Impact of Risk Assessment on Judges’ Fairness in Sentencing Relatively Poor Defendants

By

Jennifer Skeem
University of California, Berkeley

Nicholas Scurich
University of California, Irvine

John Monahan
University of Virginia School of Law

Abstract #3316266

A complete index of University of Virginia School of Law research papers is available at: Law and Economics: http://www.ssrn.com/link/U-Virginia-LEC.html
Impact of risk assessment on judges’ fairness in sentencing relatively poor defendants

Jennifer Skeem,1 Nicholas Scurich,2* & John Monahan3

1 Goldman School of Public Policy & School of Social Welfare, University of California, Berkeley. 120 Haviland Hall, Berkeley, CA 94720-7400. Email: jenskeem@berkeley.edu
2 Department of Psychological Science & Department of Criminology, Law & Society, University of California, Irvine. Email: nscurich@uci.edu
3 School of Law, University of Virginia. Email: jmonahan@law.virginia.edu

Abstract

The increasing use of risk assessment algorithms in the criminal justice system has generated enormous controversy. Advocates emphasize that algorithms are more transparent, consistent, and accurate in predicting re-offending than judges’ unaided intuition, while skeptics worry that algorithms will increase racial and socioeconomic disparities in incarceration. Ultimately, however, judges make decisions—not algorithms. In the present study, real judges (n=340) with criminal sentencing experience participated in a controlled experiment to test whether the provision of risk assessment information interacts with a defendant’s socioeconomic class to influence sentencing decisions. Results revealed that risk assessment information reduced the likelihood of incarceration for relatively affluent defendants, but the same risk assessment information increased the likelihood of incarceration for relatively poor defendants. This finding held after controlling for the sex, race, political orientation, and jurisdiction of the judge. It appears that under some circumstances, risk assessment information can increase sentencing disparities.

MAIN TEXT

Introduction

Algorithms have become ubiquitous in daily life. They identify the fastest route to your destination, reveal news stories you might find relevant, and highlight products you’re likely to buy. Algorithms also power predictive analytics that can inform decisions about people in almost every sector of public policy, including criminal justice. Data and technology are now readily available to expand the reach and impact of risk assessment—which is the well-established practice of using checklists or algorithms that summarize risk factors to estimate a person’s likelihood of future re-offending (31). Risk factors are variables like young age and criminal history that have been shown in research to predict future criminal behavior.

These advances are timely. Today, policymakers are keenly interested in using risk assessment as a tool for criminal justice reform (23). In fact, risk assessment is “the engine that drives” a federal prison reform bill that was just signed into law (11). Across the U.S., jurisdictions have been undertaking a variety of efforts to reduce absurdly high rates of incarceration (25) without compromising public safety. Risk assessment can be helpful in this regard. One way to safely reduce the human and fiscal cost of mass incarceration is to identify the people who are least likely to reoffend and release them, supervise them in the community on probation or parole, or abbreviate their period of incarceration (21; 23). Advocates argue that—when risk is a legally
relevant consideration—judges should consider risk assessment algorithms to improve the consistency, transparency, and accuracy of their decisions (e.g., 21; 26).

Judges routinely make momentous decisions in a person’s life that include consideration of the likelihood that the person will re-offend—and must make their own seat-of-the-pants judgments, without algorithms (12). At the pretrial stage, each of the 30,000 daily arrests in the U.S. (9) requires a judge to decide whether to release a defendant until their court date or keep them in jail to prevent them from absconding or reoffending before their case disposition. At the sentencing stage, each conviction requires a judge to determine an appropriate sentence. Sentencing traditionally focuses on backward-looking concerns about the defendant’s blameworthiness for a past crime, but the highly influential Model Penal Code also provides a limited role for forward-looking concerns about preventing future crimes (1). In a recent survey, 8 out of 10 judges believed that both blameworthiness and risk of re-offending should be considered at sentencing (22; see also 4).

In the pretrial context and sentencing context, risk assessment has been shown to outperform judicial intuition in predicting reoffending. Based on a sample of 758,027 arrestees, Kleinberg et al. (19) found that replacing judicial decisions about pretrial release with algorithmic decisions would reduce crime by up to 25% with no change in the incarceration rate (see also 18). In a study that assessed judicial intuition about 962 felony offenders at sentencing, Gottfredson (12) found that an algorithm (d=.90) predicted recidivism more strongly than judges’ subjective predictions (d=.54)—and that judges’ subjective predictions of recidivism strongly affected their sentencing choices. Gottfredson (12) suggested, “the use of…empirically derived methods would enhance the rationality of sentencing when risk is determined by the sentencing theory…to be a relevant and justifiable consideration” (p. 88).

Despite the clear promise of risk assessment, such suggestions have been met with intense criticism. The principal concern is that using risk assessment to inform judicial decisions will increase racial and socioeconomic disparities in incarceration. In an era of general skepticism about the fairness of algorithms (6; 27), critics assert that risk factors included in some risk assessment instruments (e.g., education level, marital status, neighborhood disadvantage) are “proxies” for minority race and poverty (34). In the view of Former Attorney General Eric Holder (17), the broad use of risk assessment “may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society.”

This concern is important—but largely untested. Whether risk assessment exacerbates, ameliorates, or has no effect on disparities in sentencing is a relative inquiry: risk assessment, compared to what existing practices (33)? Existing practices include sentencing guidelines that heavily weight criminal history and have been shown to contribute to racial disparities (10); and judges’ intuitive appraisals of risk, which are less transparent, consistent, and accurate than risk assessment (see above) are much like those of other people—largely intuitive, heuristic-based, and subject to bias (15; 29). Like other people, judges may stereotype Black men as threatening (e.g., 29; 37); and poor people as incapable, untrustworthy, and antisocial (7; 28). Currently, Black people are five times more likely to be imprisoned than White people (3) and boys born into households in the bottom 10% of earners are 20 times more likely to be imprisoned than those in the top 10% of earners (20; see also 41). Risk assessment could exacerbate these existing disparities, as Holder speculates. But risk assessment could instead have no effect on—or even reduce disparities—as others have predicted (13; 16).
To our knowledge, no studies have directly tested the effect of risk assessment on judges’ sentencing decisions. Van Wingerden, van Wilsem and Moerings (38) compared judges’ sentences of 3,059 statistically matched pairs of Dutch defendants whose presentence reports included or omitted formal risk assessment results—and found no significant increase in the probability of incarceration when risk assessment was added, even for high risk defendants. But defendant’s protected characteristics were not examined in this study. Several studies have examined the racial predictive fairness of risk algorithms alone (e.g., 5; 33), but algorithms do not determine sentences—judges do; and risk is only one consideration. Because risk assessments rarely provide dispositive answers to legal questions, it is necessary to examine how algorithms affect human judgment. From a “compared to what?” perspective, the essential question is whether adding risk assessment to other case information has a different effect on judges’ sentences, depending on the defendant’s race or socioeconomic class.

In the present study, we address this essential question. Real judges with criminal sentencing experience participated in a controlled experiment to test whether the provision of risk assessment interacts with a defendant’s socioeconomic class to change sentencing decisions. Because sentences can be influenced by a host of case characteristics and judicial tendencies, we used an experimental design to permit causal inference about the variables of interest. Judges were randomly assigned to review one of four written case vignettes that described a defendant who had been convicted of a drug offense. The case vignettes varied in only two independent factors: whether the defendant was relatively poor or affluent; and whether a set of risk assessment information was provided or omitted. After reading the case vignette, the judges then issued a sentence. If risk assessment exacerbates disparities, as Holder predicts, then providing judges with risk assessment information will increase sentencing severity significantly more for relatively poor defendants than their more affluent counterparts.

Results

To recruit judges for this study, we deliberately partnered with judiciary leaders in jurisdictions located in the Eastern, Midwestern, and Southwestern U.S. With the support of these partners, we invited judges with adult sentencing experience to anonymously participate in a study “that explores factors related to sentencing decisions.” Of judges invited to participate, 91% (N=340) did so, either at an annual judicial conference (Eastern and Midwestern jurisdictions) or online (Southwestern jurisdiction). In all three of the jurisdictions, presentence investigation reports that describe the defendant’s index offense, criminal history, and social background are submitted to the court to inform sentencing decisions. In all but the Eastern jurisdiction, these reports include the results of a risk assessment instrument.

We randomly assigned each of the 340 judges to review one of four drug case vignettes that were formatted as presentence investigation reports, and then issue a sentence. As shown in Figure 1, the four vignettes held information about the case constant, except for whether the defendant was relatively poor and whether risk assessment information was provided. Socioeconomic status was operationalized as the defendant’s occupation and level of education. For example, the defendant was either a “casual laborer in construction who dropped out of high school” or “an Apple Store employee with a bachelor’s degree in computer science.” Formal risk assessment information was either omitted or provided—and included an orientation to the instrument, the defendant’s total score, classification as “medium-to-high risk of re-arrest,” and scores on specific risk factors. The relatively poor defendant and relatively affluent defendant earned the same risk scores. To maximize ecological validity, we worked with local experts including judiciary leaders to tailor
the vignettes to law and practices in the three jurisdictions. When relevant, we used the jurisdiction’s specific risk assessment instrument. We wrote vignettes in a manner that maximized judicial discretion: Based on statutory criteria in each jurisdiction, the defendant was eligible for a sentence of probation, jail or prison. After reviewing their randomly assigned case vignette, judges sentenced the defendant to one of these three outcomes. We collapsed jail and prison sentences to focus on the contrast between probation and the more severe sentences of incarceration.

To statistically address the study aims, we conducted three binary logistic regression models using fixed effects variables (orthogonal contrasts, see 40) for defendant’s socioeconomic status (relatively poor or relatively affluent) and risk assessment information (provided or omitted), with sentences of incarceration (vs. probation) specified as the outcome variable. For all regression analyses we utilized a bootstrapping procedure with 1,000 samples to estimate the standard errors. Model 1 contained only the manipulated variables (i.e., socioeconomic status, risk assessment, and the interaction term). Model 2 contained the manipulated variables plus a variable to control for jurisdiction. Model 3 contained the manipulated variables, jurisdiction, plus variables to control for characteristics of the judge (i.e., sex, ethnicity and political orientation). The results of each model are displayed in Table 1 and reported below.

The impact of risk assessment information on judges’ sentencing severity depends on defendants’ socioeconomic status. Overall, 52.5% (n=177) of the judges sentenced the defendant to incarceration rather than probation. Model 1 tested whether the likelihood of incarceration varied as a function of the experimental treatment conditions. As shown in Table 1 (column 2), a significant crossover interaction was detected, indicating that the influence of risk assessment information differed, depending on whether the defendant was relatively poor or affluent. Providing formal risk assessment information decreased the probability of incarceration for the relatively affluent defendant but increased the probability of incarceration for the relatively poor defendant. Before interpreting this principal finding, we conducted additional tests to ensure that it was robust.

Sentencing severity also varies as a function of jurisdiction. Sentences varied by jurisdiction. Specifically, a sizeable majority (73%) of the Southwestern judges sentenced the defendants to incarceration, compared to only 31% of the Eastern judges and 20% of the Midwestern judges.

It is important to determine whether the previous results hold, after accounting for any differences across jurisdictions. To do so, we tested Model 2—which added a dummy variable for jurisdiction to the variables in the previous model. As shown in Table 1 (third column), results indicate that the aforementioned statistical interaction remained significant, after controlling for jurisdiction—and jurisdiction also predicted sentencing severity. After accounting for the interaction between socioeconomic status and risk assessment, defendants’ odds of incarceration were 11.6 (95%CI=5.77, 23.51) times higher in the Southwestern than the Midwestern jurisdiction; with no differences between the Midwestern and Eastern jurisdictions. (Logistic regression coefficients can be exponentiated to produce odds ratios).

Sentencing severity does not vary as a function of judge’s characteristics. Although we did not ask judges to identify themselves, we did ask them to indicate their sex, race, and political
affiliation. These values are shown in Table 2. One might expect individual difference variables to partly explain our finding that the likelihood of incarceration was higher in the Southwestern than Midwestern jurisdiction. For example, conservatism is correlated with “tough on crime” attitudes (e.g., Tonry, 2004), and 43% of the Southwestern judges identified as Republican compared to just 20% of Midwestern judges (see Table 2).

To determine whether the previously-detected interaction holds, after accounting for both jurisdiction and judges’ personal characteristics, we conducted Model 3, which adds judges’ sex, ethnicity and political orientation to the variables in the previous model. Since 81% of judges were White and non-Hispanic (see Table 2), all other ethnicities were combined into an “other” value for analyses. As shown in Table 1 (column 4), none of the judges’ characteristics explained additional variance in sentencing, but the jurisdiction variable remained significant, with Southwestern judges being approximately 10 times more likely (Exp(B)=9.61, 95% CI=4.52, 20.41) to incarcerate than Midwestern judges. This suggests that jurisdiction is not just a proxy for ideology, even though Southwestern judges were disproportionately conservative.

Even after controlling for jurisdiction and judges’ characteristics, the impact of risk assessment depends on defendant’s socioeconomic status. The results of Model 3 (see Table 2, column 4) principally indicate that the interaction term of interest is robust across jurisdictions and judicial characteristics. (Although the main effects of risk assessment and socioeconomic status are also statistically significant, they are not meaningful in the presence of this interaction.)

Discussion

In this era of the algorithm, the use of risk assessment to inform criminal justice decisions has never simultaneously had more widespread appeal and invoked more criticism. Advocates of reform emphasize that algorithms are more transparent, consistent, and accurate in predicting reoffending than judges—and could help reduce incarceration without jeopardizing public safety. Skeptics worry that using algorithms to inform decision-making will add a veneer of objectivity while “baking in” systemic bias that disparately impacts poor people and racial minorities. Both sides seem blind to the fact that ultimately, judges make decisions—not algorithms. Thus, we explicitly examined the interface between risk assessment and human judicial decision-making.
In a case designed to maximize judicial discretion, we found that adding risk assessment information reversed the direction of judges’ disparities in sentencing relatively poor vs. affluent defendants. This reversal held after controlling for judges’ jurisdiction and personal characteristics. We believe this reversal occurred because (a) many judges—and the Model Penal Code (1)—attempt to balance competing sentencing considerations that include the defendant’s blameworthiness and risk (35), (b) socioeconomic status has opposite effects on perceptions of blameworthiness for committing a past crime versus perceptions of risk for committing a future crime (see 24), and (c) cuing judges to focus on risk reframes how they process socioeconomic status. Providing judges with risk assessment information transformed low socioeconomic status from a circumstance that reduced the likelihood of incarceration (perhaps by mitigating perceived blameworthiness) to a factor that increased the likelihood of incarceration (perhaps by increasing perceived risk).

Specifically, without risk assessment information, judges were less likely to sentence the relatively poor defendant to incarceration than his more affluent counterpart (45.8% vs. 59.5% probabilities, respectively). In this context, judges may have implicitly processed poverty as an unfortunate circumstance that helped explain the offense and should mitigate the sentence. Arguably, a casual laborer in construction who dropped out of high school is no less blameworthy than a degree-holding computer technician, when he decides to commit a drug offense (see 24). Nevertheless, environmental deprivation occasionally has been discussed as a mitigating factor at sentencing (see 8; 35; 39). Perhaps in this context, judges processed the relatively poor defendants’ crime as the partial product of disadvantages in life, which mitigated his culpability. In contrast, the relatively affluent defendant had little excuse.

When risk assessment information was added to these cases, judges were more likely to sentence the relatively poor defendant to incarceration than his more affluent counterpart (61.2% vs. 44.4%). Adding formal risk assessment information may have cued judges to process poverty as a factor that increased the likelihood that the defendant would continue committing offenses in the future. This context may have activated stereotypes of poverty (see 7; 28) that led judges to interpret identical risk scores as signaling a much higher risk of re-arrest for the relatively poor defendant than his more affluent counterpart.

Because our study involved real judges with sentencing experience and case vignettes tailored to their jurisdictions, the results are difficult to contextualize. Still, our results are grossly consistent with past findings based on samples of students or laypeople. First, based on four different case vignettes that she assigned to 83 of her law students, Starr (35, p. 51) informally observed that students gave a poor defendant shorter sentences than a more affluent defendant in the absence of risk assessment information, but “this pattern reversed when the risk score was provided.” Second, Green and Chen (14) wrote vignettes that manipulated defendants’ race and other demographic characteristics and asked 554 online workers on Amazon’s Mechanical Turk to read the vignettes and assess the defendant’s likelihood of pretrial failure, i.e., the probability they would be rearrested or fail to appear in court. Workers were assigned to either an “algorithm” condition (where vignettes included algorithmic estimates of the defendant’s likelihood of pretrial failure) or a “control” condition (where vignettes did not include such estimates). The authors found a complex interaction: For cases in which the algorithmic estimate of re-offending was greater than the control group workers’ estimates, risk assessment had a 26% stronger average influence on increasing workers’ predictions about black defendants than white defendants. (There were no such differences by race for cases in which the algorithmic estimate was less than the control group workers’ estimate.) These findings are consistent with the notion that adding risk assessments can activate racial stereotypes (see 29; 37) that lead workers to interpret similar
risk scores as signaling greater risk for black than white defendants—but perhaps only for high
risk defendants.

None of these studies are dispositive—but taken together, sustain Holder’s concern that providing
judges with risk assessment information could increase sentencing disparities under some
conditions. The present study was designed to permit valid inferences about the cause-effect
relationship between risk assessment and socioeconomic status on judges’ sentencing decisions.
This experiment may overestimate the causal effect and findings may not generalize beyond the
specific conditions tested. First, we deliberately created cases that fell in a grey sentencing zone
(eligible for probation or incarceration) where inappropriate considerations like socioeconomic
status or race may be most likely to influence judges’ decisions (2). Results may not generalize to
cases that involve less judicial discretion. Second, it is unclear whether the present results would
generalize from a drug case to other types of offenses that may be less associated with stereotypes
of poverty (7); and from a “moderate-to high” risk cases to those at lower risk of recidivism (see
14). Finally, although we developed relatively detailed presentence vignettes tailored to local
jurisdictions to maximize ecological validity, the independent variables could have a weaker
effect in real courtroom settings where judges are exposed to a richer set of case materials and
interact with the parties involved. In future research, it will be important to test the extent to
which the present results generalize to contexts where judges have more limited discretion in
sentencing, defendants vary in their offenses and estimated risk levels, and sentencing materials
are more complete. Whether providing judges with risk assessment information increases,
reduces, or has no effect on sentencing disparities probably depends on several conditions that
are just beginning to be understood.

Fundamentally, this study demonstrates that biases can shift, as a risk assessment algorithm filters
through a judge into a sentencing decision (14). Given a medium risk defendant convicted of a
drug offense who falls in “grey” sentencing territory, providing judges with risk assessment
information transformed poverty from a mitigating circumstance that reduced the likelihood of
incarceration to a risk factor that increased the likelihood of incarceration.

In many jurisdictions, formal risk assessment information is routinely included in presentence
investigation reports. Even when judges explicitly discredit or reject risk assessment (22),
exposure to risk scores could influence how they intuitively process information about the
defendant to reach a sentence. We believe that risk assessment has an important role to play in
reducing mass incarceration in the United States, as the Model Penal Code (1) has recently
affirmed. Providing guidelines (32) or training to raise judges’ awareness about their own
intuitive biases and how they can interact with algorithms may help. Determining how to present
risk algorithms so judges can most effectively and fairly incorporate them into their decision-
making about defendants is essential.

**Materials and Methods**

**Experimental Design.** The study design is a 2 x 2 factorial experiment, as shown in Figure 1.
The goal was to achieve a high degree of control over extraneous variables to permit causal
inference about how the effect of providing formal risk assessment information on judges’
sentencing severity may depend on a defendant’s socioeconomic status. To rule out socially
desirable responding we implemented both risk assessment and socioeconomic status as between
subjects factors—making it impossible for judges to guess the goal of the study and adjust their
answers accordingly.

Electronic copy available at: https://ssrn.com/abstract=3316266
**Vignettes.** The base vignette and sentencing options were adapted from a standard vignette used in a national survey of public opinions on sentences for federal crimes (30). To maximize ecological validity and to leverage judicial discretion by ensuring that cases fell in a gray sentencing zone, we worked with local experts to tailor vignettes to each site—considering and embedding local offense designations, sentencing provisions, and any risk assessment instruments. An example base vignette appears below, with potentially identifying information redacted and the socioeconomic status manipulation shown in bold.

A 24-year old man was involved with several others in taking part over a four month period in the selling of $1,500 worth of heroin, about 10 grams. The defendant allowed his apartment to be used for drug sales. He did not carry or use any weapons or engage in violence. He thinks the charges are unfair—he says he was just hanging out with the “wrong friends.” He pled guilty to one count of [redacted felony drug offense].

The defendant works [as a casual laborer in construction and did not graduate from high school OR at the local Apple Store and has a BA in computer science]. He admits to being a heroin user himself and to feeling depressed, but seems uninterested in treatment. His relationship with his parents is strained—and he has no stable romantic relationship.

The defendant has never been imprisoned before but has prior convictions for [redacted felony burglary and misdemeanor marijuana offenses], both committed on the same occasion three years ago. He was sentenced to probation and three months in county jail—and successfully completed probation. He was adjudicated twice as a juvenile—once at age 15 for underage drinking and again at 17 for assault after a fight broke out in a bar. He was once suspended from high school.

After the base vignette, a set of risk assessment information was either omitted or added. This information consists of three parts: (1) an orientation paragraph, (2) the overall risk text and table, and (3) the risk factor text and table. An example set is shown below, with potentially identifying details modified and redacted.

As part of the presentence investigation process, a risk assessment instrument called the [redacted] was used to assess the defendant’s likelihood of re-arrest. This instrument consists of factors that research has found predict re-arrest. The court may use this instrument to help determine the appropriate sanction within the limits established by law.

**OVERALL RISK LEVEL**

The risk assessment instrument yields risk scores that range from a low of 0 to a high of 31. As shown below, the defendant obtained a score of 11—so he belongs to a group with a “moderate-to-high” risk of re-arrest.

<table>
<thead>
<tr>
<th>Offender Score</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0-4)</td>
<td>Low-Moderate (5-9)</td>
</tr>
<tr>
<td>Moderate-High (10-16)</td>
<td>High (17-31)</td>
</tr>
</tbody>
</table>

**FACTORS THAT CONTRIBUTE TO OVERALL RISK LEVEL**
Below is the calculation of the defendant’s score based on the identified risk factors. The number of possible points for each risk factor and the number of actual points received by the defendant are displayed. The defendant’s total risk score is the sum of points received across all risk factors.

<table>
<thead>
<tr>
<th>FACTOR</th>
<th>POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Possible points</td>
</tr>
<tr>
<td>Criminal History</td>
<td>8</td>
</tr>
<tr>
<td>Family Problems &amp; Antisocial Associates</td>
<td>7</td>
</tr>
<tr>
<td>Attitudes Supportive of Crime</td>
<td>7</td>
</tr>
<tr>
<td>Substance Abuse</td>
<td>5</td>
</tr>
<tr>
<td>Education &amp; Employment Problems</td>
<td>3</td>
</tr>
<tr>
<td>Young Age</td>
<td>2</td>
</tr>
<tr>
<td>Mental Health Problems</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total Risk Score</strong></td>
<td><strong>31</strong></td>
</tr>
</tbody>
</table>

Three points about this manipulation are key. First, the set of risk assessment information was identical for the relatively poor and affluent defendant—including risk scores. Even when an instrument included employment or education as risk factors, application of item definitions and scoring criteria (e.g., unemployed, less than 9th grade education, suspended) yielded equal scores across the socioeconomic manipulation. Second, to ensure that the set of risk assessment information did not introduce potential confounds, base vignettes were written to include narrative information relevant to each risk factor (e.g., all four vignettes indicated the defendant believed the charges were unfair but only the two vignettes that added the set of risk assessment information scored this as ‘attitudes supportive of crime’).

Third, because of an administrative error in the Southwestern site, partial risk assessment information appeared in the “Relatively poor, no risk assessment information” condition. Specifically, the orientation paragraph appeared, without overall risk level tables or risk factor tables. To ensure this error did not unduly affect results, we repeated Model 1 analyses with Southwestern data only—and found no material change from the results reported above: the interaction between providing formal risk assessment information and the defendants’ socioeconomic status was statistically significant $B=2.280$ (95% C.I. = 0.849, 4.102), $p=.002$.

At the end of each vignette, judges were told, “The defendant has been convicted through a plea agreement that gives the court full discretion to sentence the defendant to probation, short term jail, or prison.” Following Rossi & Berk (30), judges were asked “What sentence should be given in a case like this?” and asked to circle one of three options shown in a table: probation, jail (less than 1 year), or prison (1 year or more).

**Judges and Procedure.** We agreed not to identify jurisdictions or judges, but their essential characteristics are described in the text and Table 2. Judges from the Eastern district are appointed to the bench for renewable terms, whereas those from the Southwestern and Midwestern districts are appointed to the bench but then required to stand for election.

Judges in the Eastern and Midwestern jurisdictions completed the study at an annual educational judicial conference, just prior to a plenary session. Once judges were fully assembled, we placed in front of each judge one of four randomly selected case vignettes in the form of a questionnaire. The front page introduced the study and asked judges to await instruction before beginning. As the questionnaires were being distributed, a respected local judicial leader introduced the study.
and explained that we wanted to actively engage them in the learning process by getting their reaction to a brief description about a defendant. They were told we would aggregate their results and share them with the group at the end of the session. Judges were assured that their responses were anonymous, that their participation was voluntary, and that any judge who wanted to exclude their results from the study could do so (only two judges elected to do so). Judges were told “this is about individual decision-making” and instructed judges not to discuss the defendant with their colleagues. Above all else, they were asked to “treat this decision as if the sentence is real and applies to an actual individual.” With this backdrop, we asked judges to begin the study. The study included providing a sentence for the case vignette and then providing some basic demographic information.

Judges in the Southwestern jurisdiction completed the study online. We worked with local judiciary leaders and experts to invite judges to participate in a jurisdiction-wide “study that explores factors related to sentencing decisions.” Judges were assured that their responses would be anonymous and were informed that group results would be provided at the end of the study. Invitations and reminders were sent via email. Of eligible judges invited to participate, 85% did so (33 refused or did not respond). Qualtrics was used to randomly assign judges to one of the four case vignettes and to administer the survey. When judges clicked a link to complete the study, they were shown an introduction page that provided instructions that parallel those outlined above for judges at the other sites. Then, judges were shown the case vignette, asked to provide a “real” sentence for the individual depicted, and then provide basic demographic information.

References


Acknowledgments

General. We thank local and national judges, lawyers, and administrators for shaping the study materials and supporting data collection. We also thank [specify] for their thoughtful comments. This paper synthesizes and expands on material first presented at an annual conference of the American Psychology-Law Society, a symposium hosted by New York University’s Public Safety Lab, and a colloquium held at the University of Southern California. We thank audience members for their helpful comments.

Author contributions: JS conceptualized the study, designed and refined vignettes with stakeholders and coauthors, and collected data. JM and NS co-designed vignettes for two of the jurisdictions. NS analyzed results. JS and NS wrote the manuscript and JS edited it.

Competing interests: None.

Data and materials availability: Redacted vignettes and recruitment materials are available from the authors upon request.

Figures and Tables

Fig. 1. Experimental design. Defendants’ socioeconomic status and risk assessment were manipulated to produce four case vignettes (1, 2, 3, and 4) in the form of presentence investigation reports. Risk assessment information and scores were the same for relatively poor and affluent defendants. Judges were randomly assigned to review one report and sentence the defendant depicted.

<table>
<thead>
<tr>
<th>Risk assessment scores provided or omitted</th>
<th>Defendant’s socioeconomic status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Relatively poor</td>
<td>2. Relatively affluent</td>
</tr>
<tr>
<td>3. Relatively poor + risk assessment scores</td>
<td>4. Relatively affluent + risk assessment scores</td>
</tr>
</tbody>
</table>

Electronic copy available at: https://ssrn.com/abstract=3316266
Fig. 2. Providing judges with risk assessment scores reverses the relationship between a defendants’ socioeconomic status and probability of incarceration. Displays the predicted probability of incarceration as a function of the manipulated variables, after controlling for jurisdiction and judges’ characteristics (from logistic regression Model 3). As shown on the left side of the figure, relatively poor defendants (dark bar) are less likely to be sentenced to incarceration than those who are more affluent (light bar). As shown on the right side of the figure, this pattern reverses when formal risk assessment scores are provided for the defendants (holding those scores constant).
Table 1. Logistic regression results: Socioeconomic status interacts with risk assessment to influence judges’ sentences. Three binary logistic models were calculated, predicting sentences of incarceration (vs. probation). Model 1 includes only the manipulated variables (socioeconomic status and risk assessment); model 2 adds the jurisdiction as a control variable; and model 3 adds judges’ characteristics as control variables. In the final model, the interaction term of interest and jurisdiction predict sentences of incarceration. Notes: Bootstrapped (n=1,000 samples) raw Maximum Likelihood (ML) weights (in log odds) with standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (manipulated variables)</th>
<th>Model 2 (adds jurisdiction)</th>
<th>Model 3 (adds judges’ characteristics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.384 *</td>
<td>-1.152 ***</td>
<td>-1.139 ***</td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.410)</td>
<td>(0.501)</td>
</tr>
<tr>
<td>Relatively poor (vs. affluent)</td>
<td>-0.553 *</td>
<td>-0.577 *</td>
<td>-0.557</td>
</tr>
<tr>
<td></td>
<td>(0.325)</td>
<td>(0.352)</td>
<td>(0.394)</td>
</tr>
<tr>
<td>Risk scores provided (vs. omitted)</td>
<td>-0.608 **</td>
<td>-0.654 *</td>
<td>-0.604</td>
</tr>
<tr>
<td></td>
<td>(0.312)</td>
<td>(0.390)</td>
<td>(0.435)</td>
</tr>
<tr>
<td>Relatively poor x Risk scores provided</td>
<td>1.231 ***</td>
<td>1.442 ***</td>
<td>1.481 ***</td>
</tr>
<tr>
<td></td>
<td>(0.443)</td>
<td>(0.528)</td>
<td>(0.577)</td>
</tr>
<tr>
<td>Eastern (vs. Midwest.) jurisdiction</td>
<td></td>
<td>0.594</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.442)</td>
<td>(0.609)</td>
</tr>
<tr>
<td>Southwestern (vs. Midwest.) jurisdiction</td>
<td>2.455 ***</td>
<td>2.263 ***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.404)</td>
<td>(0.465)</td>
<td></td>
</tr>
<tr>
<td>Female (vs. male) judge</td>
<td></td>
<td></td>
<td>-0.235</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.314)</td>
</tr>
<tr>
<td>“Other” (vs. White) ethnicity judge</td>
<td></td>
<td></td>
<td>0.224</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.441)</td>
</tr>
<tr>
<td>Republican (vs. Democrat) judge</td>
<td></td>
<td></td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.399)</td>
</tr>
<tr>
<td>Other (vs. Democrat) judge</td>
<td></td>
<td></td>
<td>-0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.387)</td>
</tr>
</tbody>
</table>
Table 2. Judges’ demographic characteristics and political affiliation, by jurisdiction

Notes: Frequencies that sum to less than the relevant sample size reflect missing data.

<table>
<thead>
<tr>
<th></th>
<th>Overall (N=340)</th>
<th>Eastern Judges (n=74)</th>
<th>Southwestern Judges (n=186)</th>
<th>Midwestern Judges (n=63)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>64%</td>
<td>60%</td>
<td>64%</td>
<td>72%</td>
</tr>
<tr>
<td>Female</td>
<td>36%</td>
<td>40%</td>
<td>36%</td>
<td>28%</td>
</tr>
<tr>
<td>African- American</td>
<td>12%</td>
<td>41%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4%</td>
<td>5%</td>
<td>4%</td>
<td>0%</td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>81%</td>
<td>51%</td>
<td>88%</td>
<td>97%</td>
</tr>
<tr>
<td>Other</td>
<td>3%</td>
<td>3%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>Democrat</td>
<td>48%</td>
<td>88%</td>
<td>37%</td>
<td>39%</td>
</tr>
<tr>
<td>Republican</td>
<td>30%</td>
<td>3%</td>
<td>43%</td>
<td>20%</td>
</tr>
<tr>
<td>Other</td>
<td>22%</td>
<td>9%</td>
<td>20%</td>
<td>41%</td>
</tr>
</tbody>
</table>