Law and Human Behavior

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

Jennifer Skeem, Nicholas Scurich, and John Monahan

Online First Publication, January 13, 2020. http://dx.doi.org/10.1037/lhb0000360

American Psychological Association
American Psychology-Law Society
Law and Human Behavior

ISSN: 0147-7307
http://dx.doi.org/10.1037/lhb0000360

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

Jennifer Skeem
University of California, Berkeley

Nicholas Scurich
University of California, Irvine

John Monahan
University of Virginia

Objective: Use of risk assessment instruments in the criminal justice system is controversial. Advocates emphasize that risk assessments are more transparent, consistent, and accurate in predicting re-offending than judicial intuition. Skeptics worry that risk assessments will increase socioeconomic disparities in incarceration. Ultimately, judges make decisions—not risk assessments. This study tests whether providing risk assessment information interacts with a defendant's socioeconomic class to influence judges' sentencing decisions. ***Hypotheses:*** Tentatively, socioeconomic status was expected to have a main effect; without an interaction with risk assessment information. ***Method:*** Judges (N = 340) with sentencing experience were randomly assigned to review 1 of 4 case vignettes and sentence the defendant to probation, jail, or prison. Information in the vignettes was held constant, except the defendant's socioeconomic status and whether risk assessment information was provided. ***Results:*** Risk assessment information reduced the likelihood of incarceration for relatively affluent defendants, but the same 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. ***Conclusions:*** Cuing judges to focus on risk may re-frame how they process socioeconomic status—a variable with opposite effects on perceptions of blameworthiness for past crime versus perceptions of risk for future crime. Providing risk assessment information may have transformed low socioeconomic status from a circumstance that reduced the likelihood of incarceration (by mitigating perceived blameworthiness) to a factor that increased the likelihood of incarceration (by increasing perceived risk). Under some circumstances, risk assessment information may increase sentencing disparities.

Public Significance Statement
Risk assessment instruments and algorithms are playing an increasing role in decision making about people involved in the justice system. In this experiment, providing judges with risk assessment information about a defendant increased the severity of their sentences for relatively poor—but not affluent—defendants. It may be necessary to provide guidelines and training to help judges understand their intuitive biases and more effectively and fairly incorporate risk assessment into decision making about defendants.

Keywords: risk assessment, algorithms, decision making, judges, bias

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 are 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

Jennifer Skeem, School of Social Welfare and School of Public Policy, University of California, Berkeley; Nicholas Scurich, School of Social Ecology, University of California, Irvine; John Monahan, School of Law, University of Virginia.

This study was funded by the Mack Center on Mental Health and Social Conflict at the University of California, Berkeley. A portion of these data were previously described at annual conference of the American Psychology-Law Society (2019, Portland, Oregon) and the Conference on Empirical Legal Studies (2019, Claremont, California). The authors thank the administrators and judges in three jurisdictions who participated in this study.

Correspondence concerning this article should be addressed to Jennifer Skeem, School of Social Welfare and School of Public Policy, University of California, Berkeley, 120 Haviland Hall, Berkeley, CA 94720-7400. E-mail: jenskeem@berkeley.edu

summarize risk factors to estimate a person's likelihood of future reoffending (Scurich, 2016a). 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 (Monahan & Skeem, 2016). In fact, risk assessment is "the engine that drives" a federal prison reform bill that was just signed into law (Garrett, 2018, para. 1). Across the United States, jurisdictions have been undertaking a variety of efforts to reduce unprecedented rates of incarceration (National Conference of State Legislatures, 2017) 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 (Monahan, 2017; Monahan & Skeem, 2016). Advocates argue that—when risk is a legally relevant consideration—judges should consider risk assessment instruments (RAIs) to improve the consistency, transparency, and accuracy of their decisions (e.g., Monahan, 2017; Neufeld, 2018).

Judges routinely make momentous decisions in a person's life that include consideration of the likelihood that the person will reoffend—and must make their own intuitive judgments, without RAIs (D. M. Gottfredson, 1999). At the pretrial stage, each of the 30,000 daily arrests in the United States (Department of Justice, 2018) 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. Although sentencing traditionally focuses more on backward-looking concerns about the defendant's blameworthiness for a past crime, the Model Penal Code (American Law Institute, 2017) also provides a limited role for forward-looking concerns about preventing future crimes. In a recent survey, eight of 10 judges believed that *both* blameworthiness and risk of reoffending should be considered at sentencing (Monahan, Metz, & Garrett, 2018; see also Chanenson & Hyatt, 2016).

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, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018) 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 Jung, Concannon, Shroff, Goel, & Goldstein, 2017). In a study that assessed judicial intuition about 962 felony offenders at sentencing, D. M. Gottfredson (1999) 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 suggested that "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 (Courtland, 2018; O'Neill, 2016; Scurich & Krauss, in press), critics assert that risk factors included in some RAIs (e.g., education level, marital status, neighborhood disadvantage) are "proxies" for minority race and poverty (Starr, 2014). In the view of former Attorney General Eric Holder (2014, para. 23), 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 with what existing practices* (Skeem & Lowenkamp, 2016)? Existing practices include sentencing guidelines that heavily weight criminal history and have been shown to contribute to racial disparities (Frase, 2009); and judges' intuitive appraisals of risk, which are less transparent, consistent, and accurate than risk assessment are much like those of other people—largely intuitive, heuristic based, and subject to bias (Guthrie, Rachlinski, & Wistrich, 2007; Rachlinski, Johnson, Wistrich, & Guthrie, 2009). Like other people, judges may stereotype Black men as threatening (e.g., Trawalter, Todd, Baird, & Richeson, 2008) and poor people as incapable, untrustworthy, and antisocial (Cozzarelli, Wilkinson, & Tagler, 2001; Piff, Kraus, & Keltner, 2018). Currently, Black people are 5 times more likely to be imprisoned than White people (Carson, 2018), 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 (Looney & Turner, 2018; see also Western & Pettit, 2010). These disparities appear particularly pronounced for drug crimes (e.g., Mitchell, 2005). Risk assessment could exacerbate these existing disparities, as Holder (2014) speculated. But risk assessment could instead have no effect on—or even reduce—disparities, as others have predicted (Hoge, 2002; M. R. Gottfredson & Gottfredson, 1988).

To our knowledge, no studies have directly tested the effect of risk assessment on judges' sentencing decisions. Van Wingerden, van Wilsem, and Moerings (2014) 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 defendants' protected characteristics (e.g., race, socioeconomic status) were not examined in this study. Several studies have examined the racial predictive fairness of risk algorithms alone (e.g., Corbett-Davies & Goel, 2018; Skeem & Lowenkamp, 2016), but judges (not algorithms) determine sentences; and risk is only one consideration. Because formal risk assessments rarely provide dispositive answers to legal questions, it is necessary to examine how they affect human judgment. From a "*compared with 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—a type of offense that is common and associated with both sentencing discretion (see Rossi & Berk, 1997) and sentencing disparities (e.g., Mitchell, 2005). 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 (2014) predicted, then providing judges with risk assessment information will increase sentencing severity significantly more for relatively poor defendants than their more affluent counterparts.

Method

Design Overview

The study design is a 2 × 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 whether the effect of formal risk assessment information on judges' sentencing severity depends on a defendant's socioeconomic status. To rule out socially desirable responding, we implemented both risk assessment and socioeconomic status as between-subjects factors. Because each judge saw only one vignette for analyses described here, it was difficult or impossible for them to know what factors were manipulated across vignettes. The goal was to prevent judges from guessing the goal of the study and adjusting their answers accordingly.

We randomly assigned each judge to review one of four case vignettes that 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. 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 RAI. 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.

	Defendant's socioeconomic status	
Risk assessment scores provided or omitted	1. Relatively poor	2. Relatively affluent
	3. Relatively poor + risk assessment scores	4. Relatively affluent + risk assessment scores

Figure 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.

Participants

To recruit judges for this study, we deliberately partnered with judiciary leaders in jurisdictions located in the Eastern, Midwestern, and Southwestern United States. 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). Although we agreed not to identify jurisdictions or judges, judges' essential characteristics are described in Table 1. Judges from the Eastern jurisdiction are appointed to the bench for renewable terms, whereas those from the Southwestern and Midwestern jurisdictions are appointed to the bench but then are required to stand for election.

The sample size was sufficient to address the study aim. Specifically, we conducted an a priori power analysis using G*Power to estimate the requisite number of participants to detect an effect assuming an effect does exist. Assuming a Type I error rate of .05, power more than .80, and a small effect size (δ = .25), one would need 80 participants per cell, or 320 total participants total, to detect the interaction effect of interest. Our sample includes 340 valid responses.

Statistical comparisons of judges randomized to each of the four conditions indicated that random assignment worked. There were no significant differences among the four groups of judges in gender (p = .66), race (p = .82), or political orientation (p = .85).

Procedure

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 that we would aggregate their results and share them with the group at the end of the session (which we did). 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 that "this is about individual decision making" and were instructed 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

Table 1
Judges' Demographic Characteristics and Political Affiliation by Jurisdiction

Characteristic	Overall (N = 340)		Eastern judges (n = 87)		Southwestern judges (n = 188)		Midwestern judges (n = 65)	
	%	*n*	%	*n*	%	*n*	%	*n*
Male	64	211	60	44	64	120	72	47
African American	12	37	41	30	3	5	3	2
Hispanic	4	12	5	4	4	8	0	0
White, non-Hispanic	81	262	51	37	88	164	97	61
Other	3	11	3	2	5	9	0	0
Democrat	48	147	88	58	37	67	39	22
Republican	30	91	3	2	43	78	20	11
Other	22	66	9	6	20	37	41	23

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

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 e-mail. 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.

Vignettes

In all three 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 an RAI. To make the case vignettes as realistic and familiar as possible, we formatted them as local presentence investigation reports.

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 (Rossi & Berk, 1997). 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 RAIs. Prototype vignettes are available from the authors in a form that does not violate site confidentiality. 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 g. 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: (a) an orientation paragraph, (b) the overall risk text and table, and (c) 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, an RAI called the [*redacted*] was used to assess the defendant's likelihood of rearrest. This instrument consists of factors that research has found predict rearrest. The court may use this instrument to help determine the appropriate sanction within the limits established by law.
>
> **OVERALL RISK LEVEL**
>
> The RAI yields risk scores that range from a low of 0 to a high of 31. As shown below, the defendant obtained a score of 15—so he belongs to a group with a "moderate-to-high" risk of rearrest.

		OFFENDER SCORE = 15	
Low (0–4)	Low-Moderate (5–9)	**Moderate-High (10–16)**	High (17–31)

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.

Factor	Points	
	Possible points	Actual points
Criminal history	8	3
Family problems & antisocial associates	7	4
Attitudes supportive of crime	6	2
Substance abuse	4	3
Education & employment problems	3	1
Young age	2	1
Mental health problems	1	1
Total risk score	=31	=15

Three points about this manipulation are key. First, the set of risk assessment information was identical for the relatively poor and affluent defendants—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 ninth-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% CI [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 and Berk (1997), 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)."

Results

To statistically address the study aims, we conducted three binary logistic regression models using fixed effects variables (orthogonal contrasts; see Wendorf, 2004) for the 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 2 and reported below. Results are organized into four key findings.

Finding 1: The Impact of Risk Assessment on Judges' Sentence 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 2 (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

Table 2
Logistic Regression Results: Socioeconomic Status Interacts With Risk Assessment to Influence Judges' Sentences

Predictor	Model 1 (manipulated variables)	*p* value	Model 2 (adds jurisdiction)	*p* value	Model 3 (adds judges' characteristics)	*p* value
Constant	.384 (.232)	.079	−1.152** (.410)	.003	−1.139* (.501)	.013
Relatively poor (vs. affluent)	−.553 (.325)	.091	−.577 (.352)	.098	−.557 (.394)	.141
Risk scores provided (vs. omitted)	−.608* (.312)	.049	−.654 (.390)	.079	−.604 (.435)	.145
Relatively Poor × Risk scores provided	1.231 (.443)	.008	1.442** (.528)	.008	1.481** (.577)	.004
Eastern (vs. Midwest) jurisdiction			.594 (.442)	.158	.291 (.609)	.599
Southwestern (vs. Midwest) jurisdiction			2.455** (.404)	.003	2.263** (.465)	.001
Female (vs. male) judge					−.235 (.314)	.446
"Other" (vs. White) ethnicity judge					.224 (.441)	.584
Republican (vs. Democrat) judge					.407 (.399)	.296
Other (vs. Democrat) judge					−.057 (.387)	.884

Note. Bootstrapped ($n = 1{,}000$ samples) raw maximum likelihood weights (in log odds) with standard errors in parentheses. 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 predicts sentences of incarceration.
* $p < .05$. ** $p < .01$.

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.

Finding 2: Sentencing Severity Varies as a Function of Jurisdiction

Sentences varied by jurisdiction. Specifically, a sizable majority (73%) of the Southwestern judges sentenced the defendants to incarceration compared with 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.)

Finding 3: Sentencing Severity Does Not Vary as a Function of Judges' 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 the 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 with just 20% of Midwestern judges (see Table 1).

To determine whether the previously detected interaction holds after accounting for both jurisdiction and judges' personal characteristics, we conducted Model 3, which added judges' sex, ethnicity, and political orientation to the variables in the previous model. Because 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.

Finding 4: Even After Controlling for Jurisdiction and Judges' Characteristics, the Impact of Risk Assessment Depends on Defendants' 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.)

Figure 2 plots the predicted probability of incarceration as a function of the manipulated variables. For relatively affluent defendants (light gray bars), the probability of receiving incarceration decreased from 59.5% to 44.4% when risk assessment information was added. However, for relatively poor defendants (dark gray bars), the addition of risk assessment information increased the probability of receiving incarceration from 45.8% to 61.2%. These results suggest that providing formal risk assessment information to judges reverses the relationship between a defendant's socioeconomic status and sentencing severity. When formal risk assessment information is omitted, relatively poor defendants are less likely to be incarcerated than their more affluent counterparts. When that information is provided—holding risk scores and classifications constant—relatively poor defendants are *more* likely to be incarcerated than their richer counterparts.

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 to underappreciate the fact that, ultimately, judges make decisions—not algorithms or RAIs. 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 versus 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 (American Law Institute, 2017)—attempt to balance competing sentencing considerations that include the defendant's blameworthiness and risk (Starr, 2016), (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 Monahan & Skeem, 2016), 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

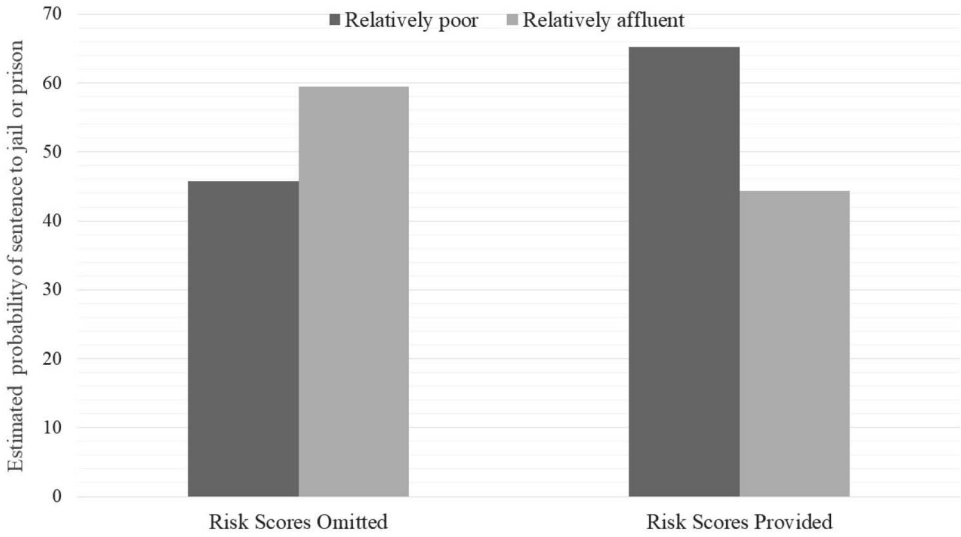

Figure 2. Providing judges with risk assessment scores reverses the relationship between a defendants' socioeconomic status and probability of incarceration. Displays the predicted probability of incarceration as a function of the manipulated variables, after controlling for jurisdiction and judges' 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).

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 Monahan & Skeem, 2016). Nevertheless, environmental deprivation has occasionally been discussed as a mitigating factor at sentencing (see Delgado, 1985; Starr, 2016; Vuoso, 1987). 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 and to process relative affluence as a factor that reduced the likelihood that the defendant would continue committing offenses. This context may have activated stereotypes of poverty and affluence (see Cozzarelli et al., 2001; Piff et al., 2018) that led judges to interpret identical risk scores as signaling a much higher risk of rearrest 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 assigned to 83 of her law students, Starr (2016, 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 (2019) 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, that is, the probability they would be rearrested or fail to appear in court. Workers were assigned to either an "algorithm" condition (in which vignettes included algorithmic estimates of the defendant's likelihood of pretrial failure) or a "control" condition (in which vignettes did not include such estimates). The authors found a complex interaction: For cases in which the algorithmic estimate of reoffending 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 Rachlinski et al., 2009; Trawalter et al., 2008) 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 are consistent with Holder's (2014) concern that providing judges with risk assessment information could exacerbate disparities in incarceration for disadvantaged defendants *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 gray sentencing zone (eligible for probation or incarceration), in which inappropriate considerations

like socioeconomic status or race may be most likely to influence judges' decisions (Baldus, Woodworth, Zuckerman, & Weiner, 1997). 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 (Courtland, 2018); and from "moderate to high" risk cases to those at lower risk of recidivism (see Green & Chen, 2019). Finally, although we developed relatively detailed presentence vignettes tailored to local jurisdictions to maximize ecological validity, the independent variables probably have a weaker effect in real courtroom settings in which 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 in which 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, decreases, 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 (Green & Chen, 2019). It is worth reiterating that there were sentencing disparities in this study, even in the absence of risk assessment information. Given a medium-risk defendant convicted of a drug offense who falls in "gray" 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 (Monahan et al., 2018), 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 (American Law Institute, 2017) has recently affirmed. Providing guidelines (Scurich, 2016b) 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 that judges can most effectively and fairly incorporate them into their decision making about defendants is essential.

References

American Law Institute. (2017). *Model Penal Code: Sentencing*. Philadelphia, PA: Author. Retrieved from https://www.ali.org/publications/show/sentencing/

Baldus, D. C., Woodworth, G., Zuckerman, D., & Weiner, N. A. (1997). Racial discrimination and the death penalty in the post-Furman era: An empirical and legal overview with recent findings from Philadelphia. *Cornell Law Review, 83,* 1638–1770.

Carson, E. A. (2018, January). *Prisoners in 2016. U.S. Department of Justice, Bureau of Justice Statistics*. Retrieved from https://www.bjs.gov/index.cfm?ty=pbdetail&iid=6187

Chanenson, E., & Hyatt, J. (2016). *The use of risk assessment at sentencing: Implications for research and policy*. Villanova University School of Law Working Paper Series. Retrieved from http://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps

Corbett-Davies, S., & Goel, S. (2018). *The measure and mismeasure of fairness: A critical review of fair machine learning*. Retrieved from https://arxiv.org/abs/1808.00023

Courtland, R. (2018). Bias detectives: The researchers striving to make algorithms fair. *Nature, 558,* 357–360. http://dx.doi.org/10.1038/d41586-018-05469-3

Cozzarelli, C., Wilkinson, A. V., & Tagler, M. J. (2001). Attitudes toward the poor and attributions for poverty. *Journal of Social Issues, 57,* 207–222. http://dx.doi.org/10.1111/0022-4537.00209

Delgado, R. (1985). Rotten social background: Should the criminal law recognize a defense of severe environmental deprivation. *Law & Social Inquiry, 3,* 9–90.

Department of Justice. (2018, Fall). 2017 Crime in the U.S: Persons arrested. *U.S. Department of Justice, Federal Bureau of Investigation*. Retrieved from https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topic-pages/persons-arrested

Frase, R. S. (2009). What explains persistent racial disproportionality in Minnesota's prison and jail populations? *Crime and Justice, 38,* 201–280. http://dx.doi.org/10.1086/599199

Garrett, B. (2018, December 27). The prison reform bill's implementation will be tricky: Here's how to ensure it's a success. *Slate*. Retrieved from https://slate.com/news-and-politics/2018/12/prison-reform-bill-success.html

Gottfredson, D. M. (1999). *Effects of judges' sentencing decisions on criminal careers*. U. S. Department of Justice, Office of Justice Programs, National Institute of Justice. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/178889.pdf

Gottfredson, M. R., & Gottfredson, D. M. (Eds.). (1988). *Decision making in criminal justice: Toward the rational exercise of discretion* (2nd ed.). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4757-9954-5

Green, B., & Chen, Y. (2019). Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In Association for Computing Machinery (Ed.), *Proceedings of the Conference on Fairness, Accountability and Transparency* (pp. 90–99). New York, NY: Author. Retrieved from https://dl.acm.org/citation.cfm?id=3287563

Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review, 93,* 1–44.

Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29,* 380–396. http://dx.doi.org/10.1177/0093854802029004003

Holder, E. (2014). *Attorney General Eric Holder speaks at the National Association of Criminal Defense Lawyers 57th annual meeting*. Retrieved from http://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th

Jung, J., Concannon, C., Shroff, R., Goel, S., & Goldstein, D. G. (2017). *Simple rules for complex decisions*. Unpublished manuscript. Retrieved from https://arxiv.org/pdf/1702.04690

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics, 133,* 237–293.

Looney, A., & Turner, N. (2018, March 14). Work and opportunity before and after incarceration. *Brookings*. Retrieved from https://www.brookings.edu/research/work-and-opportunity-before-and-after-incarceration/

Mitchell, O. (2005). A meta-analysis of race and sentencing research: Explaining the inconsistencies. *Journal of Quantitative Criminology, 21,* 439–466. http://dx.doi.org/10.1007/s10940-005-7362-7

Monahan, J. (2017). Risk assessment in sentencing. In E. Luna (Ed.), *Reforming criminal justice*. Retrieved from http://academyforjustice.org/volume4/

Monahan, J., Metz, A., & Garrett, B. (2018). Judicial appraisals of risk assessment in sentencing. *Behavioral Sciences & the Law, 36,* 565–575. http://dx.doi.org/10.1002/bsl.2380

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, 12,* 489–513. http://dx.doi.org/10.1146/annurev-clinpsy-021815-092945

National Conference of State Legislatures. (2017). *Justice reinvestment: State resources*. Retrieved from http://www.ncsl.org/research/civil-and-criminal-justice/justicereinvestment.aspx

Neufeld, A. (2018). *In defense of risk assessment tools*. Retrieved from https://www.themarshallproject.org/2017/10/22/in-defense-of-risk-assessment-tools

O'Neill, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York, NY: Crown.

Piff, P. K., Kraus, M. W., & Keltner, D. (2018). Unpacking the inequality paradox: The psychological roots of inequality and social class. *Advances in Experimental Social Psychology, 57,* 53–124. http://dx.doi.org/10.1016/bs.aesp.2017.10.002

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *The Notre Dame Law Review, 84,* 1195–1246.

Rossi, P. A., & Berk, R. (1997). *Survey: Public perceptions and the Federal Sentencing Guidelines*. United States Sentencing Commission. Retrieved from https://www.ussc.gov/research/research-reports/survey-public-perceptions-and-federal-sentencing-guidelines

Scurich, N. (2016a). An introduction to the assessment of violence risk. In J. P. Singh, S. Bjorkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 3–15). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199386291.003.0001

Scurich, N. (2016b). Structured risk assessment and legal decision making. In M. Miller & B. H. Bornstein (Eds.), *Advances in psychology and law* (pp. 159–183). Washington, DC: American Psychological Association. http://dx.doi.org/10.1007/978-3-319-29406-3_5

Scurich, N., & Krauss, D. A. (in press). Public's views of risk assessment algorithms and pretrial decision making. *Psychology, Public Policy, and Law*.

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology: An Interdisciplinary Journal, 54,* 680–712. http://dx.doi.org/10.1111/1745-9125.12123

Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review, 66,* 842–873.

Starr, S. (2016). The odds of justice: Actuarial risk prediction and the criminal justice system. *Chance, 29,* 49–51. http://dx.doi.org/10.1080/09332480.2016.1156368

Tonry, M. (2004). *Thinking about crime: Sense and sensibility in American penal culture*. New York, NY: Oxford University Press.

Trawalter, S., Todd, A. R., Baird, A. A., & Richeson, J. A. (2008). Attending to threat: Race-based patterns of selective attention. *Journal of Experimental Social Psychology, 44,* 1322–1327. http://dx.doi.org/10.1016/j.jesp.2008.03.006

van Wingerden, S., van Wilsem, J., & Moerings, M. (2014). Pre-sentence reports and punishment: A quasi-experiment assessing the effects of risk-based pre-sentence reports on sentencing. *European Journal of Criminology, 11,* 723–744. http://dx.doi.org/10.1177/1477370814525937

Vuoso, G. (1987). Background, responsibility, and excuse. *The Yale Law Journal, 96,* 1661–1686. http://dx.doi.org/10.2307/796498

Wendorf, C. A. (2004). Primer on multiple regression coding: Common forms and the additional case of repeated contrasts. *Understanding Statistics, 3,* 47–57. http://dx.doi.org/10.1207/s15328031us0301_3

Western, B., & Pettit, B. (2010). Incarceration and social inequality. *Daedalus, 139,* 8–19. http://dx.doi.org/10.1162/DAED_a_00019

Received July 4, 2019
Revision received November 20, 2019
Accepted November 22, 2019 ■