Adam Teske
12/9/2021
Cog Sci & The Law

The Pitfalls of Predicting Recidivism in Court

The American justice system is founded almost entirely on the preventive method of punishment. Rather than encouraging rehabilitation and reward, our penitentiaries keep inmates isolated for the duration of their sentence and then push them back into a world they no longer recognize. It's no wonder a large portion of convicts return to a life of crime once they get out of prison. Many studies have found that our ability to predict the likelihood of someone returning to a life of crime- often referred to more simply as *recidivism*- is marginally better than chance, yet the testimony of faux experts like Dr. James Grigson have gotten innocent people executed because they claim to know with full certainty whether or not someone is 'incurable'. Science and objective truth should always be center stage in criminal trials, but I don't believe anyone should be allowed to testify on the likelihood of recidivism in individual cases. People shouldn't be sentenced to death based on a number that the world's top scientists and algorithms can't accurately predict.

Part 1: Cognitive-Behavioral Studies on Recidivism and it's Predictability

There are practically infinite variables that can influence the likelihood of someone committing another crime post-incarceration. Many studies in the past few decades have examined this phenomenon. The vast majority of them agree that the likelihood of reoffence is difficult to pin down empirically, and socioeconomic factors that are outside the person's control play an enormous role in that dodgy figure.

Adam Teske
12/9/2021
Cog Sci & The Law

We can get fairly accurate recidivism models using modern technology and psychological/socioeconomic analyses, but most of the time they are marginally better than chance. As both our population and technological capabilities rapidly expand, our criminal justice system increasingly relies on risk assessment algorithms (Monahan & Skeem, 2014). These programs are supposedly the best way to determine the risk of recidivism because they can juggle an unfathomable amount of data and remove human bias entirely, but whether or not they actually achieve this is a topic of much debate in the world of criminal justice.

In 2018, psychologists Julia Dressel & Hany Farid examined the accuracy of a commercially used risk assessment software called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). They found that this program, which has now been used to assess well over a million ex-convicts, was no better at predicting recidivism than random people crowdsourced on the internet. Dressel & Farid also briefly mention an article that purportedly found evidence of a racial bias within COMPAS. This older study found that the algorithm produced a false positive rate (percentage of people not rearrested but classified as high-risk) of 47.7% for black defendants and a false negative rate (percentage of people rearrested but not classified as high-risk) of 44.9%. These figures were nearly twice the failure rate for white defendants (*Angwin* et al., 2016, as referred to by *Dressel & Farid*, 2018). It should be mentioned that the company that created the software has since rebuffed these claims, arguing that the recidivism rates predicted by COMPAS are a close match to the national averages.

In a related study, Skeem & Lowenkamp (2020) developed 5 alternative algorithmic methods for the Post Conviction Risk Assessment (PCRA) program that attempt to counteract any latent racial bias, specifically in relation to violent crimes. The predictive utility (PPV) of the algorithm improved with 2 of the 3 conditions, and error rates went down across the board. However, even

under the best conditions, the false positive rate for black people was 11.67% and the false negative rate was 71.19% (Skeem & Lowenkamp, 2020).

Research in this field has consistently found that race is correlated to recidivism. Black persons are statistically more likely to be rearrested and reincarcerated in a smaller time frame than their white counterparts (Langan & Levin, 2002). However, this same body of work shows that socioeconomic factors are similarly influential in the likelihood of rearrest. Those who live in poorer neighborhoods are more likely to reoffend than those who don't, and the average black person is in a worse socioeconomic position than the average white person (Lofstrom & Raphael, 2016). Black people are also more likely to be arrested for the same crime than white people (Papachristos et al., 2011), which would naturally lead to higher incarceration and reincarceration rates. These factors could very easily account for the racial discrepancy we see in American incarceration statistics.

There is also much more to the picture than just socioeconomic status and ethnicity. Some of the other highest-impact factors are age, personality, IQ/reading ability, mental health, and the nature of the original crime (Farrington & Tarling, 1985). Interestingly enough, many researchers agree that incarceration without rehabilitation has little to no effect on recidivism rates **(**Loeffler & Nagin, 2021). A study run in Pennsylvania showed that prior incarceration had no significant effect on whether felons reoffended between a year and a decade after their release (Nagin & Snodgrass, 2013), and other researchers found that prison may actually increase the chances of rearrest among adult ex-convicts (Harding et al., 2017, Mueller-Smith, 2015). Some newer studies show that jail time can successfully discourage criminal relapse, but with the stipulation that rehabilitation is a central focus. This can be observed in more progressive

countries like Norway, a country with a substantially smaller prison population and an emphasis on getting convicts back into society (Bhuller et al., 2019).

Recidivism, simply put, is difficult to predict. We have advanced machine learning programs and decades of psychological research on the subject, but even with that wealth of data at our disposal we still see error rates of 50% or higher in some predictions for rearrest. This subsect of forensic psychology must still be addressed in court though, so we should tread very carefully when we present these widely varying figures to a jury.

Part 2: The G2i Problem

A core issue in the judicial system that has only recently come to the foreground is the process of generalizing scientific findings (framework evidence) to individual cases (diagnostic testimony). This is known as the problem of G2i and is an enormously important concept in the world of criminology.

The main difference between framework and diagnostic testimony is that the latter operates under different standards than the former. The vast majority of scientific study is done "at the group level and [described] statistically" (*Faigman*, 2014), and there is an inherent hurtle in applying crowdsourced data to individual cases. Framework evidence (such as research on breast cancer or the accuracy of eyewitness testimony) can inform juries on complex topics, but information drawn from a population cannot be applied with 100% accuracy to any single situation. Two experts in the same field could have wholly conflicting opinions on certain cases depending on their niche or unique experiences.

There are currently two preeminent methods of filtering expert testimony used by American courtrooms. The first of the two was established in the court case *Frye v. United States*, in which the court set a rather vague precedent: expert testimony must be generally accepted in the field before it can be applied to the case at hand. The second set of standards that were established in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* are a bit more demanding. In short, the court must determine that the evidence in question is relevant, the expert is properly qualified (through a combination of knowledge, skill, experience, training, and education), and the testimony is supported by "good grounds." These guidelines are substantially less blurry than *Frye's* but still leave a good deal of room for interpretation.

Using models like *Daubert* and *Frye* to evaluate expert testimony only recently became an important part of the legal process. Evaluating the credibility of the relevant scientific evidence is crucial to the integrity of the system. Unfortunately, it's clear that even today these standards are not always met.

Part 3: Case Law, Expert Testimony, & Recidivism

It is common practice for psychologists to provide insight to the criminal mind during a trial, particularly ones that involve violence. This can help juries make decisions on whether or not the defendant was of sound mind when they committed the crime and can shed light on the nature of certain psychopathological disorders. These experts must undergo the same vetting process as every other scientist that is asked to present scientific findings in court. An excellent example of why the issue of G2i is so important to address is the litany of cases psychiatrist James Grigson

irresponsibly testified on that resulted in the death penalty, where no such process seemed to take place.

Case 1: Willingham v. State

In the *Willingham v. State* trial, Cameron Todd Willingham was tried for the murder of his three daughters who burned to death in their Texas home in 1992. He was denied a writ of *habeus corpus* and executed on February 17, 2004 (*Willingham v. State, 1992*).

Psychiatrist James Grigson (in)famously insisted he could predict whether a violent criminal would reoffend with 100% certainty. This case was no exception. He testified with full confidence that Willingham was an incurable sociopath who would no doubt murder again if given the chance. Without so much as shaking the man's hand, he condemned him. Grigson was later expelled from the American Psychiatric Association and the Texas Society of Psychiatric Physicians for unethical conduct. He is now known colloquially as "Dr. Death" for the overwhelming number of cases he testified on that concluded with capital punishment (Dow, 2006).

Fire marshal Manuel Vasquez violated similar legal standards when he made his statement against Willingham. He claimed there were three points of origin for the fire and evidence of accelerants on the floor. He concluded that the blaze must have been started intentionally, and most likely by Willingham. Practically all of his forensic analysis was later debunked as junk science by nine independent arson experts and a handful of investigative journalists (Dow, 2006). The Texas Forensic Science Commission also later found that the state and local arson investigators used flawed science during the trial. Despite the fact that Vasquez's testimony was not generally accepted in the field (violating the Frye standards), nor his evidence supported by

appropriate validation (violating the Daubert standards), his statement was still taken and ultimately helped put a man to death.

This case is a perfect example of why the Frye and Daubert standards exist. This potentially innocent man was executed thanks in large part to two "experts" that were put on the stand despite not having their evidence properly vetted. It also speaks to a deeper level of corruption (or gross incompetence) within the system that so much evidence came to light supporting Willingham's case, yet he was denied any sort of retrial or recourse.

Case 2: Barefoot v. State

Thomas A. Barefoot was convicted of shooting and killing Harker Heights police officer Carl Levin during the summer of 1978. The jury ultimately came to a guilty verdict and sentenced Barefoot to be executed. One of the people who testified against him was none other than disgraced psychiatrist James Grigson (*Barefoot v. State, 1980*).

Grigson claimed that Barefoot was beyond saving, would always be a menace to society, and recommended the death penalty. This eerily familiar statement eventually helped land Barefoot a death sentence. While it is true that Barefoot was a repeat offender with dangerous tendencies and the evidence against him was damning, no one can make a claim that terminal without at least a thorough psychological evaluation. Even ignoring that glaring injustice, Grigson's testimony was practically identical to the one he gave in Willingham's trial, despite the drastically different circumstances and evidence at play.

These cases (along with the other 165 Grigson testified on) lead me to doubt if we should even entertain expert testimony of this nature in the first place. If this "doctor" is making

indistinguishable prognoses on such disparate cases without even talking to the defendant, something in the system is failing.

Case 3: Anderson III v. State

On August 15, 1999, Cornealious Michael Anderson III was convicted of armed robbery after stealing $2000 from a Burger King manager at gunpoint. He tried to appeal to the court that his Baretta brochure they found had no legal relevance to the case and was therefore inadmissible. He also filed another appeal claiming he had inadequate legal representation at his trial. His original sentence was upheld both times by the appeals judge/justices (*State v. Anderson, 2002*).

This case garnered national attention in 2013 when it was discovered that Anderson had not actually served his 13-year sentence. After his initial appeals were rejected, the Missouri Department of Corrections somehow neglected to actually put the man in jail. He went on to live a crime-free life for more than a decade. He got married, founded a company, voted, and paid taxes, all while using his real name and address. The state of Missouri attempted to force Anderson to serve his sentence once the mistake was discovered. He was put away for a short period of time, but backlash from the public and an earnest statement of forgiveness from his original victim eventually bought his freedom for good.

According to modern forensic psychologists and risk assessment tools, Anderson would've fallen into several categories that would make him at high risk of recidivism (black skin, poor socioeconomic status, relatively young at the time of the robbery, etc.). However, given a second chance very few have ever been gifted, Anderson decided to live a clean life and give back to his

community. This is further evidence of why recidivism rates for large groups of people should not be applied to individual cases- anyone can turn their life around if given the opportunity.

Throughout history, innocent people have lost their lives to junk science. While we are thankfully beyond the era of burning women to death to see if they're a witch, filtering and validating expert testimony is a vital gear in the judicial process that we are still struggling with today. The chances of recidivism for any individual person cannot be confidently predicted, so experts should not be allowed to testify to that note on a diagnostic level. Informing a jury of the likelihood of rearrest among various demographics could be a reasonable route to take depending on the case, but the buck should stop at framework evidence. As we can see with some of the trials James Grigson testified on, using theories of recidivism as diagnostic evidence can be absolutely devastating. Innocent lives can and will be lost if we're not careful with how we present the data to the court.

Adam Teske
12/9/2021
Cog Sci & The Law

References

Court Cases

Anderson v. State, 76 S.W.3d 275 (2002) https://scholar.google.com/scholar_case?case=7300002629983244774&q=Cornealious+Anderson+v.+Missouri&hl=en&as_sdt=6,36

Barefoot v. State, 596 S.W.2d 875 (1980) https://scholar.google.com/scholar_case?case=2822235788528672046&q=James+grigson&hl=en&as_sdt=6,36

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993) https://supreme.justia.com/cases/federal/us/509/579/

Frye v. United States, 293 F. 1013 D.C. Cir. (1923). https://casetext.com/case/frye-v-united-states-7

Willingham v. State, 897 S.W.2d 351 (1995) https://law.justia.com/cases/texas/court-of-criminal-appeals/1995/71544-4.html

Articles & Books

Bhuller M, Dahl GB, Løken KV, Mogstad M. 2019. Incarceration, recidivism, and employment. *J. Political Econ.* 128(4):1269–324

Dow, D. R. (2006). *Executed on a technicality: Lethal injustice on America's death row*. Beacon Press.

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

Faigman, D. L., Monahan, J., & Slobogin, C. (2013). Group to individual (g2i) inference in scientific expert testimony. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2298909

Farrington, D. P., Tarling, R., & Hill, G. (1985). Predicting Recidivism Using Institutional Measures. In *Prediction in criminology* (pp. 96–105). essay, State University of New York Press.

Harding DJ, Morenoff JD, Nguyen AP, Bushway SD. 2017. Short- and long-term effects of imprisonment on future felony convictions and prison admissions. *PNAS* 114(42):111038

Langan, P. A., & Levin, D. J., Recidivism of prisoners released in 1994 (2002). Washington DC; U.S. Dept. of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Loeffler, C. E., & Nagin, D. S. (2021). The Impact of Incarceration on Recidivism. *Annual Reviews*. https://doi.org/10.1146

Lofstrom, Magnus, and Steven Raphael. 2016. "Crime, the Criminal Justice System, and Socioeconomic Inequality." *Journal of Economic Perspectives*, 30 (2): 103-26

Monahan J, Skeem JL. The evolution of violence risk assessment. CNS Spectr. 2014 Oct;19(5):419-24. doi: 10.1017/S1092852914000145. Epub 2014 Mar 28. PMID: 24679593.

Mueller-Smith M. 2015. *The criminal and labor market impacts of incarceration*. Work. Pap., Univ. Mich., Ann Arbor, MI

Nagin DS, Snodgrass GM. 2013. The effect of incarceration on re-offending: evidence from a natural experiment in Pennsylvania. J. Quant. Criminol. 29(4):601–42

Papachristos, A. V., Smith, C. M., Scherer, M. L., & Fugiero, M. A. (2011). More coffee, less crime? the relationship between gentrification and neighborhood crime rates in Chicago, 1991 to 2005. *City & Community, 10*(3), 215–240. https://doi.org/10.1111/j.1540-6040.2011.01371.x

Rhodes W, Gaes GG, Kling R, Cutler C. 2018. Relationship between prison length of stay and recidivism: a study using regression discontinuity and instrumental variables with multiple break points. *Criminol. Public Policy* 17(3):731–69

Skeem, J. L., & Lowenkamp, C. (2020). Using algorithms to address trade-offs inherent in predicting recidivism. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3578591