Algorithms should have made courts more fair. What went wrong?

Enlarge / The courts have become another frontline in the battle against climate change.

reader comments

122 with 82 posters participating

Share this story

Kentucky lawmakers thought requiring that judges consult an algorithm when deciding whether to hold a defendant in jail before trial would make the state's justice system cheaper and fairer by setting more people free. That's not how it turned out.

Before the 2011 law took effect, there was little difference between the proportion of black and white defendants granted release to await trial at home without cash bail. After being mandated to consider a score predicting the risk a person would reoffend or skip court, the state's judges began offering no-bail release to white defendants much more often than to blacks. The proportion of black defendants granted release without bail increased only slightly, to a little over 25 percent. The rate for whites jumped to more than 35 percent. Kentucky has changed its algorithm twice since 2011, but available data shows the gap remained roughly constant through early 2016.

The Kentucky experience, detailed in a study published earlier this year, is timely. Many states and counties now calculate

“risk scores” for criminal defendants that estimate the chance a person will reoffend before trial or skip court; some use similar tools in sentencing. They are supposed to help judges make fairer decisions and cut the number of people in jail or prison, sometimes as part of eliminating cash bail. Since 2017, Kentucky has released some defendants scored as low-risk based purely on an algorithm’s say-so, without a judge being involved.

How these algorithms change the way justice is administered is largely unknown. Journalists and academics have shown that risk-scoring algorithms can be unfair or racially biased. The more crucial question of whether they help judges make better decisions and achieve the tools’ stated goals is largely unanswered.

The Kentucky study is one of the first in-depth, independent assessments of what happens when algorithms are injected into a justice system. It found that the project missed its goals and even created new inequities. “The impacts are different than what policymakers may have hoped for,” says Megan Stevenson, a law professor at George Mason University who authored that study.

Stevenson looked at Kentucky in part because it was a pioneer of bail reform and algorithm-assisted justice. The state began using pretrial risk scores in 1976, a simple system that assigned defendants points based on questions about their employment status, education, and criminal record. The system was refined over time, but the scores were used inconsistently. In 2011, a law called HB 463 mandated their use for judges’ pretrial decisions, creating a natural experiment.

Kentucky’s lawmakers intended HB 463 to reduce incarceration

rates, a common motivation for using risk scores. They are supposed to make judges better at assessing who is safe to release. Sending a person home makes it easier for them to continue their work and family life and saves the government money. More than 60 percent of the 730,000 people held in local jails in the US have not been convicted, according to the nonprofit Prison Policy Initiative.

The system used in Kentucky in 2011 employed a point system to produce a score estimating the risk that a defendant will skip their court date or reoffend before trial. A simple framework translated the score into a rating of low-, moderate-, or high-risk. People tagged as low- or moderate-risk generally should be released without cash bail, the law says.

But judges appear not to have trusted that system. After the law took effect, they overruled the system's recommendation more than two-thirds of the time. More people got sent home, but the increase was small; around the same time, authorities reported more alleged crimes by people on release pending trial. Over time, judges reverted to their prior ways. Within a couple of years, a smaller proportion of defendants was being released than before the bill came into force.

Although more defendants were granted release without bail, the change mostly helped white people. "On average white defendants benefited more than black defendants," Stevenson says. The pattern held after Kentucky adopted a more complex risk-scoring algorithm in 2013.

One explanation supported by Kentucky data, she says, is that judges responded to risk scores differently in different parts of the state. In rural counties, where most defendants were white, judges granted release without bond to significantly more people. Judges in urban counties, where the defendant pool was more mixed, changed their habits less.

A separate study using Kentucky data, presented at a

conference this summer, suggests a more troubling effect was also at work. It found that judges were more likely to overrule the default recommendation to waive a financial bond for moderate-risk defendants if the defendants were black.

Harvard researcher Alex Albright, who authored that study, says it shows more attention is needed to how humans interpret algorithms' predictions. "We should put as much effort into how we train people to use predictions as we do into the predictions," she says.

Michael Thacker, risk-assessment coordinator with Kentucky pretrial services, said his agency tries to mitigate potential bias in risk-assessment tools and talks with judges about the potential for "implicit bias" in how they interpret the risk scores.

An experiment that tested how judges react to hypothetical risk scores for determining sentences also found evidence that algorithmic advice can cause unexpected problems. The study, which is pending publication, asked 340 judges to decide sentences for made-up drug cases. Half of the judges saw "cases" with risk scores estimating the defendant had a medium to high risk of rearrest and half did not.

When they weren't given a risk score, judges were tougher on more-affluent defendants than poor ones. Adding the algorithm reversed the trend: Richer defendants had a 44 percent chance of doing time but poorer ones a 61 percent chance. The pattern held after controlling for the sex, race, political orientation, and jurisdiction of the judge.

"I thought that risk assessment probably wouldn't have much effect on sentencing," says Jennifer Skeem, a UC Berkeley professor who worked on the study with colleagues from UC Irvine and the University of Virginia. "Now we understand that risk assessment can interact with judges to make disparities worse."

There is reason to think that if risk scores were implemented carefully, they could help make the criminal justice system fairer. The common practice of requiring cash bail is widely acknowledged to exacerbate inequality by penalizing people of limited means. A National Bureau of Economic Research study from 2017 used past New York City records to project that an algorithm predicting whether someone will skip a court date could cut the jail population by 42 percent and shrink the proportion of black and Hispanic inmates, without increasing crime.

Unfortunately, the way risk-scoring algorithms have been rolled out across the US is much messier than in the hypothetical world of such studies.

Criminal justice algorithms are generally relatively simple and produce scores from a small number of inputs such as age, offense, and prior convictions. But their developers have sometimes restricted government agencies using their tools from releasing information about their design and performance. Jurisdictions haven’t allowed outsiders access to the data needed to check their performance.

“These tools were deployed out of reasonable desire for evidence-based decision making, but it was not done with sufficient caution,” says Peter Eckersley, director of research at Partnership on AI, a nonprofit founded by major tech companies to examine how the technology affects society. PAI released a report in April that detailed problems with risk assessment algorithms and recommended agencies appoint outside bodies to audit their systems and their effects.

Stevenson agrees that greater transparency is needed—but also admits to feeling it may be too late to turn risk-scoring algorithms into a success, given their poor reputation and the slim gains they seem to offer. “The criminal justice system has such little good will already that I don’t want people to lose any more hope or faith at this point,” she says.

This story originally appeared on wired.com.

https://arstechnica.com/?p=1564003