stage focuses on generating a large number of decision trees and the second stage combines the results from each tree to arrive at a forecast.[[71]](#footnote-71) Each branch of the tree represent a factor, we have to set up in advance a rule for each factor and then determine if the data comply with the rule or not. Consider that we have the following rule, the subset of defendants who are unemployed males, under 30 years of age, with 2 or more prior convictions, at least 1 failure to appear and single are likely to fail in pretrial. In contrast, males who are over 50, with a stable job, long term relationship with a wife or a girlfriend, with up to one prior and no failure to appear; are unlikely to fail in pretrial. The different portraits can be visualized as “paths down branches of a tree.”[[72]](#footnote-72) The algorithm will split each one of the variables into two categories using a threshold that maximizes any association with the outcome. For example, consider the age variable, one tree will split it into below 30 and above 30, those who are under 30 will be placed on the right side of the tree and those who are more than 30, will be located on the left. In another tree, the split can be for below and above 50. In a similar way, many trees can be created, and in each one, the split of the variable will be in a different point. Other branches will represent different variables. For example, as for the branch that will represent priors, on one tree the split could be 0 prior versus one or more priors. In another tree, the split could be two or less, versus two or more.[[73]](#footnote-73) The forest consists of many classification trees that combines the results from each tree to arrive at a forecast. If the majority of the trees label defendant X as high risk, then he or she will be labeled as a high risk by the algorithm. The idea is that each tree applies a slightly different technique, that is based on the same principle, and the final result will be determined after looking at multiple mechanism and what together they predict.

### Unsupervised Learning

Unsupervised algorithms focus on the potential of the data and what we can learn from it. Under this model, we assume that there hast to be some kinds of relationships or correlation between the data we have, but data is too complex, so the role of the algorithm is to model the underlying structure or distribution in the data in order to discover the relationship between the data and the outcome.[[74]](#footnote-74) If we were to apply this technique, we will feed the algorithm with as much information as possible about defendants and we will not define in advance the factors that we think correlate best with failure in pretrial. We will let the algorithm work its magic and figure out on its own what makes defendant *X* high, medium or low risk to appear. In some instances, the model will be able to reveal the structure and point out a number of factors that it is basing the prediction on. One of the special characteristics of this type of models, is that while the model can suggest different ways to categorize or order the data, it is up to us to decide if we want to use a certain factor or not. For instance, consider that the model would suggest that the combination of race, gender, number of prior and eye color would produce the most accurate result about failure in pretrial. It is up to policy makers to set the threshold between accuracy and fairness and to determine if we will use it or not.[[75]](#footnote-75) In other instances it might be a complex connection between so many factors that is hard to reveal. The latter is called sometimes black boxes, because of the obscure nature of such algorithms. These types of algorithms are probably not the best fit for criminal justice, a sensitive policy domain that require us to understand and debate the set of factors that an algorithm considers.

In order to illustrate the difference between supervised and unsupervised machine learning algorithms, let us put aside the pretrial case for a moment and consider a simpler example of movie reviews. If we were to apply a supervised algorithm, we would collect reviews from various websites, and a human being will have to label manually the words in the review that are associated with positive reviews and those associated with negative reviews. The algorithm will develop a vocabulary list, split into positive and negative phrases. When the training process is done, the performance of the algorithm is evaluated on the validation dataset.[[76]](#footnote-76) If we were to use unsupervised algorithms, we will feed the algorithm information only about the classification of the review as positive or negative, but we will let the algorithm figure out on its own what are the word semantics that make each review positive or negative.[[77]](#footnote-77)

### Reinforcement Learning

This learning method is not commonly used in policy making yet, but it is a rapidly growing branch in machine learning that soon could be utilized for policy purposes, so grasping some knowledge about the way this method works is very valuable. Reinforcement learning is a method that is based on an agent/ an algorithm, that learns how to behave based on his interaction with the environment and the positive and negative rewards it receives.[[78]](#footnote-78) The best way to understand how reinforcement learning algorithms operate is using an example from the world of video games, the common field where this method is mainly used to date. Imagine that the goal of the agent is to collect as many apples as possible on their way to the top and to avoid being eaten by a snake. Typically, in the beginning of the game it will be easier to collect apples, they are more probable and predictable. But their value might be less than those that are closer to the top and entail more danger because they require intensive fight with the snake. The goal of the agent is to continuously interact with the environment and develop a strategy that will grant him as many points as possible and guarantying a quick and safe path to the top to collect the apples with the highest rewards.[[79]](#footnote-79) There are several ways in which the agent can be taught how to move in the game- by detailing every possible scenario and explaining what to do in each one; but since it is not easy to force all possible scenarios in advance and further complications that might arose because of them, another option is to providing a general framework and let the agent to make their own decisions based on estimation of what could be the expected value of each move.[[80]](#footnote-80) In terms of policy, it has been suggested that reinforcement learning can be used for controlling traffic lights based on traffic flow and the reward of the algorithm was granted when there was a reduction in the delay compare to previous time. In addition, reinforcement learning has been tested for news recommendations, where it could be particularly useful since in the nature of this domain, users get bored and news change rapidly.[[81]](#footnote-81) This approach might not be the most suitable for criminal justice because of its challenges. Reinforcement learning is not performing very well compare to other methods in conducting a well-defined task., it requires a huge amount of data for its training; and it is a general solution that could match to all types of problem, but in criminal justice domain specific solutions are very important.[[82]](#footnote-82)

## The Unique Characteristics of Machine Learning Relevant for Pretrial

It is important to mention that the assessment made in this chapter in regard to the impact of machine learning on pretrial is not generalizable to other stages of the criminal justice system. The pretrial stage is unique because pretrial risk measure only two specific behaviors — court appearance and re-arrest during the timeframe between the arrest and the end of the trial. The predicted task is very specific and as such, the result is more accurate.[[83]](#footnote-83) In contrast, the sentencing stage for example, require predicting a complex set of factors, such as long-term recidivism which involve way more subcategories and factors, so the impact of machine learning is different.

### Correlation does not Imply Causation

Machine learning requires no prior assumptions about the underlying relationships between the variables. It focuses on processing the data and discovering patterns. In contrast, statistical regression builds on causal relationship between the variables and the outcome.[[84]](#footnote-84) One criticism of statistical regression is that it tempts to overestimate the causal effect of the variables on the outcome. For example, those who have more education tend to have higher incomes, but that does not mean that education caused those higher incomes.[[85]](#footnote-85) A correlation quantify the statistical relationship between two data values, but unlike causation, correlation do not imply that one factor cause the other, correlations show what, not why.[[86]](#footnote-86) With correlations, there is no certainty, only probability. But if a correlation is strong, the likelihood of a link is high and vice versa. “Correlations let us analyze a phenomenon not by shedding light on its inner workings but by identifying a useful proxy for it.”[[87]](#footnote-87) There are advantages for relying mainly on correlations: (1) we can let the data speak for itself, this is particularly useful in an era where the computing capabilities and the amount of data increased significantly. Correlation can help unveil connection between factors we have not thought about before, which could encourage new research.[[88]](#footnote-88) (2) Models that are based on regression analysis depend on a relatively modest number of predictors having strong associations with the outcome. Predictors with weak associations with the outcome are usually folded into the “noise” and discarded. But when using machine learning, predictors having weak associations with the outcome can be included because a large number of weak associations can in the aggregate dramatically improve forecasting accuracy. One by one, each predictor does not matter much, but the whole is greater than the sum of its parts. Going back to the example above, by using machine learning, the newly appointed judge will not have to choose between the tattoo or the level of education, even if each one of these factors will have a small impact on the outcome, it is possible that in the aggregate the result that will be provided is much more accurate than ignoring certain factors. Despite the benefits, giving up on causality is not an easy task.

First, even if many factors will be taken into account, there is a risk that due to the complexity of the algorithm in practice very few factors will get most of the weight, but there are suggested technical solutions for this problem related to optimization methods.[[89]](#footnote-89)

Second, when factors with weak association with the outcome are also included, there is a higher chance that causality is being compromised. In the context of criminal justice and pretrial, when jeopardizing personal liberty is at stake, we expect a high certainty about the factors that the algorithm considers, and we would not want to base incarceration on random things like eye color, shoe size etc. So long that the list of factors that the algorithm claims that they are correlated with the outcome is transparent and open for debate among experts, defendants’ rights will still be guaranteed.

### Avoiding Overfitting

In all methods, both more traditional ones and advanced machine learning methods, there is the risk that we will build and adopt the model that will give us the result we wish to see. This is the problem of overfitting. Overfitting happens when the model “learns” the training dataset too well, and is not able to distinguish between the actual data and the noise in the dataset. This model will not be able to generalize and maintain the level of accuracy on new data.[[90]](#footnote-90) Consider that we split one dataset to 90% training and 10% validation. We also know in this case that the number of priors is the variable with the strongest correlation with the outcome. Imagine that we set the threshold on two priors or less, or more than two priors. When we check the performance of the algorithm on the 10% validation data, we realize that the algorithm is still failing to predict correctly who is going to fail. So, we return to the algorithm, and we change the threshold for priors, now we distinguish between violent priors and nonviolent priors. We treat them as two separate variables, and we increase the threshold to up to three priors or more than three. We are getting better prediction on the validation data, but still, we know that the algorithm can do better. At this point, we take a closer look into the second variable with a strong correlation with the outcome, it turns out to be age. Initially, the split was to groups of ten years of age (18 or under, 18-28, 28-38, 38-48; and so on). We decide to try grouping defendants into categories of 5 years of age. Finally, we are satisfied with the performance of the algorithm, we got a high accuracy rate and acceptable error rate. It is acceptable to try different combinations in order to train the algorithm. However, by tweaking the algorithm to produce the result we want we take the risk of overfitting the data because the overarching goal is to maximize the algorithm predictive accuracy on the new data points—not neces­sarily its accuracy on the training data.[[91]](#footnote-91)

one of the important ways of dealing with overfitting is being aware of its possibility and validating the algorithm often to observe the performance. Another way to avoid overfitting and to derive a more accurate estimate of model prediction performance is to apply the K-fold cross validation technique. The purpose of the technique is to maximize the benefit of the data and to allow using the whole dataset for training while escaping the tradeoff between training and validation. The technique is beneficial especially in smaller datasets, in which we do not want to “waste” data on validation; and we want the algorithm to use all the available data for training.[[92]](#footnote-92) The technique works as follows, we would randomly divide the dataset into K subsets; and we would let the algorithm perform K times. Each time, one of the K subsets is used for test and the rest for training. The method should be repeated until all the K subsets were used for training, and each piece of data was at least once in the training set and once in the validation set.[[93]](#footnote-93) To illustrate, if we have data about 1000 defendants and we want to predict the likelihood of failure in pretrial, we can divide the dataset into 10 subsets of 100 defendants each. At first, subsets 1-9 would be used for training, and subset 10 for validation. Next, subsets 1-8 and 10 would be used for training; and subset 9 for validation. In the next step, subsets 1-7 and 9-10 would be used for training; and subset 8 for validation. We would repeat the process until all subsets were used for both training and validation.[[94]](#footnote-94) Unlike the human brain, each time we perform the task, we can delete the previous knowledge from the memory of the algorithm, and that is how we avoid overfitting.

### Explainability

The biggest criticism toward adopting tools that are based on machine learning techniques is explainability. Explainability in this context has two meanings. The first one is understandability. The fact that black box algorithm will take over a task that use to be performed traditionally by a human being judge seem not acceptable from the legal perspective. Discussions of understandability sometimes suggest that human decision-makers are themselves interpretable because they can explain their actions. But as described earlier in this paper, studies show that people consciously and unconsciously weigh more than just legal factors written in a book in their decision making and that their intuitions can often be misleading.[[95]](#footnote-95)

The second meaning of explainability is what can be called interoperability. Interoperability in this context refers to the technicalities of the algorithm and the impossibility in explaining how the algorithm reached a certain result.[[96]](#footnote-96) In some cases, machine learning techniques produce a risk score on a scale of lo to high, and even the engineers who built the algorithm cannot explain which combination between which factors let to this results, let alone the judges.[[97]](#footnote-97) Understandably so, explainability in the context of pretrial will be relatively more strict compare to other policy domain since the judge should understand how to ground his/ her decision; and the defendant should be able to appeal the decision if needed. However, few points are worth keeping in mind:

First, there are no clear guidelines on what explainability in the context of criminal justice actually mean, what is the level of explainability that we expect from the algorithm, and where the line between transparency and “black box” should be drawn. Do we want to be able to trace back each step that the algorithm took until reaching a final result? Alternatively, if we can have a general idea about the workings of the algorithm, will that suffice?[[98]](#footnote-98)

Second, black box algorithms are usually algorithms that are based on unsupervised learning. But as explained above, there are other methods like supervised learning that can provide explainable results. Only if there will be an unsupervised algorithm that performs much better than a supervised algorithm, we will be faced with the question if the compromise in explainability is acceptable.

Third, as it will be explained later on, none of the existing risk assessment tools these days operates like a black box. Even the most advanced pretrial risk assessment tool use machine learning on a small scale and for a very specific task that does not harm explainability.

Fourth, there are other technical solutions for the explainability problem, such as introducing strong auditing mechanisms that focus on analyzing the fairness and level of biases in the output of the algorithm and not in the process itself. In analogy to pretrial, auditing approach would let the algorithm perform for a while and then analyze whether the algorithm treated differently black and white defendants or men and women.[[99]](#footnote-99) Auditing as a technique should complement other solutions for explainability and could not stand on its own because “justice delayed is justice denied” and discovering afterwards in an auditing process that a defendant was detained unrightfully because of a mistake made by an algorithm clashes with basic legal concepts.

Fifth, since the risk assessment tool is assisting judges and not replacing them, the risk to due process is manageable so long that judges receive adequate training and explanation of how the algorithm works.

### Proprietary Nature

Most of the opacity and lack of explainability that exist today is due to proprietary clauses written by the private companies that sells the algorithms but not due to the actual capabilities of the algorithm. Due to their increasing complexity, risk assessment tools require special technical expertise that exist mainly in the private sector.[[100]](#footnote-100) The contract between the private companies selling these tools and law enforcement agencies usually include non-disclosure agreements which prevent access to the proprietary code. Companies justify the request claiming that Intellectual Property laws in general and trade secret in particular are needed to protect the code from competitors and from alleged criminal who might tweak their actions and circumvent the technology.[[101]](#footnote-101) The claim about bypassing the technology is not very realistic since as it will be explained later in the paper, the manuals and all the operational details of the vast majority of pretrial risk assessment tools that exist today are open to the public. Only one tool is commercial and proprietary. Without access to the code, there are very limited ways to examine the validity and reliability of the tools by the agencies themselves, by defense lawyers and by third parties.[[102]](#footnote-102) In addition, since law enforcement agencies are the only customers of these tools, the market for them is dominated by very few actors. The private companies have the ability to shape the specifications of the technology and how it works, which indirectly create regulation by design.[[103]](#footnote-103) For example an algorithm that is supposed to predict failure in pretrial will include the factors that the company thought are the most relevant, and there will be no debate about the most pressing question in this regard. Policy preferences are often linked to questions of design, and the private company hold great power in shaping issues of accountability.[[104]](#footnote-104) The combination of unexplainability with proprietary nature led to portraying algorithmic risk assessment tools as the ultimate black box. But in fact, law enforcement agencies need to acknowledge the power that they have as the only purchasers of such technology and their power in negotiating the contract with the private companies. Agencies could develop a preapproval process, where local experts from the agency will examine all the available information about the product and debate its utility beyond the marketing spin of the private companies.[[105]](#footnote-105)

### Competing Notions of Fairness

It might seem intuitive that pretrial risk assessment tools have to be fair. But when the assessment is done by an algorithm the term fairness has to be defined, and this is not an easy task. Another place where regulation by design can occur is when choosing the notion of fairness that will be implemented in the algorithm. The computer science literature refers to more than 20 different notions of fairness.[[106]](#footnote-106) The notions can be divided into three main categories: notions that put the focus on the individual, notions that put special emphasis on anti-discrimination based on group affiliation and notions that focus on the causal relationship between the factors and the outcome. All notions tilt a different balance between accuracy and fairness.[[107]](#footnote-107) For example, some notions aim to equalize the type of error that the algorithm makes; and to have equal number of false positives and false negatives.[[108]](#footnote-108) Hence if the algorithm is correct in 85% of the cases, this approach will ensure that among the remaining 15% not all of those who are wrongly sent to jail are black and not all the defendants who got released and recidivated are white. This is an interesting approach, however, the challenge with equalizing false positives and false negatives is that society values them differently, and their economic cost is also different. Setting the threshold and deciding on an error rate our society is willing to tolerate is not an easy task. In deciding whom to incarcerate and whom to release, we balance between public safety and the presumption of innocence and the right to a fair trial.[[109]](#footnote-109) Balancing between false positives and false negatives will vary in severity depending on context, and translating this balance to a numeric error rate is a complicated technical and policy task that needs to be undertaken after consulting with all relevant stakeholders. Another notion of fairness calls for creating different algorithms for different groups based on their protected attribute.[[110]](#footnote-110) Imagine that it has been proven that having different algorithm for black and white defendants will improve the prediction of the algorithm, should we allow that as a society?

To conclude this point, embodying different notions of fairness in the risk assessment algorithm impact significantly the result and it requires determining complicated legal and moral questions. Each one of the 20 or so definitions of fairness is statistically valid and most of them are legally correct. However, the answer to the question which notion will achieve the best result will probably change depending on the specifications of each jurisdiction. This issue needs to be taken into account by policy makers.

### Eliminating Bad Discretion

Some judges, particularly elected judges might be motivated to be stricter in their detention decisions, since in the eye of the public they will be responsible for a crime committed by someone who was released in pretrial. The opposite outcome is less likely, and in most cases the judge’s reputation will not be impacted negatively if detaining low risk defendants.[[111]](#footnote-111) In addition, judges are less concerned with broader policy matters like the cost of pretrial detention and this too might impact their decision to be on the safe side and favor detention.[[112]](#footnote-112) in most jurisdictions, pretrial laws leave a lot of room for judges’ discretion. Discretion has some positive aspects, it allows judges to enhance individual justice and achieve a more just result that matches the facts of the case to the requirements of the law.[[113]](#footnote-113) In addition, The room for discretion allows judges to utilize their experience and professional intuition in order to weigh in on and achieve a more just result. However, discretion has also negative sides, judges can use their discriminatory power to reinforce biases, stereotypes, and prejudices.[[114]](#footnote-114) This could happen unintendedly, since the human brain is a black box, and psychology research shows that people who discriminate are usually not aware of it because there are rapid automatic responses that the brain generates before conscious can intervene.[[115]](#footnote-115) Thus, any comparison between a human judge and an algorithm should take into account the fact that many factors that judges have in mind when reaching decisions are unquantifiable and will remain unknown.[[116]](#footnote-116) When we analyze human decisions, the focus is not typically on explicit bias because of the complexity in proving biases among judges, the broad discretion given to judges and the flexibility of the legal language that is meant to adopt all scenarios. Deploying a machine learning based actuarial risk assessment tool can help filtering out some of the negative discretion, but it has to be done cautiously because actuarial tools can shift bad discretion from the judges to the engineers who build the algorithm, and this type of discretion is very hard to detect.[[117]](#footnote-117)

# The Most Commonly Used Risk Assessment Tools in Pretrial

This chapter will discuss briefly each one of the seven most commonly used risk assessment tools in pretrial. This is in order to give the reader more background information about each one of the tools separately before turning into the comparison between them.

## The Federal Instrument: Pretrial Risk Assessment [PTRA]

The Administrative Office of the U.S. Courts, have decided to develop their own actuarial risk assessment tool that is tailored to the characteristics of federal offences (for example the fact that some cases like immigration related cases can be discussed only in federal courts), and to the needs of the defendants.[[118]](#footnote-118) The PTRA was developed in 2009 in order to assist pretrial services officers, to reduce disparities in the system and to increase low risk defendants who can benefit and are willing to participate in alternative programs for detention.[[119]](#footnote-119) the PTRA was developed based on all pretrial cases processed by all federal districts (except the district of Columbia) between 2001-2007, in total 565,178 cases. The tool includes 11 factors divided into two categories – criminal history; and others. Beside each factor, the manual of the tool states how it should be treated. For example, the first factor is pending charges, the tool state that “Defendants who had one or more misdemeanor or felony charges pending at the time of arrest were 20 percent more likely to fail pending trial when compared to defendants who did not have a pending charge.”[[120]](#footnote-120) The other factors are: prior misdemeanor arrests, prior felony arrests, prior failures to appear, employment status, residence status, substance abuse, primary charge category (felony, misdemeanor or infraction), primary charge type, alcohol use and foreign ties.[[121]](#footnote-121) The factors and the estimated predictions beside them were used to create the regression analysis algorithm. Each defendant gets a raw score ranging between 0-15; and depending on the score the risk category of the defendant is being determined. PTRA category 1 include scores 0-4, PTRA 2 scores 5-6, PTRA 3 scores 7-8, PTRA 4 scores 9-10 and PTRA 5 scores 11-15.[[122]](#footnote-122) Each risk score is associated with rate of failure to appear, new criminal arrest and technical violation. The information for the tool is gathered from databases and from an interview with the defendant. The pretrial officer also includes his/ her recommendation for the release or detention, the recommendation is based on the outcome of the tool and if the officer wish to recommend something different from the tool, he/ she should consult with their supervisor.[[123]](#footnote-123)

A validation study conducted one year after the implementation of the tool in two federal district found that the tool increased recommendation for release as well as actual release.[[124]](#footnote-124) A new validation study, published in 2019 aimed to revalidate the tool and particularly examine whether it is calibrated on gender and race.[[125]](#footnote-125) This study is based on a much larger dataset, 85,369 defendants with closed cases who received PTRA assessment as part of their trial between 2009-2015. The study found that the PTRA performs well at predicting pretrial violations. For example, for defendants who were classified in risk category 1, only 5% violated their release with either failure to appear or new crime arrest. This number increased gradually as the risk categories increased; and in risk category 5 36% had a violation.[[126]](#footnote-126) As for predicting the risk to commit a violent crime, the tool was less accurate when compared with all types of crimes. In addition, the AUC-ROC value of the PTRA was found to be between 0.67-0.73, meaning that 67%-73% of the time a randomly drawn recidivist scores higher on the risk instrument than a randomly drawn non-recidivist.[[127]](#footnote-127) In terms of racial disparities, the researchers found that the PTRA has good to moderate predictive capacities of blacks and whites. In general, scores are associated with similar probabilities of rearrests among blacks and whites.[[128]](#footnote-128) Both rearrests for any offence and rearrests for violent offence increased incrementally among the two groups; and the overall accuracy rate in predicting rearrests among black and white defendants is between 0.64-0.67.[[129]](#footnote-129) as for the comparison between white and Hispanic defendants, the study found that the PTRA over estimates the likelihood of Hispanic defendants to be arrested for any offence, but for violent rearrests the predictions were similar.[[130]](#footnote-130) In terms of potential gender disparities, the PTRA was equally accurate in its predictions. It is interesting to mention that when examining racial disparities, the researchers focused on rearrests as oppose to failure to appear because it is considered according to them more objective outcome measure. It would have been interesting though to look at differences in failure to appear, because blacks and Hispanics are overrepresented in the system, they are caught on a higher rate and arrested more. Failures to appear on the other hand could also be impacted by external factors such as poverty and inability to miss working days to go to court, but there is less bias coming from the system in this regard.

## The Public Safety Assessment [PSA]

The Public Safety Assessment (PSA) is a pretrial risk assessment tool, developed in 2013 by the nonprofit the Laura and John Arnold Foundation. The tool was developed to help judges inform their decision in such an early stage if the criminal justice, and to provide them with neutral and reliable information about the defendant.[[131]](#footnote-131) The PSA was created using a very large dataset of over 750,000 cases drawn from more than 300 U.S. jurisdictions. The PSA produces two risk scores, one for failure to appear and one predicts the risk to commit a new crime. In addition, the PSA flags out defendants with high risk to commit a violent crime.[[132]](#footnote-132) After testing hundreds of factors that can potentially be included in the algorithm, the PSA’s developers decided to rely only on factors that can be obtained without an interview. The nine factors that the PSA considers are- the person’s age at the time of arrest, whether the current offence is for a violent crime, whether the person had a pending charge at the time of the current offence, whether the person has a prior misdemeanor conviction, whether the person has a prior felony conviction, whether the person has prior convictions for violent crimes, whether the person has failed to appear at a pretrial hearing in the last two years, whether the person failed to appear at a pretrial hearing more than two years ago; and whether the person has previously been sentenced to incarceration.[[133]](#footnote-133) The factors, their weights and scoring procedures are available to the public on the LJA’s website. After each one of the factors is weighted, the PSA produces two different scores on a scale of 1-6, one for failure to appear and one for new crime arrest. As for the risk to commit a violent crime, the PSA only flag it as yes or no.[[134]](#footnote-134) the PSA is designed to be a national risk assessment tool due to the general approach taken in its development, and it has been adopted so far by more than 38 jurisdictions, including the states of Arizona, Kentucky, Utah and New Jersey, and big cities like Phoenix, Chicago, and Houston.[[135]](#footnote-135) The PSA is offered for free to jurisdictions that wish to implement it and funds technical support for better implementation of the tool.

In terms of validation studies, researchers have conducted studies involving more than 650,000 cases in several jurisdictions, and many are on their way .[[136]](#footnote-136) a most recent study from 2018 that examined the validity of the PSA on a dataset from Kentucky found that in general, the predictive validity of the PSA is aligned with what is considered the norm in criminal justice. The overall predictive utility of the PSA is between 0.64-0.66, meaning that “when drawing two random cases from the dataset, one of which had the pretrial outcome and the other did not, between 64 and 66 percent of the time the case with the pretrial outcome would have a higher score than the successful case”.[[137]](#footnote-137) In terms of predictive accuracy by race, the PSA was found to be a fair predictor of new crime arrest and new violent crime arrest, but there are disparities when it comes to predicting failure to appear, the PSA assigns black defendants lower risk scores than white defendants who fail to appear.[[138]](#footnote-138) in terms of the predictive accuracy across gender, the study did not find indication of predictive bias for failure to appear or new crime arrest, but there are significant differences when it comes to predicting new violent crime.[[139]](#footnote-139) Other study shows that 79 percent of judges participated in the study reported that the PSA often informs their decision making, and that 61% of public officials that are working with the PSA often agree with its recommendation.[[140]](#footnote-140) Overall research show that jurisdictions that implemented the PSA are experiencing decrease in jail population without increasing the crime rate.[[141]](#footnote-141)

In order to further examine the validity of the PSA and its impact on actual decision making, the Arnold Foundation is funding several external organizations such as The Access to Justice Lab at Harvard, the MDRC and Research Triangle International. One interesting study, conducted by the Access to Justice Lab in Harvard is evaluating the PSA’s effectiveness in Dane County, Wisconsin. Each new pretrial case in Dane County will be assigned into one of two groups- the treatment group where the PSA score will be made available to the judge, the prosecutor and the defense attorney; or to the control group where the PSA score will not be produced. The study is ongoing, and after a couple of years, the researchers will be able to examine the impact of the PSA and the difference between decisions produced when judges were exposed to the PSA score or not.[[142]](#footnote-142)

## Virginia Pretrial Risk Assessment Instrument [VPRAI]

The Virginia Department of Criminal Justice Services developed the VPRAI in 2005, and completed its implementation in all Virginia’s pretrial services agencies in 2005. The VPRAI determines defendants’ risk of failure to appear and risk to be rearrested for committing additional crime, and it is being provided to judges as part of the investigatory report.[[143]](#footnote-143) Since the VPRAI was first implemented, Virginia has been doing a remarkable job in keeping it up to date and revalidating the tool every few years. The first revalidation study was conducted after two years of statewide use, and its purpose was to examine whether factors that can change over time such as crime patterns, law enforcement practices and demographical factors, effected the accuracy of the VPRAI. The examination confirmed the tool’s general accuracy and led to minor revisions that were implemented in early 2009.[[144]](#footnote-144) In 2014 a second and thorough revalidation study was launched, and in addition to examining again the impact of changing factors, the study also examined the race and gender neutrality of the tool. The study confirmed that the VPRAI is statistically significant in predicting failure to appear and new crime arrest. In terms of racial differences, the study found that defendants of color have higher failure to appear rate (4.5% compare to 3.6% for white defendants). White defendants on the contrary, had higher new crime arrest (6.1% compare to 5.0% for people with color). As for gender neutrality, while the rates of failure to appear were relatively equal, males had a higher rate of new crime arrest (5.8% compare to 4.5% for females) .[[145]](#footnote-145) Several changes have been implemented in the VPRAI after the last study. For example, an earlier edition of the VPRAI included the factor “lived at the same resident for less than one year”, the revalidation study found that this factor is not statistically predictor for black defendants and women, thus it has been replaced with a new factor that had better predictive value “the defendant was on active community supervision at the time of their arrest”.[[146]](#footnote-146) In addition, the employment factor was also modified based on the revalidation study, the length of the employment has been removed, and sub categories such as “primary care giver, full time student or retired” were added. Lastly, the factor “the current charge is a felony” found to be a good predictor, so it has been detailed and now includes the sub categories “felony drug, theft or fraud” .[[147]](#footnote-147)

To date, the VPRAI includes the following 8 factors: (1) active community criminal justice supervision; (2) charge is a felony drug, theft or fraud; (3) pending charge; (4) criminal history; (5) two or more failures to appear; (6) two or more violent conviction; (7) unemployed at time of arrest, primary care giver, full time student or retired; (8) history of drug abuse.[[148]](#footnote-148) The VPRAI Instruction Manual has been updated in 04/2018, it is very detailed, easy to understand and it includes for each one of the factors a comprehensive explanation on how the pretrial officer should determine the answer to each question. For example, for risk factor 3’ pending charge, the manual explains- “The defendant has a pending charge(s) when there is an open criminal case that carries the possibility of a period of incarceration, and the pending charge has an offense date that is before the offense date of the current charge. (A charge with a disposition of “deferred” is NOT counted as a pending charge.) EXCEPTION: If the current arrest is solely for a failure to appear, the underlying charge related to the failure to appear does not constitute a pending charge. In addition, if a defendant is arrested, remains incarcerated pending trial, and is served with new warrants, this does not constitute a pending charge. Select “Yes” if the defendant had one or more charges for jailable offenses pending in a criminal or traffic (not civil) court at the time of arrest. Select “No” if the defendant had no pending charge(s) at the time of arrest.”[[149]](#footnote-149) After the factors are weighted, defendants are assigned a score of 1-6, from low to high.[[150]](#footnote-150)

Accompanying the VPRAI is the Praxis, a decision grid that help in translating the VPRAI score to a practical release type and level of supervision. While the VPRAI measures the risk, the Praxis help to manage that risk.[[151]](#footnote-151) The combined process consist of four stages: calculating the VPRAI score, examining all charges and identifying the most serious charge category, based on the first two stages determining the recommendation (release, release with monitoring, release with pretrial supervision levels 1-3; and detain); lastly, if one of the charges is for failure to appear the recommendation should be increased one level.[[152]](#footnote-152)a revalidation study that examined the performance of the Praxis found that judges that have been using it tempt to release defendants 1.9 times more often than judges who were not.[[153]](#footnote-153)

As a result of the good documentation of all stages of calculating the VPRAI and coming up with the final recommendations; as well as the extensive validation studies that have been accompanying the VPRAI since its first implementation, the VPRAI has been adopted by different counties in more than 12 states, or used as a model for other jurisdictions interested in implementing a pretrial risk assessment tool.[[154]](#footnote-154)

## Colorado Pretrial Assessment Tool (CPAT)

The Colorado Pretrial Risk Assessment Tool [CPAT] was developed in 2013, as part of the Colorado Pretrial reform Act, which required pretrial agencies to “make all reasonable efforts to implement an empirically developed pretrial risk assessment tool and a structured decision-making design based on the person’s charge and the risk assessment score.”[[155]](#footnote-155) The goal of the CPAT was to improve pretrial services that had been maintained locally.[[156]](#footnote-156) In order to develop the tool, data has been collected from 10 counties that represent 81% of Colorado’s population and their local services; and factors from tools used by other jurisdictions such as Virginia and New York City were considered.[[157]](#footnote-157) Eventually, 12 items have been selected for inclusion in the tool: (1) having a home or cell phone; (2) owning or renting one’s residence; (3) contributing to residential payments; (4) past or current problems with alcohol; (5) past or current mental health treatment; (6) age at first arrest; (7) past jail sentence; (8) past prison sentence; (9) having active warrants; (10) having other pending cases; (11) currently on supervision; (12) history of revoked bond or supervision.[[158]](#footnote-158) Each one of the 12 factors is assigned a number associated with the influence of this factor on pretrial misconduct. For example, if having a past or current problem with alcohol increases the risk of pretrial misconduct by 4%, a defendant that does not have problem with alcohol will get 0 points for this factor; and a defendant who had problems will get 4 points. The sum of the total points of all factors ranges between 0-82.[[159]](#footnote-159) The total number of points is associated with a risk score on a scale of 1-4, when typically, only on those with scores 3 and 4 cash bonds will be imposed.[[160]](#footnote-160)

The pretrial officer is required to conduct an interview with the defendant in order to fill items 1-8 and to rely on available criminal records for items 9-12. In practice, pretrial officers check the criminal records also in order to verify the defendants’ answer to questions 1-8.[[161]](#footnote-161) In addition, a good amount of discretion is given to the officers in filling the questionnaire in case the answers to some of the items is not available or for bridging the gap between inconsistencies with what is in the records and the defendants’ answers.[[162]](#footnote-162) Additional interviews for determining the score and the recommendation can be conducted with the defendant’s family members or the victim

The final reporting to the judge is done in the following format- “[Defendant’s name] has a CPAT risk score consistent with other Colorado defendants whose average public safety rate is [##]% and whose average court appearance rate is [##]%.”[[163]](#footnote-163) In addition to the score, the pretrial officer includes his/ her recommendation for the suitable conditions for release or detention.

A revalidation study of the CPAT begun in January 2018, and expected to be completed in mid-2020. The revalidation study seems comprehensive and consists of three parts: survey with officers, focus groups and observations; and a pilot that compares the performance of the CPAT and alternative tool by randomly assigning cases to both.[[164]](#footnote-164) Data collected from the city of Denver show an increase in non-monetary release.[[165]](#footnote-165) However, it is hard to tell without a comprehensive revalidation study if this increase is due to the implementation of the CPTA or other parts of the pretrial reform such as the abolishment of felony bond schedules.

Critics of the CPAT pointed out recently that tool has not been validated properly for all jurisdictions it was implemented in; and potential racial bias had not eliminated.[[166]](#footnote-166) For example, one factor that can be problematic is the ownership of residence which is highly associated with class status, if a defendant does not own his home he will get 4 negative points that will be added to the final score. In addition, data from Denver County shows that in 2015 63.7% of whites in Denver own their homes and only 29.1% of blacks own their homes.[[167]](#footnote-167)

The CPAT also considers past or current mental health treatment, and this could be considered discrimination according to the Americans with Disabilities Act.[[168]](#footnote-168)

## The Ohio Pretrial Assessment Tool (PAT)

The Ohio Pretrial Assessment Tool is part of the Ohio Risk Assessment System [ORAS], a network of 10 empirical different tools that can be used throughout the criminal justice process, starting from pretrial, community supervision, in prison and in preparation for release. The development of the ORAS begun in 2006 in collaboration between the Ohio Department of Rehabilitation and Correction; and the University of Cincinnati, Center for Criminal Justice Research.[[169]](#footnote-169) The system was developed in order to better classify the risk level of defendants, to match defendants with the most useful support mechanisms, to identify criminogenic needs and to better allocate resources. Additional goal of the system was to promote consistent measurement of risk across Ohio since before the development of the system there was a good deal of discrepancy among jurisdictions.[[170]](#footnote-170) The system was developed based on data from 1834 cases gathered from 29 locations. In order to map out the factors that will be included in the system, semi structured interviews were conducted with defendants, the interviews included 26 questions, as well as 2 page self-reporting instrument that included 96 questions related to criminal history, criminal thinking, employment, education, aggression, financial stress and more.[[171]](#footnote-171) The ORAS has been implemented in a number of agencies in Ohio, as well as in few other states and local jurisdictions such as Indiana, Texas and Massachusetts.[[172]](#footnote-172)

The pretrial Assessment tool PAT was developed using 452 cases collected from 7 Ohio Counties. The interviews identified more than 100 potential factors that can be included in the tool. Eventually, 7 items were selected: (1) age at first arrest; (2) number of failure to appear warrants past 24 month; (3) three or more prior jail incarcerations; (4) employed at the time of arrest; (5) residential stability; (6) illegal drug use during past six month; and (7) severe drug use problem.[[173]](#footnote-173) The PAT scores defendants on a scale of 0-9, when scores 0-2 are considered low risk, 3-5 moderate risk and 6-9 high risk. In addition to the revalidation study conducted when the tool was developed in 2009; a new report published in 2018 reevaluated the validity and reliability of the system. The revalidation study consisted of two parts, the first part examined the inter-rater reliability and it meant to study whether professionals agree or diverge in regard to the applicable score for a certain defendant. For this purpose, the researchers surveyed professionals who worked with the tool, and gave them 4 hypothetical cases that they had to score.[[174]](#footnote-174) In terms of the PAT, the agreement rate among professionals about the seven factors was on average 87%. The participants diverged only in regards to two factors- employed at the time of the arrest and severe drug use problem, in regards to both factors the level of agreement among the participants was below 80%.[[175]](#footnote-175) One potential explanation for this divergence is the leeway and the discretion given to the pretrial officers to incorporate there impression in the assessment. As for the severe drug use problem, officers are instructed to score 0 for no problem and 1 for yes. As for the employment factors, the manual instructs officers to score 0 those who are employed full time, 1 those who are employed part time and 2 unemployed defendants.[[176]](#footnote-176) These findings are very interesting to analyze, the two factors that officers did not agree upon are very different. While employment could be considered as a more static factor with a conclusive response, severity of drug usage is a dynamic factor that is more prone to interpretation. Perhaps the study teaches us that any type of factors is open for interpretation and that it is not clear what is the ideal room for interpretation should be left in the hands of the pretrial officers. As mentioned above, the participants had given four hypothetical cases that they were asked to score. In regard to two cases out of the four, relatively high agreement was reported (for one case 89% of participants scored low risk to reoffend, and for the other case 82% agreed on moderate risk). The participants reported lower agreement rates in regards to the other two cases, only 70% of the participants scored the third case as low risk and 78% of participants scored the fourth case as moderate risk.[[177]](#footnote-177) The researchers did not attribute the differences to the gender of the officer, to their level of education; or the training level in using the tool.[[178]](#footnote-178)

The second part of the study examined the validity of the tool and any potential differences in the performance between males and females; and whites and non-whites. The study found that the average total score for males and females and for whites and non-whites is similar. However, the tool scored the majority of defendants, 58% as moderate risk, and only 24% were scored low risk and 19% high risk. In regards to one factor, severe drug use problem during the last six month, there was a significant difference between white and non-white defendants; an whites scored higher.[[179]](#footnote-179) The study found that for both males and females, there is a direct correlation between the increased risk level and the increase in new arrest, however the tool predicted weakly to moderately new conviction for male and moderately predicted new conviction for females.[[180]](#footnote-180) In terms of race, the study found that the tool predicted relatively well new arrest for white defendants but for non-white defendants no significant correlation was detected between the levels of risk and new arrest rate. Similar findings were reported for new convictions, for white defendants reconviction rate increased as level of risk increased, but for non-white defendants the rate of reconviction for low risk non-white defendants was actually higher compare to moderate risk non-white defendants.[[181]](#footnote-181)

## Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)

COMPAS] is an empirical risk and needs assessment tool, integrated in the North Point Suite, a web-based assessment and case management system for criminal justice practitioners. COMPAS was developed by the private company NorthPoint, now owned by Equivant; and it comprises of 43 different risk scale models used in different stages of a criminal trial. This chapter will focus on four models. The first two models: The General Recidivism Risk Scale and the Violent Recidivism Risk Scale are the primary risk models of COMPAS, and they are widely used. The Pretrial Release Risk [PRRS] is a risk model designed especially for pretrial; and the Recidivism Risk Screen [RRS] is a special scale designed to predict new crime arrest in the period of the next two years.[[182]](#footnote-182) The four models that will be discussed were chosen because despite the fact that there is a specialized model for pretrial, different jurisdictions are using one or combination of the other models in pretrial. The idea behind developing so many different risk models is to match better between the circumstances of the case to the appropriate tool and to provide the most accurate and relevant score.[[183]](#footnote-183) But in fact, and as it will be described later in the paper, having so many separate risk models can create confusions in terms of the appropriate tool to use in each circumstances, it leads to less unanimity among jurisdictions, and if more than one model is being used this could cause counting twice factors like criminal history which in term could unjustifiably increase the score that a defendant will get.

Accompanying COMPAS is a long questionnaire that contains 137 questions that are either answered by the defendant in an interview or collected from criminal records. The questionnaire asks defendants such things as: “Was one of your parents ever sent to jail or prison?” “How many of your friends/acquaintances are taking drugs illegally?” and “How often did you get in fights while at school?” The questionnaire also asks people to agree or disagree with statements such as “A hungry person has a right to steal” and “If people make me angry or lose my temper, I can be dangerous.”[[184]](#footnote-184) According to NorthPoint, the majority of the 137 are used to determine mainly defendants’ needs and COMPAS risk models include much shorter list of factors.[[185]](#footnote-185)

### The Pretrial Release Risk Score (PRRS)

The PRRS was developed in 2009 based on a sample of 2831 felony defendants arrested in Kent County, Michigan.[[186]](#footnote-186) The tool includes 8 factors: (1) felony top charge; (2) pending case; (3) prior failure to appear; (4) prior arrest on bail; (5) prior jail sentence; (6) drug abuse history; (7) employment status; and (8) length of residence.[[187]](#footnote-187) The tool scores defendants on a scale of 1-10 when scores 1-4 are considered low risk, 5-7 medium risk and 8-10 high risk. According to NorthPoint, the tool has been implemented in two counties in California and in terms of revalidation, no external studies other than NorthPoint Internal validation studies have been found.[[188]](#footnote-188)

### The Recidivism Risk Screen (RRS)

This tool is designed to predict the defendant’s risk to be arrested for any misdemeanor or felony offense in the period of two years. It includes 5 factors: (1) age; (2) age at first arrest; (3) number of prior arrests; (4) employment status; and (5) prior parole revocations. According to NorthPoint, this scale is not meant to replace the general assessment or the pretrial assessment, but its purpose is to help agencies flag those defendants that might need a complete and longer assessment. Since its focus is prediction in the period of 2 years it could be useful also in pretrial.[[189]](#footnote-189) Information on how where the tool was implemented and about revalidation studies was not found.

### The General Recidivism Tool and the Violent Recidivism Tool

These tools are often used simultaneously. The general recidivism tool predicts the risk to be rearrested for both misde­meanor or felony offense, and it is considered as more thorough compare to the RSS because it includes more factors and it takes longer time to calculate. The factors included in the tool are: prior arrests and prior sentences to jail, prison, and probation; vocational/educational problems, drug history, age at assessment; and age at first arrest. The Violent Recidivism Risk tool includes similar list of factors with the addition of history of violence and history and non-compliance.[[190]](#footnote-190) Each tool provides a separate score on a scale of 1-10 when scores 1-4 considered to be low, 5-7 medium risk and 8-10 is considered as high score.[[191]](#footnote-191) These tools were implemented in different jurisdictions including counties in Florida, Wisconsin and California and are used also in pretrial.

A good amount of opacity is surrounding COMPAS’s operation. The inner workings of the algorithm, as well as the way the score is calculated, are not known to the public or to defendants. This fact led to questioning COMPAS’s validity and if it is prone to bias. The news outlet ProPublica conducted an investigation into 7,000 cases of people arrested in Broward County Florida.[[192]](#footnote-192) In this county a combination of the general recidivism risk tool and the violent recidivism risk tool is being used. ProPublica looked at how many defendants were charged with new offenses after two years of their release, and concluded that COMPAS is biased because the false positive rate was much higher among black defendants. The algorithm falsely labeled black defendants as future criminals twice as much as it did so for white defendants. While among black defendants, 42% of those who were released from jail and did not commit any future crimes were wrongly labeled high-risk, among white defendants, the algorithm made the same mistake in only 22% of cases.[[193]](#footnote-193) NorthPoint dissented from the findings, and published their own investigation showing how COMPAS is equally fair to black and white defendants. NorthPoint showed how in each one of the race scores of 1-10 an equal percentage of black and white recidivated. For example, among defendants who were scored 7, 60% of white defendants and 61% of black defendants recidivated. In addition, NorthPoint claimed that COMPAS is fair since as directed by the law, it does not take into account race explicitly. Lastly, NorthPoint pointed out that, given the different base rate of black and white defendants, the gap that ProPublica referred to will always exist regardless of COMPAS (Dieterich et al. 2016).[[194]](#footnote-194) The ProPublica story attracted the attention of the media and researchers that published conflicting results.[[195]](#footnote-195) While some academics supported ProPublica’s finding and became reluctant to the use of algorithms in the criminal justice system,[[196]](#footnote-196) others attributed the gap that ProPublica found to external factors such as the different base rate among black and white defendants; and statistical errors made by ProPublica.[[197]](#footnote-197) To date, it is clear that the rebottle between ProPublica and Northpoint centered on disagreement about the definition of fairness that should be used in this case. While for ProPublica fairness meant that the algorithm should make the same type of error equally for black and white defendants, for Northpoint the algorithm was fair since it is calibrated, meaning that for each race category the same percentage of black and white defendants recidivated.[[198]](#footnote-198) As explained above, it is not possible to satisfy both notions of fairness simultaneously.[[199]](#footnote-199)

But fairness is only one challenge that a sophisticated risk assessment tool like COMPAS raises. In order to deal with the explainability problem, and to challenge the assumption that the black box algorithm COMPAS provide a better result than a transparent algorithm, a group of researchers tried to open the black box, and provide an interpretable model using the same dataset analyzed in the ProPublica article.[[200]](#footnote-200) They develop specialized tools from the fields of discrete optimization and artificial intelligence. Specifically, they introduce a special branch-and-bound algorithm, called Certifiably Optimal Rule Lists, that provides (1) the optimal solution, (2) a certificate of optimality, and (3) optionally, a collection of near-optimal solutions and the distance between each solution and the optimal one. They manage to produce certifiably optimal, inter­pretable rule lists that achieve the same accuracy as black box tools. Using only the current charge, gender, age and priors, the researchers have reached the same accuracy level as COMPAS 0.7.[[201]](#footnote-201)

## The Kleinberg et al. Study

This tool is not being used yet in practice in any jurisdiction, but due to its academic success and the unique emphasis on machine learning, it will be discussed in this section.

One prominent study applies machine learning techniques, specifically gradient boosted decision trees technique, on pretrial risk assessment tools is the study conducted by John Kleinberg and his colleagues.[[202]](#footnote-202) the algorithm is based on a large dataset of cases heard in New York City from 2008 to 2013. Judges in New York are given a release recommendation based on a six-item checklist developed by a local nonprofit, so the analysis of the researches compares the performance of the algorithm against the performance of the judges with the non-profit tool.[[203]](#footnote-203) The data include the age, the current offense, criminal record (including prior failures to appear); and the outcome of the case (release, failed while awaiting trial or rearrested while awaiting trial).[[204]](#footnote-204) The input variables for the algorithm were only three- current offence, priors and age; and the outcome variable is the likelihood that the defendant will fail to appear. They built a decision tree for each case divided through a sequence of binary splits. Starting from the bottom of the tree, the first question to be determined is if the defendant have ever been arrested before, if the answer is yes, he/ she will be located on one branch of the tree and if no on a branch in the other side. In each step, a similar split will be conducted based on the information gathered in the previous splits.[[205]](#footnote-205) In other words, using the gradient boosted decision trees enables more higher degree of interactivity between the variables and coming up with a score that is more tailored to each defendant.

Both regression analysis and machine learning are used in the tool. Regression analysis is used to identify the factors to be included in the algorithm, and the machine learning aspect of the tool is illustrated in the way the researchers trained the algorithm. The researchers did not mention that if defendant X has prior Y he or she has Z percent chance to flee or commit another crime; they let the algorithm to figure it out on its own.[[206]](#footnote-206) The algorithm focused on predicting only flight risk and not recidivism, since this is the only official factor that judges in New York are allowed to consider, but the researchers obtain qualitatively similar findings in a national dataset that examined also recidivism.[[207]](#footnote-207)

The results of the study are quite promising, they show that using machine learning, crime can be reduced up to 24.7% with no change in jailing rates, or jailing rate can be reduced in up to 41.9% with no increase in crime rates. Moreover, all categories of crime, including violent crimes, show reductions; and these gains can be achieved while simultaneously reducing racial disparities.[[208]](#footnote-208)

In addition, the researchers conclude that any additional information that judges are expose to, other than the necessary factors for prediction, acting as noise and distracting them from reaching an informed decision. They attribute some of the problem to what they call a “selective labels problem”, meaning that judges rely on many factors that are hard to measure such as mood, or specific features of the case such as the defendant’s appearance.[[209]](#footnote-209)

Following the study, the Criminal Justice Agency in New York [CJA}, the nonprofit who is responsible for risk assessment in pretrial, is redesigning its system, and it might incorporate the findings of the study.[[210]](#footnote-210) While the results of the Kleinberg et al. study are quite promising, it is important to remember that this is the first attempt to apply such a technique on pretrial, and future research that will apply similar techniques to actual data can play a significant role in designing policy in this field.

# Comparison Between the Tools

This chapter detail the most important policy considerations that jurisdictions need to take into account when implementing any risk assessment tools. The policy issues derived from a growing literature, produced by civil society organizations and academics, that attempt to aid policy makers pick the right risk assessment tool for them and to address some of the risks that those tools entail.[[211]](#footnote-211) This chapter will assess the seven risk assessment tools discussed in the previous section, in accordance with the policy considerations. More specifically, the chapter examines what is the impact of machine learning on each one of the policy considerations if added; and whether there is a difference between the traditional tools and the more advanced tools in terms of their compliance with the policy considerations.

## The Factors Used in Each Tool

Table 1 Summary of factors used in 7 different pretrial tools.

|  |  |
| --- | --- |
| **Tool** | **Policy Consideration** |
| The Federal tool (PTRA) | 11 factors: (1) pending charges; (2) prior misdemeanor arrests; (3) prior felony arrests; (4) prior failures to appear; (5) employment status; (6) residence status; (7) substance abuse; (8) primary charge category; (9) primary charge type; (10) alcohol use and (11) foreign ties. |
| The Arnold Ventures tool (PSA) | 9 factors- (1) age at current arrest; (2) whether the current arrest is for a violent crime; (3) pending charge at the current offence; (4) prior misdemeanor conviction; (5) prior felony conviction; (6) prior convictions for violent crime; (7) failure to appear in the last two years; (8) failure to appear more than two years ago; and (9) previous sentence to incarceration. | |
| The Virginia tool (VPRAI) | 8 factors- (1) active community criminal justice supervision; (2) charge is a felony drug, theft or fraud; (3) pending charge; (4) criminal history; (5) two or more failures to appear; (6) two or more violent conviction; (7) unemployed at time of arrest, primary care giver, full time student or retired; (8) history of drug abuse | |
| The Colorado tool (CPAT) | 12 factors- (1) having a home or cell phone; (2) owning or renting one’s residence; (3) contributing to residential payments; (4) past or current problems with alcohol; (5) past or current mental health treatment; (6) age at first arrest; (7) past jail sentence; (8) past prison sentence; (9) having active warrants; (10) having other pending cases; (11) currently on supervision; (12) history of revoked bond or supervision. | |
| The Ohio tool (PAT) | 7 factors- (1) age at first arrest; (2) number of failure to appear warrants past 24 month; (3) three or more prior jail incarcerations; (4) employed at the time of arrest; (5) residential stability; (6) illegal drug use during past six month; and (7) severe drug use problem. | |
| The COMPAS tool | The pretrial tool includes 8 factors- (1) number of pending charges; (2) which offense category represents the most serious current offence; (3) number of times sentenced to jail for more than 30 days; (4) number of times failed to appear for scheduled court hearing; (5) number of times arrested/ charged with a new crime while on pretrial release; (6) history of drug abuse; and (7) length of time in current community or neighborhood. The Recidivism Risk Screen includes 5 factors- (1) age; (2) age at first arrest; (3) number of prior arrests; (4) employment status; and (5) prior parole revocations. The factors included in the General Recidivism Risk tool are: prior arrests and prior sentences to jail, prison, and probation; vocational/educational problems, drug history, age at assessment; and age at first arrest. As for the Violent Recidivism Risk tool, the factors are similar with more focus on history of violence and history of non- Violent Recidivism Risk tool compliance. | |
| The Kleinberg et. al. tool | 3 factors- age, current offence and priors. The factor priors includes many sub factors, arrest, previous failure to appear, previous crimes etc. There is a difference between the way factors are laid out for the Kleinberg tool and other tools. The Kleinberg tool uses gradient boosted decision trees, a technique that enables more higher degree of interactivity between the variables. | |

The seven tools as it can be seen from Table 1 above include between 3-12 factors. The Kleinberg tool has the least number of factors, although priors is a broad category, and multiple items such as … are taken into account by the tool. As it can be seen from the table, the main focus of the tools is criminal history and its variance; and other factors include age, community ties, residential stability, employment and substance abuse. All documents that describes how each tool was developed, mention that hundreds of factors and sometimes even more were considered for inclusion in the tool. Typically, the developers will rely on long standing criminogenic theories and regression analysis to identify which factors out of the hundreds to include in the final tool. None of the tools, including the Kleinberg et al. tool, mention that they considered using machine learning to identify the factors for inclusion. As mentioned before, among the strength of machine learning is the ability to analyze a big amount of data quickly and efficiently. It will be interesting to examine whether there is any difference between the list of factors that the machine learning algorithm will identify compare to the list of factors that traditionally experts identified. There is no doubt that the list of factors identified by machine learning should be closely examined and debated by experts. If for example, the algorithm will identify eye color as a potential factor, experts most likely will not adopt this recommendation since it is not aligned with basic legal requirements of due process, and rightfully so it will not be included. In addition, it will be interesting to observe and dig deeper into the reason why certain factors were not chosen by the machine learning. This information could be very useful to policy makers that are considering a criminal justice reform.[[212]](#footnote-212)

Another task that can be performed by machine learning is to decide how it is best to split each one of the factors. This has been done by Kleinberg et al. the researchers let the algorithm to decide to which age groups to divide the dataset, and this use was proven successful. Machine learning allow researchers to try different combinations and different partitions, a task that is complicated to execute using regression analysis.

Looking simultaneously at the factors considered by each tool show us that detailing each factor and giving the pretrial officers multiple options to choose from is very important and could change the final outcome. For example, in regard to employment status, the tool used in the federal system distinct only between employed and unemployed defendants; and in Virginia also being a student, primer care giver and retired are considered as sort of employment and does not give a defendant negative points.[[213]](#footnote-213) The same for residence, in the federal tool the requirement is to own a home or being in the process of buying it, while rent is considered in other tools.[[214]](#footnote-214) Since the tools consider a relatively small number of factors, and the answer to each one is usually narrow (yes or no), providing a couple of variations to each answer can change the final prediction about the defendant.

## The Source of the Information (Interview or Solely Databases)

Table 2 Summary of the source of information used in pretrial tools.

|  |  |  |
| --- | --- | --- |
| **Tool** | **Source of Information** | |
| **Interview** | **Solely Databes** |
| The Federal tool (PTRA) | ✓ | ✓ |
| The Arnold Ventures tool (PSA) |  | ✓ |
| The Virginia tool (VPRAI) | ✓ | ✓ |
| The Colorado tool (CPAT) | ✓ | ✓ |
| The Ohio tool (PAT) | ✓ | ✓ |
| The COMPAS tool | ✓ | ✓ |
| The Kleinberg et. al. tool |  | ✓ |

As it can be seen from Table 2, all the examined tools rely on a combination of data collected from interview and databases, except the PSA and the Kleinberg tool. There is a long standing debate within the criminal justice literature about the type of factors that should be included in risk assessment tools. Studies show that criminal history is the factor with the highest correlation with recidivism and it could easily be obtained and verified through criminal records.[[215]](#footnote-215) In the context of pretrial, a study on data from Kentucky concluded that the same level of predictive accuracy could be maintained in pretrial risk assessment tools that are based on criminal justice data only. .[[216]](#footnote-216) However, some criminal justice experts point out that criminal history could be a proxy for race since minorities who has been historically disadvantaged and discriminated against often have lengthier criminal history.[[217]](#footnote-217) black defendants are overrepresented in the criminal justice system. They are over-policed, over-prosecuted, and over-convicted. Thus, if criminal history is included in the risk assessment tool, it will increase the ratchet effect.[[218]](#footnote-218) In other words, if the efforts of the police and the judges are focused on the minority groups, they will find more crime there and the balance between the offending population and the “carceral” population is interrupted.[[219]](#footnote-219) For example, consider that black defendants are responsible for 13% of all crimes (a number that is equal to their percentage in society) but they make up 40% of prison population. Considering criminal history in the tool will reinforce this imbalance over time. Other scholars are resistant to the use of dynamic factors/ items obtained typically in an interview, in risk assessment tools. The idea is that it is not fair to base the prediction on items that individuals do not have control over such as the neighborhood they were born in, gender, mental or physical health status.[[220]](#footnote-220) In addition, considering socio economic factors could distance the focus of the decision from the criminal conduct, the facts and the law.[[221]](#footnote-221)

As mentioned in the previous section, it would be interesting to examine the list of factors to be included in the tool according to a machine learning algorithm, to compare it with the current list of factors and to examine if the focus is on static or dynamic factors. It is important to clarify that although the Kleinberg tool is not relying on dynamic factor, this is not the suggestion of the algorithm, rather it is in line with the types of factors judges were exposed to in New York City under the current pretrial system. Regardless of if the factors were identified using a machine learning tool or a regression analysis tool, the debate about the type of factors to be included in the algorithm raise the issue of discretion. As of now, tools that rely on dynamic factors (the majority of the tools) leave a great deal of room for the discretion of the pretrial officer to way in. for example, in Colorado, 8 out of the 12 items are based interview, staff are instructed to mark yes or no, however the answer is not always clear. Take for instance item 4, the defendant is asked “Do you believe that you currently have or have ever had a problem with your use of alcohol?” and as mentioned, the staff can mark yes or no for this question. The term problem is not defined, even in the tips for the staff about how to fill the questionnaire, the leeway is very narrow, yes or no; and the scoring accordingly.[[222]](#footnote-222) In addition, the revalidation study conducted in Ohio about the agreement different officers reach in regards to scoring a certain factor, proves how much officers can vary in their scoring. Interestingly, the variation in the scoring occurred only in regards to dynamic factors, employment at the time of the arrest; and severe drug use problem.[[223]](#footnote-223) the question is whether we wish to leave room for discretion and how much room. The answer to this question depends on the way pretrial officers are viewed, as expert personnel that their opinion matters, or as administrators that supposed to fill the questioner in the most consistent way. None of the manuals accompanying the tools address this question despite its importance. It is possible also that the answer will change between jurisdictions. It is important to mention that even if discretion will be taken from the officers in the stage of calculating the score, in most jurisdictions the officers are also including their recommendation (detain or release on what condition) so the professional opinion will still be taken into account.

If dynamic factors will be taken out of the equation would mean less room for discretion regardless of if machine learning had been used or not since the algorithm will not interview the defendant on behalf of the officer, it will just calculate all the inputs. Another way to eliminate discretion without discarding dynamic factors completely is to rely on self-report questionnaire completed by the defendant. This method is being used in the Ohio tool, but the self-reporting questionnaire is in addition to the face to face interview and not instead.[[224]](#footnote-224) It would be interesting to examine the impact of self-reporting on the final score.

## Data Quality

Table 3 Summary of data quality assessment in pretrial tools.

|  |  |
| --- | --- |
| **Tool** | **Data Quality Assesment** |
| The Federal tool (PTRA) | 565,178 pretrial cases collected from all federal districts accept the district of Colombia, between 2001-2007. |
| The Arnold Ventures tool (PSA) | Over 750,000 cases drawn from more than 300 U.S. jurisdictions between 2001-2011. | |
| The Virginia tool (VPRAI) | The version of the tool used today was developed using a dataset of 14,383 cases of defendants arrested in all Virginia’s seven localities. | |
| The Colorado tool (CPAT) | The tool was developed based on data from 1315 cases collected during 16 month from 10 Colorado counties that represent 81% of Colorado’s population. | |
| The Ohio tool (PAT) | The tool was developed using 452 cases collected from 7 Ohio Counties between 2006-2009. | |
| The COMPAS tool | The pretrial tool was developed based on a sample of 2831 felony defendants arrested in Kent County, Michigan between 2005-2008. Information regarding the other tools was not found. | |
| The Kleinberg et. al. tool | The tool was developed based on approximately 750,000 pretrial release decisions from New York City between 2008-2013. | |

As Table 3 shows, the number of cases used to develop each tool vary dramatically and it ranges between few hundreds (Ohio) to nearly million cases (the Arnold Foundation tool and the Kleinberg tool). The development of a good and reliable actuarial pretrial risk assessment tool depends on access to high quality data about as many cases as possible; and tracking them in order to assess which are the best for inclusion in the tool.[[225]](#footnote-225) However, since the trend of using actuarial tools is relatively new, criminal justice agencies have not yet implemented the tradition of good data collection. In addition, the process of developing a protocol for data collection, adopting technological tools for collecting the data and storing it, as well as maintaining and analyzing it is a costly mission that could constrain jurisdictions from doing so.[[226]](#footnote-226) Therefore, criminal justice data is known to be notoriously poor.[[227]](#footnote-227) If high quality data is not available, jurisdictions are taking the risk of implementing a tool that is not suitable for their population which in term might also lead to inaccurate predictions.[[228]](#footnote-228) The data that is being collected needs to be accurate, full; and coming from the same population that it will be implemented on. The PSA for example was created using a diverse dataset of about 750,000 cases collected from more than 300 U.S. jurisdictions. However, there were many inconsistencies in the data since each jurisdiction was collecting different data bites.[[229]](#footnote-229) Thus, it is not just a matter of quantity but also the quality and the consistency makes a difference in the performance of the algorithm. Another problem is that if the data that is used to train the algorithm is old, it might not reflect recent legislative changes. Since 2012, more than 500 bills related to pretrial were enacted, out of them nearly 120 laws related to pretrial administration enacted in 2015.[[230]](#footnote-230) The new laws are leading to changes in decision making in pretrial, and those changes are likely to grow. If the tools are build based on data that does not reflect the new reality, there is the risk that it will cause “zombie predictions” predictions that will revive old practices that by law are not legal now.[[231]](#footnote-231)

Opponents of machine learning algorithms raise the concern that those tools will reinforce traditional biases since they are being trained on discriminatory data. According to this believe, the idea of neutrality and colorblindness associated with sophisticated algorithms is nothing more than a myth because the underlying data is already biased.[[232]](#footnote-232) The ratchet effect, mentioned in the previous section, can lead to highlighting minorities that are already over represented in the system as criminals and it will reinforce the self-fulfilling prophecy of arresting more minorities, scoring them higher by the algorithm, detaining and convicting them.[[233]](#footnote-233)

It is important to acknowledge that judges who were taking pretrial decision based on their own recognizance, were basing their decision on the same datasets that are used to train the algorithm. Thus, the algorithm is nothing but a mirror that reflects our own human biases and practices.[[234]](#footnote-234) There are different solutions for better addressing algorithmic bias.

First, as noted in the section about fairness, computer scientists have been working on different ways to deal with groups that have different base rate in the equation. Some solutions are focused on insuring that all groups are equally represented in a certain dataset, other solutions are focused on “favoring” one group over the others in order to compensate for previous discrimination; and another line of solutions focus on equalizing the types of error an algorithm makes across groups.[[235]](#footnote-235) those solutions correspond with different legal mechanisms, so it is important to pick the one that is most suitable to the situation.

Second, special attention need to be given to the potential impact on bias of each item in the tool. For example, all tools, include as a factor arrests even if they did not lead to conviction. This is probably the item that has the highest potential to reinforce biases, since the legal standard of proof (reasonable suspicion or probable cause) that the police need to establish for arresting a person is much lower than the burden of proof for conviction (beyond a reasonable doubt). Therefore, could be possible that a certain arrest will be based on prejudice or negative previous encounter of the police with other members of the group that the defendant belongs to.[[236]](#footnote-236)

Third, it is very important to ask the algorithm the right questions, meaning to design carefully the targeted outcome. ideas such as mitigating bias; and tradeoff between fairness and accuracy should be embodied in the design of the algorithm in advance.[[237]](#footnote-237)

Fourth, data for inclusion in the algorithm should be collected from diverse sources, for example, in addition to data provided by the police and the court, data could be also collected from the Bureau of Justice Statistics' National Crime Victimization Survey, which tracks crimes based on victim reports.[[238]](#footnote-238)

## Periodical Validation

Table 4 Summary of revalidation studies made for pretrial tools.

|  |  |
| --- | --- |
| **Tool** | **Revalidation Study** |
| The Federal tool (PTRA) | Validated twice, the first time was one year after the implementation and the second validation study was completed in 2019. |
| The Arnold Ventures tool (PSA) | Since the PSA is being used in more than 38 jurisdictions, different validation studies have been conducted and some are on their way. In total more than 650,000 cases have been analyzed in these studies, examining general predictive accuracy, potential race and gender bias. The LJAF is funding external organizations such as The Access to Justice Lab at Harvard for conducting cutting age revalidation studies for the tool. | |
| The Virginia tool (VPRAI) | Since the VPRAI was first implemented in 2005, Virginia has been doing a very good job in revalidating it. One revalidation study was conducted in 2007, and a second study that led to major revision in the tool was conducted in 2014. | |
| The Colorado tool (CPAT) | Revalidation study for the CPAT has begun in 2018 and expected to be completed in 2020. | |
| The Ohio tool (PAT) | In addition to the validation study conducted when the tool was implemented in 2009, a new validation study was published in 2018. The study took an interesting approach and beside validity for race and gender, it examined the agreement of different pretrial officer on the exact score that will be given to a certain defendant. | |
| The COMPAS tool | NorthPoint have been tracking and validating the pretrial tool internally but no external studies had been conducted. No revalidation studies were found for the RSS; however many revalidation studies were conducted for the general recidivism tool based on the same data used in ProPublica’s investigation. It is important to note that any external revalidation study is parcial because access to the instrument and to the way the score is being calculated is given only to jurisdictions that are using the tool and paying for it. | |
| The Kleinberg et. al. tool | The tool has not been used in practice yet so no additional revalidation study was conducted. However, the academic paper introduces the tool details how validation was done. The dataset was partitioned into training data, test data and small segment of additional test data that was not used until the study was complete. Also, the researchers repeated the study on a national dataset in order to validate the findings nationally and not just for New York. | |

As it can be seen from Table 4, the tools vary in regard to how often they are being validated, what exactly the revalidation process examine; and who conducts the validation. It is essential for any jurisdiction to validate the tool it is using and to examine its timeliness as well as its performance on the local population and on specific minority groups. in New York for example, it has been proven 20 years ago that having a landline phone in the defendant’s house is a good predictor for the ability to attend trial.[[239]](#footnote-239) However, given the increased reliance on cellphones and other connected devices that has much more advance technology, in the last decade, it is unlikely that the landline factor is still a good predictor. Thus, a tool that is not being validated often enough and that is not designed to keep up with social and technological changes, would lose its predictive ability. Periodical validation is also useful for building trust in the tool among the public, the litigants and the judges.[[240]](#footnote-240)

In some cases, there is an inherent difference between jurisdictions that require them to develop their own risk assessment tool. For example, the federal courts are the only courts that have the authority to deal with immigration related cases, thus, any pretrial risk assessment tool that will be adopted there can factor that in the final decision; but it does not have to be the case in other jurisdictions.[[241]](#footnote-241) In other cases, due to the high costs associated with developing a local risk assessment tool, many jurisdictions decide to adopt a tool that has been developed for another jurisdiction. A survey conducted by the U.S. Department of Justice found that 39% of agencies that are using pretrial risk assessment tool, adopted it from another jurisdiction. Out of those agencies, only 25% validated the tool for use on their own population.[[242]](#footnote-242)

The efforts of several jurisdictions to conduct a good and thorough revalidation studies are worth mentioning and they can serve as good examples. The Virginia tool was validated already three times since it was first developed. The last time was in 2014 and it led to major revision in the tool. The study examined the predictive accuracy of each one of the factors for the general population, across race and gender. In addition, new potential factors were included in the study in order to examine if it is worth including them. in the old addition of the tool, the answer to some of the question was yes/ no, and after the study those factors were detailed, and several alternative options were provided. For example, the employment factor now includes also the options- student, part time worker, retired or primary care giver.[[243]](#footnote-243)

The Arnold Foundation is conducting continued research into the PSA and funding external highly regarded organizations such as the Access to Justice Lab in Harvard Law School to examine the validity of the tool.[[244]](#footnote-244) The randomized trial study that the Access to Justice Lab is conducting is a very interesting example that can shed light on the true predictability of the tool compare to traditional human judgement decision making. It is recommended that other jurisdiction conduct similar studies since it is the only way to examine whether the tool in fact hold to its promise.

The Ohio example is also very interesting and teaches an important lesson. The study puts the focus on examining whether there is consistency between different officers in generating the score for a particular case; and whether there is consistency in scoring each factor.[[245]](#footnote-245) This is a very important issue to examine and to tackle it by designing the training that officers receive and instructing them on how to approach each question.

If data about the performance of the algorithm is collected in a coherent and organized way, machine learning can help in conducting the revalidation studies faster and more efficient. In addition, as illustrated by the Kleinberg study, there are different methods in machine learning for training the algorithm and validating it afterwards like for example using the *K*-fold cross validation technique (described above) which enable using the whole dataset for both training and validation and could be useful especially when the dataset is small.

## Ways of Implementation

Table 5 Summary of implementation mechanisms in pretrial tools.

|  |  |
| --- | --- |
| **Tool** | **Implementation** |
| The Federal tool (PTRA) | The score along with the pretrial officer’s recommendation is provided to the judge as part of a report. Information about training could not be found. |
| The Arnold Ventures tool (PSA) | The goal behind the PSA is to serve as a national risk assessment tool, it has been implemented so far in more than 38 jurisdictions; but the process of the implementation varies. In general the LJAF provides technical support and training for jurisdictions. | |
| The Virginia tool (VPRAI) | The tool has been implemented state wide in Virginia and in different jurisdictions across 12 states. In Virginia the VPRAI is included in the Pretrial and Community Corrections Case Management System PTCC and it is managed by the Virginia Departement of Criminal Justice Services. Implementation outside Virginia vary and depends on the needs and regulations of each jurisdiction. | |
| The Colorado tool (CPAT) | The tool has been implemented so far in 22 counties across Colorado. Training for using the tool is advised and could be done through the Colorado Association of Pretrial Services (CAPS), that also published a publicly-available training manual. | |
| The Ohio tool (PAT) | It has been implemented in different counties in Ohio as well as counties in Indiana, Texas, Massachusetts and California. Training is required for jurisdictions that purchase more than one risk assessment tool. The training has to be provided by the University of Cincinnati Correction Institute. | |
| The COMPAS tool | The pretrial tool has been implemented in two counties in California. The General Recidivism tool and the Violent Recidivism tool are used in different jurisdictions including counties in Florida, Wisconsin and California. Training is required and provided by NorthPoint - the developer of COMPAS. | |
| The Kleinberg et. al. tool | The study has not been implemented yet, but it might be considered as part of the redesign of the risk assessment tool that is currently being used in New York City. | |

As it can be seen from Table 5, the implementation of the tools varies a lot. The Virginia tool and the federal tool are similar in the sense that they have been implemented for all the population they have been developed for (statewide in Virginia and districtwide in the federal system). the Arnold Foundation tool have also been implemented statewide in Arizona, Kentucky, Utah and New Jersey. As for Ohio and Colorado, it is recommended that the tools will be implemented statewide, since the tool was created for and validated on their own population. Thus, it will probably be the most suitable tool to use and it would not make sense for a district in Ohio to adopt for example COMPAS or the Virginia tool. Adopting the same tool statewide has many advantages: first, it would ease the implementation process, it would foster more unanimity of law enforcement in the state, pretrial officers will be able to share and contribute from their experience and senior officials will be able to create guidelines and detailed manuals for using the tool. Second, research show that one of the important components of a pretrial reform is adopting evidence based actuarial risk assessment tool.[[246]](#footnote-246)

Another important implementation issue is related to how the judges approach and make use of the score produced by the tool. A research from 2017, shows that in Kentucky, within each county the risk assessment tool had similar effect on black and white defendants. However, looking at the performance of the tool state wise shows that white defendants were released pretrial in much higher rates compare to black defendants. The researcher attributed this difference to the way judges interacted and influenced by the tool. In counties with more white defendants, the judges liberalized bail practices compare to the behavior of judges in counties that had predominantly black defendants.[[247]](#footnote-247) Thus, it is very important to collect data also about the percentage of cases that judges agree or diverge from the risk score produced by the tool and the reasons for doing so.[[248]](#footnote-248)

The level of training that pretrial officers and judges receive upon using the tool also vary. Training is required only by for COMPAS and for the Ohio tool only if more than one risk assessment tool has been purchased from their package. For all the other tools training is advised. It is recommended that training will be mandatory for all tools and in all condition. Training is particularly important when implementing a machine learning based tool, in order to avoid any misconceptions. It is necessary to clarify exactly the capabilities of the tool as well as its limitation. In addition to extensive training for judges and pretrial officers who work daily with the tool, it is important also to provide the defense lawyers, prosecutors and the public with sufficient knowledge such that they can challenge it and understand how a certain score was calculated.

The implementation process should also take into account specific factors that are relevant to the jurisdiction such as current and anticipated jail dencity, attitudes toward incarceration, tolerance of misbehavior and more.[[249]](#footnote-249)

Although not mandatory, the support offered by the Arnold Foundation for jurisdiction that are implementing the PSA, is a good example. Implementation include a tailored training that focuses on collecting the needed data for implementing the tool; and setting guidelines for communicating the score to judges, prosecutors and defense lawyers.[[250]](#footnote-250) It is important to mention that this detailed implementation package is provided by the Arnold Foundation because it has been designed from the start to be suitable for implementation nationwide.

## Double Counting

Table 6 Summary of double counted factors in pretrial tools.

|  |  |
| --- | --- |
| **Tool** | **Implementation** |
| The Federal tool (PTRA) | *Potentially yes*, between item 1 pending charges and items 2-3 previous misdemeanor and felony arrest; and between item 7 substance abuse and item 10 alcohol consumption. |
| The Arnold Ventures tool (PSA) | *Potentially yes*, several items in the criminal history can be counted more than once, for example item six counts prior convictions for violent crime and item 5 counts prior felony convictions. | |
| The Virginia tool (VPRAI) | *Yes*, item 4 that refers generally to criminal history could overlap with items 3, 5 & 6 the that examine different aspects of criminal history. | |
| The Colorado tool (CPAT) | *Potentially yes*, although the factors are quite distinct, items 1-3 (having a home or cell phone; owning or renting one’s residence; and contributing to residential payments) might slightly overlap. Item 10, (other pending charges) could be a good attempt to deal with double counting. | |
| The Ohio tool (PAT) | *Most likely no*, the only two factors that are closely related but still quite distinct are items 6-7, illegal drug use during past six month; and severe drug use problem. | |
| The COMPAS tool | *Yes*, all the tools use different components of criminal history multiple times, and if jurisdictions are using more than one instrument then double counting is happening since there is a lot of overlap between the tools. | |
| The Kleinberg et. al. tool | *No*, the number of factors used in the tool is low. | |

The issue of double counting, summarized in Table 6, refers to a phenomena where the same item is being scored more than once which could in aggregate increase the final score of a certain defendant unjustifiably.[[251]](#footnote-251) When judges were taking those decisions without the aid of a risk assessment tool, they were exposed to the whole file of the defendant, evaluate it collectively and would make their decision. When actuarial tools are involved, the risk is that the impact of some factors will be multiplied which will have a negative consequences on the final outcome.[[252]](#footnote-252) as it can be seen in the table above, the majority of the tools include factors that could overlap. The main factor that is double counted is criminal history. for example, the item 5 in the PSA refers to prior felony convictions; and item 6 refers to prior convictions for violent crime. Presumably most violent crimes are felonies, so the same offence will be counted twice unless otherwise will be specified in the manual. In the Virginia tool, item 4 that refers to criminal history in general could overlap with items 3, 5-6 refer to other particular aspects of criminal history (pending charge, failures to appear and violent conviction). Item 10 in the Colorado tool could be seen as an interesting attempt to deal with the phenomena of double counting. The item refers to “other pending charges”. While other items in the tool refer to different aspects of criminal charges, the focus of item 10 is on the word “other” pending charges. Thus, this could be a solution that on one hand do not leave any item in the criminal history not counted, but on the other hand it ensures that each item is not counted more than once.

The tool that is the less prone to double counting is the Ohio tool. The 7 factors used in the tool are quite distinct and each one touches on a different aspect of the defendant life without overstepping on each other.

It is important to mention that double counting can potentially occur when the jurisdiction is using a tool that is part of a system that include many tools. This is the case for the Ohio tool that is part of the Ohio Risk Assessment System, and COMPAS that is part of the NorthPoint Suite. In this case, jurisdictions might be using two tools in the same stage of the criminal justice, for example for predicting a general pretrial score and a score for violent crime. This issue is not addressed in the manuals of the tools despite its importance.

## The Meaning of the Predicted Score

Table 7 Summary of outcome being predicted by pretrial tools.

|  |  |
| --- | --- |
| **Tool** | **Outcome** |
| The Federal tool (PTRA) | The tool produces one score on a scale of 1-5 and each risk score represent an X% risk of failure to appear, Y% risk of new criminal arrest and Z% risk for pretrial revocation. |
| The Arnold Ventures tool (PSA) | Defendants are given 2 separate scores on a scale of 1-6, one for failure to appear and one for new criminal arrest. In addition, the PSA gives a raw score (yes/ no) for the risk to commit a new violent crime. | |
| The Virginia tool (VPRAI) | The tool provides a single score on a scale of 1-6 that compounds failure to appear and new criminal arrest. The Virginia Department of Criminal Justice is considering separating the outcome into two and adding a score for violent crime. | |
| The Colorado tool (CPAT) | It provides a single score on a scale of 1-4. It has been noted that the developers of the tool considered having a separate score for failure to appear and new criminal arrest, but eventually chose a one combined score model since according to their analysis no significant differences between the prediction has been recorded. Reporting is done in the following format “[Defendant’s name] has a CPAT risk score consistent with other Colorado defendants whose average public safety rate is [##]% and whose average court appearance rate is [##]%.” | |
| The Ohio tool (PAT) | It provides a combined score on a scale of 0-9 for both failure to appear and new crime arrest. | |
| The COMPAS tool | Each sub tool provides a score on a scale of 1-10. The pretrial tool provides a score for pretrial misconduct which includes failure to appear and arrest for a new felony offence while on pretrial release. In addition there is a separate sub tool that predicts only the risk for committing a violent crime. | |
| The Kleinberg et. al. tool | The tool determines the likelihood that the defendant will fail to appear in percentage. The tool only considers failure to appear as an outcome because this is the only factor that judges in New York City are allowed to consider. | |

As it can be seen from Table 7, the final output of each tool as well as the meaning of each given score is different. Only the PSA explicitly separates between the two outcomes and provide a different score on a scale of 1-6 for new crime arrest and failure to appear. The COMPAS Pretrial tool does not separate between the outcomes, but there is additional tool, the General Recidivism Tool that focuses only on new crime arrest. Also the federal tool and the Colorado tool within the single score that they provide they estimate the probability when compared with other defendants to commit a new crime or to be rearrested.

The majority of the tools (the federal, Virginia, Colorado and Ohio tools) generate for each defendant a combined score for both failures to appear and the risk to commit a crime while awaiting trial. While both factors are important and judges have to take them into account, they have a completely different meaning.[[253]](#footnote-253) While there might be a justification for detaining a defendant that was classified as high risk for committing a new crime, there are many effective ways to avoid failure to appear such as sending several reminders, community supervision, providing low income defendants with transportation tickets valid for the date of the hearing etc. Therefore, the combined score could miss the important distinction between the factors and flag defendants as high or low score, without providing the judges with sufficient tools to understand what exactly this score means. There are legal as well as policy considerations that justifies separating the outcome into two different scores. From the legal perspective, the types of evidence and the government burden of proof for detention based danger is higher than the evidentiary standard that the courts adopted in regards to detention based flight risk.[[254]](#footnote-254) From the policy perspective, separating between the scores could reduce the reliance of judges on intuitions, because they will have to explain for example why despite a low risk for new crime arrest but high risk for failure to appear they decided to detain a certain defendant. The link between the statistical probability and the actual outcome will be clearer and hopefully this will be reflected in the judges’ opinions.[[255]](#footnote-255)

In addition, separating between flight risk and public safety will improve the ability to impose conditions of release that are actually aligned with the defendant’s needs. There could be many reasons why a certain defendant would failed to appear for a pretrial hearing. These reasons range from leaving the country to escape from serving their sentence, lack of resources to commute to court; or necessity to work and support their families. Each reason can be deterred with different means. For example, the literature shows great success for sending reminders about hearing dates via text messages instead of physical postcards.[[256]](#footnote-256) Another tool that could be particularly useful in reducing failures to appear is electronic monitoring.[[257]](#footnote-257) The same is true also for danger to commit a new crime. The array of offences in the criminal code is huge, and the risk from committing a murder is not the same as the risk from jaywalking.[[258]](#footnote-258) Indeed, the risk to public safety from a murder is much higher than jaywalking, thus, policy makers need to calculate the risk that they are willing to take in balancing between false positives and false negatives.

Having a separate score for new crime arrest and for failure to appear is the first step, but it is recommended that the tools will be more precise in their prediction about the actual pretrial misconduct. While so far none of the examined tools attempt to estimate the reasons/ motives for failure to appear, a move in this direction is taking place in regard to new crime arrest. The PSA generates a separate score for the risk to commit a violent crimes, and COMPAS has a separate tool explicitly for scoring violent crimes. It has been noted also that Virginia is considering adding additional score for violent crimes and that the new revalidation studies of the federal tool and the Ohio tool examined their ability to predict the risk to commit a violent crime, so it is possible that further changes in the tools will be added. An updated tool could ask for example, what is the likelihood that the defendant will flee if not reminded about his hearing. The risk that the tools predicts should be as specific as possible, and educating the judges and all actors in the field about the meaning of such score is very important because the statistical probability that corresponds with the actual score could be lower than what the judges think.[[259]](#footnote-259)

Machine learning can play a major role in providing a more precise prediction for the type of crime defendant X is high/ low risk to commit, and what kind of flight risk defendant Y is high/ low risk for.

## Possibility to Challenge the Outcome of the Tool

Table 8 Ability to appeal the outcome in pretrial tools.

|  |  |  |
| --- | --- | --- |
| **Tool** | **Appealability** | |
| **Easy to Appeal** | **Hard to Appeal** |
| The Federal tool (PTRA) | ✓ |  |
| The Arnold Ventures tool (PSA) | ✓ |  |
| The Virginia tool (VPRAI) | ✓ |  |
| The Colorado tool (CPAT) | ✓ |  |
| The Ohio tool (PAT) | ✓ |  |
| The COMPAS tool |  | ✓ |
| The Kleinberg et. al. tool | ✓ |  |

As it is shown in Table 8, most tools are easy to appeal if the defendant believes that the score is subject to errors. Due process requires at the minimum that the decision to detain someone will be taken by a judge who can assess the evidence and the accuracy of the information brought in front of him/ her, including the risk assessment and give a clear and detailed judgement. This way, the defendant will be able to appeal the judgement if needed and to oppose to unfair or inaccurate risk assessment.[[260]](#footnote-260) The meaning of due process differs between the pretrial stage and post-conviction. In pretrial, imposing on the defendant conditions for release or detention, is aligned with due process requirements so long that “they are reasonably related to a legitimate and nonpunitive governmental purpose.”[[261]](#footnote-261) Most criticisms toward machine learning techniques center around the fact that the decision is untraceable and therefore unappealable. But in practice a great deal of opacity is already surrounding the pretrial hearing and using a machine learning based tool, if done properly, is unlikely to cause more harm and it might be even beneficial. There is a huge variation between jurisdictions in terms of the conditions of pretrial hearings, but typically they do not last more than few minutes, they often take place through video conference and not in person, legal representation is not always provided, and the official that takes the decision could be a magistrate and not necessarily a judge. In addition, there are evidence showing that the officials spend very little time looking at the defendant’s file and determining the conditions.[[262]](#footnote-262) The procedure for appealing a pretrial decision again varies across jurisdictions, but in general it is subject to a strict standards of review.[[263]](#footnote-263) In any case, the defendant’s ability to raise substantive claims about the weight that each factor is given is quite limited.

All tools used these days in pretrial, even the COMPAS tools and the Kleinberg tool, do not generate a random score that is completely not understandable. For all the tools the list of factors that is being used to generate the score is open for the public, and this is information that the defendant can use in appealing the decision. There is no doubt that the hardest score to appeal is the score generated by COMPAS, but this is not because of the machine learning used in it, it is mainly because of its proprietary nature and the closed agreement with many proprietary clauses, between NorthPoint and the law enforcement agency. The other tools provide detailed manuals about the operation of the tools, and those manuals can provide a solid ground for defense lawyers to appeal the pretrial decision, if they think it was because of an error in the scoring. Therefore, it is recommended that when it is possible, jurisdiction will use a tool that its operation is as transparent as possible, and transparency in this context refers not just to the inner working f the algorithm, but also to the procurement of those tools.[[264]](#footnote-264) The exception to that could be if the proprietary/ nontransparent tool performs significantly better than the other tools. Nevertheless, the machine learning aspect, if exist, is mainly for deciding how to subdivide each factor used in the tool or maybe in some rare instances for choosing the list of factors, list that could eventually be exposed to the public. Thus, there is no reason why NorthPoint or any other private company will not release more information and instructions about the tool without compromising their commercial advantage.

Even when pretrial decisions were based solely on the judges’ recognizance, there was a challenge in appealing the decision and in understanding exactly what were the motives of the judge in taking a certain decision. It will be very interesting to investigate if there is any difference between the acceptance rate of appeals when risk assessment tool was used and when it was not; but until it will be proved otherwise, it seems safe to assume that there will not be a big difference.

# Conclusion

Pretrial is the “front door” of the criminal justice system, and any decision about the defendant will have an inherent impact about the rest of the trial and even on his or her future. Therefore, it is crucial that the decision taken at this stage will be fair toward the defendant, aligned with principles of law; and not biased. The goal of this paper was to examine the implications of machine learning on risk assessment tools used in pretrial, and to investigate whether as portrayed by the media and some scholars, these tools are the devil of the criminal justice system. Machine learning has a set of unique strengths and weaknesses that challenge our commitment to human judgement and basic concepts of law. Due to the way machine learning algorithms operate, they require us to adopt new means of understanding terms such as transparency, explainability and fairness. However, a comparison between machine learning and regression analysis shows us that there are more similarities than differences between the two. The comparison between the seven tools introduced in this article strengthens this conclusion. In regard to each policy consideration, the article concludes that adding a machine learning aspect will not worsen the outcome but the contrary. It might make the revalidation process much faster; it would allow detailing each factor to sub categories; and it would enable producing a more personalized score that has an actual meaning for the defendant. A successful implementation of machine learning algorithms that can improve the criminal justice depends on two main aspects. First, collaboration between the engineer’s that are creating the algorithms to the policy makers responsible for their implementation is needed in order to ensure full understanding of the capabilities of the algorithms as well as their limitation. second, a discussion and agreements about the tradeoffs between concepts such as fairness, accuracy, efficiency, transparency, justice and equity need to be part of the system design.[[265]](#footnote-265) Third, proper safeguards are needed in order to ensure that machine learning algorithms are complying with legal principles such as due process and equal protection. Such safeguards could be built in accountability mechanisms that will guaranty that the score is understandable and appealable; and negotiating the proprietary clauses in the contract between the law enforcement agency and the producer of the algorithm. Lastly, there are general recommendations that are not directly related to machine learning such as whether to rely on dynamic or static factors in the tool; and to ensure the quality of the data.

Actuarial risk assessment tools alone cannot reverse centuries of racial injustice or gender inequality, but they can land a hand. The Pretrial Justice Institute, (a leading nonprofit organization in the field), published a comprehensive report about “The State of Pretrial in America”, which gives all the 50 states a score on a scale of A-F based on their success in implementing pretrial reforms. Success is being measured according to three factors- the number of people per capita held in local jail awaiting trial; the percentage of the population living in an area where actuarial risk assessment tool has been implemented; and the percentage of people living in an area where money bond has been eliminated.[[266]](#footnote-266) The only state that got the highest score of A is New Jersey, where money bond has been eliminated statewide except in instances where no other condition is sufficient; and the Arnold Foundation tool was implemented statewide. The implementation of all the three factors has been proven successful, as pretrial detention have been dropped by 34% and public safety was improved with the reduction in all types of crime.[[267]](#footnote-267) Nine states got the score of B, in 5 of them actuarial risk assessment tool has been implemented fully across the state and in the remaining, more than 80% of the population live in an area where a tool was implemented. 10 states were classified in category C as they started some sort of pretrial reform but did not complete it yet. At the bottom of the list are 17 states classified in category F- Alabama, Alaska, Arkansas, Georgia, Idaho, Indiana, Louisiana, Mississippi, Missouri, Montana, Nebraska, North Dakota, Oklahoma, South Carolina, Tennessee, West Virginia and Wyoming.[[268]](#footnote-268) These findings already represent an improvement, they show that 25% of the American people live in a jurisdiction that implemented an actuarial tool. Four years ago, this number was approximately 10%. This progress cannot be attributed to machine learning. But what this paper argues is that we will not necessarily go backward if adopting more sophisticated algorithms. We might, on the contrary move forward faster if the right precautionary measurements are taken along the way to ensure that these algorithms are implemented properly and in a way that benefits all society.

1. Roy Walmsley, World Prison Brief, Pre-Trial/Remand Prison Population: Trend, United States, Prison Studies, (last visited Nov. 22, 2017), http://www.prisonstudies.org/country/united-states-america. [↑](#footnote-ref-1)
2. Megan Stevenson and Sandra G. Mayson, Pretrial Detention and Bail, in Reforming Criminal Justice volume 3 Pretrial and Trial Processes edited by Erik Luna, 21, 22-23 (2018). [↑](#footnote-ref-2)
3. Id. (Stevenson and Mayson, pretrial detention and bail, at 22-23) [↑](#footnote-ref-3)
4. Arthur Rizer and Caleb Watney, Artificial Intelligence can Make our Jail System More Efficient, Equitable and Just, 23 Tex. Rev. L. & Pol. 181, 183 (2018). [↑](#footnote-ref-4)
5. Cherise Fanno Burdeen, How Money Bail is a Trap for the Poor, the Atlantic (April 12, 2016), https://www.theatlantic.com/politics/archive/2016/04/the-dangerous-domino-effect-of-not-making-bail/477906/ [↑](#footnote-ref-5)
6. Wendy R. Calaway and Jennifer M. Kinsley, Rethinking Bail Reform, 52 U. Rich. L. Rev. 795, 796-797 (2015). [↑](#footnote-ref-6)
7. Angele Christin et al., Courts and Predictive Algorithms, Primer For The Data & Civil Rights, A New Era Of Policing And Justice 1, 1-2 (2015). [↑](#footnote-ref-7)
8. Kristin Bechtel, Alexander M. Holsinger, Christopher T. Lowenkamp and Madeline J. Warren, A Meta-Analytic Review of Pretrial Research: Risk Assessment, Bond Type, and Interventions, 42 Am. J. Crim. Just. 443, 446-447 (2017). [↑](#footnote-ref-8)
9. Karen Hao, AI is Sending People to Jail-and Getting it Wrong, MIT Technology Review, 01/21/2019 <https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/> [↑](#footnote-ref-9)
10. Courts are Using AI to Sentence Criminals. That Must Stop Now, Wired Magazine, 04/17/2017

    <https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/> [↑](#footnote-ref-10)
11. Rizer and Watney, Artificial Intelligence can Make Our Jail System More Efficient, supra note? at 181. [↑](#footnote-ref-11)
12. Rizer and Watney, Artificial Intelligence can Make Our Jail System More Efficient, supra note … at? Search for relatively simple goals. [↑](#footnote-ref-12)
13. Cherise Fanno Burdeen, How Money Bail is A Trap for the Poor, supra note? [↑](#footnote-ref-13)
14. Larry Schwartztol, Moving Beyond Money: A Primer on Bail Reform, Criminal Justice Policy Program (CJPP) 4 (Oct. 2016). [↑](#footnote-ref-14)
15. Kenneth Polite, Pretrial Justice, Think Justice Blog, (Nov. 13, 2015), https://www.vera.org/blog/justice-in-katrinas-wake/pretrial-justice. [↑](#footnote-ref-15)
16. Larry Schwartztol, Moving Beyond Money: A Primer on Bail Reform, Criminal Justice Policy Program (CJPP) at 5 (Oct. 2016). [↑](#footnote-ref-16)
17. Marc Levin, Pretrial Justice 101: Key Points for Policy Makers, Texas Public Policy Foundation Center for Effective Justice, 1-2 (2015), https://www.texaspolicy.com/library/doclib/PreTrialJustice-CEJ.pdf. [↑](#footnote-ref-17)
18. Stephanie Wykstra, Bail Reform, which Could Save Millions of Innocent People from Jail, Explained, Vox 10.17.2018 <https://www.vox.com/future-perfect/2018/10/17/17955306/bail-reform-criminal-justice-inequality> [↑](#footnote-ref-18)
19. The Vox article +Schwortztol Moving Beyond Money cited above, to look for Colorado [↑](#footnote-ref-19)
20. Colin Doyle, Chiraag Bains and Brook Hopkins, Bail Reform: A Guide for State and Local Policy Makers, The Criminal Justice Policy Program CJPP at Harvard Law School 2019. The beginning of appendix B. [↑](#footnote-ref-20)
21. For a thorough discussion about all needed components of a bail reform, see Colin Doyle et al. Bail Reform supra note? At the same page. [↑](#footnote-ref-21)
22. Meghan Sacks & Alissa R. Ackerman, Bail and Sentencing: Does Pretrial Detention Lead to Harsher Punishment?, 25 Crim. Just. Pol'y Rev. 59, 77 (2014). [↑](#footnote-ref-22)
23. J.C. Olsen, Marie Vannostrand, Christopher Lowenkamp, Timothy Cadigan and John Wooldredge, Pretrial Detention Choices and Federal Sentencing, 78 Fed. Probation, 12, 14 (2014). [↑](#footnote-ref-23)
24. Id. (Olsen et al. Pretrial Detention Choices, at 14). [↑](#footnote-ref-24)
25. Laura and John Arnold Foundation (LJAF), Pretrial Criminal Justice Research Brief, Pretrial Justice Institute (Nov. 2013). [↑](#footnote-ref-25)
26. Bernadette Rabuy, Pretrial Detention Cost $13.6 Billion Each Year, Prison Policy Initiative (Feb. 7, 2017); https://www.prisonpolicy.org/blog/2017/02/07/pretrial\_cost/ [↑](#footnote-ref-26)
27. Olsen et. al., Pretrial Detention Choices, supra note? at 15. [↑](#footnote-ref-27)
28. Samuel Wiseman, Pretrial Detention and the Right to be Monitored, 123 The Yale Law Journal, 1118, 1356 (2014). [↑](#footnote-ref-28)
29. Samuel R. Wiseman, Fixing Bail, 84 Geo. Wash. L. Rev. 417, 419 (2016). [↑](#footnote-ref-29)
30. Kyle Rohrer, Why Has the Bail Reform Act Not Been Adopted by the State Systems? 95 Or. L. Rev. 517, 518 (2017). [↑](#footnote-ref-30)
31. Timothy P. Cadigan & Christopher T. Lowenkamp, Implementing Risk Assessment in the Federal Pretrial Services System, 75 Fed. Probation 30, 32 (2011). [↑](#footnote-ref-31)
32. Olsen et al, Pretrial Detention Choices, supra note? at 13-14. [↑](#footnote-ref-32)
33. Marie VanNostrand and Gena Keebler, Our Journey Toward Pretrial Justice, 71 Fed. Probation, 20, 21 (2007). [↑](#footnote-ref-33)
34. Coffin v. United States, 156 U.S. 432, 545 (1895). [↑](#footnote-ref-34)
35. U.S. Const. amend. V; U.S. Const, amend, XIV. [↑](#footnote-ref-35)
36. Solesbee v. Balkcom, 339 U.S. 9, 16 (1950). [↑](#footnote-ref-36)
37. VanNostrand and Keebler, Our Journey Toward Pretrial Justice, supra note? at 21. [↑](#footnote-ref-37)
38. U.S. Const. amend. XIII. [↑](#footnote-ref-38)
39. Stack v. Boyle, 342 U.S. 1 (1951). [↑](#footnote-ref-39)
40. Mellissa Hamilton, Risk-Needs Assessment: Constitutional and Ethical Challenges, 52 Am. Crim. L. Rev. 231, 242 (2015). [↑](#footnote-ref-40)
41. Pub. L. 89-465, 80 Stat. 214 (1965). [↑](#footnote-ref-41)
42. Marie VanNostrand & Gena Keebler, Pretrial Risk Assessment in the Federal Court, 73 Fed. Probation 1, 4 (2009). [↑](#footnote-ref-42)
43. VanNostrand and Keebler, Pretrial Risk Assessment in the Federal Court, supra note? at 4. [↑](#footnote-ref-43)
44. 18 U.S.C. § 3142(e) [↑](#footnote-ref-44)
45. United States v. Salerno, 481 U.S. 739 (1987). [↑](#footnote-ref-45)
46. Id. at 750. U.S. v. Salerno [↑](#footnote-ref-46)
47. Speedy Trial Act of 1974, 18 U.S.C. § 3161 (2012). [↑](#footnote-ref-47)
48. Probation and Pretrial Services History \ United States Courts; http://www.uscourts.gov/services-forms/probation-and-pretrial-services/probation-and-pretrial-services-history [↑](#footnote-ref-48)
49. Timothy P. Cadigan, Introduction to Federal Probation’s Special Focus on the 30th Anniversary of the Passage of the Pretrial Services Act of 1982, 76 Fed. Probation 2 (2012). [↑](#footnote-ref-49)
50. Schwartztol, Moving Beyond Money, supra note? at 8. [↑](#footnote-ref-50)
51. Id. at 8-9 )Schwartztal, Moving Beyond Money). [↑](#footnote-ref-51)
52. See for example, Bill S.2185 Massachusetts, an Act relative to criminal justice reform, (2017). [↑](#footnote-ref-52)
53. Stephen D. Gottfredson and Laura J. Moriarty, Clinical versus Actuarial Judgment in Criminal Justice Decisions: Should One Replace the Other?Federal Probation 70(2) 15 (2006). [↑](#footnote-ref-53)
54. Kristin Bechtel, Alexander M. Holsinger, Christopher T. Lowenkamp and Madeline J. Warren, A Meta-Analytic Review of Pretrial Research: Risk Assessment, Bond Type, and Interventions, 42 Am. J. Crim. Just. 443, 445 (2017). [↑](#footnote-ref-54)
55. Cynthia Mamalian, State of the Science of Pretrial Risk Assessment 4-5, Pretrial Justice Institute (2011), http://www.pretrial.org/ download/risk-assessment/PJI%20State%20of%20the%20

    Science%20Pretrial%20Risk%20Assessment%20 (2011).pdf; see also, Pretrial Justice Institute, Pretrial Risk Assessment: Science Provides Guidance on Assessing Defendants 3 (2015), http://www.pretrial.org/download/advocacy/Issue%20BriefPretrial%20Risk%20Assessment%20(May%202015).pdf. [↑](#footnote-ref-55)
56. Rizer and Watney, AI can make our jail system more efficient, supra note? After fn33. [↑](#footnote-ref-56)
57. Rizer and Watney, AI can make our jail system more efficient, supra note? After fn 33. [↑](#footnote-ref-57)
58. Bernard Harcourt, Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age 2 (2007). [↑](#footnote-ref-58)
59. Sharad Goel, Ravi Shroff, Jennifer Skeem and Christopher Slobogin, The Accuracy, Equity and Jurisprudence of Criminal Risk Assessment, (2018) <https://ssrn.com/abstract=3306723>; see also, CarlyWill Sloan, George Naufal, and Heather Caspers, The Effect of Risk Assessment Scores on Judicial Behaviour and Deffendant outcomes, IZA Discussion Paper #11948, Institute of Labor Economics, (2018). [↑](#footnote-ref-59)
60. Megan Stevenson, Assessing Risk Assessment in Action,

    103 Minn. L. Rev. 303, 309 (2018). [↑](#footnote-ref-60)
61. Megan Stevenson, Assessing Risk Assessment in Action, supra note? At 309. [↑](#footnote-ref-61)
62. Tavish Srivastava, What is the Difference Between Machine Learning & Statistical Modeling, Analytics Vidhya (07.01.2015);last visited 01.15.2018 https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/#comments [↑](#footnote-ref-62)
63. Id. (Srivastava) [↑](#footnote-ref-63)
64. Ryan Calo, Artificial Intelligence Policy: A Primer and Roadmap, UC Davis Law Review, after fn 23 (2017). [↑](#footnote-ref-64)
65. Id. (Ryan Calo, Artificial Intelligence Policy, near fn 25). [↑](#footnote-ref-65)
66. See example along similar lines in, Ryan Copus, Machine Learning and the Reliability of Adjudication, PHD Dissertation, (for the page to look for “imagine a newly”) (2017) https://escholarship.org/uc/item/7t80m17d [↑](#footnote-ref-66)
67. Justine Baron, Which Machine Learning Algorithm to Choose for my Problem, Recast.AI blog, (02/02/2017) https://recast.ai/blog/machine-learning-algorithms/ [↑](#footnote-ref-67)
68. Jason Brownlee, Supervised and Unsupervised Machine Learning Algorithms, Machine Learning Mastery (03.16.2016) https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ [↑](#footnote-ref-68)
69. Compare for example, Gustavo Machado, ML Basics: Supervised, Unsupervised and reinforcement learning, Medium (10.06.2016) https://medium.com/@machadogj/ml-basics-supervised-unsupervised-and-reinforcement-learning-b18108487c5a [↑](#footnote-ref-69)
70. Jason Brownlee, Supervised and Unsupervised ML, supra note ? [↑](#footnote-ref-70)
71. Richard Berk and Jordan Hyatt, Machine Learning Forecasts of Risk to Inform Sentencing Decisions, Fed. Sentencing Reporter, 222, 223 (2015). [↑](#footnote-ref-71)
72. Berk and Hyat, Machine Learning Forecasts, supra note ? at 225-226. [↑](#footnote-ref-72)
73. Id. (Berk and Hyat, at 225-226). [↑](#footnote-ref-73)
74. Jason Brownlee Supervised and Unsupervised ML, supra note, [↑](#footnote-ref-74)
75. Gustavo Machado, ML Basics, supra note. [↑](#footnote-ref-75)
76. Pimwadee Chaovalit and Nina Lina Zhou, Movie Review Mining: A Comparason Between Supervised and Unsupervised Classification Approaches, Proceeding of the 38th Hawaii International Conference on System Sciences IEEE, (the page that refers to section 2.3) (2005). [↑](#footnote-ref-76)
77. Id. at the end of section 2.3 (Chaovalit and Zhou, Movie Review Mining). [↑](#footnote-ref-77)
78. Thomas Simonini, An Introduction to Reinforcement Learning, Free Code Camp, <https://www.freecodecamp.org/news/an-introduction-to-reinforcement-learning-4339519de419/> [↑](#footnote-ref-78)
79. Thomas Simonini, id. [↑](#footnote-ref-79)
80. Introduction to Reinforcement Learning, Algorithmia Blog 05/2018 <https://blog.algorithmia.com/introduction-to-reinforcement-learning/> [↑](#footnote-ref-80)
81. Applications of Reinforcement Learning in Real World, Towards Data Science, <https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12> [↑](#footnote-ref-81)
82. Introduction to Reinforcement Learning, Algorithmia Blog [↑](#footnote-ref-82)
83. Spike Bradford, Evidence Based, Actuarial Pretrial Risk Assessment, Pretrial Justice Institute, (Aug. 9, 2017) https://university.pretrial.org/blogs/spike-bradford/2017/08/09/evidence-based-actuarial-pretrial-risk-assessment [↑](#footnote-ref-83)
84. Aatash Shah, Machine Learning Vs. Statistics, Edvancer Eduventures (08/01/2016) last visited at Jan. 28 2018 https://www.edvancer.in/machine-learning-vs-statistics/ [↑](#footnote-ref-84)
85. Hal R. Varian, Causal Inference in Economics and Marketing, National Academy of Science, http://www.pnas.org/content/113/27/7310 [↑](#footnote-ref-85)
86. Viktor Mayer-Schonberger and Keneth Cukier, Big Data: A Revolution that will Transform how we Live, Work, and Think, 53. [↑](#footnote-ref-86)
87. Id. (Mayer-Schonberger and Kukier, at 53). [↑](#footnote-ref-87)
88. Richard Berk and Jordan Hyatt, Machine Learning Forecasts of Risk to Inform Sentencing Decisions, Fed. Sentencing Reporter, 222, 223 (2015). [↑](#footnote-ref-88)
89. P.K. Douglas, Sam Harris, Alan Yuille, Mark S. Cohen, Performance Comparison of Machine Learning Algorithms and Number of Independent Components Used in fMRI Decoding of Belief vs. Disbelief, 56(2) NeuroImage 544 (2011). [↑](#footnote-ref-89)
90. Will Koehrsen, Overfitting vs. Underfitting: A Conceptual Explanation, Towards Data Science, 01/2018 <https://towardsdatascience.com/overfitting-vs-underfitting-a-conceptual-explanation-d94ee20ca7f9> [↑](#footnote-ref-90)
91. Tom Dietterich, Overfitting and Under Computing in Machine Learning, 27 ACM Computing Surveys, 326-327 (1995). [↑](#footnote-ref-91)
92. Georgios Drakos, Cross-Validation, Towards Data Science Blog, 08/2018 <https://towardsdatascience.com/cross-validation-70289113a072> [↑](#footnote-ref-92)
93. Id. (Drakos, Cross-Validation) [↑](#footnote-ref-93)
94. Ryan Copus, Machine Learning and the Reliability of Adjudication, supra note? At section 2..1.2 K-fold cross validation. [↑](#footnote-ref-94)
95. Jennifer Doleac, (2017). Let Computers be the Judge, Medium, <https://medium.com/@jenniferdoleac/let-computers-be-the-judge-b9730f94f8c8>, published April 20th 2017 [↑](#footnote-ref-95)
96. Richard A. Berk and Jordan M. Hyatt (2015) Machine Learning Forecasts of Risk to Inform Sentencing Decisions, supra note? At 222-228. [↑](#footnote-ref-96)
97. Zachary C. Lipton (2018). The Mythos of Model Interpretability, ACM Queue Magazine, Volume 16 Issue 3, p. 30. [↑](#footnote-ref-97)
98. Zachary Lipton, The Mythos of Model Interpretability, to look for “what constitutes” [↑](#footnote-ref-98)
99. Pauline Kim, Auditing Algorithms for Discrimination, 166 U. Pen. L. Rev. 189-190 (2017). [↑](#footnote-ref-99)
100. Natalie Ram, Inovating Criminal Justice, 112 Nw. U. L. Rev. 659 (2018) after fn 338 [↑](#footnote-ref-100)
101. Elizabeth E. Joh, The Undue Influence of Surveillance Technology Companies on Policing, N.Y.U. L. Rev. (2017) near fn 17. [↑](#footnote-ref-101)
102. Natalie Ram, Inovating Criminal Justice, supra note? Before fn 139 [↑](#footnote-ref-102)
103. Elizabeth Joh, the Undue Influence of Surveillance, supra note, section b. cornering the market on… [↑](#footnote-ref-103)
104. Elizabeth Joh, the Undue Influence of Surveillance, near fn 51. [↑](#footnote-ref-104)
105. Elizabeth Joh, The New Surveillance Discretion: Automated Suspicion, Big Data and Policing, Harvard Law and Policy Review, (2015) near fn 166. [↑](#footnote-ref-105)
106. Arvind Narayanan, (2018) Translation Tutorial: 21 Fairness Definitions and their Politics, The Conference on Fairness, Accountability and transparency (FAT\*) <https://www.youtube.com/watch?v=wqamrPkF5kk> ; Sahil Verma, and Julia Rubin, (2018). Fairness Definitions Explained, Proceedings of the International Workshop on Software Fairness, IEEE/ACM pages 1-7 <https://dl.acm.org/citation.cfm?id=3194776> [↑](#footnote-ref-106)
107. Aditya Krishna Menon and Robert C. Williamson, The cost of Fairness in Binary Clasification, Section 1.1 [↑](#footnote-ref-107)
108. Moritz Hardt, Eric Price and Nathan Srebro (2016). Equality of Opportunity in Supervised Learning, ‘Advances in neural information procession systems’, published NIPS 2016, p. 3315 (the page that definition 2.1 is on) [↑](#footnote-ref-108)
109. Adam Crawford (2009). Governing Through Anti-social Behavior: Regulatory Challenges to Criminal Justice, The British Journal of Criminology, volume 49 issue 6, 810, to look for “traditional principles” [↑](#footnote-ref-109)
110. Cynthia Dwork, Nicole Immorlica, Adam Tauman Kalai, and Max Leiserson (2017). Decoupled Classifiers for Fair and Efficient Machine Learning, Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81 119-121, 2018. [↑](#footnote-ref-110)
111. Wiseman, Fixing Bail, supra note? at 422-423. [↑](#footnote-ref-111)
112. Wiseman, Fixing Bail, supra note? at 422-423. [↑](#footnote-ref-112)
113. Charles H. Jr. Koch, Judicial Review of Administrative Discretion, 54 Geo. Wash. L. Rev. 469 (before fn 16) (1985). [↑](#footnote-ref-113)
114. Stephen D. Gottfredson and Laura J. Moriarty (2006). Clinical versus Actuarial Judgment in Criminal Justice Decisions: Should One Replace the Other? Federal Probation, volume 70(2), p. 15-16. [↑](#footnote-ref-114)
115. Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein (2019). Discrimination in the Age of Algorithms, NBER Working Paper #25548. To search for psychology. [↑](#footnote-ref-115)
116. Kleinberg et al. at 5. [↑](#footnote-ref-116)
117. Ben Green and Yiling Chen, Disparate Interactions: An-Algorithm-in-the-Loop Analysis of Fairness in Risk Assessment, In FAT\* '19: Conference *on Fairness, Accountability, and Transparency (FAT\* '19) to look for “people interact”.*  [↑](#footnote-ref-117)
118. Timothy Cadigan, James Johnson and Christopher Lowenkamp, *The Re-validation of the Federal Pretrial Services Risk Assessment (PTRA)*, 76 Fed. Probation, 3, 6 (2012). [↑](#footnote-ref-118)
119. Pretrial Services Risk Assessment (PTRA) Frequently Asked Questions, <https://www.pretrial.org/download/risk-assessment/Federal%20Pretrial%20Risk%20Assessment%20Instrument%20FAQ%202010.pdf> [↑](#footnote-ref-119)
120. VanNostrand and Keebler, Pretrial Risk Assessment in the Federal Court, supra note? at 13. [↑](#footnote-ref-120)
121. VanNostrand and Keebler, Pretrial Risk Assessment in the Federal Court, supra note? at 13. [↑](#footnote-ref-121)
122. Christopher Lowenkamp and Jay Whetzel, *The Development of an Actuarial Risk Assessment Instrument for U.S. Pretrial Services*, 73 Fed. Probation 33 (2009). [↑](#footnote-ref-122)
123. Cadigan and Lowenkamp, Implementing Risk Assessment in the Federal Pretrial Services System, supra note? at 32. [↑](#footnote-ref-123)
124. Cadigan, Johnson and Lowenkamp, The Re-validation of the Federal Pretrial Services Risk Assessment, supra note? at 6. [↑](#footnote-ref-124)
125. Thomas H. Cohen and Christopher T. Lowenkamp, Revalidation of the Federal Pretrial Risk Assessment Instrument (PTRA): Testing the PTRA for Predictive Biases, Criminal Justice and Behaviour, volume 46(2) pages 234-260 )2019). [↑](#footnote-ref-125)
126. Cohen and Lowenkamp, revalidating the PTRA supra note, at? To look for 36% (PTRA five). [↑](#footnote-ref-126)
127. Cohen and Lowenkamp, Revalidating the PTRA, to look for 0.67 [↑](#footnote-ref-127)
128. Cohen and Lowenkamp, Revalidating the PTRA to look for testing for predictive bias. [↑](#footnote-ref-128)
129. Cohen and Lowenkamp, Revalidating the PTRA to look for 0.64 [↑](#footnote-ref-129)
130. Cohen and Lowenkamp, Revalidating the PTRA, to look for 0.65. [↑](#footnote-ref-130)
131. Public Safety Assessment (PSA) – Background, Laura and John Arnold Foundation, <https://www.psapretrial.org/about/background> [↑](#footnote-ref-131)
132. Id. [↑](#footnote-ref-132)
133. Laura and John Arnold Foundation (LJAF), Public Safety Assessment: Risk Factors and Formula, (2013), <http://www.arnoldfoundation.org/wp-content/uploads/PSA-Risk-Factors-and-Formula.pdf> [hereinafter Risk Factors and Formula]. [↑](#footnote-ref-133)
134. PSA risk factors formulas, supra note. [↑](#footnote-ref-134)
135. Mathew DeMichele, Peter Baumgartner, Michael Wenger, Kelle Barrick, Megan Comfort and Shilpi Misra, The Public Safety Assessment: a Re-Validation and Assessment of Predictive Utility and Differential Prediction by Race and Gender in Kentucky, +Public Safety Assessment (PSA) – Intro <https://www.psapretrial.org/about/intro> [↑](#footnote-ref-135)
136. Public Safety Assessment (PSA) – Research, The Laura and John Arnold Foundation, <https://www.psapretrial.org/about/research> [↑](#footnote-ref-136)
137. Matthew DeMichele et al. Revalidation of the PSA, supra note? To look for 0.64 [↑](#footnote-ref-137)
138. Matthew DeMichele et al. Revalidation of the PSA, supra note? To cite the discussion. [↑](#footnote-ref-138)
139. Matthew DeMichele et al. Revalidation of the PSA, to cite the discussion. [↑](#footnote-ref-139)
140. Matthew DeMichele, Peter Baumgartner, Kelle Barrick, Megan Comfort, Samuel Scaggs and Shilpi Misra, What do Criminal Justice Professionals Think about Risk Assessment at Pretrial, (to look for 61) (2018) https://www.arnoldfoundation.org/wp-content/uploads/4-Criminal-Justice-Professionals.pdf [↑](#footnote-ref-140)
141. Public Safety Assessment (PSA) – Research, The Laura and John Arnold Foundation, <https://www.psapretrial.org/about/research> [↑](#footnote-ref-141)
142. Pretrial Release, The Access to Justice Lab, (last visited Nov. 22, 2017), http://a2jlab.org/current-projects/signature-studies/pre-trial-release/ [↑](#footnote-ref-142)
143. Virginia Pretrial Risk Assessment Instrument (VPRAI) Instructions Manual, 1, 1-2 , http://www.pacenterofexcellence.pitt.edu/documents/VPRAI\_Manual.pdf [↑](#footnote-ref-143)
144. Marie VanNostrand, Pretrial Risk Assessment: Perpetuating or Disrupting Racial Bias, Pretrial Justice Institute, (last visited Nov. 22, 2017), http://www.pretrial.org/pretrial-risk-assessment-perpetuating-disrupting-racial-bias/ [↑](#footnote-ref-144)
145. Mona J.E. Danner, Marie VanNostrand, Lisa M. Sprunace, *Race and Gender Neutral Pretrial Risk Assessment, Release Recommendations, and Supervision: VPRAI and Praxis Revised*, 1, 9 (2016). [↑](#footnote-ref-145)
146. VanNostrand, Pretrial Risk Assessment: Perpetuating or Disrupting Racial Bias, supra note?. [↑](#footnote-ref-146)
147. Danner et al. Race and gender Neutral Risk Assessment, supra note? At (to look for in sum). [↑](#footnote-ref-147)
148. VPRAI Instruction Manual, supra note? at 8-10. [↑](#footnote-ref-148)
149. Virginia Departement of Criminal Justice Services, Virginia Pretrial Risk Assessment Instrument – (VPRAI) Instruction Manual Amended – Version 4.3 at 8. [↑](#footnote-ref-149)
150. Kenneth Rose, Pretrial Services Agencies: Risk-Informed Pretrial Decision Making in the Commonwealth of Virginia, (Nov. 10, 2016). [↑](#footnote-ref-150)
151. Danner et al. Race and Gender Neutral Pretrial Risk Assessment, supra note? At 1-2. [↑](#footnote-ref-151)
152. Kenneth Rose, Virginia Pretrial Risk Assessment Instrument (VPRAI) and Praxis Overview, 06/2018 pages 12-13. [↑](#footnote-ref-152)
153. Rose, VPRAI and Praxis Overview, supra note? at 16-17. [↑](#footnote-ref-153)
154. VanNostrand, Pretrial Risk Assessment: Perpetuating or Disrupting Racial Bias, supra note?; see also VPRAI fact sheet Stanford page 3 <https://www-cdn.law.stanford.edu/wp-content/uploads/2019/06/VPRAI-Factsheet-FINAL-6-20.pdf> [↑](#footnote-ref-154)
155. 16 C.R.S. § 4-106 (4) (c) (2013). [↑](#footnote-ref-155)
156. Colorado Association of Pretrial Services, The Colorado Pretrial Assessment Tool (CPAT): Administration, Scoring and Reporting Manual, version 2.1 (Jun. 2015). [↑](#footnote-ref-156)
157. The Colorado Pretrial Assessment Tool (CPAT) Revised report, 1, 6 (2012)

     [https://university.pretrial.org/HigherLogic/.../DownloadDocumentFile.ashx?...0](https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=64908e23-bf3e-9379-1a1f-f2d5b9e1702f&forceDialog=0) [↑](#footnote-ref-157)
158. Colorado Association of Pretrial Services, CPAT Manual, supra note? at 3. [↑](#footnote-ref-158)
159. Hillary Smith, Memorandum, Conditions for Supervision by Pretrial Services Programs, Colorado State Publications Library Digital Repository, 1, 9 (Apr. 16, 2012), http://hermes.cde.state.co.us/drupal/islandora/object/co%3A20977/datastream/OBJ/view. [↑](#footnote-ref-159)
160. Joshua J. Luna, *Bail Reform in Colorado: A Presumption of Release,* 88 U. Colo. L. Rev. 1067, 1097 (2017). [↑](#footnote-ref-160)
161. Colorado Association of Pretrial Services, CPAT Manual, supra note? at 4. [↑](#footnote-ref-161)
162. Id. at 4-5 (CPAT Manual) [↑](#footnote-ref-162)
163. Id. at 9 (CPAT Manual ) [↑](#footnote-ref-163)
164. Victoria Terranova and Kyle Ward, Colorado Pretrial Assessment Tool Revision, Commission on Criminal and Juvenile Justice, 12/2017. [↑](#footnote-ref-164)
165. Aubree Cote, Pretrial Release Task Force – Denver Smart Pretrial Program 01/2018 [↑](#footnote-ref-165)
166. The American Bail Coalition, Colorado’s Pretrial Risk Assessment Tool Violates the Americans with Disabilities Act, Fourteenth Amendement, 03/2019 <http://ambailcoalition.org/colorados-pretrial-risk-assessment-tool-violates-the-americans-with-disabilities-act-fourteenth-amendment/> [↑](#footnote-ref-166)
167. Id. (the American bail coaolition) [↑](#footnote-ref-167)
168. Id. (the American bail coaolition) [↑](#footnote-ref-168)
169. Edward Latessa, Paula Smith, Richard Lemke, Matthew Makarios and Christopher Lowenkamp, Creation and Validation of the Ohio Risk Assessment System, Final Report, 07/2009 [↑](#footnote-ref-169)
170. Latessa et al. Creation and Revalidation of the ORAS supra note? at 4. [↑](#footnote-ref-170)
171. Latessa et al. Creation and Revalidation of the ORAS, supra note? at 6. [↑](#footnote-ref-171)
172. Edward Latessa, Sarah Manchak, Jennifer Lux, Jamie Newsome Melissa Lugo and Jordan Papp, The Ohio Risk Assessment System (ORAS): A Revalidation & Inter-Rater Reliability Study, Final Report, 02/2018. [↑](#footnote-ref-172)
173. Latessa et al. Creation and Revalidation of the ORAS supra note at 12-13 +appendix a. [↑](#footnote-ref-173)
174. Letassa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note at 12-13. [↑](#footnote-ref-174)
175. Letassa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note at 29. [↑](#footnote-ref-175)
176. Latessa et al. Creation and Validation of the ORAS, supra note, at 45. [↑](#footnote-ref-176)
177. Latessa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note? At 30. [↑](#footnote-ref-177)
178. Id. at 31-32. (Latessa et al. Revalidation and inter-rater validation) [↑](#footnote-ref-178)
179. Latessa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note at 36-37. [↑](#footnote-ref-179)
180. Letassa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note? At 41-42. [↑](#footnote-ref-180)
181. Letassa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note? At 43-45 [↑](#footnote-ref-181)
182. Practitioners Guide to COMPAS 04/2019 at 2

     [www.equivant.com/wp-content/.../Practitioners-Guide-to-COMPAS-Core-040419.pdf](http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf) [↑](#footnote-ref-182)
183. Practitioners Guide to COMPAS, supra note? at 1-2. [↑](#footnote-ref-183)
184. Julia Angwin et al., Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And it’s Bias Against Black, ProPublica (May 23, 2016); https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [↑](#footnote-ref-184)
185. Equivant, Official Response to Science Advances, 01/18/2018 <https://www.equivant.com/official-response-to-science-advances/> [↑](#footnote-ref-185)
186. COMPAS fact sheet Stanford, <https://www-cdn.law.stanford.edu/wp-content/uploads/2019/06/COMPAS-PRRS-II-Factsheet-Final-6.20.pdf> [↑](#footnote-ref-186)
187. Practitioners Guied to COMPAS, supra note? at 31. [↑](#footnote-ref-187)
188. COMPAS fact sheet Stanford, supra note? [↑](#footnote-ref-188)
189. Practitioners guide to COMPAS at 32. [↑](#footnote-ref-189)
190. Practitioners Guide to COMPAS at 31-32 [↑](#footnote-ref-190)
191. COMPAS fact sheet Stanford, supra note? [↑](#footnote-ref-191)
192. Angwin et al. Machine Bias, supra note? [↑](#footnote-ref-192)
193. Sam Corbett-Davies, Emma Pierson, Ami Feller and Sharad Goel (2016). A Computer Program Used for Bail and Sentencing Decisions was Labeled Biased Against Blacks. It’s Actually not that Clear, The Washington Post, <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?utm_term=.ef319d030999> , published October 17, 2016 [↑](#footnote-ref-193)
194. William Dieterich, Christina Mendoza and Tim Brennan (2016). COMPAS Risks Scales: Demonstrating Accuracy Equity and Predictive Parity Performance of the COMPAS Risk Scales in Broward County, Northpoint Inc., go.volarisgroup.com/rs/430-MBX-989/.../ProPublica\_Commentary\_Final\_070616.pd.. [↑](#footnote-ref-194)
195. Tashea, Courts are Using AI to Sentence Criminals, supra note? Anthony W. Flores, Christopher Lowenkamp, and Kristin Bechtel (2016). False Positives, False Negatives and False Analyses: A Rejoinder to “Machine Bias: There’s Softeware Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks”, Federal Probation volume 80(2), p. 38. [↑](#footnote-ref-195)
196. Karen Hao, AI is Sending People to Jail-and Getting it Wrong, MIT Technology Review, 01/21/2019 <https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/> [↑](#footnote-ref-196)
197. Corbett-Davies et al. Computer Program Used for Bail, supra note? [↑](#footnote-ref-197)
198. Id. [↑](#footnote-ref-198)
199. Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan (2016). Inherent trade-offs in the fair determination of risk scores, In SIGMETRICS '18 Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems, p. 40 ; Alexandra Chouldechova (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments, Big Data, volume 5(2), p. 153-163. [↑](#footnote-ref-199)
200. Elaine Angelino et al., Learning Certifiably Optimal Rule Lists for Categorical Data, 18 Journal of Machine Learning Research, 1, (to search for ProPublica) (2018) [↑](#footnote-ref-200)
201. Id. at 1-2. [↑](#footnote-ref-201)
202. Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, Human Decisions and Machine Predictions, The Quarterly Journal Of Economics, 1 (2017). [↑](#footnote-ref-202)
203. Kleinberg, the section entitled Data and Context , +table II. [↑](#footnote-ref-203)
204. Id. at the beginning of the section entitled b. data [↑](#footnote-ref-204)
205. Kleinberg et al. the section entitled Forming the Prediction Function. [↑](#footnote-ref-205)
206. Jens Ludwig, Man Verses Machine Learning, Criminal Justice in the 21 Century, TedX Talks (May 5, 2017) https://www.youtube.com/watch?v=iAsOq-tAe\_s. [↑](#footnote-ref-206)
207. Kleinberg, Human Decisions and Machine Predictions, supra note? at 4-5. [↑](#footnote-ref-207)
208. Id. at 1. [↑](#footnote-ref-208)
209. Id. at 5. [↑](#footnote-ref-209)
210. Redesign of CJA's Risk Assessment System, http://www.nycja.org/resources/details.php?id=1388 [↑](#footnote-ref-210)
211. Christopher Bavitz, **An Open Letter to the Members of the Massachusetts Legislature Regarding the Adoption of Actuarial Risk Assessment Tools in the Criminal Justice System, Berkman Klein Center for Internet and Society, 11/2017** <https://medium.com/berkman-klein-center/the-following-letter-signed-by-harvard-and-mit-based-faculty-staff-and-researchers-chelsea-7a0cf3e925e9> see also, The Partnership on AI, Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System, <https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/> Sarah Picard-Fritsche, Michael Rempel, Jennifer A. Tallon

     Julian Adler, and Natalie Reyes, Demistifying Risk Assessment: Key Principles and Contravercies, Center for Court Inovation, 03/2017. [↑](#footnote-ref-211)
212. Joint Technology Committee, *Using Technology to Improve Pretrial Release Decision-Making*, Jtc Resource Bulletin 1, 15 (2017). [↑](#footnote-ref-212)
213. See table 1 above [↑](#footnote-ref-213)
214. See table 1 above. [↑](#footnote-ref-214)
215. Kristin Bechtel, Christopher Lowenkamp & Alex Holsinger, Identifying the Predictors of Pretrial Failure, 75 Fed. Probation, 78, 85 (2011). [↑](#footnote-ref-215)
216. Bechtel et al. A Meta-Analytic Review of Pretrial Research, supra note? at 446. [↑](#footnote-ref-216)
217. Bernard Harcourt, Risk As A Proxy for Race: the Dangers of Risk assessment, 27 Fed. Sent. R. 237, 238 (2015). [↑](#footnote-ref-217)
218. Bernard Harcourt, Against Prediction, chapter 5, University of Chicago Press (2007). [↑](#footnote-ref-218)
219. Bernard Harcourt, Against Prediction, chapter 5. [↑](#footnote-ref-219)
220. Christopher Slobogin, Proving The Unprovable: The Role Of Law, Science, And Speculation In Adjudicating Culpability And Dangerousness 113 (2007); Thomas Nilsson et al., The Precarious Practice of Forensic Psychiatric Risk Assessments, 32 Int'l J. L. & Psychiatry 400, 406 (2009). [↑](#footnote-ref-220)
221. Sonja Starr, The Risk Assessment Era: an Overdue Debate, 27 Fed. Sentencing Reporter 205, 205 (2015). [↑](#footnote-ref-221)
222. Colorado Association of Pretrial Services, CPAT Manual, supra note? at 6. [↑](#footnote-ref-222)
223. Letassa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note at 29. [↑](#footnote-ref-223)
224. Latessa et al. Creation and Validation of the ORAS, supra note? at 5-6. [↑](#footnote-ref-224)
225. Cynthia Mamalian, State of the Science of Pretrial Risk Assessment 7, Pretrial Justice Institute (2011), http://www.pretrial.org/ download/risk-assessment/PJI%20State%20of%20the%20

     Science%20Pretrial%20Risk%20Assessment%20 (2011).pdf; see also, Pretrial Justice Institute, Pretrial Risk Assessment: Science Provides Guidance on Assessing Defendants 3 (2015), http://www.pretrial.org/download/advocacy/Issue%20BriefPretrial%20Risk%20Assessment%20(May%202015).pdf. [↑](#footnote-ref-225)
226. Schwartztol, Moving Beyond Money, supra note? at 37. [↑](#footnote-ref-226)
227. Executive Office of the President, Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights 21 (May 2016), <https://www.whitehouse.gov/sites/default/files/>microsites/ostp/2016\_0504\_data\_discrimination.pdf. [↑](#footnote-ref-227)
228. See, e.g., Angele Christin et al., Courts and Predictive Algorithms, supra note? at 7-8. [↑](#footnote-ref-228)
229. Telephone Interview with Marie VanNostrand, Justice Project Manager, Luminosity, (Oct. 13, 2017). [↑](#footnote-ref-229)
230. Koepke and Robinson, Danger Ahead, supra note? At 1729-1730. [↑](#footnote-ref-230)
231. Id. at after fn 94 (Koepke and Robinson, Danger Ahead). [↑](#footnote-ref-231)
232. Ric Simmons, *Quantifying Criminal Procedure: How to Unlock the Potential of Big Data in Our Criminal Justice System*, Mich. St. L. Rev. 947 980 (2016). [↑](#footnote-ref-232)
233. Bernard Harcourt, Against Prediction, supra note? chapter 5. [↑](#footnote-ref-233)
234. Rahul Bhargava, The Algorithms Aren’t Biased, We Are, MIT Media Lab Medium 01/2018 <https://medium.com/mit-media-lab/the-algorithms-arent-biased-we-are-a691f5f6f6f2> [↑](#footnote-ref-234)
235. The Partnership on AI, Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System, 04/2019

     [https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-...](https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/) [↑](#footnote-ref-235)
236. Bail Reform: A Guide for State and Local Policy Makers, Criminal Justice Policy Program at Harvard Law School [CJPP], near fn 115. [↑](#footnote-ref-236)
237. Nicol Turner Lee, Paul Resnick and Gweie Barton, Algorithmic Bias Detection and Mitigation: Best Practises and Policies to Reduce Consumer Harms, The Brookings Institution, <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#cancel> [↑](#footnote-ref-237)
238. Simmons, Quantifying Criminal Justice, supra note? at 982-983. [↑](#footnote-ref-238)
239. Qudsia Siddiqi, Prediction of pretrial failure to appear and an alternative pretrial release risk-classification scheme in New York City: A Reassessment Study, New York City Criminal Justice Agency, (2002). [↑](#footnote-ref-239)
240. Joint Committee, Using Technology to Improve Pretrial, supra note? at 20. [↑](#footnote-ref-240)
241. Cadigan, Johnson and Lowenkamp, The Re-validation of the Federal Pretrial Services Risk Assessment, supra note? at 6. [↑](#footnote-ref-241)
242. Christopher Lowenkamp, Richard Lemke & Edward Latessa, The Development and Validation of A Pretrial Screening Tool, 72 Fed. Probation 2, 3 (2008). [↑](#footnote-ref-242)
243. See discussion above in pages 33-34. [↑](#footnote-ref-243)
244. Pretrial Release, The Access to Justice Lab, supra note? [↑](#footnote-ref-244)
245. Letassa et al. Revalidation & Inter-Rater Validation Study of the ORAS, supra note at 29. [↑](#footnote-ref-245)
246. The Pretrial Justice Institute PJI, The State of Pretrial in America, 11/2017

     [https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?...](https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=f9d452f6-ac5a-b8e7-5d68-0969abd2cc82&forceDialog=0) [↑](#footnote-ref-246)
247. Christopher Bavitz, Sam Bookman, Jonathan Eubank, Kira Hessekiel and Vivek Krishnamurthy, Assessing the Assessments: Lessons from Early State Experiences in the Procurement and Implementation of Risk Assessment Tools, between fns 13-16. [↑](#footnote-ref-247)
248. Kopke and Robinson, Danger Ahead, supra note? After fn 395 [↑](#footnote-ref-248)
249. Pretrial Release, the Access to Justice Lab, Supra note ? [↑](#footnote-ref-249)
250. Mathew Demichele, Megan Comfort, Shilpi Misra, Kelle Barrick and Peter Baumgartnar, The Intuitive Override Model: Nudging Judges Toward Pretrial Risk Assessment Instruments, At the page that fn 5 is in. [↑](#footnote-ref-250)
251. Melissa Hamilton, *Back to the Future: the Influence of Criminal History on Risk Assessments*, 20 Berkeley J. Crim. L. 75, 96-97 (2015). [↑](#footnote-ref-251)
252. Id. at 98. [↑](#footnote-ref-252)
253. Lauryn P. Gouldin, Disentangling Flight Risk from Dangerousness, B.Y.U. L. Rev. 839, 844 (2016). [↑](#footnote-ref-253)
254. Lauryn P. Gouldin, Disentangling Flight Risk from Dangerousness, B.Y.U. L. Rev. 839, 873 (2016). [↑](#footnote-ref-254)
255. Id. at 886. [↑](#footnote-ref-255)
256. Koepke and Robinson, Danger Ahead, supra note? After fn 171. [↑](#footnote-ref-256)
257. Id. after fn 186 (Koepke and Robinson, Danger Ahead). [↑](#footnote-ref-257)
258. Berk and Hyatt, Machine Learning Forecasts of Risk to Inform Sentencing Decisions, supra note? at 223. [↑](#footnote-ref-258)
259. David G. Robinson; Hannah Jane Sassaman; Megan Stevenson, Pretrial Risk Assessments: A Practical Guide for Judges, 57 Judges J. 8, after fn 2 (2018) [↑](#footnote-ref-259)
260. See Mathews v. Edridge, 424 U.S. 319 (1976). [↑](#footnote-ref-260)
261. Melissa Hamilton, Risk-Needs Assessment: Constitutional and Ethical Challenges, supra note? 267-268 (2015). [↑](#footnote-ref-261)
262. Megan Stevenson and Sandra G. Mayson, Pretrial Detention and Bail, in Reforming Criminal Justice volume 3 Pretrial and Trial Processes edited by Erik Luna, 21, 24-25 (2018). [↑](#footnote-ref-262)
263. Nick Oberheiden, Appealing a Pre-trial Detention Order (Bond), Federal Lawyer.com 2016 https://federal-lawyer.com/appealing-a-pre-trial-detention-order-bond/. [↑](#footnote-ref-263)
264. Bavitz, Bookman, Eubank, Hessekiel and Krishnamurthy, Assessing the Assessments, near fn 10. [↑](#footnote-ref-264)
265. Ryan Calo, Artificial Intelligence Policy, supra note? After fn 74. [↑](#footnote-ref-265)
266. PJI, The State of Pretrial in America, supra note? [↑](#footnote-ref-266)
267. Id. at ? to look for 34.1 [↑](#footnote-ref-267)
268. Id. at 11. [↑](#footnote-ref-268)