Evidence-Based Sentencing and the Scientific Rationalization of Discrimination

Sonja B. Starr

Forthcoming in Stanford L. Rev. 66 (2014)

Evidence-Based Sentencing and the Scientific Rationalization of Discrimination

Sonja B. Starr*

Forthcoming, STANFORD LAW REVIEW, Vol. 66 (2014)

ABSTRACT

This paper critiques, on legal and empirical grounds, the growing trend of basing criminal sentences on actuarial recidivism risk prediction instruments that include demographic and socioeconomic variables. I argue that this practice violates the Equal Protection Clause and is bad policy: an explicit embrace of otherwise-condemned discrimination, sanitized by scientific language. To demonstrate that this practice should be subject to heightened constitutional scrutiny, I comprehensively review the relevant case law, much of which has been ignored by existing literature. To demonstrate that it cannot survive that scrutiny and is undesirable policy, I review the empirical evidence underlying the instruments. I show that they provide wildly imprecise individual risk predictions, that there is no compelling evidence that they outperform judges' informal predictions, that less discriminatory alternatives would likely perform as well, and that the instruments do not even address the right question: the effect of a given sentencing decision on recidivism risk. Finally, I also present new, suggestive empirical evidence, based on a randomized experiment using fictional cases, that these instruments should not be expected merely to substitute actuarial predictions for less scientific risk assessments, but instead to increase the weight given to recidivism risk versus other sentencing considerations.

* Professor of Law, University of Michigan. Thanks to Don Herzog, Ellen Katz, Richard Primus, and participants in Michigan Law's Faculty Scholarship Brownbag Lunch for their comments, and to Grady Bridges, Matthew Lanahan, and Jarred Klorfein for research assistance.

INTRODUCTION

> *"Providing equal justice for poor and rich, weak and powerful alike is an age-old problem. People have never ceased to hope and strive to move closer to that goal. ... In this tradition, our own constitutional guaranties of due process and equal protection both call for procedures in criminal trials which allow no invidious discriminations.…[T]he central aim of our entire judicial system [is that] all people charged with crime must, so far as the law is concerned, stand on an equality before the bar of justice in every American court."*
>
> --Griffin v. Illinois, *351 U.S. 12, 16 (1956)*

Criminal justice reformers have long worked toward a system in which defendants' treatment does not depend on their socioeconomic status or demographics, but on their criminal conduct. How to achieve that objective is a complicated and disputed question. Many readers might assume, however, that there is at least a general consensus on some key "don'ts." For example, judges should not systematically sentence defendants more harshly because they are poor or uneducated, or more lightly because they are wealthy and educated. They should not follow a policy of increasing the sentences of male defendants, or reducing those of females, on the explicit basis of gender. They likewise should not increase a defendant's sentence specifically because she grew up without a stable, intact family, or because she lives in a disadvantaged and crime-ridden community.

It might surprise many readers, then, to learn that a growing number of U.S. jurisdictions are adopting policies that deliberately encourage judges to do all of these "don'ts." These jurisdictions are directing sentencing judges to explicitly consider socioeconomic variables, gender, age, and sometimes family or neighborhood characteristics—not just in special contexts in which one of those variables might be particularly relevant (for instance, ability to pay in cases involving fines), but routinely, in all cases. This is not a fringe development. At least eight states are already implementing some form of it. One state supreme court has already enthusiastically endorsed it.[1] And it now has been embraced by the American Law Institute in the draft of the newly revised Model Penal Code—a development that reflects its mainstream acceptance and, given the Code's influence, may soon augur much more widespread adoption.[2] There is a similar trend in Canada, the United Kingdom, and other foreign jurisdictions.[3] Meanwhile, the majority of states now similarly direct parole boards to take demographic and socioeconomic factors into account.

The trend is called "evidence-based sentencing" (hereinafter EBS). "Evidence," in this formulation, refers not to the evidence in the particular case, but to empirical research on factors predicting criminal recidivism. EBS seeks to help judges advance the crime-prevention objectives of punishment by equipping them with the tools of criminologists—recidivism risk prediction instruments grounded in regression models of past offenders' outcomes. The instruments give considerable weight to criminal history, which is already central to modern sentencing schemes. However, they also add something new: explicit inclusion of gender, age,

[1] Malenchik v. State, 928 N.E.2d 564 (Ind. 2010).
[2] Model Penal Code: Sentencing § 6B.09 (Discussion Draft No. 4, 2012) (hereinafter "Draft MPC").
[3] *See* James Bonta, Offender Risk Assessment and Sentencing, 49 CAN. J. CRIMINOLOGY & CRIM. JUST. 519, 519-20 (2007).

and socioeconomic factors such as employment and education (with socioeconomically disadvantaged, male, and young defendants receiving higher risk scores). Some instruments also include family background, neighborhood of residence, and/or mental or emotional disorders.

EBS has been widely hailed by judges, advocates, and scholars as representing hope for a new age of scientifically guided sentencing. The idea is to replace judges' "clinical" evaluations of defendants (that is, reliance on their own expertise) with "actuarial" risk prediction, which is purportedly more accurate. Incongruously, this trend is being pushed by progressive reform advocates, who hope it will reduce incarceration rates by enabling courts to identify low-risk offenders. In this Article, I argue that they are making a mistake. As currently practiced, EBS should be seen neither as progressive nor as especially scientific—and it is almost surely unconstitutional.

This Article sets forth a constitutional and policy case against this approach, based on analysis of both the relevant doctrine and the empirical research supporting EBS. I show that the current prediction instruments should be subject to heightened equal protection scrutiny, and that the science falls short of allowing them to survive that scrutiny. The concept of "evidence-based practice" is broad, and I do not mean to issue a sweeping indictment of all its many criminal justice applications. Indeed, I strongly endorse the general objective of informing criminal justice policy with data. Nor do I argue that actuarial prediction of recidivism is always inappropriate. My objection is specifically to the use of demographic, socioeconomic, and family status variables to determine whether and how long a defendant is incarcerated. I am concerned that a well-intentioned desire for data-driven decision-making is causing discrimination to be rationalized based on rather weak empirical evidence. I focus principally on the instruments' use in sentencing, but virtually the same case can be made against their use in parole decisions, which is now established practice in thirty states.

The technocratic framing of EBS should not obscure an inescapable truth: sentencing based on such instruments amounts to overt discrimination based on demographics and socioeconomic status. The instruments typically do not include race as a variable (even their most enthusiastic defenders have limits to their comfort with group-based punishment), but sentencing based on socioeconomic predictors will have a racially disparate impact as well. Equal treatment of all persons is a central normative objective of the criminal justice system, and EBS may have serious social consequences, contributing to the concentration of the criminal justice system's punitive impact among those who already disproportionately bear its brunt. Moreover, the expressive message of EBS—the justification of disparate treatment based on statistical generalizations about crime risk—is, when stripped of the anodyne scientific language, toxic. Group-based generalizations about dangerousness are not innocuous; they have an insidious history in our culture. And the express embrace of additional punishment for the poor conveys the message that the system is rigged.

The instruments' use of gender and socioeconomic variables should be subject to heightened constitutional scrutiny. Gender is the only equal protection issue the existing literature pays any attention to, but I show that the socioeconomic variables trigger similar scrutiny under a line of Supreme Court doctrine concerning indigent criminal defendants—doctrine that the EBS literature completely ignores. In fact, the Court has specifically (and unanimously) condemned the notion of treating poverty as a predictor of recidivism risk in sentencing, even if there is statistical evidence supporting the correlation. Finally, while other variables in the instruments

(such as age and marital status) are subject only to rational basis review under current doctrine, I also argue that they raise substantial normative concerns.

Contrary to the other commentators that have considered the gender discrimination issue, I do not think the EBS instruments can survive heightened scrutiny, nor are they justified as a policy matter. There are doubtless important and even compelling state interests at stake. But heightened scrutiny requires the state to prove a strong relationship to those interests, and the case law on wealth classifications in criminal justice also requires analysis of alternatives, as does sensible policymaking. With these principles in mind, I turn to the strength of the empirical evidence supporting EBS. It falls short for three principal reasons.

First, the instruments provide nothing close to precise predictions of individual recidivism risk. The underlying regression models estimate average recidivism rates for offenders sharing the defendant's characteristics. While some models have reasonably narrow confidence intervals for this predicted average, the uncertainty about what an *individual* offender will do is *much* greater. Individual recidivism outcomes vary for many reasons that are not captured by the models. While some uncertainty is inherent in predicting probabilistic future events, the risk prediction models also leave out many measurable variables that one might expect to be important—for instance, there are typically no variables relating to the crime of conviction or the case's facts. The individual prediction problem is constitutionally important because the Supreme Court's cases on gender and indigent defendants have consistently held that disparate treatment cannot be justified based on statistical generalizations about group tendencies, even if they are empirically supported. Instead, individuals must be treated as individuals.

Second, it is not even clear that including the constitutionally problematic variables can substantially improve risk prediction in the aggregate. A core EBS premise is that actuarial risk prediction consistently outperforms clinical predictions. I examine the literature on which that claim is based, and find it unsupportive of this claim. To be sure, meta-analyses of "clinical versus actuarial" comparisons in various fields have given an edge on average to the actuarial—but not a large edge, and not a consistent one. The specifics of the actuarial instrument matter—one cannot say that any regression model is good by definition. Only a few comparative studies actually concern recidivism, and those have had mixed results. If anything, the studies support actuarial instruments that are very different from the crude ones that are actually being used—suggesting less discriminatory alternatives that could more effectively serve the state's penological interests. Another alternative is simply to drop the constitutionally problematic variables, perhaps to be replaced with crime characteristics. The empirical research gives no reason to believe that including these variables offers any nontrivial predictive improvement.

Third, even if the instruments predicted individual recidivism perfectly, they do not even attempt to predict the thing that judges need to know to use recidivism information in a utilitarian sentencing calculus. What judges need to know is not just how "risky" the defendant is in some absolute sense, but rather how the sentencing decision will *affect* his recidivism risk. For example, if a judge is deciding between a one-year and a two-year prison sentence for a minor drug dealer, it is not very helpful to know that the defendant's characteristics predict a "high" recidivism risk, absent additional information that tells the judge how much the additional year in prison will reduce (or increase) that risk. Current risk prediction instruments do not provide that additional information. Future research might be able to fill that gap, but it will not be easy. Estimating the causal relationship between sentences and recidivism is challenging, in part because sentencing judges take recidivism risk into account, introducing reverse causality

concerns. Some researchers have used quasi-experimental methods to tease out these causal pathways, but so far their estimates of incarceration's effects have not been demographically and socioeconomically specific.

Finally, I consider two interrelated counterarguments that defend EBS essentially by saying that it doesn't do much. The first is the claim that the instruments are innocuous because they do not directly specify a resulting sentence. Rather, they merely provide information—and what kind of obscurant would prefer sentencing to be ill-informed? This argument is not persuasive. The EBS instruments are meant to be *used* by judges, and to the extent they are used, they will systematically, and by design, produce disparate sentences across groups. The fact that the instruments do not exclusively determine the sentence might help in a "narrow tailoring" inquiry, but it is not enough alone to establish their constitutionality, nor their desirability.

The second counterargument might be labeled the "So what else is new?" defense. Risk prediction has always been central to sentencing, implicating its incapacitation, rehabilitation, and specific deterrence objectives. EBS advocates thus often argue that judges will inevitably predict risk, and may well rely on demographic and socioeconomic factors, even if they do not say so expressly. The instruments, on this view, are merely there to improve this assessment's accuracy. I argue, however, that EBS is not likely *merely* to replace one form of risk prediction with another. Rather, it will probably place a thumb on the sentencing scale in favor of more judicial emphasis on recidivism prevention relative to other sentencing goals. In many contexts, judges and other decision-makers tend to defer to "expert" assessments, especially with respect to scientific methods that they do not really understand. Moreover, providing risk predictions may simply increase the salience of crime prevention in judges' minds.

On this point, I also provide some new empirical evidence, based on a small experimental study that presented subjects with two fact patterns involving slight variations on the same crime. The two defendants varied sharply on several dimensions considered by risk prediction instruments. All subjects were presented with both scenarios and asked to recommend sentences; the experimental variation was that half the subjects were also presented with actuarial risk prediction scores. The effects of providing the scores were statistically significant and large. Subjects who did not receive the scores tended to give higher sentences to the lower-risk defendant, apparently focusing on small differences in the fact pattern that rendered that defendant more morally culpable. This pattern reversed when subjects received the scores, suggesting that the scores encouraged them to emphasize recidivism risk over moral desert. These results are tentative; judges in real cases might act differently. But the experiment adds to the existing empirical evidence that decision-making is affected by quantification and claims of scientific rigor.

Part I of this Article introduces the EBS instruments, describes their rise, and reviews the literature. Part II sets forth the disparity concern and makes the case for heightened constitutional scrutiny, and Part III applies that scrutiny to the empirical evidence underlying EBS. Part IV considers the above-described counterarguments. Finally, I offer some conclusions. Ultimately, in my view, the equality concerns are so serious that aggravating sentences on the basis of demographics and poverty would be bad policy even if the instruments advanced the state's interests far more substantially than they do. Likewise, the Supreme Court's case law on statistical discrimination may simply preclude deeming people dangerous on the basis of gender or poverty even if those generalizations were sufficiently well-supported that doing so would advance important state interests. But the fact that the instruments, and the use

of the problematic variables therein, do *not* advance those interests strongly (if at all) means that there is no defense of them available. This approach does not satisfy heightened constitutional scrutiny, and courts and policymakers should not embrace it.

I. Actuarial Risk Prediction and the Movement Toward Evidence-Based Sentencing

"Evidence-based sentencing" (EBS) refers to the use of actuarial risk prediction instruments to guide the judge's sentencing decision. The instruments are based on past regression analyses of the relationships between various offender characteristics and recidivism rates. Criminologists have developed a wide range of such instruments.[4] All incorporate criminal history variables, such as number of past convictions, past incarceration sentences, and number of violent or drug convictions.[5] Surprisingly, almost none include the crime of conviction in the case at hand. A few include very basic information such as whether it was a drug crime or a violent crime; others include no crime information.[6]

Most of the instruments include gender, age, and employment status; many also include education, and some include composite socioeconomic variables like "financial status."[7] Although risk prediction instruments used by some parole boards included race until as late as the 1970s, the modern EBS instruments overwhelmingly do not. One exception is a "sentencing support" software program promoted by an Oregon state judge, Michael Marcus,[8] but this not been formally adopted by any state. There appears to be a general consensus that using race would be unconstitutional. In 2000, the Supreme Court granted certiorari in a capital case to consider whether "a defendant's race or ethnic background may ever be used as an aggravating circumstance"; the issue was not a judicial sentencing instrument, but problematic testimony by a prosecution expert.[9] Before oral argument, the State of Texas conceded error and granted a new sentencing hearing, mooting the case.[10]

Most instruments now in sentencing use are limited to fairly objective factors, such as demographics, employment status, and criminal history.[11] But others include much more abstract, conceptual variables, which are meant to be coded by experienced evaluators. For instance, the Indiana Supreme Court in 2010 upheld against a state law challenge, and endorsed enthusiastically, use in sentencing of the Level of Services Inventory-Revised (LSI-R), which is also used by at least eight states elsewhere in the corrections process.[12] In addition to objective factors, the instrument also requires "subjective evaluations on … performance and interactions at work, family and marital situation, accommodations stability and the level of crime in the neighborhood, participation in organized recreational activities and use of time, nature and extent

---

[4] *See* J.C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 S.M.U. L. REV. 1329, 1399 (2011) (listing variables in 19 different instruments); Malenchik, 928 N.E.2d at 571-73.
[5] *See* Oleson, *supra*.
[6] *Id.*
[7] *Id.*
[8] *See* Draft MPC § 6B.09, cmt. (i) (discussing and criticizing this system); Michael H. Marcus, *Conversations on Evidence Based Sentencing*, 1 CHAP. J. CRIM. JUST. 61 (2009).
[9] See Saldano v. State, 70 S.W.3d 873, 875 (Tex. Crim. App. 2002) (describing the case's history); Monahan, *A Jurisprudence of Risk Assessment,* 92 VA. L. REV. 391, 392-93 (2006).
[10] Monahan, *supra*, at 393.
[11] Oleson, *supra*.
[12] *See* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 78-84 (2007) (describing the LSI-R's uses).

of social involvement with companions, extent of alcohol or drug problems, emotional/psychological status, and personal attitudes.[13]

The instruments are mechanical: each possible value of each variable corresponds to a particular increase or reduction in the risk estimate *in every case*. The variables' weights are not determined based on each case's circumstances—for instance, men will *always* receive higher risk scores than otherwise-identical women (because, averaged across all cases, men have higher recidivism rates), even if the context is one in which men and women tend to have similar recidivism risks or in which women have higher risks.[14] This is a feature of the simple underlying regression models, which generally have no interaction terms. Moreover, in practice the instruments use even simpler point systems, in which the "high risk" answer to a yes-or-no question results in a point or two being added to the defendant's score, based only quite loosely on the underlying regression.[15]

Demographic variables and socioeconomic variables receive substantial weight. For instance, in Missouri, presentence reports include a score for each defendant on a scale from -8 to 7, where "4-7 is rated 'good,' 2-3 is 'above average,' 0-1 is 'average', -1 to -2 is 'below average,' and -3 to -8 is 'poor.'"[16] Unlike most instruments in use, Missouri's does not include gender. However, an unemployed high school dropout will score three points worse than an employed high school graduate—potentially making the difference between "good" and "average," or between "average" and "poor."[17] Likewise, a defendant under age 22 will score three points worse than a defendant over 45.[18] By comparison, having previously served time in prison is worth one point; having four or more prior misdemeanor convictions that resulted in jail time adds one point (three or fewer adds none); having previously had parole or probation revoked is worth one point; and a prison escape is worth one point.[19] Meanwhile, current crime type and severity receive no weight.

Recidivism risk prediction instruments have been developed in various forms by criminologists over nearly a century,[20] and their use in parole determinations dates back decades,

---

[13] Malenchik, 928 N.E.2d at 572.

[14] For instance, medical studies suggest that women are on average more vulnerable to addiction and relapse than men are, so it may be that for some drug crimes women are more likely to recidivate. *See, e.g.*, Jill B. Becker & Ming Hu, *Sex Differences in Drug Abuse*, 29 FRONT NEUROENDOCRINOL 36 (2008). Recidivism studies do not break down gender effects like this, however.

[15] The point additions are at best crude roundings of regression coefficients. Moreover, the instrument does not track the regression's functional form. The underlying studies typically use logistic regression models, in which the coefficients translate nonlinearly into changes in probability of recidivism. When the instruments translate the coefficients into fixed, additive increases on a point scale, they are "linearizing" the variables' effects, and the resulting instrument will be only loosely related to the underlying nonlinear model, especially (because of the probability curve's shape) for very high-risk or very low-risk cases.

[16] Michael A. Wolff, *Missouri's Information-Based Discretionary Sentencing System*, 4 OHIO ST. J. CRIM. L. 95, 113 (2006).

[17] *Id.* at 112-13.

[18] *Id.*

[19] *Id.* A defendant with *every possible* criminal history risk factor (four or more misdemeanors resulting in jail, two or more prior felonies, prior imprisonment, prior prison escape, convictions within five years, revocation of probation and parole, and past conviction on the same offense as the current charge) will score eight points higher than one with no criminal history--just two points more than the combined effect of age, employment status, and high school graduation. *Id.*

[20] *See* HARCOURT, *supra*, at 1-2, 39-92 (reviewing this history).

although it has expanded sharply beginning in the 1980s.[21] Their use in sentencing is newer, however, and other than the state-specific instruments, none were initially designed for use in sentencing. For instance, the LSI-R manual specifically states that it "was never designed to assist in establishing the just penalty," which did not discourage the Indiana Supreme Court from endorsing its use for that purpose.[22] The first state to incorporate such an instrument in sentencing was Virginia in 1994, but the trend has taken off nationwide much more recently. Judge Roger Warren, the President Emeritus of the National Center for State Courts, argues that two developments in 2007 catalyzed this acceleration: a formal resolution of the Conference of Chief Justices and the Conference of State Court Administrators[23] and a report by the NCSC, the Crime and Justice Institute, and the National Institute of Corrections.[24] Another factor may be the recent shift toward discretionary sentencing after *Blakely v. Washington* and *United States v. Booker*. Tight sentencing guidelines leave little room for considering the defendant's individual risk, but in discretionary systems, judges are expected to assess it.[25]

Whatever the reasons, in recent years increasing number of states have followed Virginia's lead.[26] In fact, Douglas Berman states that "[i]n some form, nearly *every* state in the nation has adopted, or at least been seriously considering how to incorporate, evidence-based research and alternatives to imprisonment into their sentencing policies and practices."[27] EBS has many enthusiastic advocates in academia,[28] the judiciary and sentencing commissions,[29] and think tanks and advocacy organizations.[30] The National Center on State Courts has advocated using risk instruments to guide decision-making at all process stages, including training prosecutors and defense counsel to identify high- and low-risk offenders and thereby shaping

---

[21] *Id.* at 9, 77-80.
[22] Malenchik, 928 N.E.2d at 572-73.
[23] Conference of Chief Justices and Conference of State Court Administrators, Resolution 12 in Support of Sentencing Practices that Promote Public Safety and Reduce Recidivism, August 1, 2007; *see* Roger K. Warren, *Evidence-Based Setencing: Are We Up to the Task?*, 23 FED. SENT. R. 153, 153 (2010).
[24] Nat'l Inst. Of Corr. and Crime & Justice Inst, Evidence-Based Practice to Reduce Recidivism (2007).
[25] *See* Steven L. Chanenson, *The Next Era of Sentencing Reform*, 53 EMORY L.J. 377 (2005).
[26] Warren, *supra*, usefully reviews national and state policies promoting EBS.
[27] Douglas A. Berman, *Editor's Observations: Are Costs a Unique (and Uniquely Problematic) Kind of Sentencing Data?*, 24 FED. SENT. R. 159 (2012).
[28] *E.g.*, Jordan M. Hyatt, Mark H. Bergstrom, & Steven Chanenson, *Follow the Evidence: Integrate Risk Assessment into Sentencing*, 23 FED. SENT. R. 266 (2011); Lynn S. Branham, *Follow the Leader: The Advisability and Propriety of Considering Cost and Recidivism Data at Sentencing*, 24 FED. SENT. R. 169 (2012; Richard E. Redding, *Evidence-Based Sentencing: The* Science *of Sentencing Policy and Practice*, 1 CHAP. J. CRIM. JUST. 1, 1 & n.4 (reviewing articles praising EBS, and stating that failure to employ EBS "constitutes sentencing malpractice and professional incompetence").
[29] E.g., Marcus, *supra*; Warren, *supra*; Justice Michael Wolff (Chair, Missouri Sentencing Advisory Commission), *Evidence-Based Judicial Discretion: Promoting Public Safety through State Sentencing Reform*, 83 N.Y.U. L. Rev. 1389 (2008); Chief Justice William Ray Price, State of the Judiciary Address, Feb. 3, 2010, available at http://www.courts.mo.gov/page.jsp?id=36875; Mark H. Bergstrom (Pa. Commission on Sentencing) & Richard P. Kern (Va. Criminal Sentencing Commission), *A View from the Field: Practitioner's Response to "Actuarial Sentencing: An 'Unsettled' Proposition*, 25 FED. SENT. R. 185 (2013).
[30] *E.g.*, Pamela M. Casey, Roger K. Warren, & Jennifer K. Elek, USING OFFENDER RISK AND NEEDS INFORMATION AT SENTENCING 14 (Nat'l Ctr for State Courts 2011); PEW Ctr. on the States, Arming the Courts with Research: 10 Evidence-Based Sentencing Initiatives to Control Crime and Reduce Costs, 8 Pub. Safety Policy Brief 2-3 (2009); NAT'L CTR. FOR STATE COURTS, EVIDENCE-BASED SENTENCING TO IMPROVE PUBLIC SAFETY AND REDUCE RECIDIVISM: A MODEL CURRICULUM FOR JUDGES (2009); Matthew Kleiman, *Using Evidence-Based Practices in Sentencing Criminal Offenders,* in THE BOOK OF THE STATES (Council of State Gov'ts 2012).

plea-bargaining decisions.[31] Other academics have offered more cautious takes, but have ultimately offered qualified endorsements.[32]

The new Model Penal Code, currently undergoing its first revision since its adoption in 1962, embraces this new movement. This is a serious development, both because it reflects an emerging academic consensus and because of the MPC's influence. The original MPC was "one of the most successful law reform projects in American history," producing "revised, modernized penal codes in a substantial majority of the states."[33] Section 6B.09 of the new Code not only endorses use of "actuarial instruments or processes, supported by current and ongoing recidivism recidivism, that will estimate the relative risks that individual offenders pose to public safety," but also their formal incorporation into presumptive sentencing guidelines.[34] It also provides that when particularly low-risk offenders can be identified, otherwise-mandatory minimum sentences should be waived.[35] While parts of the revision are still being drafted, the American Law Institute has already approved Section 6B.09.[36]

The official Commentary to the MPC revision illustrates the core argument for EBS: recidivism risk prediction is inevitably part of sentencing, and should be guided by the best available scientific research:

> Responsible actors in every sentencing system—from prosecutors to judges to parole officials—make daily judgments about…the risks of recidivism posed by offenders. These judgments, pervasive as they are, are notoriously imperfect. They often derive from the intuitions and abilities of individual decisionmakers, who typically lack professional training in the sciences of human behavior. …. Actuarial—or statistical—predictions of risk, derived from objective criteria, have been found superior to clinical predictions built on the professional training, experience, and judgment of the persons making predictions.[37]

Most EBS advocates frame it as a strategy for reducing incarceration and its budgetary costs and social harms.[38] These advocates argue, or assume, that the prediction instruments will primarily allow judges to identify low-risk offenders whose sentences can be reduced, not high-risk offenders whose sentences must be increased. Some suggest that, absent scientific information on risk, judges probably already err on the side of longer sentences.[39] Others suggest that the instruments should categorically only be used in mitigation.[40]

---

[31] Casey et al., *supra*, at 23-26.
[32] *E.g.*, Margareth Etienne, *Legal and Practical Implications of Evidence-Based Sentencing by Judges*, 1 CHAPMAN J. CRIM. JUST. 43 (2009).
[33] Gerald Lynch, *Revising the Model Penal Code: Keeping It Real*, 1 OH. ST. J. CRIM. L. 219, 220 (2003) (also observing that the Code's classroom use makes it "the document through which most American lawyers come to understand criminal law").
[34] Draft MPC § 6B.09 (2).
[35] *Id.* at § 6B.09 (2).
[36] *See id.* at 133.
[37] Draft MPC, § 6B.09(2), cmt. (a). *See also, e.g.*, Wolff, *supra*, at 1406 (emphasizing superiority of actuarial prediction).
[38] *E.g.*, Nat'l Ctr. for State Courts, *Using Offender Risk and Needs Assessment Information at Sentencing* 2-3 (2011); Price, *supra* (citing EBS as a way to "move from anger-based sentencing" toward reduced incarceration); Wolff, *supra*, at 1390; PEW Ctr. on the States, *supra*, at 1; Michael Marcus, *MPC—The Root of the Problem: Just Deserts and Risk Assessment*, 61 FLA. L. REV. 751, 751 (2009).
[39] *E.g.*, Bonta, *supra*, at 524.
[40] *E.g.*, Etienne, *supra*.

In this spirit, the draft MPC Commentary asserts that "Section 6B.09 takes an attitude of skepticism and restraint concerning the use of high-risk predictions as a basis of elongated prison terms, while advocating the use of low-risk predictions as grounds for diverting otherwise prison-bound offenders to less onerous penalties." However, despite this "attitude," the actual *content* of Section 6B.09 endorses incorporation of risk assessment procedures into sentencing guidelines, including for the purpose of increasing sentences. The Commentary expresses hope that moving risk instruments from parole (the MPC would abolish parole) to sentencing will effectively constrain their "incapacitative" use, because access to counsel and greater transparency at sentencing would allow the defendant a chance to argue his case.[41] But the Commentary never explains how these procedural protections will ameliorate the instruments' substantive consequences for defendants whose objective characteristics render them "high risk." Even the best counsel will have trouble contesting the defendant's age, gender, education level, employment status, and past criminal convictions.[42] Moreover, if state legislatures adopt Section 6B.09 but not the MPC's recommendations concerning abolition of parole, the claim that parole-stage use is worse would be irrelevant.

Although most of the EBS literature is positive, or even celebratory, a few scholars have criticized it. The most thorough critique of risk prediction in criminal justice more broadly has come from Bernard Harcourt in his book AGAINST PREDICTION.[43] Some of Harcourt's arguments center on law enforcement profiling, but others apply to sentencing and parole. In particular, he argues that prediction instruments contravene punishment theory, because punishment turns on *who the defendant is* (and what he is therefore expected to do in the future), rather than just what he has done.[44] Although Harcourt's book primarily focuses on actuarial risk prediction, his theoretical objection is applicable to clinical prediction too—he seeks to "make criminal justice determinations blind to predictions of future dangerousness."[45] Likewise, advocates of purely retributive punishment have always held that a defendant's future risk is morally irrelevant to the state's justification for punishment.[46] Indeed, beyond mere irrelevance, there may be direct conflict (raising practical dilemmas for defense counsel): some factors that heighten a defendant's predicted recidivism risk, from young age to mental illness to socioeconomic disadvantage, are frequently considered *mitigating* factors from a retributive perspective.[47]

Other commentary on EBS has raised similar theoretical objections.[48] John Monahan, while advocating actuarial prediction in other contexts (such as civil commitment), has argued

---

[41] *Id.*

[42] Because the MPC draft advocates *mandatory* sentencing guidelines, it points out that the Sixth Amendment would require aggravating factors (but not mitigating factors) to be found by juries. *Id.* cmt. (e). This constraint, if anything, seems likely to discourage states from including difficult-to-prove dynamic factors like "antisocial attitudes" in the instruments. For factors like gender, age, and employment, the jury trial requirement seems essentially irrelevant.

[43] HARCOURT, *supra* note 12.

[44] *Id.* at 31-34, 188-89. Another of Harcourt's arguments is discussed below in Part III.C.

[45] *Id.* at 5; *see id.* at 237-38 (arguing that clinical judgment is just as vulnerable to his critique); Yoav Sapir, *Against Prevention? A Response to Harcourt's* Against Prediction *on Actuarial and Clinical Predictions and the Faults of Incapacitation*, 33 LAW & SOC. INQUIRY 253, 258-61 (2008) (arguing that the problem with the instruments is really a broader problem with incapacitation as a punishment objective, including via clinical judgment).

[46] *E.g.*, Paul Robinson, *Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114 HARVARD L. REV. 1429 (2001).

[47] *E.g.*, Graham v. Florida, 560 U.S. 48 (2010) (explaining mitigating role of young age).

[48] *See* Oleson, *supra*, at 1388-92 (reviewing literature).

against the current instruments' use in sentencing.[49] His view is that, while recidivism risk may be a legitimate sentencing consideration, blameworthiness is nonetheless the central question, and thus the only risk factors that should be considered are those that *also* bear on the defendant's moral culpability: past and present criminal conduct.[50] Some critics protest the probabilistic nature of risk prediction, ensuring "false positives" when those deemed high-risk do not, in fact, recidivate.[51] Others draw an unfavorable analogy to the science fiction movie "Minority Report," in which the government punishes "pre-crime," suggesting that even if the future could be known with certainty, punishing people for future acts is fundamentally unfair.[52] Many commentators raise such criticisms but do not treat them as dispositive, but merely as cautionary notes.[53] For others, like Harcourt, they are more fundamental flaws.

I do not seek to answer foundational sentencing-philosophy questions here. I accept EBS advocates' premise that recidivism prevention will inevitably play at least *some* role in the sentencing process in many cases (although I argue below that adoption of actuarial instruments will probably increase this role). The Supreme Court has affirmed the relevance of recidivism risk to sentencing, for example permitting judges to hear expert testimony concerning the defendant's dangerousness.[54]

Instead, this Article's central question is about discrimination and disparity: whether risk prediction instruments that classify defendants by demographic, socioeconomic, and family characteristics can be constitutionally or normatively justified. One could, after all, predict risk in other ways—for instance, based only on past or present criminal behavior, or based on individual assessment of a defendant's conduct, mental states, and attitudes. Current literature's treatment of the disparity concern is surprisingly limited; the MPC Commentary, for instance, barely mentions it. Among scholars who do raise the issue, most treat it as a policy concern, rather than (also) a constitutional one. For example, Harcourt, addressing the instruments' use in early release decisions, has argued that "risk is a proxy for race," observing that the instruments give heavy weight to criminal history, which is highly correlated with race.[55] He argues that this strategy will "unquestionably aggravate the already intolerable racial imbalance in our prison populations."[56] Kelly Hannah-Moffat has similarly critiqued the criminal history variables on grounds of racially disparate impact, and further emphasizes that criminal history may be influenced by past discriminatory decision-making.[57]

---

[49] In the civil commitment literature, scholars have focused on whether expert testimony predicting dangerousness is admissible evidence, rather than on the constitutionality or desirability of a particular judicial decision-making process. *E.g.*, Alexander Scherr, Daubert *and Danger: The 'Fit' of Expert Predictions in Civil Commitments*, 55 HASTINGS L.J. 1, 5-28 (2003) (reviewing case law and literature). I do not focus on the evidence law issues here.

[50] Monahan, *supra*, at 427-28.

[51] The MPC Commentary raises, but ultimately is unswayed by, this objection; see *infra* note 62 and accompanying text.

[52] *E.g.*, Oleson, *supra*, at 1390; Etienne, *supra*, at 59; Peter Moskos, Book Review, *Against Prediction*, 113 AM. J. SOCIOLOGY 1475, 1477 (2008).

[53] *E.g.*, Oleson, *supra*, at 1397-98 (concluding simply that EBS "raises excruciatingly difficult questions" and that "judges and jurists must determine" how to answer them).

[54] *Barefoot v. Estelle*, 463 U.S. 880 (1983); *see also Jurek v. Texas*, 428 U.S. 262 (1976) (holding that "prediction of future criminal conduct is an essential element in many" criminal justice-related decisions).

[55] Bernard Harcourt, *Risk as a Proxy for Race*, CRIM. & PUBLIC POL'Y (forthcoming), draft available at http://www.law.uchicago.edu/files/file/535-323-bh-race.pdf.

[56] *Id.*

[57] Kelly Hannah-Moffat, *Actuarial Sentencing: An 'Unsettled' Proposition*, at 17, available at http://www.albany.edu/scj/documents/Hannah-Moffatt_RiskAssessment.pdf.

The existing constitutional analyses, meanwhile, have focused on gender (and the hypothetical use of race), and have been limited in their doctrinal analysis.[58] The most extensive such analysis, by J.C. Oleson, concludes that the instruments survive even strict scrutiny.[59] Similarly, Monahan, while opposing use of demographic variables in sentencing on punishment-theory grounds, defends the constitutionality of their use in civil commitment, arguing that only race and gender raise constitutional issues at all, and that gender survives intermediate scrutiny because the gender differences are real and the state interests are substantial.[60]

In my view, the existing literature has seriously understated both the breadth and the gravity of the constitutional concern. There is a strong case that most or all of the risk prediction instruments now in use are unconstitutional, and current literature has not made that case or even seriously examined it. I seek to fill that gap, comprehensively analyzing the relevant case law and empirical research. I show both that the use of gender cannot be defended on the statistical bases that other authors have offered and that the problem goes well beyond gender—the socioeconomic variables, at least, should also receive heightened constitutional scrutiny. And if such scrutiny is applied, the empirical evidence is not currently strong enough to sustain the instruments, and it likely never will be.

In the criminological literature on the instruments, there is considerable debate over issues of reliability, validity, and precision. Current EBS scholarship often notes these concerns but ultimately advocates the instruments' use anyway.[61] The MPC Commentary is a striking example. It states that "error rates when projecting that a particular person will engage in serious criminality in the future are notoriously high" and that "most projections of future violence are wrong in significant numbers of cases," and yet concludes:

> Although the problem of false positives is an enormous concern—almost paralyzing in its human costs—it cannot rule out, on moral or policy grounds, all use of projections of high risk in the sentencing process. If prediction technology shown to be reasonably accurate is not employed, and crime-preventive terms of confinement are not imposed, the justice system knowingly permits victimizations in the community that could have been avoided.[62]

In my view, for all their apparent agonizing, the MPC drafters and other EBS advocates are missing the legal import of the methodological concerns: If the instruments don't work well, their use in sentencing is almost surely unconstitutional, and terribly unwise as well. As I show in Part II, the Supreme Court has warned against disparate treatment based on generalizations about (at least) gender and poverty, *even if* the generalizations have statistical support. If the statistical support is shoddy, there is simply no defending them.

It is curious that the EBS literature has not taken the constitutional concern more seriously. EBS scholars have occasionally asserted that actuarial prediction is obviously

---

[58] *E.g.*, Christopher Slobogin, *Risk Assessment and Risk Management in Juvenile Justice*, 27-WTR CRIM. JUST. 10, 13-14 (2013); Pari McGarraugh, Note, *Up or Out: Why "Sufficiently Reliable" Statistical Risk Assessment is Appropriate at Sentencing and Inappropriate at* Parole, 97 Minn. L. Rev. 1079, 1102 (2013).

[59] Oleson, *supra*, at 1388-92; *see also* Slobogin, *supra*, at 13-14 (briefly stating that gender discrimination probably survives intermediate scrutiny).

[60] Monahan, *supra*, at 429-432.

[61] *E.g.*, Slobogin, *supra*, at 16-17; McGarraugh, *supra*, at 1105-07; *see also* Hannah-Moffat, *supra* (raising various concerns but reaching an ambivalent conclusion: "Arguably, we should pause to reflect on the complexities of risk-needs assessments and concordant calls for and against evidence-based risk jurisprudence.").

[62] MPC Draft §6B.09, cmt. (e).

constitutional because the Supreme Court has approved, against a due process challenge, admission of even-less-reliable expert *clinical* predictions of risk in sentencing proceedings.[63] This assertion is wrong. The equal protection issue is not presented in those cases, and in general is not presented by individualized clinical assessments of risk *per se*; it is presented by punishment of group membership, which is explicit in the actuarial instruments. And even assuming actuarial predictions are more accurate than clinical ones, a question to which I return in Part III, the fact that evidence is reliable enough to be admissible does not mean that it establishes a strong enough relationship to an important government interest to withstand heightened scrutiny.[64] In the next Part, I show that such scrutiny applies.

II. The Disparate Treatment Concern

The most distinctive feature of EBS is that it formally incorporates discrimination based on socioeconomic status and demographic categories into sentencing. In this Part, I set forth the basic constitutional and policy objections to this practice. I begin with the constitutionality of gender-based sentencing in Section A (setting aside race because the current instruments do not include it).[65] Although it is uncontroversial that gender classifications are subject to heightened scrutiny, I examine the gender case law in some detail because it illuminates a core feature of the Supreme Court's equal protection jurisprudence that will make it very hard for EBS to *survive* heightened scrutiny: otherwise-unconstitutional discrimination cannot be justified by statistical generalizations about groups, even if the generalizations have empirical support. In Section B, I show that that the constitutional concern goes beyond gender: a form of heightened scrutiny (and a similar prohibition on group generalizations) also applies to socioeconomic discrimination in the criminal justice context. And in Section C, I articulate reasons policymakers should take the disparity concern seriously even if courts were to sustain EBS against constitutional challenges. This Part does not *complete* either the constitutional or the normative analysis; rather, it establishes the seriousness of the disparity concern and the resulting need at least for a very strong empirical justification for EBS. In Part III, I address whether such a justification exists.

Note that I frame my constitutional argument within existing doctrine, and thus do not argue for heightened scrutiny of certain other variables in the model—for instance, age and marital status are routine government classifications that are subject to rational basis review. There is, however, a plausible broader argument for strict scrutiny of group-based sentencing discrimination more generally, grounded in the "fundamental rights" branch of equal protection jurisprudence rather than the "suspect classifications" branch. Incarceration, after all, profoundly interferes with virtually every right the Supreme Court has deemed fundamental, and EBS makes these rights interferences turn on identity rather than criminal conduct. Although I would be happy to see the Supreme Court adopt such an approach, it is presently foreclosed to lower courts by language the Court used in a case called *Chapman v. United States*, and I do not focus

[63] *E.g.*, Slobogin, *supra*, at 15; Steven J. Morse, *Mental Disorders and Criminal Law*, 101 J. CRIM. L. & CRIMINOLOGY 885, 944 (2011); *see* Barefoot v. Estelle, 463 U.S. 880 (1983).

[64] In *Barefoot*, the Court made clear that the defects in evidence would have to be extreme before their admission would be barred by the Due Process Clause on the grounds of sheer unreliability. 463 U.S. at 898-99.

[65] The instruments do include socioeconomic variables that are highly correlated with race, a point I return to in § C, but they would be hard to challenge constitutionally on that basis. The Supreme Court has consistently held that absent a racially disparate purpose, policies that are facially neutral as to race cannot be challenged merely on the grounds of a racially disparate impact. *E.g.*, Washington v. Davis, 426 U.S. 229 (1976).

on it.[66] Certain variables used in some models might also merit new recognition as quasi-suspect—particularly variables relating to an offender's family background or family members' criminal history, which are closely analogous to illegitimacy, a quasi-suspect classification—but again, I do not rely on this possibility.[67] The policy critique in Section C thus applies more broadly, to more variables, than the constitutional arguments in Sections A and B do.

A. Gender Classifications and the Problem with Statistical Discrimination

Virtually every risk prediction instrument in use incorporates gender. Because the coefficient on gender is the same for all defendants, every single male defendant will, due to gender alone, be assigned a higher risk score than an otherwise-identical woman. Gender classifications are subject to heightened constitutional scrutiny, requiring an "exceedingly

---

[66] *Chapman v. United States*, 500 U.S. 453, 464-65 (1991). In *Chapman*, the defendant challenged the U.S. Sentencing Guidelines' method of calculating LSD weight, which included the carrier medium; the claim was that this method created unfair distinctions between people who carried the same amount of actual LSD. The Court rejected the notion that fundamental rights analysis should apply to sentencing distinctions within the statutory sentencing range, reasoning that once convicted, the offender no longer has a fundamental right to any sentence below the statutory maximum. Note that this holding does not preclude a challenge to a sentencing decision based on the nature of the classification; it speaks only to the "fundamental rights" branch. As I show below, both gender and poverty-based discrimination have triggered successful challenges to sentences within the statutory range.

Although *Chapman*'s holding is not entirely surprising (the Court in general is quite reluctant to apply constitutional scrutiny to sentences, *see* Carissa Byrne Hessick & F. Andrew Hessick, *Recognizing Constitutional Rights at Sentencing*, 99 Cal. L. Rev. 47, 49 (2011), and presumably worried that doing so in that case would require the extension of strict scrutiny to virtually every sentencing distinction), its reasoning, in my view, fails to take seriously the tremendous stakes of sentencing choices within statutory ranges. Those ranges are often very broad (say, zero to 20 years), and it is hard to imagine any government decision that would have a more drastic impact on a defendant's exercise of fundamental liberties than the choice between, say, 5 and 20 years' incarceration. Moreover, the Court's characterization of the right at issue was unduly narrow; the question is not whether the defendant had a right to a sentence below the statutory maximum. Rather, underlying, clearly established fundamental rights are being taken away (including the defendant's most basic physical liberty, which is directly and deliberately retracted by the incarceration decision, plus iadditional rights as procreation, communication, and voting). *Cf.* Lawrence v. Texas, 539 U.S. 558, 567 (2003) (critiquing the Court's past, overly narrow characterization of the right to sexual intimacy as a "right to homosexual sodomy").

The *outcome* in *Chapman* is perfectly defensible, but it could have been reached with a different rationale. The drug-weighting rule was a classification of *criminal conduct*, not persons, and thus (absent evidence of some discriminatory motive) raised no equal protection concern at all; all persons are prospectively subject to the same weighting rules, and have an equal chance to conduct their activities to avoid the rule. Applying fundamental rights analysis to EBS thus would not imply that routine sentencing distinctions between crimes are also subject to strict scrutiny. One could likewise defend sentencing distinctions based on criminal history as also being conduct-based and universally applicable—all persons who commit crimes are subjecting themselves to potential higher sentences for subsequent crimes. But when the state systematically gives different sentences to different groups of people for the same crime, with the same past criminal conduct, the Constitution *should* demand a compelling justification.

[67] Such variables are outside the defendant's control, unchangeable, generally unrelated to legitimate state policy, and often—especially in the case of familial incarceration or time in foster care—the basis for considerable social stigma and disadvantage. *See* Miriam J. Aukerman, *The Somewhat Suspect Class: Towards a Constitutional Framework for Evaluating Occupational Restrictions for People With Criminal Records*, 7 J. L. Society 18, 51 (2005) (reviewing case law and identifying factors that often trigger heightened scruting); John Hagan & Ronit Dinovitzer, *Collateral Consequences of Imprisonment for Children, Communities, and Prisoners*, 26 CRIME & JUST. 121 (1999) (reviewing literature on effects of parental incarceration); United States v. Sprei, 145 F.3d 528, 535 (2d Cir. 1998) (describing stigma and reduced marital prospects as an "inevitable result" of a parent's incarceration); Daniel Pollack et. al., *Foster Care as a Mitigating Circumstance in Criminal Proceedings*, 22 TEMP. POL. & CIV. RTS. L. REV. 43, 59 (2012) (quoting Sandra Stukes Chipungu & Tricia B. Bent-Goodley, *Meeting the Challenges of Contemporary Foster Care*, 14 FUTURE CHILD. 75, 85 (2004)).

persuasive justification"—that is, the state must prove "that the classification serves important governmental objectives and that the discriminatory means employed are substantially related to the achievement of those objectives."[68] Given this well-established doctrine, one might have thought that gender's inclusion in the instruments would have occasioned considerable concern and debate. And yet most scholarship ignores this concern, or else briefly asserts that the state's interests are important.[69] The draft Model Penal Code recommends excluding race, and the Commentary notes that sentencing based on race would be unconstitutional.[70] And yet the MPC drafters recommend including gender, and offer no commentary defending this on constitutional grounds, as though its constitutionality is self-evident.[71]

In the rare cases in which the issue has been presented, modern courts have consistently held (outside the EBS context) that it is unconstitutional to base sentences on gender.[72] There is, to be sure, considerable statistical research suggesting that judges (and prosecutors) *do* on average treat women defendants more leniently than men.[73] But it is virtually unheard of for modern judges to *say* that they are taking gender into account,[74] and demonstrating gender bias would usually be challenging. Before the past few decades, explicit consideration of gender as well as race was common, but few today defend that practice.[75] The U.S. Sentencing Guidelines, for example, expressly forbid the consideration of both race and sex.[76] Outside the literature on EBS, scholars have likewise mostly treated the gender gap as "unwarranted" sentencing disparity.[77]

Given this widespread consensus against sentencing based on gender, there is a certain surreal quality to the EBS literature's mostly untroubled embrace of it. The justification offered (if any) is that women in fact pose substantially lower recidivism risk than men do.[78] Some scholars add that to fail to account for this fact is unfair to women, essentially punishing them for men's recidivism risk.[79] More generally (referring to "gender, ethnicity, age, and disability"),

---

[68] *United States v. Virginia*, 518 U.S. 515 (1996).

[69] *E.g.*, Slobogin, *supra*, at 14. McGarraugh, *supra* at 1102, states that gender should be removed from the instruments to preserve their constitutionality, but does not develop the legal reasoning for this point.

[70] Draft MPC, *supra*, Sec 6B.09 cmt. (i).

[71] *Id.*

[72] Carissa Byrne Hessick, *Race and Gender as Explicit Sentencing Factors*, 14 J. GENDER RACE & JUST. 127, 137 (2010); *United States v. Maples*, 501 F.2d 985, 989 (4th Cir. 1974); *Williams v. Currie*, 103 F. Supp. 2d 858, 868 (M.D.N.C. 2000).

[73] *E.g.*, Sonja B. Starr, *Estimating Gender Disparities in Federal Criminal Cases* (under review) (2013) (finding large gender gaps at multiple procedural stages that are unexplained by observable variables, and also reviewing other studies).

[74] Hessick, *supra*, at 128.

[75] *Id.* at 129-36.

[76] U.S.S.G. Sec 5H1.10.

[77] *E.g.*, Oren Gazal-Ayal, A Global Perspective on Sentencing Reforms, 76 LAW & CONTEMP. PROBS. I, iii-iv (2013); Mona Lynch, Expanding the Empirical Picture of Sentencing: An Invitation, 23 FED. SENT. R. 313 (2011). Some scholars criticize increasing female incarceration rates, but do not generally argue that women should receive lower sentences based on gender per se. Rather, they argue that the system should take more account of certain mitigating factors that are more often present in female defendants' cases. E.g., Phyllis Goldfarb, *Counting the Drug War's Female Casualties*, 6 J. GENDER RACE & JUST. 277, 291-93 (2002); Leslie Acoca & Myrna S. Raeder, *Severing Family Ties: The Plight of Nonviolent Female Offenders and their Children*, 11 STAN. L. & POL'Y REV. 133 (1999).

[78] *E.g.*, Monahan, *supra*, at 431.

[79] *See* Margareth Etienne, *Sentencing Women: Reassessing the Claims of Disparity*, 14 J. GENDER, RACE, & JUSTICE 73, 82 (2010).

Judge Michael Marcus states: "We are not treating like offenders alike if we insist on ignoring factors that make them quite unalike in risk."[80]

But this argument, which embraces a concept of "actuarial fairness,"[81] stands on unsound constitutional footing. The Supreme Court has consistently rejected defenses of gender classifications that are grounded in statistical generalizations about groups—even those with empirical support. In *Craig v. Boren*, for instance, the Court considered a challenge to a law subjecting men to a higher drinking age for certain alcoholic beverages than women. The state had defended the law with statistical evidence, including a study showing that young men were arrested for drunk driving at more than ten times the rate of young women (2% versus 0.18%). The Court noted observed that "prior cases have consistently rejected the use of sex as a decisionmaking factor even though the statutes in question certainly rested on far more predictive empirical relationships than this." That is, what is prohibited is not *just* "outdated misconceptions" and merely "hypothesized" gender differences.[82] What is prohibited is inferring an individual tendency from group statistics. Note that the government's argument in *Craig* could easily have been framed in "actuarial fairness" terms: arguably it would have been unfair to bar young women from drinking based on a drunk driving risk that came almost entirely from males. But the Court's approach to equal protection means that individuals are neither entitled to a favorable statistical generalization based on gender, nor subject to unfavorable ones.

Examples of this principle abound. For instance, the Court has repeatedly held that government cannot base benefits policies on the assumption that wives are financially dependent on their husbands—even though, when the cases were decided in the 1970s, that presumption was usually correct.[83] The Court explained that "such a gender-based generalization cannot suffice to justify the denigration of the efforts of women who do" support their families.[84] Likewise, the Court has struck down gender-based peremptory challenges in jury selection, holding that the state cannot make assumptions about jurors based on gender, "even when some statistical support can be conjured up."[85] And in *United States v. Virginia*, the Court ordered the Virginia Military Institute to admit women, rejecting its arguments about "typically male or typically female 'tendencies.'" The Court observed: "The United States does not challenge any expert witness estimation on average capacities or preferences of men and women. … It may be assumed, for purposes of this decision, that *most* women would not choose VMI's adversative method." But, the Court emphasized, the point is not what *most* women would choose. "[W]e have cautioned reviewing courts to take a hard look at generalizations or 'tendencies' of the kind pressed by Virginia… [T]he State's great goal [of educating soldiers] is not substantially advanced by women's categorical exclusion, in total disregard of their individual merit, from the State's premier 'citizen soldier' corps."[86]

---

[80] Marcus, *supra*, at 769.

[81] This is a concept that has traditionally (although subject to some limitations) dominated insurance law—the idea is that it is fair for insurers to tailor rates to the risks posed by particular groups, and unfair to expect groups to cross-subsidize one another's risks. *See, e.g.*, Tom Baker, *Health Insurance, Risk, and Responsibility After the Patient Protection and Affordable Care Act*, 159 U. PA. L. REV. 1577, 1597-1600 (2011).

[82] *See* Monahan, *supra*, at 432-433 (defending gender-based risk prediction for civil commitment).

[83] Frontiero v. Richardson, 411 U.S 677 (1973); Weinberger v. Wiesenfeld, 420 U.S. 636, 645 (1975).

[84] Wiesenfeld, 420 U.S. at 645.

[85] J.E.B. v. Alabama ex rel. T.B., 511 U.S. 127, 140 n.11 (1994).

[86] In *City of Los Angeles v. Manhart*, 432 U.S. 702 (1978), the Court similarly struck down, on Title VII grounds, a requirement that female employees pay higher pension plan premiums because of their higher actuarial life expectancy. The Court stated:

In short, the Supreme Court has squarely rejected "statistical discrimination"—use of group tendencies as a proxy for individual characteristics—as a permissible justification for otherwise constitutionally forbidden discrimination. Economists often defend statistical discrimination as efficient, arguing that if a decision-maker lacks detailed information about an individual, relying on group-based averages (or even mere stereotypes, if the stereotypes have a grain of truth to them) will produce better decisions in the aggregate. But the Supreme Court has held that this defense of gender and race discrimination offends a core value embodied by the equal protection clause: that people have a right to be treated as individuals.

Individualism, indeed, is at the very heart of the Supreme Court's equal protection case law.[87] Many scholars have criticized this characteristic, arguing that it renders the Court's jurisprudence overly formalistic and too inattentive to substantive inequalities. On this view, the primary purpose of the Equal Protection Clause is to dismantle group-based subordination, not to ensure that government will treat individuals in ways that are blind to group identity; the latter approach may actually undermine the former if it prevents government from recognizing and acting to rectify societally entrenched inequalities.[88] I am sympathetic to this view myself, in fact, but I frame this Article within the approach that dominates current doctrine. In any event, an antisubordinationist approach to equal protection law would hardly be friendlier to EBS, an approach that amplifies inequality in the criminal justice system's impact by inflicting additional criminal punishment on the poor and, via disparate impact, on people of color. In Section D, I explore further EBS's social and distributive impacts, and explain why (even though men, in general, are not a subordinated class) its inclusion of gender can be expected to exacerbate this social impact on disadvantaged groups.

Thus, although gender discrimination is not wholly constitutionally forbidden, EBS proponents are going to face tough sledding if their defense of it depends on statistical generalizations about men and women. And it does—EBS is *all about* generalizing based on statistical averages, and its advocates defend it on the basis that the averages are right. At least in the gender context, that probably will not convince courts. The statistical relationship would at the very least have to be so strong that courts could deem the resulting *individual* predictions noticeably more sound than those the Supreme Court has rejected in the past, and could accordingly hold that an "exceedingly persuasive justification" for the classification was present. But this requirement sets a high bar—in *United States v. Virginia*, for instance, the Court's only example of sex differences that the government *could* (within constraints) consider was the irreducible *physical* differences between men and women.[89]

Beyond gender, the Court's emphasis on individualism and rejection of statistical discrimination should inform our thinking about the constitutionality of other variables as well.

> This case … involves a generalization that the parties accept as unquestionably true: Women, as a class, do live longer than men….It is equally true, however, that all individuals in the respective classes do not share the characteristic that differentiates the average class representatives.….. [Title VII] precludes treatment of individuals as simply components of a racial, religious, sexual, or national class. … Even a true generalization about the class is an insufficient reason for disqualifying an individual to whom the generalization does not apply.

Id. at 707-08; *see* Metro Broadcasting, Inc. v. F.C.C., 497 U.S. 547, 620 (1990) (O'Connor, J., dissenting) (citing this passage to inform the interpretation of the Equal Protection Clause).

[87] *See* Richard Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 493, 553 (2002).

[88] *See id.* at 554-59; Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIAMI L. REV. 9, 9-10 (2004) (reviewing antisubordinationist scholarship).

[89] 518 U.S. at 533.

To be sure, it is not *always* forbidden for the government to rely on statistical generalizations; it would be hard to imagine government functioning if it did not, since it would have to tailor every action it takes to every individual. Government sometimes has to draw clear lines that may overgeneralize—for instance, the state sets a maximum blood-alcohol content for driving, rather than requiring each individual's fitness to drive to be individually assessed. Frederick Schauer has made this point forcefully, offering a fairly broad defense of reliance on statistically supported generalizations.[90] But as Schauer emphasizes, this practice properly has limits—certain kinds of generalizations (including those based on gender) are particularly socially harmful, or expressively invidious, even if they have statistical support.[91] The practice of applying more demanding equal protection scrutiny to some government classifications than to others is grounded in similar reasoning.

Note that the problem with EBS could be framed either as excess generalization (failure to treat people as individuals whose risk varies for reasons particular to them) or as *insufficient* generalization (failure to treat all those with the same criminal conduct the same way). Schauer, for instance, defended the then-mandatory U.S. Sentencing Guidelines, and particularly their bar on demographic and socioeconomic considerations, along the latter lines: "Ignoring real differences in sentencing -- sentencing socially beneficial heart surgeons to the same period of imprisonment for murder as socially parasitic career criminals -- may well serve the larger purpose of explaining that at a moment of enormous significance … we are all in this together."[92] In my view, the problem with EBS cannot be simply described in terms of generality versus particularity; the problem is not that the instruments generalize, but that they employ particular kinds of generalizations that are insidious, in a context that has huge consequences for individuals and communities.

B. Wealth-Related Classifications in the Criminal Justice System

The constitutional problem with EBS goes beyond gender. In this Section, I show that current doctrine also supports application of heightened scrutiny to variables related to socioeconomic status, such as employment status, education, or income. The Supreme Court's case law in *other* contexts has consistently held that similar wealth-related classifications are not constitutionally suspect,[93] and perhaps this is why EBS scholars have completely ignored the potential constitutional concerns with these variables. But this case law is not dispositive in the sentencing context. Many criminal defendants have challenged policies and practices that effectively discriminate against the indigent, including discrimination in punishment. These defendants have often succeeded, and the Supreme Court and lower courts have applied to their claims a special form of heightened constitutional scrutiny, citing intertwined equal protection and due process considerations.

The treatment of indigent criminal defendants has for more than a half-century been a central focus of the Supreme Court's criminal procedure jurisprudence. Indeed, the Court has often used very strong language concerning the importance of eradicating wealth-related

[90] FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES (2003).
[91] *Id.* at 38-41.
[92] *Id.* at 261-62. Although I am uncomfortable with group-based sentencing distinctions, I do not favor mandatory sentencing, because offenses are often defined too broadly to capture real differences in criminal conduct and culpability. Also, mandatory sentencing laws generally do not eliminate individualization of punishment, but shift the power to individualize toward prosecutors (a possibility Schauer acknowledges, *id.* at 256).
[93] *E.g.*, Maher v. Roe, 432 U.S. 464, 471 (1977).

disparities in criminal justice; in *Griffin v. Illinois*, for instance, it called this objective "the central aim of our entire judicial system."[94] *Griffin* struck down the requirement that defendants pay court costs before receiving a trial transcript, which they need to prepare an appeal. The Court held that "[i]n criminal trials a State can no more discriminate on account of poverty than on account of religion, race, or color," and that "[t]here can be no equal justice where the kind of trial a man gets depends on the amount of money he has."[95]

Numerous other cases also stand for the principle that both equal protection and due process concerns require that indigent criminal defendants not be subject to special burdens. Principally, these cases have focused on access to the criminal process: "the belief that justice cannot be equal where, simply as a result of his poverty, a defendant is denied the opportunity to participate meaningfully in a judicial proceeding in which his liberty is at stake."[96] Notably, these cases have applied heightened scrutiny even when the wealth-based classification did not deprive the defendant of something to which he otherwise would have had a substantive right—the cases relating to appeal procedures, for instance, reiterated the then-established principle that a State need not provide an appeal as of right at all. Rather, *Griffin* and its progeny involved a special "equality principle" motivated by "the evil [of] discrimination against the indigent."[97] For this reason, a challenge to EBS need not establish that the defendant has some free-standing constitutional entitlement to a lower sentence than he received.

For our purposes, the most on-point Supreme Court case is *Bearden v. Georgia*, in which the district court had revoked the probation of an indigent defendant who had been unable to pay his court-ordered restitution.[98] The Court unanimously reversed, holding that incarcerating a defendant merely because he was unable to pay amounted to unconstitutional wealth-based discrimination.[99] Importantly, the Court in *Bearden* squarely rejected the state's argument that poverty was a recidivism risk factor that justified additional incapacitation:

> [T]he State asserts that its interest in rehabilitating the probationer and protecting society requires it to remove him from the temptation of committing other crimes. This is no more than a naked assertion that a probationer's poverty by itself indicates he may commit crimes in the future. …[T]he State cannot justify incarcerating a probationer who has demonstrated sufficient bona fide efforts to repay his debt to society, solely by lumping him together with other poor persons and thereby classifying him as dangerous. This would be little more than punishing a person for his poverty.[100]

The Court's resistance to "lumping [the defendant] together with other poor persons" is very similar to its reasoning concerning statistical discrimination in the gender cases. The Court

---

[94] 351 U.S. 12, 16 (1956).

[95] *Id.* at 19. *Accord* Mayer v. City of Chicago, 404 U.S. 189 (1971).

[96] Ake v. Oklahoma, 470 U.S. 68, 76 (1985); *see also* Gideon v. Wainwright, 372 U.S. 335, 344 (1963) (citing the goal of achieving a justice system in which, regardless of finances, "every defendant stands equal before the law").

[97] Lewis v. Casey, 518 U.S. 343, 369-72 & nn. 2-3 (1996) (Thomas, J., concurring) (reviewing case law); *see* Douglas v. California, 372 U.S. 353, 355-57 (1963); *United States v. MacCollom,* 426 U.S. 317, 331 (1976) (Brennan, J., dissenting) (referring to the "*Griffin* equality principle").

[98] 461 U.S. 660 (1983).

[99] *Id. Bearden* built on *Williams v. Illinois*, 399 U.S. 235 (1970), in which the Court had similarly reversed a revocation of probation for failure to pay restitution. In *Williams*, the resulting incarceration sentence exceeded the statutory maximum for the crime, and the Court stated in dictum that absent that problem, no constitutional concern would have been raised. *Id.* at 243 In *Bearden*, however, the incarceration sentence did *not* exceed the statutory maximum, and the Court nonetheless held it unconstitutional, apparently rejecting the *Williams* dictum.

[100] 461 U.S. at 671.

observed that the state had cited "several empirical studies suggesting a correlation between poverty and crime," but it was not persuaded by this appeal to a statistical generalization.[101]

Bearden does not establish that financial background is *always* irrelevant to sentencing. Although the Court decisively rejected the use of poverty to predict crime risk, it took more seriously a different defense of the provocation revocation. The Court emphasized one reason it may be permissible to consider ability to pay (and related factors such as employment history) when choosing between incarceration and restitution sentences:

> The State, of course, has a fundamental interest in appropriately punishing persons--rich and poor--who violate its criminal laws. A defendant's poverty in no way immunizes him from punishment. Thus…the sentencing court can consider the entire background of the defendant, including his employment history and financial resources.[102]

That is, the State may consider financial factors as necessary to ensure the poor do not *avoid* punishment—as they would if sentenced only to pay a fine or restitution that they then cannot pay. But with EBS, poverty is not being considered to *enable equal punishment* of rich and poor, but to trigger extra, unequal punishment.[103] The Court further held that even when probation revocation is necessary to ensure that the poor do not avoid punishment, it is only permitted after an inquiry to determine if there are viable alternatives, such as "a reduced fine or alternate public service…Only if the sentencing court determines that alternatives to imprisonment are not adequate in a particular situation to meet the State's interest in punishment and deterrence may the State imprison a probationer who has made sufficient bona fide efforts to pay."[104]

This requirement that less restrictive alternatives be considered is a hallmark of strict scrutiny. However, the Court resisted expressly categorizing its analysis within any particular tier of scrutiny. Indeed, reviewing the case law on indigent criminal defendants, the Court expressed ambivalence as to whether the key constitutional provision was really the Equal Protection Clause at all, as opposed to the Due Process Clause. As the Court explained, these constitutional concerns are intertwined in these cases, and in any event,

> "[w]hether analyzed in terms of equal protection or due process, the issue cannot be resolved by resort to easy slogans or pigeonhole analysis, but rather requires a careful inquiry into such factors as 'the nature of the individual interest affected, the extent to which it is affected, the rationality of the connection between legislative means and purpose, [and] the existence of alternative means for effectuating the purpose ....'"[105]

This language suggests an unconventional, perhaps somewhat flexible balancing test: a stronger legislative purpose and connection to that purpose might be required depending on the individual

---

[101] *Id.* at 671 n.11.

[102] *Id.* at 669-70.

[103] *See also* Williams v. Illinois, 388 U.S. at 244 (stating that ability to pay can be considered to avoid "inverse discrimination"); United States v. Altamirano, 11 F.3d 52, 53 (5th Cir. 1993) (discussing the circumstances in which courts can consider indigency). A defendant, indeed, is constitutionally entitled to a judicial inquiry into her ability to pay a fine. *See, e.g.*, Powers v. Hamilton County Public Defender Comm'n, 501 F.3d 592, 608 (6th Cir. 2007).

[104] 461 U.S. at 671-72. Similarly, Justice White wrote that because "poverty does not insulate those who break the law from punishment," the poor may be imprisoned if they cannot pay fines, but only "if the sentencing court makes a good-faith effort to impose a jail sentence that in terms of the state's sentencing objectives will be roughly equivalent to the fine and restitution that the defendant failed to pay." That is, the magnitude of the punishment must be the same, even if the means is not. Bearden, 461 U.S. at 675 (White, J., concurring in the judgment).

[105] 461 U.S. at 666-67; *see Evitts v. Lucey*, 469 U.S. 387 (1985) (discussing the interrelationship between due process and equal protection concerns in these cases).

interest at stake and the extent to which it is effected. But in requiring a "careful inquiry" into each factor, including the existence of alternatives, it is clear that the Court means to require *some* form of heightened scrutiny, considerably more assertive than mere rational basis review.

Although *Bearden* involved revocation of probation, lower courts have treated it as a constraint on initial sentencing decisions. For instance, the Ninth Circuit has cited *Bearden* to reverse a district court's decision to treat inability to pay restitution as an aggravating sentencing factor, explaining that "the court improperly injected socioeconomic status into the sentencing calculus" and that "the authority forbidding such an approach is abundant and unambiguous."[106] Conversely, citing the same disparity concern, the Ninth Circuit has also reversed (as "unreasonable" under *United States v. Gall*) a decision to *reduce* a defendant's sentence due to ability to pay restitution, holding: "Rewarding defendants who are able to make restitution in large lump sums…perpetuates class and wealth distinctions that have no place in criminal sentencing."[107] Even before *Bearden*, several circuits had already held that equal protection entitles an indigent defendant who was unable to make bail to credit against the eventual sentence for time served, to avoid impermissible wealth-based distinctions in sentencing.[108]

The Supreme Court and lower courts have recognized a divergence between the Supreme Court's treatment of indigent criminal defendants and its normally deferential review of wealth-based classifications: "legislation which has a disparate impact on the indigent defendant should be subject to a more searching scrutiny than requiring a mere rational relationship."[109] The Supreme Court itself has repeatedly noted this divergence. In *United States v. Kerr*, a district court reasoned that special scrutiny is justified by a combination of the serious stakes and the nature of the class: "At stake here is not mere economic or social welfare regulations but deprivation of a man's liberty. The courts 'will squint hard at any legislation that deprives an individual of his liberty—his right to remain free.' Moreover, the indigent, though not a suspect class, have suffered unfair persecution."[110]

Outside the context of inability to pay fines and restitution, there is relatively little case law focusing on use of wealth classifications to determine substantive sentencing outcomes. This dearth should not be taken to suggest judicial approval—the issue likely rarely arises because the practice is rare. The criminal justice system has been rife with procedural obstacles to equal treatment of the indigent, and there are no doubt many subtle or *de facto* ways in which

---

[106] United States v. Burgum, 633 F.3d 810, 816 (9th Cir. 2011); *accord* United States v. Parks, 89 F.3d 570, 572 (1996) ("[The defendant] may be receiving an additional eight months on this sentence due to poverty. Such a result is surely anathema to the Constitution."); *see also* United States v. Ellis, 907 F.2d 12, 13 (1st Cir. 1990) (stating that "the government cannot keep a person in prison solely because of indigency"); *but see* State v. Todd, 147 Idaho 321, 323 (2009) (upholding inability to pay as an aggravating factor).

[107] United States v. Bragg, 582 F.3d 965, 970 (9th Cir. 2009).

[108] *See, e.g.*, King v. Wyrick, 516 F.3d 321, 323 (8th Cir. 1975); *Ham v. North Carolina,* 471 F.2d 406, 407, 408 (4th Cir.1973); *Johnson v. Prast,* 548 F.2d 699, 703 (7th Cir.1977); *but see* Vasquez v. Cooper, 862 F.2d 250 (10th Cir. 1988) (finding no constitutional violation because the court considered inability to pay when setting bail).

[109] *U.S. v. Luster*, 889 F.2d 1523 (6th Cir. 1989); *see also* Maher v. Roe, 432 U.S. 464, 471 n.6 (1977); Kadrmas v. Dickinson Public Schools, 487 U.S. 450, 461 n.*(1988) (rejecting heightened scrutiny in a non-criminal case because "the criminal-sentencing decision at issue in *Bearden* is not analogous to the user fee … before us"); Dickerson v. Latessa, 872 F.2d 1116, 1119-1120 (1st Cir. 1989) (observing that classifications implicating appeal rights receive heightened scrutiny only if they are wealth-based); United States v. Avendano-Camacho, 786 F.2d 1392, 1394 (9th Cir. 1986) (Kennedy, J.) (same); *United States v. Kerr*, 686 F. Supp. 1174 (W.D. Pa. 1988).

[110] Kerr, 686 F. Supp. at 1178 (quoting Williams v. Illinois, 399 U.S. at 263 (Harlan, J., concurring) and citing Papachristou v. City of Jacksonville, 405 U.S. 156 (1972)).

poverty might influence the sentence. But the practice of actually treating poverty as an aggravating factor in sentencing has not been prevalent (before EBS) and has been considered illegitimate. For instance, the formerly mandatory U.S. Sentencing Guidelines forbid consideration of socioeconomic status.[111] It is true that, now that the guidelines are merely advisory, federal courts do occasionally refer to education or employment when discussing the offender's circumstances (as do state courts—in contrast to gender, which is essentially never cited).[112] Such cases might well also be constitutionally problematic, unless such factors are used in service of the "equal punishment" principle discussed above; I do not focus here on the factors that can be considered in individualized judicial assessments of offenders. But at least such cases do not necessarily reflect a generalization that unemployed or uneducated people are categorically more dangerous, in the mechanical way that the EBS instruments do. Instead, the court can assess what each factor means in the context of a particular case—considering, for instance, whether the offender is making an effort to find employment or otherwise pursue rehabilitation, rather than simply blindly adding a given number of points based on current employment status or past educational attainment.

The federal Guidelines do include an enhancement for offenders with a "criminal livelihood,"[113] and defendants have occasionally challenged that enhancement as disparately affecting the poor, because the same criminal revenue would constitute a larger share of a low-income person's livelihood. Soon after the guideline's adoption, a least one district court held (citing *Bearden*) that, to avoid this potential constitutional concern, it should be interpreted to focus on the absolute amount of criminal income, rather than the share of total income, and the Sentencing Commission amended the guideline to come closer to this view.[114] After the amendment, the Sixth Circuit upheld the new guideline against a similar challenge, holding that although *Bearden* required heightened scrutiny of sentencing burdens on the poor, the amended guideline appropriately targeted "professional criminals" who have "chosen crime as a livelihood" and that any disproportionate effect on the poor did not reflect disparate treatment, but rather was "an incidental effect of the statute's objective."[115]

This rationale, however, cannot be applied to EBS, in which poverty indicators are themselves treated as recidivism risk factors—exactly the statistical generalization that the Supreme Court squarely condemned in *Bearden.* As the district court put it in *Kerr*, even though *Bearden* recognized "a correlation between poverty and crime,…a person cannot be punished solely for his poverty. As a matter of constitutional belief, the presumption that the indigent will act criminally 'is too precarious for a rule of law.'"[116]

It is difficult to see how the socioeconomic variables in EBS can avoid *Bearden*-like heightened scrutiny. Unemployment and education, the most common such variables, cannot

---

[111] U.S.S.G. 5H1.10; *see also* Joan Petersilia & Susan Turner, *Guideline-Based Justice: Prediction and Racial Minorities*, 9 CRIME & JUST. 151, 153-154, 160 (1987) (describing sentencing reformers' objective of eliminating role of "status" factors like employment).

[112] *E.g.*, United States v. Trimble, 2013 WL 1235510 (11th Cir. 2013).

[113] U.S.S.G. 4B1.3.

[114] United States v. Rivera*,* 694 F.Supp. 1105 (S.D.N.Y.1988); *see* United States v. Luster, 889 F.2d 1523 (6th Cir. 1989) (describing the amendment). The amended guideline's quantitative inquiry concerns only the amount of criminal income; there is also a qualitative inquiry into whether crime was the defendant's "primary occupation."

[115] Luster, 889 F.2d at 1530.

[116] 686 F. Supp. at 1179. Cf. *Edwards v. California*, 314 U.S. 160, 177 (striking down a vagrancy law and holding that it could not be "seriously contended that because a person is without employment and without funds he constitutes a 'moral pestilence'. Poverty and immorality are not synonymous.").

meaningfully be distinguished from the ability to pay, nor can composite variables like "financial status." All are proxies for poverty, and the case law in the *Bearden-Griffin* line makes interchangeable references to "wealth," "poverty," "class," and so forth without fine distinctions. For instance, the Court has always treated "ability to pay" as being equivalent to poverty, even though the two are not identical—ability to pay also depends on what one's other expenses are, whether one can borrow money from someone, and so forth. *Bearden* directly addresses, and limits, the circumstances under which courts can consider "employment history and financial resources," specifically rejecting the consideration of such factors as recidivism predictors.[117] Indeed, the argument the Court was rejecting in that passage turned fundamentally on employment status; the empirical studies that Georgia had cited in *Bearden* to support its recidivism-risk argument were mainly studies of the relationship between unemployment and recidivism, and the state emphasized that the defendant's recent job loss made him a higher recidivism risk.[118] Meanwhile, the point of including education in the recidivism instrument is that it is a proxy for the defendant's future prospects for employment and legitimate earnings; it would be hard to defend the use of this factor using logic that clearly distinguished it from past, present, or future poverty. Neighborhood characteristics could potentially also be considered socioeconomic variables, since they are also very closely related to poverty, although this example is more disputable because these variables operate at a geographic level and do not draw distinctions among persons within the neighborhood.[119]

While there are limits to the courts' efforts to protect indigent defendants, those limits have been found in cases testing what affirmative assistance the state must provide in order to level the criminal justice playing field. EBS, in contrast is a deliberate effort to *unlevel* that field. As with gender, its defenders will be fighting an uphill battle to overcome heightened scrutiny, because if, as *Bearden* holds, one cannot impute individual risk based on the average risk posed by poor defendants, the rationale for EBS disappears.

C. The Social Harm of Demographic and Socioeconomic Sentencing Discrimination

EBS's use of demographic, socioeconomic, and family-related characteristics is also highly troubling on public policy grounds. As noted above, EBS advocates frequently emphasize its potential to help reduce incarceration rates.[120] But what they do not typically emphasize is that the mass incarceration problem in the United States is drastically disparate in its distribution. This unequal distribution is a core driver of its adverse social consequences, because it leaves certain neighborhoods and subpopulations decimated. Black men, for instance, are 52 times as likely to be incarcerated as white women are.[121] Young black men are especially at risk: one in nine black men under 35 are currently behind bars,[122] and one in three will be at some point in

[117] 461 U.S. at 671.
[118] Brief of the Respondent, *Bearden v. Georgia*, 1982 U.S. Sup. Ct. Briefs LEXIS 438, at 32-35.
[119] Given fairly high levels of residential segregation, *see generally* U.S. Census Bureau, Racial and Ethnic Residential Segregation in the United States: 1980-2000, available at http://www.census.gov/hhes/www/housing/housing_patterns/pdf/censr-3.pdf, neighborhood might also be a racial proxy, but challengers would likely have trouble proving a racially discriminatory purpose.
[120] *See supra* note 38 and accompanying text.
[121] *See* HEATHER C. WEST, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, PRISON INMATES AT MIDYEAR 2009—STATISTICAL TABLES, 21 tbl.18 (2010).
[122] PEW CTR. ON THE STATES, ONE IN 100: BEHIND BARS IN AMERICA 2008, at 3 (2008), *available at* http://www.pewstates.org/research/reports/one-in-100-85899374411.

their lives.[123] And the concentration of incarceration's effects is even more dramatic when one takes into account socioeconomic and neighborhood-level predictors. High school dropouts, for example, are 47 times as likely to be incarcerated as college graduates are, and young black male dropouts are incarcerated at a rate of approximately 22% at any given time.[124] An ample literature documents these disparities and their effects on communities.[125]

The EBS instruments produce higher risk estimates, other things equal, for the same subgroups that are already disproportionately incarcerated, and so it is reasonable to predict that EBS will exacerbate these disparities. Although we do not know whether EBS will reduce incarceration on balance, the most intuitive expectation is that it will increase incarceration for some people (those deemed high-risk) and reduce it for others (those deemed low-risk). If so, it will further concentrate mass incarceration's impact demographically.

This is likely to include concentrating its racial impact. I have ignored race in my constitutional analysis, because the instruments do not include it. But the socioeconomic, family, and neighborhood variables that they do include are highly correlated with race, as is criminal history, so they are likely to have a racially disparate impact.[126] Although the courts have not recognized equal protection claims grounded in disparate impact, policymakers should care about the consequences of their policies, and not just about the facial distinctions that they draw. Ample literature documents mass incarceration's severe consequences for African-American communities in particular. If EBS exacerbates this problem, it would be particularly hard to defend it as a progressive strategy for responding to the mass incarceration crisis.

The demographic concentration problem is one reason to worry about the gender and age variables, in addition to socioeconomic status. In other contexts, discrimination based on young age is often treated as not particularly morally troublesome. Young age is not a significant social disadvantage, nor is it even really a discrete group trait; everyone has it and then loses it. Likewise, many advocates no doubt worry less about gender discrimination that adversely affects men because men, taken as a whole, have dominant political and economic power. But the likely impact of EBS is not centered on "men taken as a whole," nor on young people generally. Rather, it will principally affect a subgroup of young men—those involved in the criminal justice system, mostly poor men of color—who are highly disadvantaged. The age and gender criteria exacerbate the extent to which incarceration's impact targets a particular slice of disadvantaged communities, effectively resulting in a substantial part of a generation of men being absent from communities and exacerbating the socially distortive effects of mass incarceration. A broad literature explores the effects of high, demographically concentrated incarceration rates on everything from marriage rates to overall community cohesion.[127]

---

[123] THOMAS BONCZAR, U.S. DEP'T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, PREVALENCE OF IMPRISONMENT IN THE U.S. POPULATION, 1974-2001 (2003).

[124] Center for Labor Market Studies, *The Consequences of Dropping Out of School* (2009), available at http://hdl.handle.net/2047/d20000596; *see also* Robert J. Sampson & Charles Loeffler, *Punishment's Place: The Local Concentration of Mass Incarceration,* DAEDALUS (Summer 2010) (discussing neighborhood effects).

[125] *E.g.*, MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS (2011); TODD R. CLEAR, IMPRISONING COMMUNITIES: HOW MASS INCARCERATION MAKES DISADVANTAGED NEIGHBORHOODS WORSE (2007); IMPRISONING AMERICA: THE SOCIAL EFFECTS OF MASS INCARCERATION (Mary Patillo et al. eds., 2004).

[126] *See* Harcourt, *supra* note 55 (arguing that "prior criminal history has become a proxy for race").

[127] Todd R. Clear, *supra*, at 97; William A. Darity, Jr. & Samuel L. Myers, Jr., *Family Structure and the Marginalization of Black Men: Policy Implications*, *in* THE DECLINE IN MARRIAGE OF AFRICAN AMERICANS 263,

Another serious disadvantage is the expressive message sent by state endorsement of sentencing based on group traits. Consider specifically the traits associated with socioeconomic disadvantage. Though many Americans no doubt already suspect that the criminal justice system is biased against the poor, EBS ends any doubt on the matter. It involves the state telling judges explicitly that poor people should get longer sentences because they are poor—and, conversely, that socioeconomic privilege should translate into leniency. That is a message that, I suspect, many state actors would find embarrassing to defend in public. Doing so would require pointing to a justification that hardly improves matters: that the poor are dangerous. Generalizing about groups based on crime risk is a practice with a pernicious social history.[128] Dressing up that generalization in scientific language may have succeeded in forestalling public criticism, but mostly because few Americans understand these instruments or even are aware of them. If the instruments were better understood (and as EBS expands, perhaps they will be), they would send a clear message to disadvantaged groups: the system really is rigged. Further, if that message undermines the criminal justice system's legitimacy in disadvantaged communities, it could undermine EBS's crime prevention aims.[129]

Some EBS advocates propose that it should be used only to *mitigate* sentences, and such proposals have, at first glance, a seductive appeal—reducing incarceration rates is an important objective.[130] But there is no persuasive reason to believe access to risk predictions would tend to reduce sentences rather than increasing them (or doing both in different cases). Some advocates blame a retributivist approach to sentencing for the rise in incarceration, and suggest that EBS would help to make sentencing more moderate by encouraging a practical focus on crime prevention instead.[131] This line of argument is curious, however, because much of the political "tough on crime" movement over the past several decades has in fact been accompanied by public safety language, responding to the public's (oft-exaggerated) perceptions of crime risk.[132]

One could attempt to force unidirectional use of risk assessments, but it may be difficult. If judges are given the risk assessments before they choose the sentence, even if they are told to only use them for mitigation, it is difficult to expect them to completely ignore high-risk assessments.[133] And even if the risk score is not provided until an initial sentence is chosen, judges who know that subsequent mitigation will be available if it turns out that the defendant is

---

286 (M. Belinda Tucker & Claudia Mitchell-Kernan eds., 1995); Bruce Western et al., *Incarceration and the Bonds Between Parents in Fragile Families*, *in* IMPRISONING AMERICA, *supra*, at 21-45; Elizabeth I. Johnson & Jane Waldfogel, *Children of Incarcerated Parents*, in IMPRISONING AMERICA, *supra*, at 98; James P. Lynch & William J. Sabol, *Effects of Incarceration on Informal Social Control in Communities*, *in* IMPRISONING AMERICA, *supra*, at 135-164.

[128] For a recent, prominent reflection on the way such generalizations about black men have affected African-American communities, *see* Barack Obama, Remarks by the President on Trayvon Martin (July 19, 2013), available at http://www.whitehouse.gov/the-press-office/2013/07/19/remarks-president-trayvon-martin.

[129] *See* William Stuntz, *Race, Class, and Drugs*, 98 COLUM. L. REV. 1795, 1825-30 (1998) (discussing the effects of community perceptions of unfairness on law compliance).

[130] *E.g.*, Etienne, *supra*; J. Richard Couzens, Evidence-Based Practices: Reducing Recidivism to Increase Public Safety: A Cooperative Effort by Courts and Probation 10 (June 27, 2011), available at http://www.courts.ca.gov/documents/EVIDENCE-BASED-PRACTICES-Summary-6-27-11.pdf; Kleiman, *supra*, at 301 (explaining that Virginia's EBS program diverts 25% of nonviolent prison-bound offenders to probation).

[131] Marcus, *supra*, at 751.

[132] Rachel Barkow, *Federalism and the Politics of Sentencing*, 105 COLUM. L. REV. 1276, 1278-81 (2005).

[133] Analogously, limiting instructions to juries—instructions to consider evidence for one purpose but not another—are "notoriously ineffective" and "may be counterproductive because they call jurors' attention to the evidence that is supposed to be ignored." Prescott & Starr, *supra*, at 323 (citing studies).

low risk might err on the side of higher preliminary sentences. Likewise, the risk scores could affect the parties' strategies; in particular, prosecutors might push for longer sentences for higher-risk offenders. Even if the scores are withheld at first from the parties, given that the instruments are quite simple, one would expect the parties to calculate the scores themselves and plan accordingly, and not to wait for the official report.

But let us hypothesize that it could be guaranteed that risk scores would only reduce sentences. Would such an approach be justified? I am loath to resist strategies for reducing unnecessary incarceration. But the key question here is not whether low-risk defendants should be diverted from incarceration—it is whether those low-risk diversion candidates should be identified based on constitutionally problematic demographic and socioeconomic characteristics (instead of past or present criminal conduct or other personal, behavioral assessments).

I conclude that such an approach raises the same problems as does EBS generally. As a constitutional matter, policies that benefit only the lowest-risk offenders may actually be more objectionable because they are less flexible and narrowly tailored—more like quotas than "plus factors." Those with sufficiently unfavorable demographic and socioeconomic characteristics will never qualify as "low risk," no matter how favorable their other characteristics. Consider the Missouri instrument described in Part I. A 20-year-old high school dropout with no job loses six points for those characteristics alone, and can never score higher than 1 on the scale ("average" risk), even if he has no criminal history and no other risk factors and has committed a relatively minor offense. Other instruments that consider gender and a wider variety of socioeconomic and family traits could be even more strongly driven by those factors.[134]

Special exceptions for the privileged cut against the foundational principle that the justice system should treat everyone equally. Moreover, one likely driver of the growth of incarceration is that the relatively privileged majority of the population has been spared its brunt.[135] Those who are primarily incarcerated—poor young men of color—are not politically well represented, and most other citizens have little reason to worry about the growth of incarceration. Progressives should hesitate before endorsing policies that give them another reason not to worry, even if those policies will have the immediate effect of somewhat restraining that growth.

Merely raising the potential policy concerns associated with discrimination and disparity does not necessarily end the argument, just as the constitutional inquiry is not ended by establishing that EBS merits heightened constitutional scrutiny. One must consider how strongly EBS advances competing state interests. In the next Part, then, I turn to the question whether the studies support EBS advocates' optimism.

III. Assessing the Evidence for Evidence-Based Sentencing

Protecting society from crime while avoiding excessive incarceration is no doubt an important interest, even a "compelling" one. But the Constitution and good policy also require

[134] The mitigation-only approach also would not deprive defendants of standing to challenge EBS; a defendant who would have received diversion to probation had the risk instrument not considered his gender, for instance, is harmed by that consideration. The Supreme Court has often considered equal protection challenges in which the plaintiff claims she was denied a government benefit (such as university admission) on the basis of some improper consideration. *E.g.*, Fisher, __ S.Ct. at __.

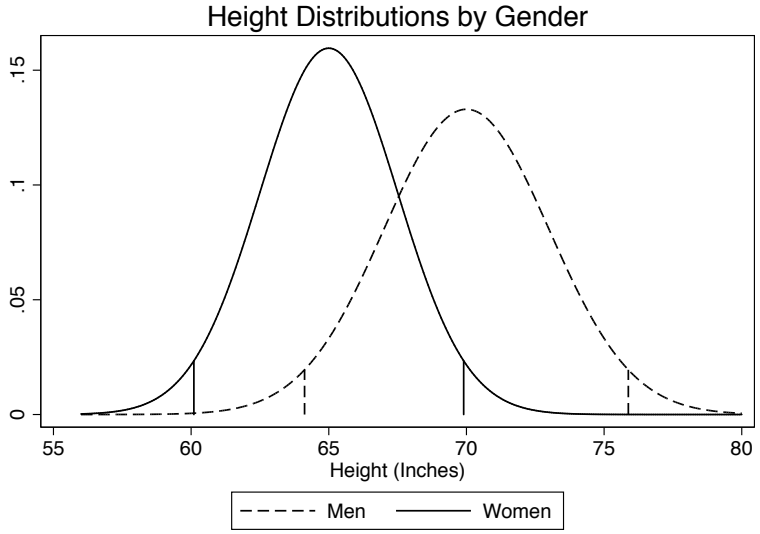
[135] James Forman, Jr., *Why Care About Mass Incarceration?*, 108 MICH. L. REV. 993, 1001 (2010); William J. Stuntz, *Unequal Justice*, 121 HARV. L. REV. 1969, 1974 (2008).

assessing the strength of the relationship between EBS and that interest. When heightened scrutiny applies, it is the state's burden to provide convincing evidence establishing that relationship. In this Part, I show that the current empirical evidence does not suffice.

A. Precision, Group Averages, and Individual Predictions

The instruments' first serious limitation is that they do not provide anything even approaching a precise prediction of an individual's recidivism probability. At best, what they predict with reasonable precision is the *average* recidivism rate for all offenders who share with the defendant whichever characteristics are included as variables in the model. If the model is well specified and based on a sample that is representative of the population to which the results are extrapolated, then it might perform this task well. But that does not necessarily make it particularly useful for individual predictions. Individual vary much more than groups do, and even a relatively precisely estimated model will often not do well at predicting individual outcomes in particular cases.[136] Social scientists sometimes refer to the broader ranges attached to individual predictions as "prediction intervals" (or sometimes as "forecast" uncertainty or "confidence intervals for a forecast") to distinguish them from the "confidence intervals" that are estimated for the group mean or for the effect of a given variable.

To illustrate this point, let's start with an example that involves predicting a continuous outcome rather than a binary future event. To simplify, we will consider only one explanatory variable (sex) and one normally distributed outcome variable (height), which are quite strongly related. The height distributions of the U.S. male and female populations look approximately like Figure 1 below, which is based on average heights of 70 inches for males (standard deviation 3 inches) and 65 inches for females (standard deviation 2.5 inches).



But suppose one did not know the true population distributions, and one had to estimate them by taking a random sample. If one takes a large enough sample, it is easy to obtain quite precise estimates of the average male height and the average female height (as well as the average additional height associated with being male, which is just the difference between the

[136] *See* David J. Cooke & Christine Michie, *Limits of Diagnostic Precision and Predictive Utility in the Individual Case*, 34 LAW & HUM. BEHAV. 259, 259 (2010) ("It is a statistical truism that the mean of a distribution tells us about everyone, yet no one.").

group means). This point is illustrated in Table 1. I created simulated data for a "true population" of men and women that has the height distributions shown in Figure 1. Then I drew from that population random samples with sample sizes 20, 200, and 400, regressed height on gender within each sample, and recorded the predicted mean heights for men and women and the confidence intervals for those means.

Table 1. Precision of Predicted Means versus Individual Forecasts: An Illustration

Sample Size	Male Height in Inches			Female Height in Inches		
	Mean (& Forecast)	95% Conf. Int. for the Mean	95% Pred. Int. for Indiv. Forecast	Mean (& Forecast)	95% Conf. Int. for the Mean	95% Pred. Int. for Indiv. Forecast
20	69.8	[68.2, 71.4]	[64.4, 75.1]	64.8	[63.2, 66.4]	[59.4, 70.1]
200	69.8	[69.3, 70.4]	[64.3, 75.4]	64.6	[64.0, 65.1]	[59.0, 70.1]
400	70.0	[69.6, 70.4]	[64.6, 75.4]	64.9	[64.5, 65.3]	[59.5, 70.3]
Samples are drawn from a simulated "true population" with population means and standard deviations of 70.0 (3.0) for men and 65.0 (2.5) for women.						

Notice that even the smallest sample quite closely approximates the true population means of 70 and 65 inches, while the largest sample comes even closer. Exactly how close each sample comes involves chance (different random samples of the same sizes would have different means), but in general chance plays a smaller role the larger the sample is; as the sample grows the estimates should converge on the true population values. This expectation is captured in the estimation of confidence intervals for the mean, which get narrower as the sample gets larger. Confidence intervals are a way of accounting for chance in sampling. For the 400-person sample, one can express 95% condidence in quite a precise estimate: for males, between 69.6 inches and 70.4 inches, and for females, between 64.5 inches and 65.3 inches.[137] If you keep drawing more and more 400-person samples, they don't tend to differ very much; with that sample size, you will generally do quite a good job approximating the underlying population, which is why the confidence interval is narrow. Meanwhile, the 20-person sample gives you wider 95% confidence intervals, each spanning more than three inches—a much rougher estimate.

But what if you wanted to use your 400-person sample not to estimate the averages for the population, but to predict the height of just the next random woman you meet? Your single best guess—the one that is statistically expected to err by the lowest margin—would be the group mean from your sample, which is 64.9. But you wouldn't be *nearly* as confident in that prediction as you would in the prediction for the group mean. In fact, within that sample, only 13.5% of women have heights that are between 64.5 inches and 65.3 inches, which was your 95% confidence interval for the group mean. If you wanted to give an individual forecast for that next woman that you could be 95% confident in, it would have to be much less precise—you could predict that she would be somewhere between 59.5 inches and 70.3 inches, the 95% prediction interval for the individual forecast that is shown in Table 1. That's a range of nearly

[137] To describe something as a 95% confidence interval for an estimated group mean is to express confidence that 95% of the time, when one draws a random sample and uses the same estimation procedure, the interval one estimates will contain the true group mean for the underlying population.

eleven inches—in other words, you don't know much at all about how tall to expect the next woman to be.[138]

One could make the example much more complicated, with multiple variables and more irregular distributions of outcomes, but the prediction interval for an individual forecast is always wider than the confidence interval for the mean—generally *much* wider.[139] Note that while the confidence intervals for the means gets much narrower as the sample gets larger, the prediction interval does not. The underlying uncertainty that it reflects is not mainly the possibility of having gotten an unusual sample; it's the variability in the underlying population that we saw in Figure 1. One *could* narrow the prediction interval by adding variables to the regression that help to explain this underlying variability—for example, the heights of the individual's parents.

The same basic intuition also applies to models of binary outcomes, like whether a defendant will recidivate—the expected outcome for an individual is much less certain than the expected rate for a group. Some of the recidivism risk prediction instruments include confidence intervals for the probabilities they predict. Indeed, some scholars have urged that confidence intervals should always be provided (rather than mere point estimates) so that judges can get an idea of how precise the instruments are.[140] But given that judges are using the instruments for the purpose of predicting a specific individual's probability of recidivism, providing them a confidence interval for the *group* recidivism rate might even be more misleading than not providing any at all. For instance, if judges are told "The estimated probability that Defendant X will recidivate is 30%, and the 95% confidence interval for that prediction is 25% to 35%," that may well sound to the judge like a reasonably precise individual prediction, but it is not. It is merely a reasonably precise estimate of an average recidivism rate.[141] If the underlying study has a large sample size, such a prediction could be very precise even if the model's variables do not capture much of the variation in individual probabilities at all.

With binary outcomes, though, while the confidence interval for the mean may be misleading, the "prediction interval" is not a very useful alternative way of expressing the precision of an individual forecast, because it does not tell you anything that was not already

[138] Note that the estimated uncertainties in Table 1 are based on a regression of height on gender using standard Stata postestimation prediction commands. By construction, the uncertainties are the same for men and women. Another way to estimate a 95% prediction interval for the height of the next woman you meet would be to just ignore the men and give the range within which the middle 95% of the women in your sample fall. Because female height has a slightly narrower distribution, your interval would then be a bit narrower (about 10 inches), but this method would produce a wider interval for the next male's height (about 12 inches). These ranges are marked on Figure 1.

[139] *See* Cooke & Michie, *supra*, at 271 (illustrating this point using simulated data on violence risk among psychiatric patients, and showing how measurement error for subjective criteria amplifies the uncertainty of individual predictions).

[140] *E.g.*, McGarraugh, *supra*, at 1095-96.

[141] This problem has some similarities to the broader problem of assessing scientific evidence of causation in legal contexts, in which "the law is interested not simply in whether a particular variable causes a particular effect [in general], but, ultimately, in whether a particular variable did cause the effect [in the specific case]." David L. Faigman, *A Preliminary Exploration of the Problem of Reasoning from General Scientific Data to Individualized Legal Decision-Making*, 75 BROOK. L. REV. 1115, 1119 (2010). But this issue is not identical, and my objection here is not that the models cannot establish "individual-level causation," McGarraugh, *supra*, at 1101. The models are predictive, and make no causal claims, so their advocates cannot be accused of confusing correlation with causation. And they aim to predict future probabilistic events, not to prove what caused a particular past event. When one's goal is merely to predict, correlations can be useful, even if the causal pathway is uncertain. For instance, how one voted in the 2012 presidential election is no doubt a very strong predictor of how one will vote in 2016—information campaign strategists can use even if the former does not cause the latter.

made clear by the point estimate itself. Unless the predicted probability is extremely low or extremely high, a 95% confidence interval for an individual prediction will by nature always run from 0 to 1.[142] Recidivism is rarely nearly certain or nearly impossible. So even a good recidivism prediction model could produce prediction intervals of [0,1] for essentially *every* defendant: that is, the only prediction that can be made with 95% confidence about any given individual is that she will either recidivate or not. This fact does not reflect poorly on the design of the prediction instruments or the quality of the underlying research. It reflects the inherent uncertainty of this predictive task and the binary nature of the outcome.

In order to assess how well a model predicts recidivism risk for individuals, some other metric is necessary.[143] There is no single, agreed-upon method for assessing the individual predictive accuracy of a binary model, but there are several possibilities. One common metric used in the recidivism prediction literature is called the "area under the curve" (AUC) approach.[144] This method pairs each person who ended up recidivating with a random person who did not; the score is the fraction of these pairs in which the recidivist had been given the higher predicted risk score. A perfect, omniscient model would rank all eventual recidivists higher than all eventual non-recidivists, and the AUC score would be a 1, while coin flips would on average produce a score of 0.5. The best published scores for recidivism prediction instruments appear to be around 0.75, and these are rich models that include various dynamic risk factors, including detailed psychological assessments, rather than the simple point systems based on objective factors.[145] Many studies have reported AUC scores closer to 0.65.[146] By comparison, a prominent meta-analysis of studies of psychologists' clinical (non-actuarial) predictions of violence found a mean AUC score of 0.73, which the author characterized as a "modest, better than chance level of accuracy."[147] As another point of comparison, if one turns height into a binary variable called "tall" (which denotes being above the median height of the sample), our basic, one-variable model does much better at predicting who will be tall than any

---

[142] *See* R. Karl Hanson & Philip D. Howard, *Individual Confidence Intervals Do Not Inform Decision-Makers about the Accuracy of Risk Assessment Evaluations*, 34 L. & HUM. BEHAV. 275, 276 (2010)

[143] *See* Hanson & Howard, *supra*, at 276. Stephen D. Hart et al., *Precision of Actuarial Risk Assessment Instruments*, 174 BRIT. J. PSYCHIATRY s60 (2007), offer an alternative way of calculating a prediction interval for an individual. They use a traditional method for estimating the confidence interval for a probability prediction given a point estimate for the probability and a sample size, and calculate it for each risk-level category in two common violence prediction instruments, using a sample size of 1. *See* E.B. Wilson, *Probable Inference, The Law of Succession, and Statistical Inference*, 22 J. AM. STAT. ASSOC. 209 (1927). The intervals Hart et al. calculate do not always run from 0 to 1, but they are always very wide, ranging between 79 and 89 percentage points in width. The authors conclude that it is "impossible to make accurate predictions about individuals using these tests."

Hart et al. interpret their intervals as follows: "Given an individual with an ARAI score in this particular category, we can state with 95% certainty that the probability he will recidivate lies between the upper and lower limit.'" This is a slightly odd interpretation, given that, as the authors state, Wilson's confidence intervals are normally interpreted as expressing an interval within which one is confident that the *actual observed rate* for the new sample (not the *ex ante* probability) will fall. The actual observed binary outcome for one individual always must be 0 or 1, however, so I agree with Hanson and Howard, *supra*, that the prediction interval for all but the extreme cases should be 0, 1 (rather than, say, .10 to .94). But either way, it is wide.

[144] *See* Douglas Mossman, *Assessing Predictions of Violence: Being Accurate About Accuracy*, 62 J. CONSULTING & CLINICAL PSYCH. 783 (1994) (describing the method as well as competing approaches).

[145] *See* Mairead Dolan & Michael Doyle, *Violence Risk Prediction*, 177 BRIT. J. PSYCH. 303, 305-07 (2000); AOUSC, *supra*, at 9.

[146] Dolan & Doyle, *supra*, at 305-07.

[147] Mossman, *supra*, at 788.

actuarial model does at predicting who will recidivate—it has an AUC score of 0.825.[148] This is despite the fact that, as we saw, that model gives only rather wide bounds for individual predictions of height—gender is actually quite a strong predictor of height (most men are taller than most women), but it still leaves considerable individual variation unexplained.[149]

Another simple measure of prediction accuracy is the linear correlation between the predicted probabilities and the actual outcomes for offenders; this measure will be 0 if the instrument explains nothing more than chance and 1 if it predicts perfectly.[150] In 1994, a prominent meta-analysis of studies comparing several actuarial recidivism prediction instruments found that the LSI-R (the instrument that the Indiana Supreme Court upheld) had the highest reported correlation with outcomes, at 0.35.[151] By comparison, the gender-only model of the binary "tall" variable has a correlation coefficient of 0.65 (in the same sample used above).

All in all, these metrics suggest that the prediction models do have individual predictive value, but they do not make a resounding case for them. Again, this should not be seen as an indictment of the quality of the science—it is just that even given all the best insights of decades of criminological and psychological research, recidivism remains an extremely difficult outcome to predict at an individual level, much more difficult than height. The models improve considerably on chance, which for some policy purposes (or for the purpose of mental health treatment decisions, which is what many of the models were originally developed for) is no doubt quite valuable. But to justify group-based discrimination in sentencing, both the Constitution and good policy require a much more demanding standard for predictive accuracy. Moreover, note that the accuracy measures discussed here assess the *total* predictive power of each recidivism model, combining all its variables, and are thus overly generous for the purpose of assessing whether particular variables should be included in the model. The marginal predictive power added by just the constitutionally problematic variables is even less, as discussed in the next Section.

The basic difference between individual and group predictions has been pointed out by some scholars in the empirical literature surrounding the risk prediction instruments.[152] But it is lost in much of the EBS legal and policy literature, and more importantly, it may be lost on judges and prosecutors, who may have an inflated understanding of the estimates' precision. Hannah-Moffat explored this issue by interviewing lawyers and probation officers in Canada,

[148] This is estimated in the same 400-person sample used above, pairing each "tall" person with one "short" person, scoring the prediction as correct (i.e., 1) if the tall person was male (i.e., predicted to be taller) and the short person was female, incorrect in the reverse case (0), and as 0.5 if the two had the same gender (i.e., predicted to have the same height), following the standard tie-breaking procedure used to calculate AUC scores. Conversely, if one pairs 200 random women with one random man each (eliminating the possibility of "tied" gender), the man is taller 89% of the time—much better than the chance level of 50%.

[149] Note that a 95% prediction interval for an individual forecast of the binary variable "tall" would run from 0 to 1 for both men and women—one could not be anywhere close to 95% confident that any given woman would be short, or that any given man would be tall. In the sample, 17.5% of women and 82.5% of men were "tall."

[150] The square of this correlation coefficient is one variant on the "fit" measure "pseudo R-squared." This and several other variants could be used to assess a model's ability to explain individual variation, although none should be interpreted as a measure of the overall quality of the model. For a concise summary, see Institute for Digital Research & Education, *FAQ: What are pseudo R-squareds?*, http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm.

[151] Paul Gendreau et al., *A Meta-Analysis of the Predictors of Adult Recidivism: What Works!*, 34 CRIMINOLOGY 575, tbl. 4 (1994).

[152] *See, e.g.*, Hart et al., *supra*; Cooke & Michie, *supra*.

where risk instruments are common. She found that even if caveats about the difference between group and individual predictions are provided, the message often does not get through:

> [F]ew understand and appropriately interpret probability scores. Despite receiving training on these tools and their interpretation, practitioners tended to struggle with the meaning of the risk score….Rather than understanding that an individual who obtains a high risk score *shares characteristics* of an aggregate group of high-risk offenders, the individual is likely to become *known* as a high-risk offender. Instead of being understood as correlations, risk scores are misconstrued in court submissions, pre-sentence reports, and the range of institutional file narratives that ascribe the characteristics of a risk category to the individual.[153]

Advocates of actuarial methods, in this and other contexts, have often sharply criticized the claim that it is not safe to draw conclusions about individuals based on group averages. Mark Cunningham and Thomas Reedy argue that the "distinction between individualized as opposed to group methods is a false dichotomy," contending, essentially, that truly individualized methods do not exist; the discipline of psychology, and its sub-discipline of violence prediction, draws its fundamental scientific character from its willingness to draw insights from data collected on groups and apply them to individuals.[154] Likewise, EBS advocate Richard Redding quotes Paul Meehl, an early pioneer in actuarial prediction in psychology: "If a clinician says 'This case is different' or 'It's not like the ones in your [actuarial] table,'…the obvious question is 'Why should we care whether you think this one is different or whether you are surer?"[155] Jennifer Skeem and John Monahan, quoting Grove and Meehl, argue:

> Our view is that group data can be, and in many cases empirically are, highly informative when making decisions about individual cases….[C]onsider the revolver analogy of Grove and Meehl:
>
> > …Two revolvers are placed on the table, and you are informed that one of them has five live rounds with one empty chamber, the other has five empty chambers and one live cartridge, and you are required to play Russian roulette….Would you seriously think 'Well, it doesn't make any difference what the odds are. Inasmuch as I'm only going to do this once, there is no aggregate involved, so I might as well pick either one of these two revolvers; it doesn't matter which?"[156]

These responses strike me as off base. I do not argue, nor could anybody, that group averages have nothing to do with individual behavior. Of course group averages will *on average* predict outcomes for the individuals in the group—that much is a tautology—and thus provide some information that could guide individual decision-making. But that does not always mean that the group average tells us *much* about what to expect for any given individual. One does not have to be naïve to think that an individual case may be different from the average if it's a situation in which individual outcomes in fact vary widely. The question is how much individual variation there is in a given population, and how much of that variation the variables in the

---

[153] Hannah-Moffat, *supra*, at 12-13.

[154] Mark D. Cunningham & Thomas J. Reidy, *Violence Risk Assessment at Federal Capital Sentencing*, 29 CRIM. JUST. & BEHAV. 512, 517 (2002); *accord* Jessica M. Tanner, *"Continuing Threat" to Whom?: Risk Assessment in Virginia Capital Sentencing Hearings*, 17 CAP. DEF. J. 381, 402-05 (2005).

[155] Redding, *supra*, at 12 n.52 (quoting Paul E. Meehl, CLINICAL VERSUS STATISTICAL PREDICTION (1954)).

[156] Jennifer L. Skeem & John Monahan, *Current Directions in Violence Risk Assessment*, U. Va. School of Law Public Law and Legal Theory Research Paper No. 2011-13, 9-10.

model explain. In the recidivism context (unlike, for instance, the Russian roulette context), the variables included in the instruments leave most of the variation unexplained.[157]

One could defend the instruments on the ground that the precision of individual predictions does not matter from an efficiency perspective. If the group average estimates are good, then the model will, averaged across cases, improve judges' predictions of recidivism, leading more efficient use overall of the state's incarceration resources to prevent crime.

There are two main problems with this response. First, it almost certainly does not suffice for constitutional purposes, at least with respect to any variable triggering heightened scrutiny. The argument amounts to the claim that it doesn't matter whether an instrument has any meaningful predictive power for individuals, so long as the group generalizations have some truth to them. But this is exactly the kind of statistical discrimination defense that the Supreme Court has repeatedly rejected. This point is one reason the Russian roulette analogy is inapt. I would, of course, choose the gun with just one bullet. And if the same dictator forced me to choose between driving on a highway on which 2% of the drivers were drunk and one in which 0.18% of the drivers were drunk, I would choose 0.18% every time. But just that disparity did not suffice, in *Craig v. Boren*, to justify a gender-discriminatory alcohol law. When demographic and socioeconomic characteristics are used to justify the state's serious adverse treatment of individuals, the Constitution requires more than a statistical generalization. Nobody would worry that choosing the gun with one bullet is unfair or harmful to the gun with five. But it is not harmless to base an individual's incarceration on a statistical inference that, based on his poverty or gender, treats him as the human equivalent of a loaded gun.

Second, the "efficient discrimination" argument is not even necessarily correct in terms of efficiency. It is not true that any model with *any* improved predictive power over chance will provide efficiency gains, because EBS isn't replacing chance. If the actuarial instruments don't capture much of the individual variation in recidivism probability, then there is certainly a possibility that the thing EBS is meant to displace—judges' "clinical" prediction of risk—might actually be more efficient because it captures more of that variation. This point is explored further in the next Section.

B. Do the Instruments Outperform Clinical Prediction and Other Alternatives?

The *Bearden* test requires assessment of whether other available and nondiscriminatory (or less discriminatory) alternatives could accomplish the state's penological objectives. Here, I consider two such alternatives: actuarial methods that *do not* rely on constitutionally troubling variables; and judges' exercise of their professional judgment ("clinical" prediction). Even if analysis of alternatives were not constitutionally required, if EBS does not improve at least on the clinical method that it seeks to replace, it does not substantially advance the state's penological interests, and is also undesirable on policy grounds.

EBS advocates have concluded that it is superior to available alternatives, but they have had to stretch the existing evidence quite far to support this claim. J.C. Oleson, for instance, argues that even inclusion of race would be constitutionally permissible, and concludes that it is

[157] In the Russian roulette hypothetical, the decision-maker is given the only variable that matters. The number of bullets quite strongly predicts the individual's probability of dying; it would explain most of the individual variation, with the remaining variation being pure chance. The recidivism models are not in the same ballpark.

"straightforward" to show that no less restrictive means is available.[158] To support this conclusion, he cites just a single study from 1987, by Joan Petersilia and Susan Turner, for the proposition that "omitting race-correlated factors from a model to predict recidivism reduced the accuracy of the model by five to twelve percentage points." Even taking this at face value, it hardly seems obvious that a statistical advantage this modest would justify explicit sentencing discrimination based on race; the Supreme Court has rejected gender discrimination based on stronger statistical evidence than that. And given the Supreme Court's disparate impact jurisprudence, it is odd to justify including race itself based on the predictive power of *race-correlated* factors from the model.

More importantly for present purposes, the Petersilia and Turner study actually suggests that demographic and socioeconomic factors could be excluded from risk prediction instruments without losing any significant predictive value. The "race-correlated factors" in their study *included criminal history and crime characteristics,* which accounted for *all* the additional explanatory value provide by correlates of race (and which no sentencing scheme ignores).[159] Once those factors were already included, adding "demographic" and "other" variables—which included employment, education, marital status, substance abuse, and mental health variables—did not significantly improve the model's predictive power. This is presumably because conduct is generally a better predictor of future conduct than static characteristics are, a point other studies corroborate. For instance, Douglas Mossman's 1994 meta-analysis of studies concerning violence prediction found that "the average accuracy of predictions based on past behavior is higher" than either mental health professionals' clinical judgments or actuarial instruments.[160]

More recent studies of risk prediction instruments have typically not broken down the extent to which adding socioeconomic and demographic variables improves the overall predictive power of the model (a distinct question from the *coefficients* on those variables). But Peterilia's and Turner's results, at least, suggest that a viable alternative is to base actuarial prediction only on crime characteristics and criminal history. Of course, existing sentencing schemes already incorporate those variables, so perhaps providing judges with risk predictions based on them would be redundant. It would be more sensible to have the sentencing commission or legislature incorporate the instruments' insights when determining sentencing ranges. But the fact that an instrument like this might not be terribly useful to judges does not mean that the instruments with the additional variables are *more* useful; the Petersilia and Turner study, at least, suggests that they are not.

Even setting aside the possibility of using *different* actuarial instruments, what about the basic question whether the instruments outperform clinical prediction? It is gospel in the EBS literature that they do. But while scores of studies have found that actuarial prediction methods outperform clinical judgment, this finding is not universal, the average accuracy edge is not drastic, and the vast majority of studies are from wholly different contexts (such as medical diagnosis or business failure prediction). In one widely cited meta-analysis, Grove et al. evaluated all the studies addressing the actuarial versus clinical comparison that were published

---

[158] Oleson, *supra*, at 1386; *see id.* at 1387 (also concluding that "[o]nce the constitutional door is open to race, all other sentencing factors can pass through: gender, age, marital status, education, class, and so forth.").

[159] Petersilia & Turner, *supra*, at 171 (showing, in the table for "All Convicted Defendants," that 57% of outcomes could be accurately predicted by chance, 60% when racially noncorrelated factors were added, 67% when crime characteristics were added, 70% when criminal history variables were added, and still 70% when demographic and "other" variables were added).

[160] Mossman, *supra*, at 789-90.

between 1945 and 1994 and that met certain quality criteria; just five criminal recidivism studies made the cut, plus 131 other studies.[161] Overall, actuarial prediction performed on average about 10% better, but the authors warned: "However, our results qualify overbroad statements in the literature opining that such superiority is completely uniform; it is not. In half of the studies we analyzed, the clinical method is approximately as good as mechanical prediction, and in a few scattered instances, the clinical method was notably more accurate."[162]

If the actuarial advantage does not exist in half of studied contexts, then it is obvious that the specifics matter. And the EBS literature often cites research on far more complicated instruments than the simple ones (like Missouri's, described above) that states actually use. Take, for instance, a study by Grant Harris, Marie Rice, and Catherine Cormier testing an instrument called the Violence Risk Appraisal Guide, which has been cited by EBS advocates.[163] The VRAG consists of twelve variables, the first and most heavily weighted of which is itself a composite of twenty variables: "conning, lying, manipulation, callousness, lack of remorse, proneness to boredom, shallow affect, irresponsibility, impulsivity, poor behavior controls, criminal versatility, juvenile delinquency, sexual promiscuity, and parasitic lifestyle."[164] Assessing these factors requires an elaborate psychological profile, which in the study was carried out by groups of mental health clinicians who "knew the patients well."[165] Nothing like this is typically involved in EBS. Even in the case of sentencing instruments that try to use somewhat nuanced personality characteristics, like the LSI-R, it is not at all obvious that a probation officer filling out a presentence report can carry out a comparable analysis. The VRAG's success simply says nothing about the potential success of a totally different instrument and assessment process. Moreover, the comparability of the populations is also dubious; the VRAG studies involved Canadian psychiatric patients.[166]

Indeed, the past success of instruments that rely on elaborate personality profiles may, if anything, suggest a *disadvantage* of the EBS instruments. The studies show that ideally, after a trained clinician collects all the relevant information and makes the numerous required qualitative assessments, her ultimate predictions will be better informed if she then uses an actuarial model to tell her how much weight to give each factor. This result is unsurprising. But it is a far cry from saying that a different actuarial model that relies on far less overall information (completely ignoring all of the qualitative personality factors) will outperform the judgment of a judge who has had a chance to assess the individual defendant and the complete facts of the case. The relevant comparison, in short, is not just between actuarial versus clinical weighting of variables. It is between actuarial weighting of a few variables versus clinical weighting of a much wider range of variables.[167] It is possible that the actuarial instruments would win that comparison, but we cannot conclude that based on existing research.

---

[161] W.M. Grove et al, *Clinical vs. Mechanical Prediction: A Meta-analysis*, 12 PSYCH. ASSESSMENT 19, 22-24 (2000) (listing studies.

[162] *Id.* at 22-24.

[163] Grant T. Harris et al., *Prospective Replication of the "Violence Risk Appraisal Guide" in Predicting Violent Recidivism Among Forensic Patients*, 26 LAW & HUM. BEHAV. 377 (2002); *see* Wolff, *supra*, at n.73.

[164] Harris et al., *supra*, at 378.

[165] *Id.* at 379.

[166] *Id.* at 381.

[167] Psychologist Stephen Hart states that similar simplified instruments for predicting sexual violence arguably do not deserve even the label "evidence-based" because "scientific and professional literature would not consider [it] informed, guided, or structured since they only include a relatively small set of risk factors." Stephen D. Hart, *Evidence-Based Assessment of Risk for Sexual Violence*, 1 CHAPMAN J. CRIM. JUST. 143, 155, 164 (2009).

A review of each of the five older recidivism studies that Grove et al. included in their meta-analysis likewise does not produce any meaningful support for the modern EBS instruments. Two of the five studies found no discernable advantage for actuarial prediction.[168] Glaeser (1955), one of two studies that found a substantial advantage, involves an archaic prediction instrument in which the most strongly predictive variable was the offender's (clinically assessed) "social development pattern": "Respected Citizen," "Inadequate," "Fairly Conventional," "Ne'er-Do-Well," "Floater," "Dissipated," and "Socially Maladjusted."[169] It also involved very few clinical decisionmakers (four psychiatrists and four sociologists who worked in a parole system in the 1940s), so one possible explanation for the results is that a couple of these people might have not have been terribly good at their jobs.[170] A study by Wormith and Goldstone (1984) evaluates an instrument with more objective criteria and also found that it predicted recidivism better than did the parole board's actual (clinical) decisions. But the study relied on a small Canadian sample that the authors warned "should not be construed as being representative of incarcerated offenders either nationally or internationally."[171] The authors also warned that their measures of clinical and actuarial judgment were not really fairly comparable, in that the "clinical prediction" was not actually a risk prediction at all (instead, it was a binary parole decision), whereas the actuarial prediction was.[172] Finally, a study by Sacks (1974) includes a brief analysis of the clinical versus actuarial comparison, but the comparison it draws is nonsensical (the clinical measure is a parole decision, but only those granted parole are included in the sample) and the purported actuarial advantage is in any case small and not tested for significance.[173]

Nor are more recently published studies more compelling. Oleson et al. (2011) purport to compare the accuracy of clinical and actuarial judgment in federal probation officers' assessment of a probationer's recidivism risk.[174] The study included over a thousand decision-makers (but only one individual's case) and used a modern instrument recently developed by the Administrative Office of the U.S. Courts, called the Federal Post-Conviction Risk Assessment

---

[168] Terrill L. Holland et al., *Comparison and Combination of Statistical and Clinical Predictions of Recidivism Among Adult Offenders*, 68 J. APPLIED PSYCH. 203 (1983) (finding that individual decisionmakers better predict violent recidivism, but actuarial prediction better predicts some measures of overall recidivism); James Smith & Richard I. Lanyon, *Prediction of Juvenile Probation Violators*, 32 J. CONSULTING & CLINICAL PSYCH. 54 (1968) (finding that a juvenile recidivism base expectancy table was slightly more accurate than the predictions of two clinical assessors, but was less accurate than simply predicting that everyone would recidivate would have been).

[169] Daniel Glaser, *The Efficacy of Alternative Approaches to Parole Prediction*, 20 AM. SOC. REV. 283, 285(1955).

[170] *Id.* Problems like this recur in other actuarial versus clinical studies as well—they state a sample size consisting of the number of subjects, and calculate statistical significance as though all of the observations were independent. This approach is misleading because there are usually a far smaller number of clinical decision-makers involved in the study (standard errors should instead be calculated with clustering on the decision-maker).

[171] J. Stephen Wormith & Colin S. Goldstone, *The Clinical and Statistical Prediction of Recidivism*, 11 CRIM. JUST. & BEHAV. 3 (1984).

[172] *Id.* at 20. A general issue with studies that compare real-world "clinical" parole decisions to recidivism risk prediction instruments is that the predictive value of a *prediction* is being compared to that of a *decision*. Wormith et al. explain that it is unsurprising that the parole decision does not predict recidivism as well as an actuarial prediction does, because the parole decision might be affected by factors unrelated to risk prediction, and by the desire to err on the side of caution. *Id.*

[173] Howard R. Sacks, *Promises, Performances, and Principles: An Empirical Study of Parole Decisionmaking in Connecticut*, 9 CONN. L. REV. 347, 402-403 (1977).

[174] J.C. Oleson et al., *Training to See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessments Among Federal Probation Officers*, 75-SEP FED. PROBATION 52 (2011).

(PCRA).[175] The researchers asked officers to watch a video about an individual and predict his risk, and then to redo the exercise after being given the individual's PCRA score and training in the PCRA method. The researchers concluded that the officers were "more accurate" when they had the PCRA.[176] But their only evidence for that claim is that officers' risk scores after being given the PCRA and instructed on its implementation were *more consistent with the PCRA*. That is, in a study purporting to assess whether the PCRA improved prediction accuracy, the researchers assumed the PCRA was perfectly accurate; there was no other measure of what the "accurate" score was.[177]

In sum, the shibboleth that "actuarial prediction outperforms clinical prediction" is—like the actuarial risk predictions themselves—a generalization that is not true in every case. Its accuracy depends on the outcome being evaluated, the actuarial prediction instrument, the clinical predictors' skills, the information on which each is based, and the sample. There is little evidence that the recidivism risk prediction instruments offer any discernable advantage over the status quo, and even if they did, that does not mean particular contested variables need to be included in the model. Alternative models might work as well or better.

C. Do the Risk Prediction Instruments Address the Right Question?

Even if the instruments *could* identify high-risk offenders, does that mean that using them would substantially advance the state's interests? EBS's advocates have typically taken this for granted, but the answer may well be no. The instruments tell us, at best, who is at the highest risk of recidivism. They do not tell us whose risk of recidivism will be the *most reduced* by incarceration. The two questions are not the same, and only the latter directly pertains to the state's penological interests.

At the outset, let's precisely identify the state interest that EBS is designed to serve. Its advocates generally refer either to crime prevention, reduction of incarceration, or both. These can be seen as two sides of the same coin: EBS is meant to help the state balance these interests, which are at least potentially in tension. I agree that this objective is compelling. Crime inflicts great harm on society, and so does excessive incarceration. Striking an appropriate balance between these concerns is an enormous and vital challenge.[178]

But that does not necessarily mean actuarial prediction of recidivism—even if it were perfect—substantially advances that interest. Suppose a judge is considering whether to sentence a defendant to five years in prison versus three. Assuming that the costs of incarceration are the same across defendants,[179] the question is whether the additional two years' incarceration will reduce enough crime to justify those costs. The EBS prediction instruments do not seek to answer that question. Their predictions are not conditional on the sentence. The

[175] This instrument includes qualitative and dynamic factors plus objective factors like age and education. It is in use for planning probation supervision and treatment interventions, not sentencing. Admin. Office of the U.S. Courts, Office of Probation and Pretrial Services, *An Overview of the Federal Post Conviction Assessment* 1 (Sept. 2011).

[176] Oleson et al, *supra*, at 54-55.

[177] The AOUSC's other validation studies for the PCRA did not compare its effectiveness to clinical prediction, and did not find anything close to *perfect* accuracy. AOUSC, *supra*, at 9.

[178] One could frame the state interest as being about the efficient use of finite incarceration resources to maximize crime prevention effects. Unless states have reached their prison capacities and cannot expand, though, I assume that the incarceration rate isn't fixed, so sentencing judges don't think about incarcerating one defendant as trading off with incarceration of another. Instead, they think about whether that particular sentence is worth its costs.

[179] This assumption may not be true. Some defendants have families that are affected, for instance.

samples in the underlying studies include people given all kinds of sentences. They measure recidivism within a particular period, measured from the time of release or (for probationers) from sentencing, but there are no variables relating to the sentence in the regressions. The judge accordingly cannot use the instrument to answer the question "How much crime should I expect this defendant to commit if I incarcerate her for five years?", or three years, or any other potential length. The judge only knows how "risky" she is in the abstract.[180]

This point has been ignored by the EBS literature. Bernard Harcourt, however, makes a similar point about the general deterrence consequences of police profiling and criminal history-based sentencing enhancements.[181] Some have argued that it is efficient for police to focus on groups that commit crimes at greater rates because it concentrates the deterrent effect of policing on the more dangerous groups. Harcourt responds that the fact that members of a particular group commit more crimes on average does not mean that that group is more readily deterred by policing. In fact, high-risk, socially disadvantaged groups may be less willing to cooperate with police, or less deterred by the marginal increase in detection risk, meaning that policing in their communities may actually deter *fewer* crimes than policing in other communities. The relevant question, Harcourt argues, is not rate of crime commission; it is "elasticity" to policing.[182]

Harcourt's argument focuses on general deterrence effects on community crime rates, but a similar problem arises when one considers the effects of marginal changes in incarceration specifically on the defendant's own future crime risk—that is, the very thing that the risk prediction instruments are ostensibly there to help judges minimize. If we are going to base incarceration length on group averages with the objective of reducing crime, then surely the relevant group characteristic is how much incarcerating its members reduces crime—its elasticity to incarceration. And that question is not the same as the question of recidivism probability. There is no particular reason to believe that groups that recidivate at higher rates are also more responsive to incarceration. EBS advocates presumably think that point is intuitive: lock up the people who are the riskiest, and you will be preventing more crimes. But that intuition oversimplifies the relationship between incarceration and recidivism.

Incarceration's effect on an individual's subsequent offending has two components. First, there is an *incapacitation effect*: while behind bars, he cannot commit crimes that he would have committed outside.[183] If the incapacitation effect were the *only* effect that incarceration has on subsequent crime, then it would be logical to assume that the state's incarceration resources are best targeted at the highest-risk offenders. But the situation is not that simple, because of the second component: the effect on the defendant's *post-release* crimes. I will refer to this as the "specific deterrence" effect, but it is really more complicated—it includes on the one hand specific deterrence (fear of reincarceration) plus any rehabilitative effect of prison programming, and on the other hand potentially criminogenic effects of incarceration (interfering with

[180] A related concern is that the length of incarceration may be a confounding variable in the underlying predictive model. If the people who have one set of characteristics tend to get longer sentences than those with other characteristics, then the comparison of their recidivism rates could be apples-to-oranges, because one group's rate is the average after, say, an average of 3 years of incarceration and the other group's rate is the average after 5. We thus don't even know from the models who is the riskiest *today*, much less who is the riskiest X or Y number of years from now.

[181] HARCOURT, *supra* note 12, at 122-36.

[182] *Id.*; Bernard E. Harcourt, *A Reader's Companion to* Against Prediction, 33 LAW & SOCIAL INQUIRY 265, 269 (2008).

[183] This incapacitation effect should be discounted for crime in prison, a complication I will bracket for simplicity.

subsequent employability, building criminal networks, and so forth). There is no intuitive reason to assume that the specific deterrence effect is determined by, or even correlated with, the defendant's recidivism risk level. It is very possible that higher-risk defendants (or some of them, anyway) might be more *inelastic* to specific deterrence and rehabilitation, and might be more vulnerable to the possible criminogenic effects of incarceration. If so, lengthening high-risk offenders' sentences might be more likely to increase the risk they pose after they get out, or at least to lower that risk less than locking up some low-risk offenders might.

If so, this disadvantage has to be weighed against the incapacitation advantage. Implicitly, the current EBS instruments (by ignoring the elasticity question) embrace the premise that only incapacitation matters, but this is not obvious. Most incarceration sentences are fairly short: in 2006, the median prison sentence in state courts was 1.7 years (and that is excluding jail sentences, which are shorter).[184] Moreover, EBS advocates often emphasize its value in determining whether a person should be incarcerated at all, versus probation; presumably, in cases on the incarceration margin, the incarceration sentence being considered is quite short. So, suppose a judge is considering whether to incarcerate a person for one year, versus zero. In that case the potential incapacitation effect lasts a year—a one-year slice of the defendant's offending is taken away. But all the other effects of the judge's choice may last, at least to some degree, the rest of the defendant's lifetime after that year.

There is simply no reason to assume the incapacitation effect is the most important factor, much less the only important factor—and if it is not, then the correspondence between risk prediction and crime-elasticity prediction may well be wholly lost. And this complication arises even if one assumes the relevant state interest only relates to reducing the *defendant's* crime risk. If we also consider effects on *other* individuals' crime commission, there are many more factors to consider, none of which have any intuitive connection to recidivism risk scores: general deterrence, expressive effects on social norms, future crime risk from the defendant's family members, substitution effects in criminal markets, and so forth.

While much of the current EBS literature totally ignores the question of responsiveness of recidivism risk to incarceration, some advocates have taken the general position that incarceration *increases* recidivism risk, citing as evidence simply the fact that persons released from prison recidivate at higher rates than probationers.[185] But this reasoning relies on an apples-to-oranges comparison. It is unsurprising that prisoners recidivate more often than probationers, because prisoners are usually more serious offenders with more prior criminal history. Also, the claim that incarceration generally increases recidivism would make the entire premise of EBS dubious: unless one is considering a life sentence, why identify the most dangerous criminals in order to incarcerate them if incarceration will only make them *more* dangerous? Risk prevention is only a plausible justification for incarceration if the sign on incarceration's effects goes the other way for at least some offenders—and a truly useful risk prediction instrument would try to identify who those offenders are.

---

[184] Bureau of Justice Statistics, *Felony Sentences in State Courts, 2006—Statistics* tbl. 1.3 (2009), http://www.bjs.gov/content/pub/pdf/fssc06st.pdf.

[185] *E.g.*, McGarraugh, *supra* , at 1107; Roger K. Warren, *Evidence-Based Sentencing: The Application of Principles of Evidence-Based Practice to State Sentencing Practice and Policy*, 43 U.S.F. L. REV. 585, 594 (2009); Michael A. Wolff, *Lock 'Em Up and Throw Away the Key? Cutting Recidivism by Analyzing Sentencing Outcomes*, 20 FED. SENT. R. 320, 320 (2008).

Drawing solid causal inferences in this area is difficult. Some studies have used regression or matching methods to compare recidivism rates after controlling for observed characteristics like crime type and criminal history. [186] But while this approach is better than a raw comparison of means, it still does not produce strong causal identification. Causal inference based on regression depends on the assumption that all the important potentially confounding variables have been observed and controlled for. This assumption is often not valid, so one has to be very cautious not to interpret regression results to mean more than they do.

A particular concern arises when the treatment variable of interest (here, incarceration) might itself be influenced by a decision-maker's anticipation of the outcome of interest (here, recidivism). Measuring a statistical association between the two variables provides no way to disentangle which component comes from incarceration causing recidivism, which from anticipated recidivism risk causing incarceration, and which from other confounding variables that affect both sentencing decisions and recidivism outcomes. Regression does not solve the reverse causality problem unless the control variables in the regression account for *all* the reasons that a judge might think a defendant poses a higher risk. As we have seen already, though, even the best recidivism models do not even come close to accounting for all of the sources of individual variation in risk. They surely do not account for all of the sources of variation in judicial anticipation of risk, either—for instance, judges' appraisal of the detailed facts of the case or defendants' courtroom demeanor.

Some recidivism studies have used more rigorous quasi-experimental methods to assess causation, seeking to exploit an exogenous source of variation in incarceration length—that is, a source of variation that is not itself affected by anticipated recidivism risk or by any of the other various factors that affect recidivism risk. [187] Several studies take advantage of the random assignment of judges or public defenders. The intuition is that getting randomly assigned to a particularly harsh judge, or to a less capable public defender, will tend to increase a defendant's sentence in a way unrelated to the defendant's characteristics—thus, while the sentence is not entirely random, it has an effectively random component. Instrumental variables methods are used to estimate the effect of this exogenous increase in sentences on subsequent recidivism. Other studies take advantage of legal reforms that introduce sentencing variation.[188]

These studies have fairly consistently found that increased sentence length on average reduces subsequent offending, although the effect seems to be nonlinear—the marginal effect of increasing sentence lengths declines and eventually disappears as sentence lengths get longer.[189] Thus, specific deterrence lengths on average cut in the same direction as incapacitation effects do.[190] Reported incapacitation effects typically appear larger,[191] but the results of the two types of studies are hard compare. Incapacitation studies generally estimate the number of crimes

---

[186] *See, e.g.*, Oregon Dep't of Corrections, *The Effectiveness of Community-Based Sanctions in Reducing Recidivism* 18-19 (Sept. 2002).

[187] For a useful recent review of this literature, see David A. Abrams, *The Imprisoner's Dilemma: A Cost-Benefit Approach to Incarceration*, 98 IOWA L. REV. 905, 929-36 (2013).

[188] *E.g.*, Shawn D. Bushway & Emily G. Owens, *Framing Punishment: A New Look at Incarceration and Deterrence* (Jan. 2010) (unpublished manuscript), www.human.cornell.edu/pam/people/upload/Framing-Jan-2010.pdf; Ilyana Kuziemko, *Going Off Parole: How the Elimination of Discretionary Prison Release Affects the Social Cost of Crime* 13-22 (Nat'l Bureau of Econ. Research, Working Paper No. 13380, 2007), available at http://www.nber.org/papers/w13380.pdf?

[189] *See* Abrams, *supra*, at 936.

[190] *Id.* at 936-39 (reviewing incapacitation studies)..

[191] *Id.*

avoided during each "person-year" of incarceration,[192] measuring incapacitation's full effect, whereas specific deterrence studies of subsequent recidivism do not estimate the full specific deterrence effect (that is, the change in crime commission over the defendant's whole remaining lifetime). Instead, such studies mostly have quite short follow-up periods, and generally measure not number of crimes committed but recidivism "survival," i.e., whether an offender makes it through the study period without being rearrested or reconvicted, and if not, how long he lasts.[193] Moreover, incapacitation studies sometimes use reported crime as their measure,[194] whereas recidivism studies use the more underinclusive measures of rearrest or reconviction.

Regardless, what the existing research on causal effects has *not* done is to estimate either specific deterrence or incapacitation elasticities that are conditional on the kinds of characteristics that are included in the EBS risk prediction instruments. Instead, the research has focused on estimating the causal relationship between incarceration and crime at a more general level, perhaps subdivided by broad crime category or by deciles of the sentencing-severity distribution, but not by detailed socioeconomic, demographic, and family characteristics. One Urban Institute study, by Avi Bhati, does estimate incapacitation elasticities that are gender, race, and state-specific, but not specific deterrence elasticities, and not broken down by socioeconomic status. It finds no major differences in the total number of crimes averted by either gender or race.[195] Notably, variations by state were far more dramatic, suggesting the need to worry about another problem with the risk prediction instruments: extrapolation from the sample on which they were developed to different offender pools in different jurisdictions. A study by Ilyana Kuziemko on specific deterrence effects finds that incarceration length increases have a "much stronger deterrent effect for older offenders than younger ones, for whom time served actually weakly increases recidivism." That is, young age—one of the most heavily weighted predictors of increased recidivism risk in the current instruments—actually appears to correspond to a lower effectiveness of incarceration length increases in deterring post-release recidivism. This suggests that the EBS instruments are weighing this factor in the wrong direction.

Perhaps future research will improve matters. To effectively inform the state's pursuit of its penological objectives, the research underlying future instruments would have to satisfy the following criteria:

(1) the use of valid causal identification methods, e.g., exploiting random assignment of judges;

(2) application of those methods to obtain estimates for incarceration's effects that are interacted with the variables that the state seeks to include in the instrument;

(3) accounting for nonlinear effects of incarceration length (e.g., the effect of a tenth year of incarceration is probably not the same as the effect of a first);

---

[192] *E.g.*, Rucker Johnson & Steven Raphael, *How Much Crime Reduction Does the Marginal Prisoner Buy?* 28 (Oct. 2010) (unpublished manuscript), http://ist-socrates.berkeley.edu/~ruckerj/johnson_raphael_crimeincarcJLE.pdf.

[193] *E.g.*, Kuziemko, *supra*, at 22.

[194] *E.g.*, Johnson & Raphael, *supra*, at 24

[195] *E.g.*, Avi Bhati, *An Information Theoretic Method for Estimating the Number of Crimes Averted by Incapacitation*, Urban Institute Research Report 24 tbl. 2 (July 2007) (showing estimated male elasticities that were slightly greater in most states, but not in every state and by very small margins). Expressed as a *percentage* reduction in crime rate, rather than an absolute number of crimes averted, females were actually more responsive to incarceration in every state studied. *Id.* at 27 tbl. 4.3.

(4) long enough follow-up periods to allow researchers to meaningfully approximate the change in an individual's lifetime recidivism risk;[196]

(5) incorporation of both incapacitation and specific deterrence effects, with comparable outcome measures;

(6) testing of the instrument within the jurisdiction in which it will be used, on a representative sample; and

(7) evidence of substantial *additional* explanatory power for each constitutionally problematic variable that the state seeks to include.

The current instruments do not do anything like this, and I am not optimistic that this research challenge will be overcome soon. And even it is, the above-discussed problems concerning the uncertainty of *individual* predictions would still apply to the prediction of individual elasticities.

Finally, it might also be objected that it would be unfair to treat an individual's greater expected responsiveness to incarceration as the basis for incarcerating her for longer—offenders might be penalized for *not* being incorrigible. I am sympathetic to this objection. But once sentencing is based on predicting future actions on the basis of demographic and socioeconomic considerations, "fairness" is no longer a decisive sentencing criterion anyway. I do not really advocate it, but at least an elasticity-prediction sentencing instrument would be connected to the state's penological interests. The current instruments are not.

IV. Will Risk Prediction Instruments Really Change Sentencing Practice?

Advocates of EBS sometimes defend it against disparity and retributive justice objects by arguing that it will not really change very much at all. These "defenses" come in two forms. The first is to observe the risk prediction instruments don't directly determine the sentence--they merely provide information to judges. The second defense is that minimization of the defendant's future crime risk already plays an important role in sentencing, so perhaps EBS merely replaces judges' individual judgments of that risk with more accurate actuarial predictions. I address these points in Sections A and B, respectively.

A. Does EBS Merely Provide Information?

One response to disparity concerns is to defend the instruments as innocuous insofar as they only provide information, rather than completely controlling the sentence.[197] The judge can take or leave the information, supplement it with her own clinical assessments of risk, and weigh other, non-recidivism-related factors. As a constitutional defense of EBS, this point could be framed in two ways. The strong form of the argument would assert that the state's adoption of

[196] Collecting data on an offender's entire life is unrealistic, but follow-up periods substantially longer than the typical one or two years are needed. Most people eventually desist from crime, and people who have not recidivated for 7 or 8 years (after release, if they were incarcerated) have quite low subsequent recidivism rates. *E.g.*, Megan C. Kurlychek, Robert Brame, & Shawn D. Bushway, *Scarlet Letters and Recidivism: Does an Old Criminal Record Predict Future Offending?*, 5 CRIMINOLOGY & PUB. POL'Y 483 (2006). Thus, to study the effect of a first year of incarceration (versus none), eight or ten years of outcome data would probably be fine. The study should simply estimate total crime by each individual over a fixed period of time beginning at sentencing, conditional on (among other things) the share of that time that is spent in prison—that measure would incorporate both incapacitation and specific deterrence effects.

[197] *E.g.*, Malenchik v. State, 928 N.E.2d 564, 573 (Ind. 2010); David E. Patton, *Guns, Crime Control, and a Systemic Approach to Federal Sentencing*, 32 CARDOZO L. REV. 1427, 1456 (2011); Kleiman, *supra*, at 301.

the risk prediction instrument does not itself amount to disparate "treatment" at all. Rather, it merely provides social scientific information to a government decision-maker, and surely the Constitution does not require sentencing judges to be ill-informed.

The problem with this framing, however, is that the point of evidence-based sentencing is for the sentence to be *based* on the statistical "evidence," at least in part. The risk score is not calculated for academic purposes. Even if the instrument itself is "only information," the sentencing process that incorporates it is not. Sentencing law already tells judges to consider recidivism risk,[198]and the instrument tells the judge how to calculate that risk. Inescapably, unless judges completely ignore the instruments (rendering them pointless), some defendants will receive longer sentences than they would have but for their group characteristics, such as youth, male gender, or poverty. And that, indeed, is the whole point: if the state did not want unemployed people to be, on average, given longer sentences than otherwise-identical employed people, why put unemployment in the risk prediction instrument? Moreover, arguably even the information provision itself is constitutionally troubling: it represents state endorsement of statistical generalizations like those that, in the gender and poverty contexts, the Supreme Court has condemned.

To be sure, for any individual defendant, each factor included in the risk prediction models is not the *only* determinant of the sentence—it is merely one determinant of the risk score. If a court were looking for ways to distinguish *Bearden*, it could seize on this difference. That case involved revocation of probation, and the Court emphasized that because the trial court had initially chosen probation, it was clear that "the State is seeking here to use as the *sole* justification for imprisonment the poverty of a probationer."[199] This distinction is unpersuasive, however. Anything treated as a sentencing factor will at least sometimes solely trigger a change in the sentence relative to what it would otherwise have been. To give a simple illustration, if a sentence is based on crime severity plus gender, and these factors together produce a 10-year sentence for a male when an otherwise identical woman would have received seven years, male gender is not solely responsible for the sentence; crime severity establishes the baseline of seven years. But male gender is solely responsible for the extra three years.

If this point is slightly more obscured in EBS cases than in *Bearden* itself, it is only because judges won't routinely state what alternative sentence they would have given if the defendant had had different characteristics. In *Bearden* the dispositive role of poverty could not be hidden because of the posture of the case: the defendant had been sentenced to probation and restitution until he failed to pay. But surely if a court's decision-making is unconstitutional in substance, it cannot become constitutional through obscurity of reasoning. In any event, here the use of the discriminatory factor is not obscure, even if its specific consequence for any given defendant is not transparent. A defendant subjected to an unconstitutional decision-making process should be entitled to resentencing.[200] Notably, the Supreme Court has often applied heightened constitutional scrutiny to the mere *consideration* of constitutionally suspect factors. In *Fisher v. University of Texas at Austin*, for instance, the Supreme Court applied strict scrutiny to the use of race as one of many factors in university admissions—indeed, as Justice Ginsburg

[198] *E.g.*, 18 U.S.C. 3553(a).
[199] *Bearden*, 461 U.S. at 671.
[200] *See* Chapman v. California, 386 U.S. 18 (1967).

characterized it in dissent, as a "factor of a factor of a factor of a factor" that very likely was not the reason that the plaintiff in the case was denied admission.[201]

The claim that "it's just information" thus should not enable EBS to *avoid* heightened equal protection scrutiny. A weaker, and more persuasive, version of this claim is that it should make it easier for EBS to *survive* such scrutiny under a "narrow tailoring" requirement. Analogously, in the affirmative action cases, the Court has held that race may be used as a "plus factor" (if there is no race-neutral alternative that will suffice), but it has squarely rejected the use of racial quotas.[202] But the fact that the risk prediction instruments do not completely displace all other sentencing factors is a point in its favor when assessing narrow tailoring, but it is hardly dispositive, as *Fisher* suggests. One must also consider the extent to which they advance the state's interests as well as the availability of alternatives.

Moreover, although *Fisher* made narrow tailoring somewhat challenging to demonstrate even in the affirmative action context, it should be even harder to show in the EBS context. Educational affirmative action involves a state interest that is itself defined in race-conscious terms: student body diversity, of which "racial or ethnic origin" is an "important element," although not the only one.[203] It is more than plausible that considering race as one admissions factor is narrowly tailored to the objective of ensuring racial diversity, and that no totally race-blind alternative will suffice to achieve that objective. In the EBS context, however, the state's penological interests are not defined in group-conscious terms, and the problematic classifications in the instruments are not so closely linked to those interests.

B. Does EBS Merely Replace One Form of Risk Prediction With Another?

Another response to the disparity concern (and to the retributivist objection raised by other critics) is to say that none of this is new: risk prediction is already part of sentencing.[204] If judges are *not* given statistical risk predictions, many will predict risk on their own, perhaps relying implicitly on many of the same factors that the statistical instruments use, such as gender, age, and poverty; actuarial instruments will merely allow them to do so more accurately.[205] One could take this argument further: Conceivably, judges' current clinical assessments could *overweight* some of those variables relative to the weights assigned by the actuarial instruments.[206] These possibilities not been empirically tested and cannot be ruled out.

As a constitutional matter, this "substitution" defense is not very persuasive. It is not likely that courts would uphold an across-the-board state policy explicitly endorsing an otherwise impermissible sentencing criterion on the rationale that the same variables *might* sometimes

---

[201] Fisher v. University of Texas at Austin, 570 U.S.__, ___ (2013) (Ginsburg, J., dissenting).
[202] *Fisher*, 570 U.S. at __.
[203] *Id.* at __.
[204] *See, e.g.*, 18 U.S.C. 3553.
[205] *See, e.g.*, Oleson, *supra*, at 1373; Patton, *supra*, at 1456; Jennifer Skeem, *Risk Technology in Sentencing: Testing the Promises and Perils*, 30 JUSTICE Q. 297 (2013); Bergstrom & Kern, *supra*, at 2; Commentary to Draft MPC § 6B.09; *Michael H. Marcus, MPC--The Root of the Problem: Just Deserts and Risk Assessment*, 61 FLA. L. REV. 751, 757 (2009); Branham, *supra*, at 169.
[206] This is perhaps a particularly realistic possibility with respect to race, because of its absence from the instruments: if judges currently implicitly take race into account in predicting recidivism risk, it is possible that giving them a statistical prediction that is not race-specific could cause them to stop doing so. Thus, even if EBS increases the weight given to socioeconomic variables that are correlated with race, it could reduce the weight given to race itself, offsetting or even reversing its expected effect on racial disparity.

already have been used *sub rosa*. In general, the difficulty of eradicating subtle unconstitutional discrimination does not justify codifying or formally endorsing it.

Moreover, the "substitution" defense depends on a questionable empirical premise. Do the EBS instruments really merely substitute one form of risk prediction for another? Or does providing judges with statistical estimates of recidivism risk increase the salience of recidivism prevention in their decision-making vis-à-vis other punishment objectives? Notably, some EBS advocates affirmatively express the hope that EBS will lead to an expanded emphasis on recidivism prevention.[207] If it does, it will almost surely increase the role of the individual demographic and socioeconomic characteristics used in the EBS instruments. Those characteristics are not relevant to retributive motivations for punishment (or may even cut the other direction).

There are logical reasons to suspect that EBS might increase the emphasis judges place on risk prediction. Most judges no doubt recognize that predicting recidivism risk is difficult, and that difficulty might well lead many of them to discount this factor. If such a judge is presented with a quantified risk assessment framed as scientifically established, they may well give it more weight.[208] In many other legal, policy, and other decision-making contexts, scholars have observed that judges and other decision-makers often defer to scientific models that they do not really understand, and to "expert" viewpoints.[209] Moreover, sentencing is high-stakes, complex decision-making that many judges describe as weighing heavily on their emotions,[210] rendering the use of a simple, seemingly objective algorithm potentially appealing.[211] For elected judges, research has shown that political considerations influence sentences,[212] and reliance on risk predictions might provide political cover for release decisions while making it

[207] *E.g.*, Hyatt, Bergstrom, & Chanenson, *supra*, at 266.

[208] *See* Hannah-Moffat, *supra*, at 7 ("Risk scores impart a moral certainty and legitimacy into the classifications that they produce, 'allowing people to accept them as normative classifications and therefore as scripts for action."); Harcourt, *supra*, at 273 (describing the "pull of prediction").

[209] *E.g.*, Janine Pearson, *Construing* Crane: *Examining How State Courts Have Applied its Lack-of-Control Standard*, 160 U. PA. L. REV. 1527, 1550-53 (2012) (discussing jury overreliance on expert testimony of dangerousness in civil commitment hearings); Michael H. Shapiro, *Updating Constitutional Doctrine: An Extended Response to the Critique of Compulsory Vaccination*, 12 YALE J. HEALTH POL'Y, L. & ETHICS 87, 128-29 (discussing the problem of judicial overreliance on expert claims of causation); Kathryn M. Campbell, *Expert Estimates from 'Social' Scientists*, 16 CAN. CRIM. L. REV. 13 (2011); Robert L. Kane, *Creating Practice Guidelines: The Dangers of Over-Reliance on Expert Judgment*, 23 J.L. MED. & ETHICS 62, 63 (1995); Robert E. Schween & Steven P. Larson, 32 ROCKY MTN. MINERAL L. INST. PROC. 22 (1986) (describing courts' and policymakers tendency to overrely on models and perceived expertise in the environmental context); Case, *Problems in Judicial Review Arising From the Use of Computer Models and Other Quantitative Methodologies in Environmental Decision Making*, 10 B.C. L. REV. 251, 256 (1982) (same).

[210] *See* Oleson, *supra*, at 1330 & n.2 (citing sources); D. Brock Hornby, *Speaking in Sentences*, 14 Green Bag 2d 147, 157 (2011); Judge Robert Pratt, *The Implications of* Padilla v. Kentucky *on Practice in United States District Courts*, 31 ST. LOUIS UNIV. PUBLIC L. REV. 169, 169 (2011) ("Sentencing is unquestionably the most difficult job of any district court judge."); Judge Thomas M. Hardiman, *Foreword*, 49 DUQ. L. REV. 637, 637 (2011) ("Any preconceived notions that a judge may have about sentencing upon taking the bench are quickly dwarfed by the awesome responsibility it entails.").

[211] This point may help to explain the continuing heavy weight federal judges give to the sentencing guidelines that they are not required to follow.

[212] Gregory A. Huber & Sanford C. Gordon, *Accountability and Coercion: Is Justice Blind When It Runs for Office?*, 48 AM. J. POL. SCI. 247, 261 (2004) (finding that judges increase sentences as elections approach).

politically difficult to release offenders rated as high-risk.[213] Prosecutors might similarly feel political pressure to push for harsh sentences for offenders rated high-risk, but free to offer better deals to those rated low-risk.[214]

To be sure, some of the research on clinical versus actuarial prediction has suggested that clinicians may resist reliance on actuarial instruments, but that research comes from medical and mental health diagnosis settings in which the clinician may be much more confident in their professional diagnostic skills than judges are in their ability to foresee a defendant's future.[215] Even if a particular judge does not really trust the instrument, its prediction might influence her thinking through anchoring.[216] And presenting the judge with a risk prediction instrument may simply remind her that risk is a central basis on which the state expects her to base punishment.

All of this is speculative; no empirical research documents how risk prediction instruments affect judges' weighting of recidivism risk versus other factors. To provide some suggestive evidence informing the question, I carried out a small experimental study, with 83 law students as subjects. All subjects were given the same fact patterns describing two criminal defendants and told to recommend a sentence for each. The key experimental variation was that for half the subjects, the descriptions also included a paragraph with the defendant's score on a Recidivism Risk Prediction Instrument (RRPI) and a brief explanation of what the RRPI was.

The cases involved the same conviction (grand larceny of $100,000 worth of jewelry) and the same minimal criminal history (one misdemeanor underage-drinking conviction). Both defendants were male, and no race was mentioned. Beyond that, their characteristics varied sharply. Robert was a middle-aged, married, college-educated executive in a jewelry store chain, and was motivated to steal from the chain's stores by concern about the cost of his daughters' college education. William was a 21-year-old, single, unemployed, alcoholic high school dropout with incarcerated siblings, recently evicted from his parents' home, who was visiting a mall looking for retail work when he saw a jewelry display case open and spontaneously grabbed a bunch of items. These fact patterns allowed some possible distinctions between the defendants' criminal conduct. William's crime was spontaneous, while Robert's involved an extended course of conduct, elaborate deceptive behavior (replacement of the jewels with fakes), and arguably more victims (buyers of the fakes). These distinctions allowed subjects primarily motivated by retribution to have a possible basis for distinguishing the two—likely in William's favor—whereas those inclined to rely on a defendant's characteristics to assess future dangerousness would likely give William a longer sentence.[217] Subjects were given a wide statutory sentencing range (zero to 20 years) and not told what punishment theories to prioritize.

All subjects were given all these underlying facts; the difference was whether they were also translated into an RRPI score. Robert's probability of recidivism was rated "low risk" while William's was "moderate-to-high risk." Although the RRPI is fictional, these ratings realistically approximate the difference that one would see using real instruments. For instance, on the Missouri instrument's -8 to 7 scale, Robert would have a perfect score of 7, while William

---

[213] Hannah-Moffat, *supra*, at 30 ("The use of risk scores can have considerable cache[t] with 'elected' judges and prosecutors who must defend their decisions to an electorate concerned with security.").
[214] *Id.*
[215] *E.g.*, Atul Gawande, THE CHECKLIST MANIFESTO (2009).
[216] *See* Prescott & Starr, *supra*, at 325-30 (reviewing anchoring research); Cass R. Sunstein et. al., *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153, 1170 (2002).
[217] Students' comments after completing the exercise supported this interpretation.

would score -1 ("below average"). Subjects considered the scenarios in a prescribed, randomized order.

The results in Table 2 suggest that the RRPI score sharply affected the relative sentences some subjects gave to Robert and William. Among the 43 students who were not given the RRPI score, 17 gave Robert (the "low-risk" defendant) the higher sentence, 13 gave them the same sentence, and 13 gave William the higher sentence. Among the 40 students who received the RRPI score, only 8 gave Robert the higher sentence, 9 gave them the same sentence, and 23 gave William the higher sentence.

Table 2: An Experiment: Risk Prediction Instruments and Relative Sentence Outcomes

	(1) Robert Higher	(2) William Higher	(3) Which Higher? (William, Same, Robert)	(4) Sentence
	(Probit)	(Probit)	(Ordered Probit)	(OLS)
RRPI	-0.603*	0.710*	0.662**	-0.871
	(0.305)	(0.284)	(0.257)	(0.733)
William				-0.711
				(0.473)
william*RRPI				1.67*
				(0.61)

Cols. 1 & 2 show probit regressions of indicators for giving the "low-risk" or "high-risk" defendant, respectively, a higher sentence. Col. 3 shows an ordered probit regression of a variable valued at 2 if the high-risk defendant's sentence was higher, 1 if they received the same sentence, and 0 if the low-risk defendant's sentence was higher. Col. 4 is an OLS regression with sentence in years as the outcome. An indicator for which case the subjects considered first was also included. Standard errors in parentheses. *p<0.05, **p<0.01

I assessed the size and statistical significance of this shift toward higher sentences for William in several ways, using different definitions of the outcome variable. First, I used probit regressions to estimate the change in the probabilities that Robert would be given a higher sentence (Col. 1) or that William would (Col. 2.). These two are not just mirror-image inquiries, since there is a third option of giving both the same sentence. I next used an ordinal probit regression to assess change in the relative probability of each of these three possible outcomes (Col. 3). Next, I used the recommended sentence, in years, as the outcome variable, an approach that takes into account the magnitude and not just the direction of the sentencing distinctions (Col. 4). The results are statistically significant, and fairly sizeable, in all specifications. The use of the RRPI instrument is associated with an increase in William's sentence, relative to Robert's, of about 1.67 years (that is, 20 months), or about one-third of the overall average sentence (5 years). The average sentence given to William was about 0.8 years higher in the RRPI condition; the average sentence given to Robert was about 0.9 years lower.[218]

A reasonable interpretation of these results is that receiving the RRPI score caused at least some subjects to emphasize recidivism risk more, relative to other sentencing considerations, than they would have otherwise. Moreover, the instrument's apparent effect on sentences was not unidirectional—the instrument's estimated effect on the difference between

[218] Subjects who were given William's case first gave significantly higher sentences to both defendants than those who were given Robert's case first. But order did not significantly affect the *relative* sentences given nor the effect of the RRPI.

the two defendants reflected a combination of an increase in the high-risk defendant's sentence and a reduction in the low-risk defendant's sentence.

These results provide a piece of suggestive evidence that quantified risk assessments might affect the weight placed on different sentencing considerations. However, the study is small, and moreover, although much experimental research on decision-making uses student subjects, one has to be cautious in extrapolating the results of such studies to "real world" settings. A real criminal case is not a four-paragraph vignette, and judges are not law students—their experience and expertise may make them less suggestible. Still, it cannot be assumed that judges are wholly resistant to attempts to influence their sentencing decision-making. After all, judges tend to defer to non-binding sentencing guidelines, and research from other legal settings suggests that courts tend to defer to scientific expertise.[219] While it remains an unsettled question, for now there is no empirical evidence pointing the other way, and little reason to believe that EBS will *merely* substitute one form of risk prediction for another.

CONCLUSION

The inclusion of demographic and socioeconomic variables in risk prediction instruments that are used to shape incarceration sentences is normatively troubling and, at least with respect to gender and socioeconomic variables, very likely unconstitutional. As the EBS movement charges full steam ahead, advocates have minimized the first concern and almost wholly ignored the second. This is a mistake. To be sure, EBS has an understandable appeal to those seeking a politically palatable way to cut back on the United States' sprawling system of mass incarceration. It is difficult to persuade policymakers to reduce incarceration at the cost of increased crime, and EBS offers a technocratic solution to this normative dilemma: just identify the people who can be released without increasing crime. But this identification is not that easy, and moreover, there is no reason to assume, and no good way to ensure, that EBS will only lead to sentences being reduced. Even if it does, there is something troubling, at best, about using group identity and socioeconomic privilege as a basis for reducing defendants' sentences.

Note that while I have focused on sentencing, essentially the same arguments apply to use of actuarial instruments in - decisions, which is now routine in thirty states, including almost all of those that have not abolished discretionary parole.[220] This practice has been given little attention by legal scholars or the public,[221] and has rarely been challenged in court, perhaps because of the absence of counsel in parole proceedings or because parole decision-making is not very transparent. Many prisoners may not even know of the existence of the risk prediction instruments, much less understand how they work or their constitutional infirmities.[222] But while risk prediction unquestionably is properly central to the parole decision,[223] the use of

[219] *See supra* note 209.

[220] HARCOURT, *supra*, at 78-80.

[221] Scholarly criticism has focused on procedural concerns—mainly on the prisoner's lack of counsel at parole hearings. For this reason, the MPC claims to "'domesticate[]' the use of risk assessments by repositioning them in the open forum of the courtroom"—that is, by using them in sentencing instead of in parole (which the MPC seeks to abolish entirely). Draft MPC § 6B.09 cmt. (a). *See also* McGarraugh, *supra* (advocating barring the instruments at parole but using them in sentencing).

[222] In some states, the basis for the parole decision is confidential by law, so the parole board may refuse the prisoner's request to see the risk assessment. McGarraugh, *supra*, at 1079 & n.5.

[223] Indeed, risk is arguably the *only* legitimate parole consideration, because considerations such as retributive justice or general deterrence have already been considered by the sentencing judge. The only reason to leave the

demographic and socioeconomic variables to predict risk raises the same disparate treatment concerns that EBS does.[224] Moreover, the parole context may offer additional available alternatives to the constitutionally objectionable variables. For instance, rather than basing parole decisions on a prisoner's prior education or employment, one could consider his efforts while in prison to improve his future prospects, such as participation in job training or education programs. Such factors would speak to the prisoner's individual efforts to achieve rehabilitation, rather than to his socioeconomic background.

In contrast, it is easier to defend the use of risk prediction instruments in assignment of prisoners, probationers, and parolees to correctional and reentry programming (*e.g.*, job training), and to shape conditions of supervised release (*e.g.*, drug tests).[225] In this context, risk assessment is often combined with instruments assessing "criminological needs" and predicting "responsivity" to various such interventions. The empirical merits of such instruments are beyond this paper's scope, though I note that the responsivity instruments at least address the right question: what can be gained by treating an offender in a certain way? In any event, such uses of actuarial instruments raise less serious constitutional and policy concerns. To be sure, supervision conditions may be burdensome, especially if they affect the likelihood that probation or parole will be revoked, and programming decisions can affect access to valuable services. Still, the stakes are not as high as they are in sentencing, and therefore there is less reason to apply heightened scrutiny to socioeconomic classifications and other traits that are not treated as suspect outside the criminal justice context. Distributing access to correctional programming based on risk, needs, and responsivity assessments is not particularly different from distributing access to non-correctional social services and government benefits to those populations who most need them, which is a routine government function, subject to rational basis review unless suspect or quasi-suspect classifications are involved.

In sentencing, however, the defendant's most fundamental liberties and interests are at stake, as are the interests of families and communities. EBS advocates have not made a persuasive case that this crucial decision should turn on a defendant's gender, poverty, or other group characteristics. The risk prediction instruments offer little meaningful guidance as to each individual's recidivism risk, and they do not even attempt to offer guidance as to the way in which sentencing choices affect that risk. The instruments, and the problematic variables, advance the state's penological interests weakly if at all, and there are alternatives available. Risk prediction is here to stay as part of sentencing, and perhaps actuarial instruments can play a legitimate role. But they should not include these problematic variables, which do not offer much additional predictive value once crime characteristics and criminal history are taken into account. The current instruments simply do not justify the cost of state endorsement of express discrimination in sentencing.

---

sentence indeterminate is to account for the fact that recidivism risk may evolve over time; those who believe risk prediction is an improper basis for punishment should simply oppose indeterminate sentencing. *See, e.g.*, Christopher Slobogin, *Prevention as the Primary Goal of Sentencing: The Modern Case for Indeterminate Sentencing*, 48 SAN DIEGO L. REV. 1127, 1128-30 (2011).

[224] Note that while the Supreme Court once labeled parole an "act of grace," the deprivation of which a prisoner could not contest, this theory is now considered "long-discredited." Samson v. California, 547 U.S. 843, 864 n.5 (2006). States have no obligation to provide a system of parole, but once they do, its operation is constrained by the Constitution. Board of Pardons v. Allen, 482 U.S. 369, 378 (1987); Morrissey v. Brewer, 408 U.S. 471, 482 (1982).

[225] *See* Nat'l Ctr for State Cts, *supra*, at 16-20; Warren, *supra*; Melissa Aubin, *The District of Oregon Reentry Court: An Evidence-Based Model*, 22 Fed. Sent. R. 39 (2009) (discussing evidence-based practices in federal "reentry courts").