16

ALL MODELS ARE WRONG, BUT ARE RISK ASSESSMENTS USEFUL?

Colin Doyle

Introduction

This chapter examines the usefulness of a recent development in American pretrial law and policy: actuarial risk assessment tools. To address the United States' crisis of mass pretrial incarceration, actuarial risk assessments have been proposed as a technical, bipartisan reform that may lower jail populations and reduce racial disparities in pretrial incarceration—all without compromising public safety. These algorithmic tools assign defendants risk scores based on the rate at which people with similar characteristics missed court dates, were arrested, or were arrested for a violent crime while on pretrial release. Judges are supposed to rely on these risk scores when they decide to incarcerate or release a person pretrial (*Developing a National Model for Pretrial Risk Assessment*, 2013).

Risk assessments have proven controversial. The main criticism is that the tools are racially inequitable. Risk assessments are built using biased datasets, and the errors that they make disproportionately harm Black defendants (The Leadership Conference on Civil and Human Rights, 2018). This controversy has helped to launch the field of study of algorithmic fairness, which has brought in scholars from a variety of disciplines, including computer science, economics, law, philosophy, psychology, anthropology, and sociology. The first wave of the study of algorithmic fairness focused on formal metrics for quantifying and calibrating equitable outcomes across racial groups (Chouldechova & Roth, 2020). A second wave now seeks to expand that discussion beyond the internal workings of the algorithms to critique the systems in which the algorithms are imbedded (Pasquale, 2019).

As part of that second wave, this chapter conducts a functional analysis of pretrial risk assessments. Rather than study whether risk assessments can produce a fair distribution of outcomes, this chapter examines risk assessments' job performance in the tasks that they have been assigned. As it turns out, risk assessments cannot predict what matters, and what they can predict does not matter.

Actuarial risk assessments' predictions of serious crime are wrong almost all of the time. Under the prevailing legal theory of pretrial incarceration on general dangerousness, the state should preventively lock up people who would commit serious crimes on pretrial release and release people who would not. Accordingly, risk assessments assign defendants a risk score for "new criminal activity." Some tools also flag a defendant as high risk for "new violent criminal activity" (*Public Safety Assessment: Risk Factors and Formula*, n.d.). To generate "new criminal activity" scores, risk assessments predict whether someone will be arrested. But arrests are a poor proxy for predicting serious crime, because arrest data is both over- and under-inclusive. "New criminal activity" predictions capture only a fraction of

DOI: 10.4324/9781003149842-17

serious crimes and mostly predict actions that would not justify preventive incarceration. To generate "new violent criminal activity" scores, risk assessments predict whether someone will be arrested for a violent crime. But predicting violence is hard, the pretrial period is short, and these predictions are rarely correct.

Risk assessments fare better at predicting who will miss future court dates. But this prediction doesn't add valuable information to the pretrial decision-making process. Judges and algorithms can both look at a person's history of missed court dates to predict if the person will miss future court dates. Missing a court date is not the same as flight. People miss court dates for many innocuous reasons that do not warrant preventive incarceration: lack of transportation, childcare and work requirements, necessary medical treatment, mental health problems, to name a few. If a court system wants to invest resources to stop people from missing court dates, it's better off investing in interventions instead of predictions.

This chapter starts with a primer on pretrial risk assessment tools, continues with a usefulness critique, and concludes by addressing counterarguments.

Primer On Actuarial Pretrial Risk Assessment Tools

Before examining actuarial pretrial risk assessment tools' usefulness, it's helpful to understand how these tools work, why they've been adopted, and how they have been critiqued.

How They Work

Although dozens of different actuarial pretrial risk assessments exist, they are quite uniform in design. Risk assessments are built to predict when a person released pretrial will (1) miss a court date, (2) be arrested, and (3) be arrested for a violent crime (*Public Safety Assessment: Risk Factors and Formula*, n.d.). To make these predictions, the tools' developers build statistical models based on personal characteristics that correlate with these outcomes. These characteristics often include age, history of arrest, history of convictions, and time spent in jail or prison. Age tends to be the most predictive factor (Stevenson & Slobogin, 2018). Some tools consider only a person's age and criminal history. Other tools are more eclectic and include personal information such as owning a cellphone or renting, rather than owning, a home (*The Colorado Pretrial Assessment Tool Revised Report*, 2012).

After someone is arrested but before an initial bail hearing, a court officer administers the risk assessment. Depending on the tool, an interview with the accused person may or may not be required. After the court officer inputs the person's background information, the risk assessment assigns that person a risk score and recommends jailing, releasing, or conditionally releasing the person. The court officer gives the assessment's results to the judge and attorneys with the rest of defendant's file.

Why They Have Been Adopted

Risk assessments have appealed to state and local governments as a bipartisan, technical solution to the ongoing crisis in American bail. The pretrial incarceration levels in this country are staggering: on any given day, American jails detain nearly half a million people who have been accused but not convicted of a crime (Zeng, Zhen, 2020). The United States has four percent of the global population but 20 percent of the global pretrial jail population (Lee, 2015). The astonishing truth is that there are more legally innocent people behind bars in America today than there were convicted people in jails and prisons in 1980 (*Prisoners in 1980*, 1981).

Pretrial laws vary from state to state, but there are typically two ways that judges are able to jail someone before their case has been decided: either directly, through an order of pretrial incarceration, or indirectly, by imposing unaffordable money bail.

Nearly every state legally allows judges to incarcerate some people pretrial based on their apparent threat to public safety. Most jurisdictions allow judges to incarcerate only people facing certain, serious violent charges. In some states, judges can incarcerate someone only to prevent grave physical harm. The U.S. and state constitutions require significant procedural protections before someone can be incarcerated without a conviction. The prosecution must prove the defendant's dangerousness by clear and convincing evidence at a special hearing in which defense counsel can cross-examine witnesses and present evidence (Doyle, Bains, & Hopkins, 2019).

Incarceration via money bail is a more frequently used, looser, often unconstitutional process. At an initial hearing following arrest, a judge—or often a magistrate—can condition a person's release from jail upon the payment of a certain bond amount. Those who can afford bond, or who can at least afford a bail bond company's fee, leave jail. Those who can't pay are locked up until their case is over. Bail hearings are brief, often less than a minute long. In many places bond amounts are predetermined based on the criminal charge, and defendants have no right to court-appointed counsel. Because people are jailed with minimal procedure and remain in jail only because they can't afford a bond amount, these practices violate people's constitutional right to due process. In recent years, a series of civil rights lawsuits have challenged bail practices in states across the country, resulting in preliminary injunctions and pushing many states and counties to consider changing their laws and policies (Civil Rights Corps, 2020).

Money bail represents some of the worst excesses of American criminal law. The process is discriminatory on its face, locking up the poor and letting the rich go free. Black and Hispanic communities are disproportionately harmed because they tend to be less wealthy than white communities. But the inequities don't end there. Judges tend to overestimate the risk of Black defendants committing crimes on pretrial release and underestimate the risk of white defendants committing crimes on pretrial release (Arnold, Dobbie, & Yang, 2018). Accordingly, money bail is imposed more often on Black defendants than white defendants. Black defendants receive higher bail amounts than white defendants and are more likely to be incarcerated pretrial. People who are jailed because they cannot afford to post bail may lose their jobs, their homes, and custody of their children (Criminal Justice Policy Program, 2016). Even when someone is able to post bail, their community often suffers because bail bond companies require family or friends to co-sign bail bonds and put up their homes or cars as collateral.

Money bail has become the darling of criminal justice reform. Although there's now a broad bipartisan consensus among the public and policymakers that the current bail system is unjust, no particular solution has gathered similarly universal support. Think tanks, scholars, community groups, and non-profits have proposed reforms that include the elimination of money bail, heightened procedural protections for pretrial incarceration, reform of pretrial services, and the elimination of pretrial incarceration altogether (Doyle, Bains, & Hopkins, 2019). Actuarial risk assessment tools are, by far, the most commonly adopted bail reform measure.

In the span of just a few years, actuarial pretrial risk assessments have spread across the country to over 1,000 counties in all but four states (*National Landscape*, 2020). Some states, like New Jersey, have adopted a uniform risk assessment for every court in the state (*Public Safety Assessment New Jersey Risk Factor Definitions*, 2018). But this is the exception. The decision to adopt pretrial risk assessments is more often made at the county level, which can result in a patchwork of risk assessments in use across a single state. In California alone, over a dozen different pretrial risk assessment tools are used, while some counties do not use pretrial risk assessments at all (*National Landscape*, 2020).

Risk assessments are pitched as an objective, neutral reform that harnesses the power of big data to "moneyball" pretrial decision-making ("The Moneyball Effect," 2014). If Major League baseball could adopt statistical algorithms that outperformed baseball scouts' assessments of minor league prospects, so too might local governments adopt algorithms that outperform judges' assessments of people accused of crimes, or so the thinking goes. As the developers of one tool explain, switching to a system "in which judges have access to scientific, objective risk assessment tools could further our central goals of increasing public safety, reducing crime, and making the most effective, fair, and efficient use of public resources" (*Developing a National Model for Pretrial Risk Assessment*, 2013). The providers of risk assessment instruments assure local governments that the tools are objective and race-neutral. In the academic literature, some economists have also been enthusiastic about resolving the crisis in bail on technical terms, arguing that replacing judges with machine learning tools could dramatically lower both racial disparities and incarceration rates without any apparent tradeoffs (Kleinberg et al., 2017).

Some of the popularity of actuarial risk assessment tools can be explained by their low cost and easy implementation. Unlike, say, the demand to eliminate money bail—which for most states would require statutory changes and a legislative battle with a powerful bail bond lobby—risk assessments can be adopted at the behest of a judicial committee, or even a single chief judge or bureaucrat. The acquisition and administrative costs are low. Many risk assessments have been created by non-profits that do not charge for the tools and offer free technical assistance. Although touted as a bail reform measure, adopting risk assessments does not require changes to existing money bail procedures. Many localities have adopted risk assessment tools without changing any other part of their pretrial systems. Existing court staff, often probation officers, can administer the risk assessment. Some tools do not require an interview with the defendant and can be run automatically with data that courts already collect. Even the change within bail hearings can be minimal: the judge and attorneys, if present, receive a copy of the accused person's risk score and proceed as usual.

Pretrial risk assessments have become a divisive issue. Today, advocates for civil rights and civil liberties oppose both the adoption and continued use of these tools. More than 100 civil rights and community groups, including the ACLU and the NAACP, have signed a statement opposing pretrial risk assessment tools because they "threaten to further intensify unwarranted discrepancies in the justice system and to provide a misleading and undeserved imprimatur of impartiality for an institution that desperately needs fundamental change" (*The Leadership Conference on Civil and Human Rights*, 2018). This commitment is beginning to play out politically. In 2020, both Ohio and Massachusetts considered and rejected adopting pretrial risk assessments statewide. In Ohio, the ACLU ran a campaign advocating for bail reform without risk assessment that swayed the state supreme court (Krouse, 2020). In Massachusetts, a special commission consulted with experts and chose not to move forward with risk assessments (*Final Report of the Special Commission to Evaluate Policies and Procedures Related to the Current Bail System*, 2019). But by and large, pretrial risk assessments have continued to spread across the country, particularly at the county level, and they have their ardent defenders.

Why They Are Controversial

Pretrial risk assessments have become the centerpiece of an ever-growing discourse about algorithmic fairness, accountability, and transparency. They garnered the attention of the field after a 2016 ProPublica report investigated risk assessment outcomes in a Florida county and concluded: "There's software used across the country to predict future criminals. And it's biased against blacks" (Angwin et al., 2016). After obtaining data on pretrial decisions and outcomes through a public records request, ProPublica reporters found that a pretrial risk assessment tool had different error rates for different racial groups. When the algorithm made mistakes, it disproportionately overestimated the dangerousness of Black people and disproportionately underestimated the dangerousness of white people. In statistical language, the false-positive and false-negative rates were different for Black and white populations. This article ignited a discourse on data science accountability and potential measures for algorithmic fairness. The dataset, which ProPublica shared for free online, became a popular intellectual sandbox for testing measures of algorithmic fairness.

The first wave of algorithmic fairness research attempted to answer questions of fairness and equity through adjustments to the statistical models used to build these assessments (Pasquale, 2019). As a case study, the ProPublica report prompted a series of important research questions. Could pretrial risk assessment algorithms be redesigned to eliminate racial disparities in error rates? Are disparities in error rates the right way to evaluate algorithmic fairness? If there are multiple definitions of algorithmic fairness, can they each be met? The goal of the first wave was to find racial balance within an algorithm's inputs, outputs, and errors.

Researchers found that pretrial risk assessments could be redesigned to eliminate racial disparities in error rates. But this could only be achieved by explicitly including race as a factor that a risk assessment considers. The technical challenge is that when different populations have different behaviors, a predictive algorithm will not generate evenly distributed errors. Across any domain, a predictive algorithm will have more false positives for the group that has more positive outcomes and will have more false negatives for the group that has more negative outcomes. It is not possible to create a predictive algorithm with equal error rates for different groups, unless the algorithm explicitly considers and adjusts for group differences (Berk et al., 2017; Chouldechova, 2016; Kleinberg, Mullainathan, & Raghavan, 2016).

Other researchers contended that ProPublica did not identify a problem with risk assessments at all. According to this view, algorithmic fairness does not require error rates to be equal. Instead, risk assessments should be judged by their "predictive parity" across race (Flores, Bechtel, & Lowenkamp, 2016). A statistical tool achieves predictive parity if two people with identical risk profiles but different racial identities would be assigned the same risk score.

Predictive parity is an anemic, hyper-technical conception of fairness for pretrial risk assessments. It appeals, rightly, to the intuition that people with similar risk profiles should be judged similarly. But it assumes, incorrectly, that the risk profiles are reliable and fair. For actuarial risk assessment tools, risk profiles are built using arrest records and criminal records, which are not neutral, objective data sources. These data sources track the activity of police and courts. When risk assessments use this data to predict civilian behavior, the tools use comprehensive information about government action as though it were comprehensive information about people's actions (Elliott, 1995). More heavily policed communities will appear riskier in these datasets, not only because of actually higher crime rates but because of policing practices themselves. Police practices, like stop-and-frisk, combined with stringent enforcement of low-level offenses, like marijuana possession, create arrest records for people in these communities while people who engage in similar behavior but live in less heavily policed areas will not have arrest records. The implicit and explicit biases of police, prosecutors, and judges further distorts the data. Decades of research have shown that, for the same conduct, Black and Hispanic people are more likely to be arrested, prosecuted, convicted, and sentenced to harsher punishments than their white counterparts. People of color are treated more harshly than similarly situated white people at each stage of the criminal legal process, which distorts the data used to develop these risk profiles.

To further complicate things, other research revealed that risk assessments cannot be optimized to satisfy competing definitions of fairness (Kleinberg, Mullainathan, & Raghavan, 2016). Risk assessments cannot achieve both predictive parity and equality across error rates. At a certain point, even conflicts between technical definitions of fairness must be resolved with normative commitments.

A second wave of algorithmic fairness discourse takes a different tack, looking less to the internal structure of the algorithms and more to the background societal structures in which the algorithms operate and the ways that they are implemented. Rather than try to change how the algorithms work, this scholarship asks whether an algorithm is needed and what interests it serves. Second wave research into pretrial risk assessments has questioned whether these tools can justifiably be relied on to promote criminal justice reform or make incarceration decisions, often concluding that these algorithms entrench and obscure harmful penal ideologies (Barabas, n.d.; Barabas et al., 2020; Green, 2018; Richardson, Schultz, & Crawford, 2019).

Usefulness Critique

This chapter conducts a functional analysis of pretrial risk assessments. The study of algorithmic fairness typically explores whether a tool like risk assessments can produce a fair distribution of outcomes. This chapter takes a different approach, examining risk assessments' job performance in the tasks that they have been assigned.

Usefulness and a fair distribution of outcomes are not opposing definitions of algorithmic fairness but are complementary concerns. This can be demonstrated with an example outside the charged context of criminal law. Let's turn to a plausible medical hypothetical: an algorithm that processes mammogram images to detect breast cancer. If the algorithm had different error rates for Black and white populations, there would be a justified concern over distributional fairness. But distributional fairness should not be the only concern. It would also matter whether the algorithm could *actually* detect breast cancer. If the algorithm was ineffective at detecting breast cancer, that would be a problem independent of concerns about the fair distribution of errors and outcomes. So too with risk assessments: Fair distribution and usefulness are independent, important considerations. It matters both that pretrial risk assessments' errors and outcomes are fairly distributed and also that risk assessments can do their job in their first place.

Risk Assessments Cannot Predict What Matters

Under the legal theory of pretrial incarceration on general dangerousness, the state should preventively lock up people who would commit serious crimes on pretrial release and release people who would not. Accordingly, the function of pretrial risk assessments is to identify who should be incarcerated pretrial, based on their likelihood of committing future crime, particularly future violent crime. Under an ideal system of pretrial incarceration on general dangerousness, a judge, an algorithm, or some combination of the two would correctly predict an accused person's future actions if released pretrial. If the person was going to commit a serious enough crime to warrant preventive incarceration, the judge would send them to jail. If not, they would be released. Those who would commit serious crimes would be incapacitated, and those who would not commit serious crimes would be free.

Risk assessment instruments have been introduced because policymakers think judges are overestimating risk and locking up too many people. At bail hearings and pretrial detention hearings, judges decide whether to send someone to jail or release them to the community. One question, above all others, motivates a judge's decision to release or jail someone before trial: Will this person commit a violent crime? Judges understandably do not want to be responsible for releasing someone back into the community who will hurt others. But this fear has led judges to routinely overestimate pretrial violence. Implicit, and sometimes explicit, biases have also led judges to overestimate the dangerousness of Black people. The result is high rates of pretrial incarceration, particularly within Black communities.

The job of algorithmic risk assessments is to correct judges' prediction errors by using data to identify who will commit serious crimes, particularly violent crimes, on pretrial release. Consider a modern risk assessment tool like the Public Safety Assessment: Every defendant receives a "new criminal activity" risk score between 1 (lowest risk) and 6 (highest risk) and is either flagged or not flagged for "new violent criminal activity" (*Public Safety Assessment: Risk Factors and Formula*, n.d.). Most pretrial risk assessments also recommend that a judge incarcerate or release a person based on these scores. After calculating a person's risk score, that score is filtered through a "decision-making framework" or "decision-making matrix" that encourages judges to lock up people with high-risk scores and free most of the rest (*Pretrial Release Recommendation Decision Making Framework (DMF)*, 2018). In at least one state, judges must provide a written explanation if they depart from a risk assessment's recommendation to incarcerate someone. Some jurisdictions only provide a risk score and do not include an explicit recommendation to incarcerate or release someone. But judges can connect the dots: low-risk people should go free and high-risk people should go to jail.

Both "new criminal activity" and "new violent criminal activity" risk scores provide weak justifications for preventive pretrial incarceration. To generate "new criminal activity" scores, risk assessments predict whether someone will be arrested. But arrests are a poor proxy for predicting serious crime, because arrest data is both over- and under-inclusive. To generate "new violent criminal activity" scores, risk assessments predict whether someone will be arrested for a violent crime. But predicting violence is hard, and these predictions are wrong almost all of the time.

By relying upon arrest data, "new criminal activity" risk scores predict a broad range of behavior, including many future actions that would not justify preventive incarceration. Unfortunately for the makers of risk assessments, there is no data source that provides the ground truth for who commits crimes. Risk assessments' predictions of "new criminal activity" are instead predictions of arrest for any reason. Arrest records are over-inclusive because people are wrongly arrested and arrested for minor offenses, including technical violations, petty misdemeanors, status offenses, and acts that should not be criminalized at all (Natapoff, 2018). Some risk assessments define a risk to "public safety" as any "new criminal filing," including for traffic stops and municipal offenses (*The Colorado Pretrial Assessment Tool Revised Report*, 2012). Overall, less than 5 percent of arrests in the United States are for violent offenses, and in many places the most common reason for an arrest is a non-violent traffic offense, such as driving with a suspended license (*Arrests*, 2018). Arrest records are also under-inclusive because they chart only law enforcement activity, and many crimes do not result in arrest. Less than half of all reported violent crimes result in an arrest, and less than a quarter of reported property crimes result in an arrest (*Clearances*, 2019). If the goal of pretrial incarceration is to predict and prevent serious and violent crime, a "new criminal activity" score captures only a sliver of those crimes and predicts mostly other actions entirely.

Risk assessments are not effective at predicting "new violent criminal activity." Predicting the future is always difficult but is nearly impossible when predicting very rare, interpersonal events, like violence, within a short timeframe, like the pretrial period.

A look under the hood at risk assessments' training data can reveal how risk assessments are not expected to predict violence well. Training data consists of the datasets that are used to build a risk assessment tool. A risk assessment's performance with training data is indicative of its performance in the field, but results vary when the tools is actually deployed. With the Public Safety Assessment (PSA), every defendant is flagged or not flagged for "new violent criminal activity." Within the training data, of those who were flagged for "new violent criminal activity" (NVCA), only 7 percent went on to commit a violent crime on pretrial release (*14B Measuring and Managing Pretrial Risk with the Public Safety Assessment: Assessor Training*, n.d.). In other words, the "new violent criminal activity" flag can be expected to correctly identify who will commit violence about 7 per cent of the time and can be expected to incorrectly identify who will commit violence about 93 per cent of the time. This pattern has proven to be fairly consistent in places where the PSA has been adopted, despite regional variations in pretrial incarceration rates, policing practices, crime rates, and more. Over different time periods, the percent of people flagged for NVCA who go on to be arrested for a violent crime on pretrial release has been 14 per cent in New Jersey, 3 per cent in Kentucky, and only 1 per cent in Cook County, Illinois, which includes Chicago (Corey, 2019). Risk assessments make errors when predicting who will not commit violence, too. Risk assessments do not flag the overwhelming majority of people who go on to commit violence on pretrial release. The result is that a jurisdiction that uses the "New Violent Criminal Activity" flag to determine whether people are incarcerated pretrial will jail mostly people who would not commit violence and will not prevent most instances of pretrial violence.

There is little reason to expect the quality of these predictions to improve. Recent decades have seen little improvement in our ability to predict violence, among any populations. Although pretrial risk assessments have only recently come to prominence, they are far from the first actuarial predictive tools used to predict violence. Most states have long used risk assessments in other parts of their criminal legal systems including prison management, probation, and parole. Pretrial risk assessments are an offshoot of a broader field of risk prediction in the social sciences, particularly psychology, which has developed a rich literature on the subject. A leading meta-analysis of risk assessment tools concludes that "the ceiling of predictive efficacy may have been reached with the available technology" (Yang et al., 2010). Rare, interpersonal events are among the hardest to predict, and historical background information about people can only tell us so much about what they will do in the future. Given these limits, the authors conclude that risk assessments, "should not be used as the sole or primary means for clinical or criminal justice decision-making that is contingent on a high level of predictive accuracy, such as preventive detention."

Incarcerating people based on actuarial predictions of violence runs counter to foundational principles of criminal law. Incarcerating people flagged for "new violent criminal activity" requires primarily incapacitating people who do not need to be incapacitated. A risk assessment optimized to produce the most accurate results would not flag for anyone for violence. For any person in a dataset, the most likely outcome is that they will not commit violence on pretrial release. Therefore, to produce predictions of violence and encourage incarceration, risk assessments must sacrifice accuracy. In doing so, they generate substantially more false positives—people who are flagged for violence but do not go on to commit a violent crime—than true positives—people who are flagged for violence and do go on to be arrested for a violent crime. The necessary result is jailing the many people labeled "high risk" to prevent the violence of a few of their members. Consider a group of 100 people flagged for "new violent criminal activity." According to the PSA, roughly seven of those 100 would be expected to commit a violent crime on pretrial release. and ninety-three other people must also be incarcerated to prevent the crimes of those seven.

At the trial stage, a person can only be convicted and subject to incarceration based upon proof beyond a reasonable doubt that the person committed a crime. But under this model, at the pretrial stage a person can be subject to incarceration for resembling people who commit violent crimes. And the resemblance need not be that strong. Consider one of the central maxims in our legal tradition: William Blackstone's "[B]etter that ten guilty persons escape, than that one innocent suffer," echoed by John Adams following the Boston Massacre, "[W]e are to look upon it as more beneficial, that many guilty persons should escape unpunished, than one innocent person should suffer" (Volokh, 1997). Pretrial incarceration based on these predictions generates—in some cases precisely—the opposite ratio: better for ten people to be incarcerated than for one to commit a violent crime on pretrial release.

What Risk Assessments Can Predict Does Not Matter

Actuarial risk assessments are somewhat better at predicting who will miss a future court date. But it's not that hard to predict missed court dates. Risk assessments do little more than check to see if someone has missed court dates in the past and project that same behavior into the future (*Public Safety Assessment: Risk Factors and Formula*, n.d.). Judges do not need cutting-edge statistical tools to make the connection—they can easily see who has and hasn't missed court dates in the past and arrive at a similar conclusion.

Algorithms might be slightly better at predicting missed court dates than judges, but that marginal improvement in prediction shouldn't affect pretrial decision-making. By itself, a prediction of a missed court date is not grounds for pretrial incarceration. People miss court dates for many reasons that do not warrant preventive incarceration: lack of transportation, childcare and work requirements, necessary medical treatment, mental health problems, and substance abuse problems, to name a few. Only rarely does a missed court date indicate that someone has fled flight from the jurisdiction (Gouldin, 2018).

If a court system wants to invest resources to ensure people make their court dates, it is better off investing in interventions over predictions (Barabas et al., 2017). In states across the country, phone and text reminders for upcoming court dates have proven to increase court appearance rates. To use one example, Multnomah County, Oregon ran a pilot program that placed automatic calls to pretrial defendants to alert them of upcoming court dates. The program increased appearance rates by 31 percent and saved the county over a million dollars in eight months (Doyle et al., 2019). Jurisdictions can also help people with vouchers for public transit or allow people to waive some in-person appearance requirements.

Counterarguments

Aren't Risk Assessments Better than Judges?

Risk assessments should be able to modestly outperform judges at making the same predictions of arrest or arrest for a violent crime. But by itself this comparison is not enough to justify using risk assessments to send people to jail for predictions of serious crime. Neither judges nor algorithms can predict general dangerousness well—judges just happen to be the worst of the two.

In the literature, risk assessments are commonly evaluated against human decision-making. Statistical models can modestly outperform unaided human prediction for predicting arrest and arrest for a violent offense. A meta-analysis of over half a century of psychology research concludes that "statistical prediction methods are, in general, more accurate than clinical prediction methods." (Ægisdóttir et al., 2006). But statistical methods are no crystal ball. This same meta-analysis acknowledges that these tools are, in general, only slightly more accurate than humans. For pretrial risk assessments, a recent empirical paper arrives at the opposite conclusion. (Dressel & Farid, 2018). Within a study framework that gave people specific factors to consider and provided feedback on the accuracy of their predictions, human subjects recruited on the internet were able to slightly outperform the COMPAS pretrial risk assessment tool. Some have criticized the study's framework for being too divorced from real-world circumstances, given that judges have more information in front of them and rarely know whether their predictions were correct or not. (Goel et al., 2018). Perhaps the problem is not so much with the study design but with a legal system that doesn't provide judges feedback on their predictions. But whether this criticism is valid or not, the point remains that statistical tools are, at best, a modest improvement over human prediction.

One might expect that a combination of judges and algorithms would fare better than algorithms on their own, but that's not case. Judicial overrides are likely to produce worse outcomes than algorithms acting alone. Studies of judges, probation officers, and other criminal justice professionals all reveal that human overrides decrease accuracy in predicting arrest. (Goel et al., 2018).

Some of the recent literature oversells the superiority of machine prediction for pretrial decision-making. A number of scholars have constructed risk assessment models and simulated their risk assessment's performance against historical records of judges' bail decisions. (Baradaran & McIntyre, 2012; Berk et al., 2016; Kleinberg et al., 2017). The risk assessments tend to outperform the judges. But these papers aren't perfect. As Megan Stevenson argues, an ideal study design "would be explicitly set up as a horse race between the two approaches" rather than a comparison between an algorithm and a historical record. (Stevenson, 2018). Instead, these papers assume that judges' objectives when making past bail decisions universally align with the risk assessment algorithm's objectives. This is unlikely to be the case, as Stevenson and Slobogin have shown using the example of age. (Stevenson & Slobogin, 2018). A person's age tends to be one of the most predictive characteristics for future arrest. All things being equal, younger people are arrested more frequently than older people. From a purely preventive incarceration standpoint, judges should be more inclined to lock up young people pretrial because they are more likely to commit crimes. But in practice, judges may look at age quite differently. Young people may have less culpability because of their stage of mental development, have greater capacity to change if given the opportunity, or suffer more within the harsh conditions of jail. Papers that compare algorithms to the historical record assume that judges are making mistaken predictions when they release "high-risk" people, when judges may often be acting from a different set of values.

Comparisons between risk assessments and humans making the same predictions are relevant. It would be foolish to replace human predictions with statistical predictions that are less accurate. All things being equal, risk assessments are preferable to judges' unaided predictions of general dangerousness. But all other things don't need to be equal. Myopically comparing human and machine predictions ignores the possibility that neither machines nor humans can make accurate enough predictions of general dangerousness to justify incarcerating people on those grounds.

Couldn't Risk Assessments Be Used Only for Release?

There's an option for risk assessments to be use early in the pretrial process as a means of releasing "low-risk" people without the need for a hearing before a judge. Under this line of reasoning, high-risk scores for "new criminal activity" and "new violent criminal activity" may be too weak to justify detention, but low-risk scores on both counts could justify release. At a preliminary stage, such as booking at a police station, a risk assessment could be run. People who are lower risk could be released immediately without spending any time in jail awaiting their day in court. Indeed, the state of New Jersey has implemented just such a program, which can be credited, in part, for reducing the state's pretrial jail population and reducing the number of people who spend any amount of time in jail pretrial. (Doyle, Bains, & Hopkins, 2019).

Actuarial risk assessments could serve this role, but there are potential drawbacks. Under the principle of parsimony, the state should avoid imposing any unnecessary jail time. If someone ought to be released pretrial, it's better for that person not to be stuck in jail until a hearing can be scheduled for a judge to order their release. The parsimony principle counsels for *some* mechanism of early, automatic pretrial release, not necessarily through actuarial assessment. Bear in mind that protocols for early release are never just protocols for release. By releasing some people and not others, these protocols also determine who to keep in jail, at least until a hearing can be scheduled. Given that risk assessments' predictions of serious crime are weak, a prediction of serious crime may not be the right filtering rule. As a simple alternative, a jurisdiction could have a policy of who should be released pretrial. Rather than rely on an actuarial tool's generalized predictions about future behavior, a jurisdiction may find other concerns more salient—like the nature of the alleged offense, the age and health of the accused person, and their caregiving or work responsibilities. Or as an alternative to actuarial risk assessment, a much simpler, more straightforward predictive tool could be built. An algorithm could be trained to predict who judges in the jurisdiction release pretrial. The people whom judges tend to release could be automatically released early.

Second, the racial inequities of risk assessments would continue to be a concern under this new arrangement. As discussed earlier in the chapter, the risk profiles that these tools construct are distorted because they rely upon biased data. Risk assessments build risk profiles using policing and court data. Because this data is biased against poor communities and communities of color, people from those communities will be considered higher risk than similarly risky people from other communities. If risk assessments use these risk profiles to grant and deny early release, the outcomes will be inequitable. Many people from poor communities and communities of color will remain in jail while similarly risky people from other communities are released.

Conclusion

Actuarial risk assessments dominate the landscape of American bail reform, largely on a promise to fix the money bail crisis with technical wizardry. No doubt, data science and statistical tools have made incredible strides in recent decades. But it is hard to predict how people will behave in short timespan based on limited information about their past. With pretrial risk assessments, the challenge becomes even greater because of poor data, rare outcomes, and high stakes. Risk assessments cannot meet the challenge and are unlikely to do so any time soon.

There may be a constructive lesson here. Risk assessments have always been bargain-basement bail reform. They demand little from government coffers, slide in with existing procedures, and often let the money bail system—and its attendant harms—continue unabated. Sometimes you can't afford cheap. Mass pretrial incarceration is a serious, moral crisis that demands transformative change. Resolving the money bail crisis will require meaningful, sometimes difficult changes to law, policy, and culture. Technology may help the process, but it cannot carry the burden.

References

14B Measuring and Managing Pretrial Risk with the Public Safety Assessment: Assessor Training. (n.d.). Arnold Ventures.

Ægisdóttir et al. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist*, 359.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. ProPublica. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932.

Arrests. (2018, September 28). Arrest Trends. https://arresttrends.vera.org/arrests

Barabas, C. (n.d.). Beyond Bias: Re-imagining the Terms of "Ethical AI" in Criminal Law. *Georgetown Journal of Law and Modern Critical Race Perspectives*, 12(2), 40-42.

Barabas, C., Dinakar, K., Ito, J., Virza, M., & Zittrain, J. (2017). Interventions Over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment. ArXiv:1712.08238 [Cs, Stat]. http://arxiv.org/abs/1712.08238

Barabas, C., Doyle, C., Rubinovitz, J., & Dinakar, K. (2020). *Studying Up: Reorienting the study of algorithmic fairness around issues of power*. 9. FAT★ '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. https://dl.acm.org/doi/abs/10.1145/3351095.3372859

Baradaran, S. & McIntyre, F. L. (2012). Predicting Violence. *Texas Law Review*, *90*, 75.

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2017). Fairness in Criminal Justice Risk Assessments: The State of the Art. ArXiv:1703.09207 [Stat]. http://arxiv.org/abs/1703.09207

Berk, R., Susan, B. S., & Barnes, G. (2016). Machine Learning Risk Assessment at Preliminary Arraignments for Domestic Violence. University of Pennsylvania, Department of Criminology, Working Paper No. 2016-2.0.

Chouldechova, A. (2016). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. ArXiv:1610.07524 [Cs, Stat]. http://arxiv.org/abs/1610.07524

Chouldechova, A. & Roth, A. (2020). A Snapshot of the Frontiers of Fairness in Machine Learning. *Communications of the ACM*, *63*(5), 82–89. https://doi.org/10.1145/3376898

Civil Rights Corps. (2020). *Challenging the Money Bail System*. www.civilrightscorps.org/work/wealth-based-detention

Clearances. (2019). FBI. https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/clearances

Corey, E. (2019, August 8). How a Tool to Help Judges May Be Leading Them Astray. The Appeal. https://theappeal.org/how-a-tool-to-help-judges-may-be-leading-them-astray/

Criminal Justice Policy Program. (2016). Moving Beyond Money: A Primer on Bail Reform. http://cjpp.law.harvard.edu/assets/FINAL-Primer-on-Bail-Reform.pdf

Developing a National Model for Pretrial Risk Assessment. (2013). The Laura And John Arnold Foundation. https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/LJAF-research-summary_PSA-Court_4_1.pdf

Doyle, C., Bains, C., & Hopkins, B. (2019). Bail Reform: A Guide for State and Local Policymakers (p. 106).

Dressel, J. & Farid, H. (2018). The Accuracy, Fairness, and Limits of Predicting Recidivism. *Science Advances*, *4*(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Elliott, D. (1995). *Lies, Damn Lies and Arrest Statistics*, 1.

Final Report of the Special Commission to Evaluate Policies and Procedures Related to the Current Bail System. (2019). https://d279m997dpfwgl.cloudfront.net/wp/2020/01/0102_bail-reform-report.pdf

Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks." *80*(2), 9.

Goel, S., Shroff, R., Skeem, J. L., & Slobogin, C. (2018). The Accuracy, Equity, and Jurisprudence of Criminal Risk Assessment. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3306723

Gouldin, L. P. (2018). Defining Flight Risk. *University of Chicago Law Review*, *85*, 677.

Green, B. (2018). "Fair" Risk Assessments: A Precarious Approach for Criminal. *Justice Reform*, 5.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). Human Decisions and Machine Predictions*. *Quarterly Journal of Economics*. https://doi.org/10.1093/qje/qjx032

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores. ArXiv:1609.05807 [Cs, Stat]. http://arxiv.org/abs/1609.05807

Krouse, P. (2020, January 24). Ohio Supreme Court Proposes Bail Reforms That Don't Include Risk Assessments. Cleveland.Com. www.cleveland.com/news/2020/01/ohio-supreme-court-proposes-bail-reforms-that-dont-include-risk-assessments.html

Lee, M. Y. H. (2015, April 30). Does the United States Really Have 5 Percent of the World's Population and One Quarter of the World's Prisoners? *Washington Post*. /www.washingtonpost.com/news/fact-checker/wp/2015/04/30/does-the-united-states-really-have-five-percent-of-worlds-population-and-one-quarter-of-the-worlds-prisoners/

Natapoff, A. (2018). *Punishment Without Crime: How Our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal* (First edition). New York: Basic Books.

National Landscape. (2020). Mapping Pretrial Injustice. https://pretrialrisk.com/national-landscape/

Pasquale, F. (2019, November 25). The Second Wave of Algorithmic Accountability. *Law and Political Economy*. https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/

Pretrial Release Recommendation Decision Making Framework (DMF). (2018). New Jersey Courts.

Prisoners in 1980. (1981). U.S. Department of Justice, Bureau of Justice Statistics. www.bjs.gov/content/pub/pdf/p80.pdf

Public Safety Assessment New Jersey Risk Factor Definitions. (2018). New Jersey Courts. https://njcourts.gov/courts/assets/criminal/psariskfactor.pdf?c=99i

Public Safety Assessment-Risk Factors and Formula. (n.d.). The Laura and John Arnold Foundation.

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review*, *94*, 42.

Stevenson, M. T. (2018). Assessing Risk Assessment in Action. *Minnesota Law Review*, *103*. https://doi.org/10.2139/ssrn.3016088

Stevenson, M. T. & Slobogin, C. (2018). Algorithmic Risk Assessments and the Double-Edged Sword of Youth. *Washington University Law Review*, *96*, 26.

The Colorado Pretrial Assessment Tool Revised Report. (2012).

The Leadership Conference on Civil and Human Rights. (2018). More than 100 Civil Rights, Digital Justice, and Community-Based Organizations Raise Concerns About Pretrial Risk Assessment. https://civilrights.org/2018/07/30/more-than-100-civil-rights-digital-justice-and-community-based-organizations-raise-concerns-about-pretrial-risk-assessment/

The Moneyball Effect: How Smart Data is Transforming Criminal Justice, Healthcare, Music, and Even Government Spending. (2014, January 28). *TED Blog*. https://blog.ted.com/the-moneyball-effect-how-smart-data-is-transforming-criminal-justice-healthcare-music-and-even-government-spending/

Volokh, A. (1997). N Guilty Men. *University of Pennsylvania Law Review*, *174*.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools. *Psychological Bulletin*, *136*(5), 740–767. https://doi.org/10.1037/a0020473

Zeng, Zhen. (2020). Jail Inmates in 2018. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. www.bjs.gov/content/pub/pdf/ji18.pdf