EVIDENCE-BASED SENTENCING AND THE SCIENTIFIC RATIONALIZATION OF DISCRIMINATION

Sonja B. Starr*

This Article 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 raises serious constitutional concerns, I comprehensively review the relevant case law, much of which has been ignored by existing literature. To demonstrate that the policy is not justified by countervailing state interests, 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 empirical evidence, based on a randomized experiment using fictional cases, suggesting 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.

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."
>
> —Justice Hugo Black, *Griffin v. Illinois* (1956)[1]

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

---

    1. 351 U.S. 12, 16-17 (1956) (plurality opinion) (footnote omitted) (quoting Chambers v. Florida, 309 U.S. 227, 241 (1940)).

all of these "don'ts." These jurisdictions are directing sentencing judges to explicitly consider a variety of variables that often include socioeconomic status, gender, age, family, and 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. Courts in at least twenty states are already implementing some form of it.[2] One state supreme court has already enthusiastically endorsed it.[3] And it now has been embraced by the American Law Institute in the draft of the newly revised Model Penal Code[4]—a development that reflects its mainstream acceptance and may soon augur much more widespread adoption. There is a similar trend in Canada, the United Kingdom, and other foreign jurisdictions.[5] Meanwhile, the majority of states now similarly direct parole boards to consider demographic and socioeconomic factors.[6] This Article critiques this new trend on constitutional and policy grounds.

The trend is called "evidence-based sentencing" (EBS). "Evidence," in this formulation, refers not to the evidence in the particular case but to empirical research on factors predicting criminal recidivism. Based on that research, EBS provides sentencing judges with risk scores for each defendant based on variables that, in addition to criminal history, often include gender, age, marital status, and socioeconomic factors such as employment and education. EBS has been widely hailed by judges, advocates, and scholars as representing hope for a new age of scientifically guided sentencing.[7] 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, methodological, and policy case against this approach, based on analysis of both the relevant doctrine and the empirical research supporting EBS. I show that several of the variables that many of the instruments use raise serious constitutional and normative concerns, and I review the empirical literature to show that the instruments do not advance state interests sufficiently to overcome those concerns. 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

---

2. *See infra* note 11 and accompanying text (listing examples).

3. *See* Malenchik v. State, 928 N.E.2d 564, 573-75 (Ind. 2010).

4. MODEL PENAL CODE: SENTENCING app. A at 133, 135 (Discussion Draft No. 4, 2012).

5. *See* James Bonta, *Offender Risk Assessment and Sentencing*, 49 CAN. J. CRIMINOLOGY & CRIM. JUST. 519, 519-20 (2007).

6. *See* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 78-80 (2007) (reviewing state parole practices).

7. *See infra* Part I.B (reviewing the literature).

the objective of informing criminal justice policy generally, and sentencing specifically, with data. My objection is specifically to the use of demographic, socioeconomic, family, and neighborhood variables to determine whether and for how long a defendant is incarcerated. I focus principally on the instruments' use in sentencing, but virtually the same case can be made against their use in parole decisions.

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' use of gender and socioeconomic variables, in particular, raises serious constitutional concerns, and yet, surprisingly, no existing scholarship sets forth the constitutional case against this practice. Gender is the only equal protection issue the existing literature pays any attention to, but those who have discussed it have reached the wrong conclusions, as I show here. I also show that the socioeconomic variables raise equally serious concerns under a line of Supreme Court doctrine concerning indigent criminal defendants, and in fact the Court has specifically condemned the notion of treating poverty as a predictor of recidivism risk.

Beyond the constitutional concerns, the use of these and other variables, such as family and neighborhood characteristics, is also troubling on policy grounds. Equal treatment of all persons is a central objective of the criminal justice system, and EBS as currently practiced may have serious social consequences. It can be expected to contribute to the concentration of the criminal justice system's punitive impact among those who already disproportionately bear its brunt, including people of color. And the expressive message of this approach to sentencing is, when stripped of the anodyne scientific language, toxic. Group-based generalizations about dangerousness 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.

To be sure, the state has an important (even compelling) interest in reducing crime without unduly increasing incarceration. But contrary to other commentators, I do not think this interest can justify the use of demographic and socioeconomic variables in EBS. A careful review of the empirical evidence and methods underlying the instruments shows that their use does not substantially advance the state's penological interests and that less discriminatory alternatives would likely perform at least as well. This is true for three major reasons.

First, the instruments provide nothing close to precise predictions of individual recidivism risk. The underlying regression models may provide reasonably precise estimates of the *average* recidivism rates for the group of offenders sharing the defendant's characteristics, but the uncertainty about what an *individual* offender will do is much greater, and when it comes to predicting individual behavior, the models offer fairly modest improvements over chance. While EBS literature sometimes acknowledges this limitation, most advocates have downplayed its seriousness, and existing scholarship has not recognized

its legal import. 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. Rather, individuals are entitled to be treated as individuals.

Second, there is no persuasive evidence that the instruments' predictive power exceeds that of either the current system (in which judges use their individual "clinical" judgment to assess risk) or less discriminatory alternative instruments. A core EBS premise, ubiquitously repeated, is that actuarial risk prediction consistently outperforms "clinical" predictions.[8] I examine the literature on which that claim is based and find it unsupportive of this generalization. Instead, it shows that the specifics of the actuarial instrument matter—and the few comparative studies that specifically involve recidivism prediction have had mixed results and largely involve instruments that do not look much like the ones actually being used in sentencing. The literature also indicates that the constitutionally problematic variables add little marginal predictive power, suggesting the alternative of instruments that do not include those variables

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. Current risk prediction instruments do not even attempt to provide this information. Future research might be able to fill that gap, but it will require careful research design to tease out the causal relationship between sentences and recidivism. There has been some promising recent research along those lines, but so far its findings have not been finely tailored to the characteristics of individual offenders.

Finally, I consider two interrelated counterarguments that defend EBS by essentially 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 sentences might help in a "narrow tailoring" inquiry, but it is not enough alone to establish their constitutionality or 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,

---

8. *E.g.*, MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a at 55 (Tentative Draft No. 2, 2011); *see* sources cited *infra* note 43 (providing additional examples).

usually covertly, on demographic and socioeconomic factors. The instruments, on this view, are merely there to improve this assessment's accuracy.

If it were true that EBS will not actually increase the role played by the problematic variables, then some of the policy arguments against its use would be weakened (although there would still be expressive costs associated with the state's explicit labeling of groups as dangerous, and the constitutional concerns would not dissipate simply because the pre-EBS system also raises them). But this counterargument relies on an empirical premise that is unlikely to be true. Although judges certainly engage in risk prediction now, EBS is likely to expand the role that risk prediction (and thereby the problematic variables) plays in sentencing relative to other sentencing factors such as moral desert. I review literature from psychology and other fields suggesting that providing judges with risk predictions that are framed as scientific and data driven will likely increase the weight placed on them. 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; while all subjects were given the same underlying facts, half were also provided with the defendants' actuarial risk prediction scores. The effects of providing the scores were large and statistically significant: subjects given the scores gave higher sentences to the "riskier" defendant, even though the facts suggested he was less morally culpable, whereas those who did not receive them gave higher sentences to the other defendant. These results are tentative; judges in real cases might well act differently. But the experiment adds to the existing empirical evidence that decisionmaking is affected by quantification and claims of scientific rigor.

In short, this Article seeks to make several contributions to the existing critical literature on EBS and risk prediction: a new constitutional critique, an exploration of the legal consequences of concerns others have raised about individual predictive accuracy, additional new criticisms of the underlying empirical research and of the instruments' relevance to state interests, responses to advocates' counterarguments, and a small experiment that is suggestive of EBS's possible effects. Part I introduces the EBS instruments, describes their rise, and reviews the literature. Part II sets forth the disparity concern. It begins with constitutional analysis, first reviewing the Supreme Court's case law proscribing "statistical discrimination" in the gender context, and then turning to the use of socioeconomic variables in sentencing. It then raises policy concerns about EBS's expressive message and social consequences. Part III then considers whether these constitutional and policy objections can be overcome by countervailing state interests, closely scrutinizing the empirical evidence underlying EBS and considering the availability of alternatives. Part IV considers the counterarguments described above. Finally, I offer some conclusions.

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 a judge's sentencing decision. The instruments are designed to assist judges in their pursuit of several traditional utilitarian sentencing objectives—incapacitation, specific deterrence, and rehabilitation—each of which centers on the reduction of the defendant's future crime risk. In Subpart A, I provide an overview of the instruments' content and trace their recent expansion. In Subpart B, I outline advocates' core arguments for EBS. In Subpart C, I review the limited existing scholarly criticisms of the instruments, including those grounded in retributive-justice perspectives. I also discuss gaps in the current literature that this Article seeks to fill.

A. *The Actuarial Instruments*

Recidivism risk prediction instruments have been developed by criminologists over nearly a century and used for a variety of correctional purposes;[9] their use in parole determinations dates back decades, although it expanded markedly beginning in the 1980s.[10] Virginia was the first state to adopt a risk prediction instrument, doing so in 1994, but this practice has rapidly expanded much more recently. A review of case law, sentencing commission websites, and relevant legislation indicates that at least twenty states' courts are now, in some or all cases, incorporating actuarial risk assessments into the determination of the defendant's sentence.[11] That figure does not include the widespread

---

9. *See* HARCOURT, *supra* note 6, at 1-2, 39-92 (reviewing this history); J.C. Oleson, *Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing*, 64 SMU L. REV. 1329, 1399-402 (2011) (listing variables in eighteen different instruments).

10. HARCOURT, *supra* note 6, at 9, 77-80.

11. At least twelve states have integrated or are currently integrating some form of actuarial risk assessment into sentencing throughout the state pursuant to legislation, state sentencing commission policy, or state supreme court decision. *See Presentence Report*, ARIZ. JUD. BRANCH, http://www.azcourts.gov/apsd/EvidenceBasedPractice/PresentenceReport.aspx (last visited Mar. 26, 2014) (announcing the inclusion of "evidence based criminogenic factors" in presentence reports in response to an order of the Arizona Supreme Court); *Malenchik v. State*, 928 N.E.2d 564, 571-73 (Ind. 2010) (endorsing risk assessment in sentencing); KY. REV. STAT. ANN. § 532.007 (West 2013) (requiring all sentencing judges to consider "the results of a defendant's risk and needs assessment included in the presentence investigation"); *MCOLES—COMPAS at Sentencing—What Every Defense Attorney Needs to Know*, MICH. STATE APP. DEFENDER OFFICE, http://www.sado.org/Calendar/Details/146 (last visited Mar. 26, 2014) (announcing trainings in anticipation of Michigan's decision to begin using risk prediction instruments at sentencing on June 1, 2014); *Offense Risk Score*, MO. SENT'G ADVISORY COMMISSION, http://www.mosac.mo.gov/page.jsp?id=45493 (last visited Mar. 26, 2014) (describing Missouri's state-specific instrument); OHIO REV. CODE ANN. § 5120.114 (LexisNexis 2013) (mandating the adoption of a new unified risk assessment tool for use in sentencing as well as for parole and other corrections purposes); OKLA. STAT. tit. 22, § 988.19 (2013) (making eligibility for diversion to "community sentences" conditional

use of specialized risk assessment instruments in sentencing sex offenders.[12]

---

in most cases on the defendant's score on the Level of Services Inventory (LSI-R)); 42 PA. CONS. STAT. § 2154.7 (2012) (ordering the sentencing commission to develop a risk assessment tool for sentencing use); John Monahan & Jennifer L. Skeem, *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning*, 26 FED. SENT'G REP. 158, 159-60 (2014) (describing Utah sentencing system that requires the incorporation of LSI-R scores into presentence reports and mandates that judges consider them); VA. CRIMINAL SENTENCING COMM'N, 2012 ANNUAL REPORT 6 (2012), *available at* http://www.vcsc.virginia.gov/ 2012VCSCAnnualReport.pdf (describing program that, since 1994, has used "an empirically-based risk assessment to select 25% of the lowest risk, incarceration-bound, drug and property offenders for placement in alternative (non-prison) sanctions"); WASH. REV. CODE § 9.94A.500 (2013) (providing for risk assessment before sentencing if requested by the court and mandating it in all cases involving sex offenses); Press Release, Supreme Court of Appeals, State of W. Va., Court Issues Pre-Sentencing Testing Mandate for All Felons (Jan. 22, 2013), *available at* http://www.courtswv.gov/public-resources/press/releases/2013-releases/jan22_13.pdf (requiring risk assessments to be provided to sentencing judges in every felony case).

    In at least three additional states, local pilot programs for EBS or risk-assessment-based diversion programs exist in particular judicial districts. *See* ADULT REDEPLOY ILL. OVERSIGHT BD., 2012 ANNUAL REPORT TO THE GOVERNOR AND GENERAL ASSEMBLY ON THE IMPLEMENTATION AND IMPACT OF ADULT REDEPLOY ILLINOIS 29-48 (2013), *available at* http://www.icjia.state.il.us/public/redeploy/pdf/annualreports/2012_adult_redeploy_illinois_ annual_report.pdf (describing various incarceration diversion pilot programs in ten Illinois counties, almost all of which use the LSI-R instrument for assignment purposes); BERNALILLO CNTY., NEW MEXICO 2ND JUDICIAL DISTRICT CRIMINAL JUSTICE STRATEGIC PLAN, at ii (2012), *available at* http://www.bernco.gov/upload/images/commission/dist5/ Bernalillo%20County%20Criminal%20Justice%20Strategic%20Plan.pdf   (recommending that Bernalillo County adopt "an automated objective risk assessment tool known as 'COMPAS'" for a variety of purposes, including presentence reports); N.D. DEP'T OF CORR. & REHAB., BIENNIAL REPORT FOR THE PERIOD COVERING JULY 1, 2007-JUNE 30, 2009, at 83-84 (2009), *available at* http://www.nd.gov/docr/media/biennial/archive/BiennialReport%20 2007-09.pdf (describing a pilot program in the Bismarck District that provides "[s]entencing recommendations . . . based on recidivism risk as measured by the LSI-R").

    Finally, in at least five additional states, case law or official reports indicate that risk assessment instruments are, at least in practice, sometimes used in sentencing. *See, e.g.*, State v. Gauthier, 939 A.2d 77, 81, 85-86 (Me. 2007) (upholding a sentence that was based in part on the defendant's high LSI-R risk score); MINN. DEP'T OF CORR., STUDY OF EVIDENCE-BASED PRACTICES IN MINNESOTA: 2011 REPORT TO THE LEGISLATURE 5 (2010), *available at* http://archive.leg.state.mn.us/docs/2013/mandated/130241.pdf (reporting that risk assessment tools are "widely used" by supervision agencies in Minnesota "in making sentencing/disposition recommendations"); N.C. SENTENCING & POLICY ADVISORY COMM'N, RESEARCH FINDINGS AND POLICY RECOMMENDATIONS FROM THE CORRECTIONAL PROGRAM EVALUATIONS, 2000-2008, at 15 (2009), *available at* http://www.nccourts.org/Courts/CRS/ Councils/spac/Documents/correctionalevaluation_0209.pdf ("Sentencing Services utilizes [the LSI-R] to develop sentencing plans for approximately 2,000 felony offenders each year at the request of the court or defense counsel."); Brand v. State, 414 S.W.3d 854, 856 (Tex. App. 2013) (noting the inclusion of defendant's LSI-R score in the presentence investigation report); State v. Duchay, 647 N.W.2d 467, 2002 WL 862458, at *1-2 (Wis. Ct. App. May 7, 2002) (unpublished table decision) (rejecting challenge based on presentence investigation report's reliance on the LSI-R score).

    12. The most common such instrument is the Static-99. *See Static-99/Static-99R*, STATIC99 CLEARINGHOUSE, http://www.static99.org (last visited Mar. 26, 2014) ("It is the

Judge Roger Warren, the president emeritus of the National Center for State Courts (NCSC), 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 and a joint report by the NCSC, the Crime and Justice Institute, and the National Institute of Corrections.[13] Another factor may be the shift toward discretionary sentencing after *Blakely v. Washington*[14] and *United States v. Booker*.[15] Tight sentencing guidelines leave little room to consider the defendant's individual risk, but in discretionary systems, judges are expected to assess it. Whatever the reasons, there is good reason to expect the trend to continue. In fact, Douglas Berman states: "In 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."[16]

The EBS instruments are loosely based on past regression analyses of the relationships between various offender characteristics and recidivism rates. The instruments generally incorporate criminal history variables, such as number of past convictions, past incarceration sentences, and number of violent or drug convictions.[17] Surprisingly, almost none include the crime for which the defendant was convicted in the case at hand.[18] Most of the instruments include demographic variables, such as age, gender, and marital status, as well as socio-economic variables, such as employment status and education.[19] Although risk prediction instruments used by some parole boards included race until as late as the 1970s, modern EBS instruments overwhelmingly do not. One exception is a "sentencing support" software program promoted by an Oregon state judge,

---

most widely used sex offender risk assessment instrument in the world, and is extensively used in the United States, Canada, the United Kingdom, Australia, and many European nations.").

13. *See* Roger K. Warren, *Evidence-Based Sentencing: Are We Up To the Task?*, 23 FED. SENT'G REP. 153, 153 (2010); *see also* CONFERENCE OF CHIEF JUSTICES, RESOLUTION 12 (2007), *available at* http://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/council_reports_and_resolutions/August%202012%20Council%20Open%20Session%20Materials/2012_ccj_resolution_re_doe_recognition.authcheckdam.pdf; ROGER K. WARREN, NAT'L CTR. FOR STATE COURTS, EVIDENCE-BASED PRACTICE TO REDUCE RECIDIVISM: IMPLICATIONS FOR STATE JUDICIARIES (2007), *available at* http://static.nicic.gov/Library/023358.pdf.

14. 542 U.S. 296 (2004) (holding that Washington State's system of sentencing guidelines violated the Sixth Amendment right to trial by jury).

15. 543 U.S. 220 (2005) (holding that the Federal Sentencing Guidelines violated the Sixth Amendment right to trial by jury).

16. Douglas A. Berman, *Are Costs a Unique (and Uniquely Problematic) Kind of Sentencing Data?*, 24 FED. SENT'G REP. 159, 160 (2012) (emphasis added).

17. *See* Oleson, *supra* note 9, at 1399-402.

18. A few include very basic information such as whether it was a drug crime or a violent crime; others include no crime information at all. *Id.*

19. *Id.*

Michael Marcus,[20] but this has not been formally adopted by any state. There appears to be a general consensus that using race would be unconstitutional.[21]

While many instruments now in use are limited to fairly objective factors,[22] others include much more abstract, conceptual variables, which are coded by experienced evaluators. For instance, the Indiana Supreme Court in 2010 upheld against a state-law challenge, and endorsed enthusiastically, the use in sentencing of the Level of Services Inventory-Revised (LSI-R).[23] The LSI-R is the most popular prediction instrument in use among states that have not adopted their own, state-specific instruments, despite the fact that its manual specifically states that it "was never designed to assist in establishing a just penalty."[24] 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 of social involvement with companions, extent of alcohol or drug problems, emotional/psychological status, and personal attitudes.[25]

The family variables included in the LSI-R include not just the defendant's current living situation but also history variables outside the defendant's control; for instance, a defendant will be considered higher risk if his parents had criminal backgrounds.[26] Similarly, Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), another leading instrument, includes variables related to parental convictions, incarceration, and drug use, as well as

---

20. *See* MODEL PENAL CODE: SENTENCING § 6B.09 reporter's note at 62 (Tentative Draft No. 2, 2011) (discussing and criticizing this system); Michael H. Marcus, *Conversations on Evidence-Based Sentencing*, 1 CHAP. J. CRIM. JUST. 61, 107 (2009).

21. *See, e.g.*, MODEL PENAL CODE: SENTENCING § 6B.09 reporter's note at 62. In 2000, the Supreme Court granted certiorari in a capital case to consider "[w]hether a defendant's race or ethnic background may ever be used as an aggravating circumstance in the punishment phase of a capital murder trial in which the State seeks the death penalty"; the issue was not a judicial sentencing instrument but problematic testimony by a prosecution expert. *See* Saldano v. State, 70 S.W.3d 873, 875 (Tex. Crim. App. 2002) (internal quotation marks omitted) (describing the case's history). Before oral argument, the State of Texas conceded error and granted a new sentencing hearing, mooting the case. *See* John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 392-93 (2006).

22. *See* Oleson, *supra* note 9, at 1399-402.

23. Malenchik v. State, 928 N.E.2d 564, 572-74 (Ind. 2010).

24. DEP'T OF CORRECTIVE SERVS., LSI-R TRAINING MANUAL 8 (2002); *see* HARCOURT, *supra* note 6, at 78-84 (describing the LSI-R's uses); *see also supra* note 11 (citing examples of the LSI-R's use in sentencing).

25. *Malenchik*, 928 N.E.2d at 572.

26. The full LSI-R is a proprietary commercial product, but a sample report listing certain risk factors, including "Criminal-Family/Spouse" is publicly available. *See* MULTI-HEALTH SYS. INC., LSI-R PROFILE REPORT FOR REX DARLINGTON (2001), *available at* http://downloads.mhs.com/lsir/lsi-r-5-profile.pdf.

whether family members have been crime victims.[27] The LSI-R and COMPAS also include more complex and subjective socioeconomic inquiries, including such factors as "dependence on social assistance" (LSI-R), high school grades (COMPAS), chances of finding work above minimum wage (COMPAS), and neighborhood crime rates (both).

Once the scores on each question or composite variable are assessed, the calculation of the risk score is mechanical: each possible value of each variable corresponds to a particular increase or reduction in the risk estimate in every case. For instance, if the instrument includes gender, 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.[28] 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, which are based only quite loosely on the underlying regression.[29]

Demographic 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.'"[30] 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."[31] Likewise, a defendant under age twenty-two will score three points worse than a defendant over forty-five.[32] By comparison, having previously served time in prison subtracts one point; having four or

---

27. COMPAS may be requested from Northpointe Inc. *See Northpointe Software Suite*, NORTHPOINTE, http://www.northpointeinc.com/products/northpointe-software-suite (last visited Mar. 26, 2014).

28. 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 FRONTIERS NEUROENDOCRINOLOGY 36, 36 (2008). Recidivism studies typically do not break down gender effects like this, however.

29. The point additions for particular risk factors are at best crudely rounded approximations of regression coefficients. Moreover, the instruments do 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.

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

31. *Id.* at 112-13.

32. *Id.*

more prior misdemeanor convictions that resulted in jail time subtracts one point; having previously had parole or probation revoked subtracts one point; and a prison escape subtracts one point.[33] Meanwhile, current crime type and severity receive no weight.

B. *The Arguments for Evidence-Based Sentencing*

EBS has many enthusiastic advocates in academia,[34] the judiciary and sentencing commissions,[35] and think tanks and advocacy organizations.[36] The NCSC has advocated using risk instruments to guide decisionmaking at all pro-

---

33. *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 lower than one with no criminal history—just two points more than the combined effect of age, employment status, and education level. *Id.*

34. Jordan M. Hyatt et al., *Follow the Evidence: Integrate Risk Assessment into Sentencing*, 23 FED. SENT'G REP. 266, 266 (2011); *see, e.g.*, Lynn S. Branham, *Follow the Leader: The Advisability and Propriety of Considering Cost and Recidivism Data at Sentencing*, 24 FED. SENT'G REP. 169, 169 (2012); Richard E. Redding, *Evidence-Based Sentencing: The* Science *of Sentencing Policy and Practice*, 1 CHAP. J. CRIM. JUST. 1, 1 & n.4, 5-6 (2009) (reviewing articles praising EBS and stating that failure to employ EBS "constitutes sentencing malpractice and professional incompetence").

35. *See, e.g.*, Richard P. Kern & Mark H. Bergstrom, *A View from the Field: Practitioners' Response to Actuarial Sentencing; An "Unsettled" Proposition*, 25 FED. SENT'G REP. 185 (2013) (containing a promotion of EBS written by directors of the Virginia Criminal Sentencing Commission and Pennsylvania Commission on Sentencing); Marcus, *supra* note 20, at 61-63; Warren, *supra* note 13, at 153; Michael A. Wolff, *Evidence-Based Judicial Discretion: Promoting Public Safety Through State Sentencing Reform*, 83 N.Y.U. L. REV. 1389 (2008) (containing a promotion of EBS written by the chair of the Missouri Sentencing Advisory Commission); William Ray Price, Jr., Chief Justice, Supreme Court of Mo., State of the Judiciary Address (Feb. 3, 2010) (transcript available at http://www.courts.mo.gov/page.jsp?id=36875).

36. *See, e.g.*, PAMELA M. CASEY ET AL., NAT'L CTR. FOR STATE COURTS, USING OFFENDER RISK AND NEEDS ASSESSMENT INFORMATION AT SENTENCING: GUIDANCE FOR COURTS FROM A NATIONAL WORKING GROUP (2011), *available at* http://www.ncsc.org/~/media/Files/PDF/Services%20and%20Experts/Areas%20of%20expertise/Sentencing%20Probation/RNA%20Guide%20Final.ashx; NAT'L CTR. FOR STATE COURTS, EVIDENCE-BASED SENTENCING TO IMPROVE PUBLIC SAFETY & REDUCE RECIDIVISM: A MODEL CURRICULUM FOR JUDGES (2009), *available at* http://www.ncsc.org/sitecore/content/microsites/csi/home/~/media/Microsites/Files/CSI/Education/Faculty_Handbook.ashx; ROGER K. WARREN, PEW CTR. ON THE STATES, PUBLIC SAFETY POLICY BRIEF NO. 8, ARMING THE COURTS WITH RESEARCH: 10 EVIDENCE-BASED SENTENCING INITIATIVES TO CONTROL CRIME AND REDUCE COSTS 1-3 (2009), *available at* http://www.colorado.gov/ccjjdir/Resources/Resources/Ref/PEW_ArmingTheCourtWithResearch.pdf; Matthew Kleiman, *Using Evidence-Based Practices in Sentencing Criminal Offenders*, *in* 44 THE BOOK OF THE STATES 299 (Council of State Gov'ts ed., 2012), *available at* http://knowledgecenter.csg.org/drupal/system/files/matthew_kleiman_2012.pdf.

cess stages, including training prosecutors and defense counsel to identify high- and low-risk offenders, thereby shaping plea bargaining decisions.[37] Some academics have offered more cautious takes but have ultimately given qualified endorsements.[38]

The Model Penal Code (MPC), currently undergoing its first revision since its adoption in 1962, has embraced 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 "modernized penal codes in a substantial majority of the states."[39] Section 6B.09 of the revised MPC endorses not only use of "actuarial instruments or processes, supported by current and ongoing recidivism research, that will estimate the relative risks that individual offenders pose to public safety," but also their formal incorporation into presumptive sentencing guidelines.[40] It also provides that when particularly low-risk offenders can be identified, otherwise-mandatory minimum sentences can be waived.[41] While parts of the revision are still being drafted, the American Law Institute has already approved section 6B.09.[42]

The official commentary to the MPC revision illustrates the core argument for EBS:

> 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.[43]

---

37. CASEY ET AL., *supra* note 36, at 23-25.

38. *See, e.g.*, Margareth Etienne, *Legal and Practical Implications of Evidence-Based Sentencing by Judges*, 1 CHAP. J. CRIM. JUST. 43 (2009).

39. Gerard E. Lynch, *Revising the Model Penal Code: Keeping It Real*, 1 OHIO ST. J. CRIM. L. 219, 220 (2003) (observing, in addition, that the MPC's classroom use makes it "the document through which most American lawyers come to understand criminal law").

40. MODEL PENAL CODE: SENTENCING § 6B.09(2) (Tentative Draft No. 2, 2011).

41. *Id.* § 6B.09(3).

42. *See* MODEL PENAL CODE: SENTENCING app. A at 133, 135 (Discussion Draft No. 4, 2012).

43. MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a at 53, 55 (Tentative Draft No. 2, 2011); *see also, e.g.*, Oleson, *supra* note 9, at 1342 & n.84 (emphasizing the superiority of actuarial prediction over human judgment and listing other sources making the same claim); Wolff, *supra* note 35, at 1406 & n.73 (same); Stephen D. Gottfredson & Laura J. Moriarty, *Clinical Versus Actuarial Judgments in Criminal Justice Decisions: Should One Replace the Other?*, FED. PROBATION, Sept. 2006, at 15, 15 ("In virtually all decision-making situations that have been studied, actuarially developed devices outperform human judgments."); Patri-

In short, recidivism risk prediction is inevitably part of sentencing, and rather than being guided by judges' unreliable "clinical" assessments of offenders, it should be guided by the best available scientific research.

Most advocates of EBS frame it as a strategy for reducing incarceration and the resulting budgetary costs and social harms.[44] 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 about risk, judges probably already err on the side of longer sentences.[45] Others propose that the instruments should be categorically limited to being used in mitigation.[46]

In this spirit, the commentary to the MPC revision asserts that "[s]ection 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."[47] 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.[48] The commentary expresses hope that moving risk instruments from parole (which the MPC would abolish) to sentencing will effectively constrain their "incapacitative" use, because access to counsel and greater transparency at sentencing allows the defendant a better chance to argue his case.[49] 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.[50] Moreover, if state legislatures adopt section 6B.09 but

---

cia M. Harris, *What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment*, FED. PROBATION, Sept. 2006, at 8, 9-11 (summarizing literature supporting superiority of actuarial approaches).

   44. *See, e.g.*, CASEY ET AL., *supra* note 36, at 2-3; WARREN, *supra* note 36, at 1; Michael Marcus, *MPC—The Root of the Problem: Just Deserts and Risk Assessment*, 61 FLA. L. REV. 751, 751 (2009); Wolff, *supra* note 35, at 1390; Price, *supra* note 35 (citing EBS as a way to "move from anger-based sentencing" toward reduced incarceration).

   45. *E.g.*, Bonta, *supra* note 5, at 524.

   46. *E.g.*, Etienne, *supra* note 38, at 49-51.

   47. MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a at 54 (Tentative Draft No. 2, 2011).

   48. *Id.* § 6B.09(1)-(2).

   49. *Id.* § 6B.09 cmt. a at 54.

   50. Because the revised MPC 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.* § 6B.09 cmt. e at 57. This constraint, if anything, seems likely to discourage states from including difficult-to-prove dynamic factors such as "antisocial atti-

not the MPC's recommendations concerning abolition of parole, the claim that parole-stage use is worse would be irrelevant.

C. *Scholarly Criticisms*

Although most of the literature on EBS is positive, or even celebratory, a few scholars have criticized it. Many commentators raise criticisms but treat them only as cautionary notes, rather than as dispositive.[51] Others see more fundamental flaws. Here, I briefly review the existing critical literature as well as one of its major gaps: a serious investigation of EBS's constitutionality.

The most thorough critique of the use of risk prediction in criminal justice more broadly has come from Bernard Harcourt in his book *Against Prediction*.[52] 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.[53] Other commentary on EBS has raised similar theoretical objections.[54] John Monahan, while advocating actuarial prediction in other contexts (such as civil commitment[55]), has argued against the current instruments' use in sentencing. 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.[56] Some critics protest the probabilistic nature of risk prediction, ensuring some "false positives" when those deemed high risk do not, in fact, recidivate.[57] Others draw an unfavorable analogy to the science fiction movie *Minority Report*, in which the government punishes "pre-crime," suggesting that even

---

tudes" in the instruments. For factors such as gender, age, and employment, the jury trial requirement seems essentially irrelevant.

51. *E.g.*, Oleson, *supra* note 9, at 1397-98 (concluding simply that EBS "raises excruciatingly difficult questions" and that "[j]udges and jurists must determine" the answers).

52. HARCOURT, *supra* note 6.

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

54. *See* Oleson, *supra* note 9, at 1388-93 (reviewing literature).

55. Monahan, *supra* note 21, at 429. 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 decisionmaking process. *E.g.*, Alexander Scherr, Daubert *& 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.

56. Monahan, *supra* note 21, at 427-28.

57. The commentary to the revised MPC raises, but ultimately is unswayed by, this objection. *See infra* note 71 and accompanying text.

if the future could be known with certainty, punishing people for future acts is fundamentally unfair.[58]

These critiques are not necessarily limited to the use of *actuarial* risk prediction. For example, although Harcourt's book primarily focuses on actuarial risk prediction, he notes that his theoretical objection is also applicable to clinical prediction; he seeks to "mak[e] criminal justice determinations blind to predictions of future dangerousness."[59] The fact that retributive objectives of sentencing may conflict with utilitarian ones is of course not a new dilemma: advocates of purely retributive punishment have always held that a defendant's future risk is morally irrelevant to the state's justification for punishment.[60] EBS may, however, exacerbate the problem from retributivists' perspective if it expands the role that risk prediction plays in sentencing decisions. Moreover, some of the specific factors that heighten a defendant's predicted risk according to the instruments, from young age to mental illness to socioeconomic disadvantage, are frequently considered *mitigating* factors from a retributive perspective.[61]

I do not seek to answer foundational sentencing-philosophy questions here. My critique is not grounded in a purely retributivist perspective on sentencing, and I do not argue that judges should never attempt to predict or to mitigate a defendant's future crime risk. 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 in Part IV that adoption of actuarial instruments will probably increase this role). The Supreme Court has affirmed the relevance of recidivism risk to sentencing, for example by permitting judges to hear expert testimony concerning the defendant's dangerousness.[62] And recidivism-related objectives will continue to be balanced—and sometimes to conflict—with retributive concerns as well as other utilitarian punishment objectives that do not focus on the defendant's future crime risk: general deterrence and expressive efforts to shape social norms.

---

58. *E.g.*, Oleson, *supra* note 9, at 1390; Peter Moskos, Book Review, 113 AM. J. SOC. 1475 (2008).

59. HARCOURT, *supra* note 6, at 5; *see also id.* at 237-38; 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, 258-62 (2008) (arguing that the problem with the instruments is really a broader problem with incapacitation as a punishment objective, including its pursuit via clinical judgment).

60. *See, e.g.*, Paul H. Robinson, Commentary, *Punishing Dangerousness: Cloaking Preventive Detention as Criminal Justice*, 114 HARV. L. REV. 1429, 1429-32 (2001).

61. *See, e.g.*, Graham v. Florida, 560 U.S. 48, 76-77 (2010) (assuming the mitigating role of young age).

62. *See* Barefoot v. Estelle, 463 U.S. 880, 896-99 (1983); *see also* Jurek v. Texas, 428 U.S. 262, 275 (1976) (plurality opinion) ("[P]rediction of future criminal conduct is an essential element in many of the decisions rendered throughout our criminal justice system.").

Instead, this Article's central question is about discrimination and dispari-ty: whether risk prediction instruments that classify defendants by demograph-ic, socioeconomic, and family characteristics can be constitutionally or norma-tively justified. One could, after all, predict risk in other ways—for instance, based only on past or present criminal behavior, or based on individual assess-ment of a defendant's conduct, mental states, and attitudes.

The current literature's treatment of the disparity concern is surprisingly limited; the commentary to the MPC revision, for instance, barely mentions it.[63] 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.[64] He argues that this strategy "will un-questionably aggravate the already intolerable racial imbalance in our prison populations."[65] Kelly Hannah-Moffat has similarly critiqued the criminal histo-ry variables on the grounds of their racially disparate impact and further em-phasized that criminal history may be influenced by past discriminatory decisionmaking.[66] These critiques of EBS's racial impacts have not been framed in constitutional terms.

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—a gap this Article seeks to fill. The existing consti-tutional analyses have focused narrowly on gender (and the hypothetical use of race) and have typically not been fully developed.[67] The most extensive such analysis, by J.C. Oleson, concludes that the instruments survive even strict scrutiny.[68] Similarly, Monahan, while opposing use of demographic variables in sentencing on punishment-theory grounds, defends the constitutionality of

---

63. *See* MODEL PENAL CODE: SENTENCING § 6B.09 reporter's note at 62 (Tentative Draft No. 2, 2011).

64. Bernard E. Harcourt, *Risk as a Proxy for Race*, CRIMINOLOGY & PUB. POL'Y (forthcoming) (manuscript at 2), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1677654.

65. *Id.*

66. Kelly Hannah-Moffat, *Actuarial Sentencing: An "Unsettled" Proposition*, 30 JUST. Q. 270, 279-84 (2013).

67. *See, e.g.*, Pari McGarraugh, Note, *Up or Out: Why "Sufficiently Reliable" Statisti-cal Risk Assessment Is Appropriate at Sentencing and Inappropriate at Parole*, 97 MINN. L. REV. 1079, 1101-02 (2013) (briefly asserting that gender as well as race "must be purged from the list of inputs" but providing little explanation); Christopher Slobogin, *Risk Assess-ment and Risk Management in Juvenile Justice*, CRIM. JUST., Winter 2013, at 10, 13-15 (very briefly addressing possible equal protection objections and concluding that while use of race, ethnicity, or religion is probably impermissible, gender and age are "less open to constitu-tional challenge").

68. Oleson, *supra* note 9, at 1385-88; *see also* Slobogin, *supra* note 67, at 13-14 (sug-gesting that gender discrimination probably survives intermediate scrutiny).

their use in civil commitment. He argues 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.[69] By contrast, I argue below that the use of gender cannot be defended on the statistical bases that other authors have offered, that the constitutional problem goes beyond gender, and that the empirical evidence is not strong enough to sustain the instruments against heightened constitutional scrutiny.

In the criminological literature on the instruments, there is considerable debate over issues of reliability, validity, and precision. Current EBS legal and policy scholarship often notes these concerns but ultimately advocates the use of the instruments anyway.[70] The commentary to the MPC revision 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.[71]

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 as well as terribly unwise. 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 constitutional because the Supreme Court has approved, against a due process challenge, admission of even less reliable clinical predictions of risk provided by experts at sentencing hearings.[72] This assertion is wrong. The equal protection issue is not presented in those cases and, in gen-

---

69. Monahan, *supra* note 21, at 429-32.

70. *See, e.g.*, McGarraugh, *supra* note 67, at 1105-07; Slobogin, *supra* note 67, at 16-17; *see also* Hannah-Moffat, *supra* note 66, at 290-91 ("It is vital to examine the nuances and complexities of risk assessments and concordant calls for and against evidence-based risk jurisprudence in further detail.").

71. MODEL PENAL CODE: SENTENCING § 6B.09 cmt. e at 56-57 (Tentative Draft No. 2, 2011).

72. *E.g.*, Stephen J. Morse, *Mental Disorder and Criminal Law*, 101 J. CRIM. L. & CRIMINOLOGY 885, 944 (2011); Slobogin, *supra* note 67, at 15; *see also* Barefoot v. Estelle, 463 U.S. 880, 896-99 (1983).

eral, is not presented per se by assessments of risk; it is presented by punishment based on 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 governmental interest to withstand heightened scrutiny.[73] 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 reasons to be seriously concerned about this practice. Subpart A outlines constitutional objections. In Subpart B, 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.

A. *Equal Protection*

This Subpart sets forth constitutional objections to the EBS instruments' use of gender and socioeconomic variables. Although it is uncontroversial that gender classifications are subject to heightened scrutiny, I begin in Subpart A.1 by examining the gender case law in some detail because it illuminates a core doctrinal principle that will make it very hard for EBS's constitutionality to be sustained: otherwise-unconstitutional discrimination cannot be justified by statistical generalizations about groups, even if the generalizations have empirical support. In Subpart A.2, I show that this constitutional concern goes beyond gender, examining a line of Supreme Court case law relating to socioeconomic classifications in the criminal justice context.

Note that I frame my constitutional argument within existing doctrine and thus focus the discussion in this Subpart on the use of gender and socioeconomic status. It bears briefly noting that the argument could plausibly be pushed further, however. To be sure, certain variables included in most instruments, like age and marital status, are routine government classifications that are subject to rational basis review. But certain others might well merit new recognition as quasi-suspect—particularly variables relating to an offender's family

---

73. 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.

background or family members' criminal history. These variables are outside the defendant's control, unchangeable, and often the basis for considerable social stigma and disadvantage, which makes them closely analogous to illegitimacy, a quasi-suspect classification.[74] Moreover, there is a defensible broader argument for strict scrutiny of group-based sentencing discrimination, 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 analytical approach, it is presently foreclosed to lower courts by language the Court used in *Chapman v. United States*,[75] and I do not focus on it. The policy

---

74. *See* 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); Miriam J. Aukerman, *The Somewhat Suspect Class: Towards a Constitutional Framework for Evaluating Occupational Restrictions Affecting People with Criminal Records*, 7 J.L. SOC'Y 18, 51 (2005) (reviewing case law and identifying factors that often trigger heightened scrutiny of a classification, including the use of immutable characteristics, associated social disadvantage, and irrelevance to legitimate state objectives); 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); Sandra Stukes Chipungu & Tricia B. Bent-Goodley, *Meeting the Challenges of Contemporary Foster Care*, FUTURE CHILD., Winter 2004, at 75, 85 (describing stigma and educational and social disadvantages experienced by children in foster care); *see also* Mathews v. Lucas, 427 U.S. 495 (1976) (establishing illegitimacy as a quasi-suspect classification).

75. 500 U.S. 453, 464-65 (1991). In *Chapman*, the defendant challenged the Federal 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. *Id.* at 455-56. 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. *Id.* at 464-65.

Chapman's holding is not entirely surprising; the Court is generally quite reluctant to apply constitutional scrutiny to sentences, *see* Carissa Byrne Hessick & F. Andrew Hessick, *Recognizing Constitutional Rights at Sentencing*, 99 CALIF. L. REV. 47, 49 (2011), and it presumably worried that doing so in that case would require the extension of strict scrutiny to virtually every sentencing distinction. However, the Court's reasoning fails to take seriously the tremendous stakes of sentencing choices within statutory ranges. Those ranges are often very broad (say, 0 to 20 years), and it is hard to imagine any government decision that more drastically impacts a defendant's exercise of fundamental liberties than the choice between, say, five and twenty years of 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, the sentencing decision directly interferes with underlying fundamental rights (including the defendant's most basic physical liberty). *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 engage in consensual sodomy").

The *outcome* in *Chapman* is perfectly defensible, but it could have been reached with a different rationale. The drug-weighing rule was a classification of *criminal conduct*, not per-

critique in Subpart B, by contrast, applies more broadly to more variables than the constitutional arguments in this Subpart do.

        1.   *Gender classifications and the problem with statistical discrimination*

    Many of the risk prediction instruments now used for sentencing and parole decisions incorporate gender.[76] Gender classifications require an "exceedingly persuasive justification."[77] In *United States v. Virginia*, the Court elucidated this requirement as follows:

> The burden of justification is demanding and it rests entirely on the State. The State must show at least that the [challenged] classification serves important governmental objectives and that the discriminatory means employed are substantially related to the achievement of those objectives. The justification must be genuine, not hypothesized or invented *post hoc* in response to litigation. And it must not rely on overbroad generalizations about the different talents, capacities, or preferences of males and females.[78]

    Given this well-established doctrine, one might have thought that gender's inclusion in EBS instruments would have occasioned considerable concern and debate. Yet most scholarship ignores this concern or else briefly asserts that the

---

sons, and thus (absent evidence of discriminatory motive) raised no equal protection concerns at all; all persons are prospectively subject to the same weighing rules. Applying fundamental rights analysis to EBS thus would not require routine sentencing distinctions between crimes to be subject to strict scrutiny. One could likewise defend sentencing distinctions based on criminal history as conduct based and universal—all who commit crimes subject themselves to higher sentences for subsequent crimes. But when the state systematically gives different sentences to different groups of people, with no distinctions in current or past criminal conduct, the Constitution *should* demand a compelling justification.

    76. *See, e.g.*, Monahan & Skeem, *supra* note 11, at app. 2 (listing the factors in the Pennsylvania statewide instrument); *see also* EDWARD LATESSA ET AL., CREATION AND VALIDATION OF THE OHIO RISK ASSESSMENT SYSTEM: FINAL REPORT 26 (2009), *available at* http://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/ORAS_Final_Report.pdf (describing Ohio's statewide system, which uses different risk category cutoffs for men and women, though its various risk scales vary in terms of which cutoff is higher); BRIAN J. OSTROM ET AL., NAT'L CTR. FOR STATE COURTS, OFFENDER RISK ASSESSMENT IN VIRGINIA 5, 123 (2002), *available at* http://www.vcsc.virginia.gov/risk_off_rpt.pdf (describing the role of gender in Virginia's instrument and acknowledging that young male offenders have a very difficult time qualifying for diversion from incarceration); WASH. STATE DEP'T OF CORR., STATIC RISK ASSESSMENT (2007), *available at* http://www.ofm.wa.gov/sgc/meetings/2008/06/SGCmeeting_20080613_StaticRiskAssessment.pdf (describing factors in Washington's instrument). The COMPAS software suite includes a completely separate version designed for female offenders. *See Women*, NORTHPOINTE, http://www.northpointeinc.com/solutions/women (last visited Mar. 26, 2014).

    77. United States v. Virginia, 518 U.S. 515, 531 (1996) (internal quotation marks omitted).

    78. *Id.* at 533 (alteration in original) (citations omitted) (internal quotation marks omitted).

state's interests are important.[79] The revised MPC recommends excluding race, and the commentary notes that sentencing based on race would be unconstitutional.[80] But the MPC drafters recommend including gender and offer no commentary defending this on constitutional grounds,[81] as though its constitutionality is self-evident.

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.[82] There is, to be sure, considerable statistical research suggesting that judges (and prosecutors) *do* on average treat female defendants more leniently than male defendants.[83] But it is virtually unheard of for modern judges to *say* they are taking gender into account,[84] and demonstrating gender bias would usually be challenging. Until the past few decades, explicit consideration of gender as well as race was common, but few today defend that practice.[85] The Federal Sentencing Guidelines, for example, expressly forbid the consideration of both race and sex.[86] Outside the literature on EBS, scholars have likewise mostly treated the gender gap as an "unwarranted" sentencing disparity.[87]

---

79. *See, e.g.*, Slobogin, *supra* note 67, at 14. McGarraugh states that gender should be removed from the instruments to preserve their constitutionality but does not develop the legal reasoning for this point. McGarraugh, *supra* note 67, at 1102.

80. *See* MODEL PENAL CODE: SENTENCING § 6B.09 reporter's note at 62 (Tentative Draft No. 2, 2011).

81. *See id.*

82. *See, e.g.*, United States v. Maples, 501 F.2d 985, 987 (4th Cir. 1974); Williams v. Currie, 103 F. Supp. 2d 858, 868 (M.D.N.C. 2000); Carissa Byrne Hessick, *Race and Gender as Explicit Sentencing Factors*, 14 J. GENDER RACE & JUST. 127, 137 & n.68 (2010).

83. *See, e.g.*, Sonja B. Starr, *Estimating Gender Disparities in Federal Criminal Cases* 3-4, 17 (Univ. of Mich. Law Sch. Law & Econ. Research Paper Series, Paper No. 12-018, 2012), *available at* http://ssrn.com/abstract=2144002 (finding large gender gaps at multiple procedural stages that are unexplained by observable variables and also reviewing other studies).

84. *See* Hessick, *supra* note 82, at 127-28 (observing that "modern sentencing systems do not permit the explicit consideration of race or gender," although consideration of both was once common).

85. *See id.* at 129-37.

86. U.S. SENTENCING GUIDELINES MANUAL § 5H1.10 (2012).

87. *See, e.g.*, Oren Gazal-Ayal, *Foreword: A Global Perspective on Sentencing Reforms*, 76 LAW & CONTEMP. PROBS., at i, iii-iv (2013); Mona Lynch, *Expanding the Empirical Picture of Federal Sentencing: An Invitation*, 23 FED. SENT'G REP. 313, 314 (2011). Some scholars criticize increasing female incarceration rates, but they 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. *See, e.g.*, Leslie Acoca & Myrna S. Raeder, *Severing Family Ties: The Plight of Nonviolent Female Offenders and Their Children*, 11 STAN. L. & POL'Y REV. 133, 135, 141 (1999); Phyllis Goldfarb, *Counting the Drug War's Female Casualties*, 6 J. GENDER RACE & JUST. 277, 291-93 (2002).

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.[88] Some scholars add that to fail to account for this fact is unfair to women, essentially punishing them for men's recidivism risk.[89] More generally (referring to "gender, ethnicity, age, and disability"), Judge Michael Marcus states: "We are not treating like offenders alike if we insist on ignoring factors that make them quite unalike in risk."[90]

But this argument, which embraces a concept of "actuarial fairness,"[91] stands on unsound constitutional footing. It offers what *United States v. Virginia* specifically stated could never amount to an "exceedingly persuasive justification" for a gender classification—namely, "overbroad generalizations" about the tendencies of males and females.[92] 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.[93] 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%).[94] The Court noted that "prior cases ha[d] 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."[95] That is, what is prohibited is not *just* "outdated misconceptions"[96] and merely "hypothesized" gender differences.[97] What is prohibited is

---

88. *See, e.g.*, Monahan, *supra* note 21, at 431.

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

90. Marcus, *supra* note 44, at 769. Similarly, some criminologists have criticized the LSI-R, the most notable example of a prediction instrument that does *not* include gender, for failing to accurately predict women's risk and have advocated the development of separate "gender-calibrated" instruments that give different weights to a variety of variables. *See* Kristy Holtfreter & Rhonda Cupp, *Gender and Risk Assessment: The Empirical Status of the LSI-R for Women*, 23 J. CONTEMP. CRIM. JUST. 363, 363 (2007). *But see* Paula Smith et al., *Can 14,737 Women Be Wrong? A Meta-Analysis of the LSI-R and Recidivism for Female Offenders*, 8 CRIMINOLOGY & PUB. POL'Y 183, 183 (2009) (finding that the LSI-R is just as accurate for women as for men).

91. 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, 1597-600 (2011).

92. *See* 518 U.S. 515, 533-34 (1996).

93. 429 U.S. 190, 191-92 (1976).

94. *Id.* at 200 & n.8, 201.

95. *Id.* at 202.

96. J.E.B. v. Alabama *ex rel.* T.B., 511 U.S. 127, 135 (1994).

inferring an individual tendency from group statistics. Note that the government's argument in *Craig* could easily have been framed in "actuarial fairness" terms: it arguably would have been unfair to bar young women from drinking based on a risk of drunk driving that came almost entirely from males. But the Court's approach to equal protection means that individuals are neither entitled to favorable statistical generalizations based on gender nor subject to unfavorable ones. *Craig* illustrates that the proscription on gender-based statistical generalizations applies even when the generalization favors women and even when the state has a weighty public-safety interest at stake, such as the prevention of drunk driving.

Examples of this principle abound. For instance, the Court has repeatedly held that the 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.[98] 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.[99] Likewise, in *J.E.B. v. Alabama ex rel. T.B.*, the Court struck down gender-based peremptory challenges in jury selection, notwithstanding a "plethora of studies" showing that gender is predictive of juror voting patterns in sexual assault cases.[100] The Court held that it is impermissible to make assumptions about individual jurors based on gender, "even when some statistical support can be conjured up."[101] To do so would have been to accept, as a justification for discrimination, "the very stereotype the law condemns."[102] 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."[103] 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 adversarial method."[104] 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 Commonwealth's great goal [of educating soldiers] is not substantially advanced by women's categorical exclusion, in total disregard of

---

97. *Virginia*, 518 U.S. at 533; *see also* Monahan, *supra* note 21, at 429-32 (defending gender-based risk prediction for civil commitment).

98. *See* Weinberger v. Wiesenfeld, 420 U.S. 636, 645 (1975); Frontiero v. Richardson, 411 U.S. 677, 690-91 (1973).

99. *Wiesenfeld*, 420 U.S. at 645.

100. 511 U.S. at 148-49 (O'Connor, J., concurring).

101. *Id*. at 139 n.11 (majority opinion).

102. *Id*. at 138 (quoting Powers v. Ohio, 499 U.S. 400, 410 (1991)) (internal quotation marks omitted).

103. 518 U.S. 515, 541 (1996) (internal quotation marks omitted).

104. *Id*. at 541-42 (emphasis added).

their individual merit, from the Commonwealth's premier 'citizen-soldier' corps."[105]

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 decisionmaker 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: 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.[106] 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 socially entrenched inequalities.[107] From this perspective, perhaps statistical discrimination should be prohibited when it is of a sort that tends to entrench subordination and not otherwise.[108] I am generally sympathetic to the antisubordination approach, in fact, but I frame this Article within the approach that dominates current doctrine: anticlassification. In any event, an antisubordination approach to equal protection law would

---

105. *Id.* at 541, 545-46. In *City of Los Angeles Department of Water & Power v. Manhart*, 435 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:

> 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 Broad., Inc. v. FCC, 497 U.S. 547, 620 (1990) (O'Connor, J., dissenting) (citing this passage to inform the application of the Equal Protection Clause).

106. *See* Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 Harv. L. Rev. 493, 553 (2003).

107. *See* Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. Miami L. Rev. 9, 9-10 (2003) (reviewing antisubordinationist scholarship); Primus, *supra* note 106, at 557.

108. *See* Samuel R. Bagenstos, *"Rational Discrimination," Accommodation, and the Politics of (Disability) Civil Rights*, 89 Va. L. Rev. 825, 855-59 (2003).

hardly be friendlier to EBS, a practice that amplifies the inequality of the criminal justice system's impact by inflicting additional criminal punishment on the poor and, via disparate impact, on people of color. In Subpart B, I further explore 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 its 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—even if they are otherwise persuaded that the state's interests are strong and that the policy is narrowly tailored and substantially advances these interests (issues to which I return in Part III). The Court's language in *United States v. Virginia* suggests that the bar on "overbroad generalizations" is an *additional* constitutional restriction in the gender context; such generalizations cannot count as "exceedingly persuasive justifications," because they contravene a distinct and fundamental constitutional value. Notably, as David Strauss has pointed out when discussing the same principle's application in the race context, the Constitution often forbids the use of generalizations even when their avoidance has "innocent victims."[109] Indeed, it is always the case that when courts strike down the use of "efficient" group generalizations as unconstitutional, "additional costs will have to fall on someone."[110] The claim of the commentary to the MPC revision that even statistical instruments that get a large number of cases wrong should be used if doing so has the chance of preventing "victimizations in the community"[111] implicitly embraces a type of utilitarian calculus that this line of case law rejects.

To avoid being characterized as an "overbroad generalization," the statistical relationship would at the very least have to be so strong that courts could deem the resulting *individual* predictions about recidivism noticeably more sound than the generalizations the Supreme Court has rejected in the past. 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.[112] The Court picked up on this possibility in *Nguyen v. INS*, upholding a law that gave preferential treatment to children of U.S. citizen mothers (but not fathers) born abroad out of wedlock.[113] Its rationale turned on mothers'

---

109. David A. Strauss, *The Myth of Colorblindness*, 1986 SUP. CT. REV. 99, 109-11.

110. *Id.* at 110.

111. MODEL PENAL CODE: SENTENCING § 6B.09 cmt. e at 56-57 (Tentative Draft No. 2, 2011).

112. 518 U.S. 515, 533 (1996).

113. 533 U.S. 53, 73 (2001).

unique capacity to give birth to children—a "[p]hysical difference[]" that the Court differentiated from stereotyping.[114] Whatever the merits of *Nguyen*'s reasoning (which is questionable), a generalization about a behavioral tendency like criminal recidivism is simply not comparable to a physical difference, especially since—as we shall see in Part III—the EBS instruments' predictions do not tell us much about what to expect for any given individual.

To be sure, the government is not categorically forbidden from relying on statistical generalizations of *any* sort; 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. The government sometimes has to draw clear lines that may overgeneralize; for instance, it sets a maximum blood-alcohol content for driving rather than requiring that each individual's fitness to drive while intoxicated be individually assessed. Frederick Schauer has made this point forcefully, offering a fairly broad defense of reliance on statistically supported generalizations.[115] But as Schauer emphasizes, this practice properly has limits; certain kinds of generalizations (including those concerning gender) are particularly socially harmful or expressively invidious, even if they have statistical support.[116] The Supreme Court's doctrine on statistical discrimination reflects a similar view, tolerating statistical generalizations in some contexts but not as a defense of classifications triggering heightened constitutional scrutiny.[117]

---

114. *Id.* at 68 (quoting *Virginia*, 518 U.S. at 533); *see also id.* at 62-66 (elaborating on the relevance of this physical difference).

115. *See* FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES (2003).

116. *Id.* at 38-41. 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 Federal 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." *Id.* at 261-62. Although I share the concern that group-based sentencing distinctions are socially divisive, I do not think the "all in this together" rationale can justify mandatory sentencing laws, which tend to merely shift the power to individualize toward prosecutors (a possibility Schauer acknowledges, *id.* at 256) and which are often defined too broadly to capture real differences in criminal conduct and culpability. 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, and constitutionally suspect, in a context that has huge consequences for individuals and communities.

117. *See Nguyen*, 533 U.S. at 76 (O'Connor, J., dissenting); *see also Virginia*, 518 U.S. at 533.

2. *Wealth-related classifications in the criminal justice system*

The constitutional problem with EBS goes beyond gender. Current doctrine also calls into serious question the variables related to socioeconomic status, such as employment status, education, income, dependence on government assistance, and job skills. The Supreme Court's case law in *other* contexts has consistently held that similar wealth-related classifications are not constitutionally suspect,[118] 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 in the context of punishment. These defendants have often succeeded, and the Supreme Court and lower courts have applied a demanding form of scrutiny in these cases, 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 disparities in criminal justice; in *Griffin v. Illinois*, for instance, the plurality called this objective "the central aim of our entire judicial system."[119] *Griffin* struck down the requirement that defendants pay court costs before receiving a trial transcript, which was needed 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."[120]

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, under "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."[121] Notably, these cases have applied a demanding form of 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

---

118. *See, e.g.*, Maher v. Roe, 432 U.S. 464, 471 (1977) ("[T]his Court has never held that financial need alone identifies a suspect class for purposes of equal protection analysis.").

119. 351 U.S. 12, 17 (1956) (plurality opinion).

120. *Id.* at 17, 19; *accord* Mayer v. City of Chi., 404 U.S. 189, 193-97 (1971).

121. 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").

need not provide an appeal as of right at all. Rather, *Griffin* and its progeny involved a special "equality principle"[122] motivated by "the evil [of] . . . 'discrimination against the indigent.'"[123] For this reason, a challenge to EBS need not establish that the defendant has some freestanding constitutional entitlement to a lower sentence than he received.

For our purposes, the most pertinent Supreme Court case is *Bearden v. Georgia*, which assessed the constitutionality of a trial court's decision to revoke the probation of an indigent defendant who had been unable to pay his court-ordered fine and restitution.[124] The Court unanimously reversed, holding that incarcerating a defendant merely because he was unable to pay amounted to unconstitutional wealth-based discrimination.[125] Importantly, the Court in *Bearden* squarely rejected Georgia'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.[126]

The Court's resistance to "lumping [the defendant] together with other poor persons"[127] is very similar to its reasoning concerning statistical discrimination in the gender cases. The Court 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.[128]

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 probation revocation. The Court emphasized one reason it may be permissible to consider ability

---

122. Lewis v. Casey, 518 U.S. 343, 369-70, 371 & n.2, 372 (1996) (Thomas, J., concurring) (quoting United States v. MacCollom, 426 U.S. 317, 331 (1976) (Brennan, J., dissenting)) (internal quotation mark omitted).

123. *Id*. at 372; *see also* Douglas v. California, 372 U.S. 353, 355 (1963).

124. 461 U.S. 660, 663 (1983).

125. *Id.* at 661-62. *Bearden* built on *Williams v. Illinois*, 399 U.S. 235 (1970), in which the Court had similarly held unconstitutional the imprisonment of an indigent criminal defendant for failure to pay a fine. 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.

126. 461 U.S. at 671 (footnote omitted).

127. *Id.*

128. *Id.* at 671 n.11.

to pay (and related factors such as employment history) when choosing be-
tween incarceration and restitution sentences: "The State, of course, has a fun-
damental interest in appropriately punishing persons—rich and poor—who vio-
late 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."[129]

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.[130] But with EBS, poverty is being consid-
ered not to *enable equal punishment* of rich and poor but to trigger extra, une-
qual punishment. The *Bearden* 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 reduction of the fine. "Only if the sentencing court determines that al-
ternatives to imprisonment are not adequate in a particular situation to meet the
State's interest in punishment and deterrence may the State imprison a proba-
tioner who has made sufficient bona fide efforts to pay."[131]

This requirement that less restrictive alternatives be considered is a hall-
mark 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 wheth-
er the key constitutional provision was really the Equal Protection Clause at all,
as opposed to the Due Process Clause. As the Court explained, these constitu-
tional concerns are intertwined in these cases, and in any event,

> [w]hether analyzed in terms of equal protection or due process, the issue can-
> not be resolved by resort to easy slogans or pigeonhole analysis, but rather re-
> quires a careful inquiry into such factors as "the nature of the individual inter-
> est affected, the extent to which it is affected, the rationality of the connection

---

129. *Id.* at 669-70.

130. *See also Williams*, 399 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). Indeed, several cir-
cuits have held that a defendant is constitutionally entitled to a judicial inquiry into her abil-
ity to pay a fine. *See, e.g.*, Powers v. Hamilton Cnty. Pub. Defender Comm'n, 501 F.3d 592,
608 (6th Cir. 2007) (citing Alkire v. Irving, 330 F.3d 802, 816-17 (6th Cir. 2003)).

131. *Bearden*, 461 U.S. at 671-72. Similarly, Justice White wrote that because
"[p]overty 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 ef-
fort to impose a jail sentence that in terms of the State's sentencing objectives will be rough-
ly equivalent to the fine and restitution that the defendant failed to pay." *Id.* at 675 (White, J.,
concurring in the judgment). That is, the magnitude of the punishment must be the same,
even if the means are not.

between legislative means and purpose, [and] the existence of alternative means for effectuating the purpose . . . ."[132]

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 interest at stake and the extent to which it is affected. The approach does not quite correspond to the "tiers of scrutiny" framework. 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 searching review and justification.

Although *Bearden* involved revocation of probation, lower courts have treated the decision 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."[133] Conversely, citing the same disparity concern, the Ninth Circuit has also reversed (as an abuse of the discretion afforded under *Gall v. United States*[134]) 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."[135] Even before *Bearden*, several circuits had already held that, in order to avoid impermissible wealth-based distinctions in sentencing, equal protection entitles an indigent defendant unable to make bail to credit against the eventual sentence for time served.[136]

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: "[L]egislation

---

132. *Id.* at 666-67 (second and third alterations in original) (footnote omitted) (quoting *Williams*, 399 U.S. at 260 (Harlan, J., concurring in the result)); *see* Evitts v. Lucey, 469 U.S. 387, 405 (1985) (discussing the interrelationship between due process and equal protection concerns in these cases).

133. United States v. Burgum, 633 F.3d 810, 816 (9th Cir. 2011); *accord* United States v. Parks, 89 F.3d 570, 572 (9th Cir. 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) ("[T]he government cannot keep a person in prison solely because of indigency."). *But see* State v. Todd, 208 P.3d 303, 305-06 (Idaho Ct. App. 2009) (upholding inability to pay as an aggravating factor).

134. 552 U.S. 38 (2007).

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

136. *See, e.g.*, Johnson v. Prast, 548 F.2d 699, 703 (7th Cir. 1977); King v. Wyrick, 516 F.2d 321, 323 (8th Cir. 1975); Ham v. North Carolina, 471 F.2d 406, 408 (4th Cir. 1973). *But see* Vasquez v. Cooper, 862 F.2d 250, 251-52 (10th Cir. 1988) (finding no constitutional violation in the court's refusal to credit the defendant for time he served in custody prior to trial because he could not make bail).

which has a disparate impact on the indigent defendant should be subject to a more searching scrutiny than requiring a mere rational relationship."[137] 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."[138]

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 long been rife with procedural obstacles to equal treatment of the indigent, and there are no doubt many subtle or de facto ways in which poverty might influence sentences. 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 Federal Sentencing Guidelines forbid consideration of socioeconomic status.[139] 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)[140]—in contrast to gender, which is essentially never cited. Such cases might well also be constitutionally problematic, unless these 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 con-

---

137. United States v. Luster, 889 F.2d 1523, 1530 (6th Cir. 1989); *see also* Kadrmas v. Dickinson Pub. Sch., 487 U.S. 450, 461 & n.* (1988) (rejecting heightened scrutiny in a noncriminal case because "the criminal-sentencing decision at issue in *Bearden* is not analogous to the user fee . . . before us"); Maher v. Roe, 432 U.S. 464, 471 n.6 (1977); Dickerson v. Latessa, 872 F.2d 1116, 1119-20 (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) ("At least where the classification at issue is not based on wealth, the right to appeal is not a fundamental right."); United States v. Kerr, 686 F. Supp. 1174, 1178 (W.D. Pa. 1988).

138. 686 F. Supp. at 1178 (quoting Williams v. Illinois, 399 U.S. 235, 263 (1970) (Harlan, J., concurring in the result)).

139. U.S. SENTENCING GUIDELINES MANUAL § 5H1.10 (2012); *see also* Joan Petersilia & Susan Turner, *Guideline-Based Justice: Prediction and Racial Minorities*, in 9 CRIME AND JUSTICE: A REVIEW OF RESEARCH 151, 153-54, 160 (Don M. Gottfredson & Michael Tonry eds., 1987) (describing sentencing reformers' objective of eliminating role of "status factors" such as employment).

140. *E.g.*, United States v. Trimble, 514 F. App'x 913, 915 (11th Cir. 2013).

text 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 Sentencing Guidelines do include an enhancement for offenders with a "criminal livelihood,"[141] and defendants have occasionally challenged that enhancement as disparately affecting the poor, because the same amount of criminal revenue constitutes a larger share of a low-income person's livelihood than that of a higher-income person. Soon after the Guidelines' adoption, at least one district court held (citing *Bearden*) that, to avoid this potential constitutional concern, this enhancement should be interpreted to focus on the absolute amount of criminal income, rather than the share of total income, and the United States Sentencing Commission ultimately amended the Guidelines to come closer to this view.[142] After the amendment, the Sixth Circuit upheld the new version of the enhancement against a similar challenge, holding that, while *Bearden* required searching scrutiny of sentencing burdens on the poor, the amended enhancement 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."[143]

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.'"[144]

It is difficult to see how the socioeconomic variables in EBS can avoid invalidation under *Bearden*. Unemployment and education, the most common such variables, cannot meaningfully be distinguished from the ability to pay, nor can other variables like financial status, dependence on government assistance, or access to jobs paying above minimum wage. All are proxies for pov-

---

141. U.S. SENTENCING GUIDELINES MANUAL § 4B1.3.

142. *See* United States v. Rivera, 694 F. Supp. 1105, 1106-07 (S.D.N.Y. 1988); *see also* United States v. Luster, 889 F.2d 1523, 1528-29 (6th Cir. 1989) (describing the amendment). The current quantitative inquiry concerns only the amount of criminal income; there is also a qualitative inquiry into whether crime was the defendant's "primary occupation." U.S. SENTENCING GUIDELINES MANUAL § 4B1.3 cmt. application notes (italics omitted).

143. *Luster*, 889 F.2d at 1529-30.

144. United States v. Kerr, 686 F. Supp. 1174, 1179 (W.D. Pa. 1988) (quoting Papachristou v. City of Jacksonville, 405 U.S. 156, 171 (1972)); *cf.* Edwards v. California, 314 U.S. 160, 177 (1941) ("[W]e do not think that it will now be seriously contended that because a person is without employment and without funds he constitutes a 'moral pestilence.' Poverty and immorality are not synonymous.").

erty and income, and the *Griffin* line and *Bearden* make interchangeable refer-
ences to "wealth," "poverty," "class," and related terms without fine distinc-
tions. 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 circum-
stances under which courts can consider "employment history and financial re-
sources," specifically rejecting the consideration of such factors as recidivism
predictors.[145] 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 stud-
ies of the relationship between unemployment and recidivism, and the State
emphasized that the defendant's recent job loss made him a higher recidivism
risk.[146] Meanwhile, the point of including education in the recidivism instru-
ment 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. Neigh-
borhood characteristics could potentially also be considered socioeconomic
variables, since they are also very closely related to poverty, although this ex-
ample is more disputable because these variables operate at a geographic level
and do not draw distinctions among persons within the neighborhood.[147]

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 satisfy the demanding constitutional stand-
ards established in the *Griffin* line and *Bearden*, because if, as *Bearden* holds,
one cannot impute individual risk based on the average risk posed by poor de-
fendants, the rationale for EBS disappears.

B.   *The Social Harm of Demographic and Socioeconomic Sentencing Discrimination*

EBS's use of demographic, socioeconomic, and family- and neighborhood-
related characteristics is also highly troubling on public policy grounds—a con-

---

145. 461 U.S. 660, 671 (1983).

146. Brief for Respondent at 32-35, *Bearden*, 461 U.S. 660 (No. 81-6633), 1982 U.S. S.
Ct. Briefs LEXIS 438, at *35-39.

147. Given fairly high levels of residential segregation, *see generally* JOHN ICELAND ET
AL., U.S. CENSUS BUREAU, SERIES CENSR-3, RACIAL AND ETHNIC RESIDENTIAL
SEGREGATION IN THE UNITED STATES: 1980-2000 (2002), *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.

cern that reaches beyond the constitutionally problematic variables. As noted above, advocates of EBS frequently emphasize its potential to help reduce incarceration rates.[148] 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 about fifty times as likely to be incarcerated as white women are.[149] Young black men are especially at risk: one in nine black men between the ages of twenty and thirty-four are behind bars,[150] and a 2003 study projected that one in three young black men would be incarcerated at some point in their lives.[151] And the concentration of mass incarceration's effects is even more dramatic when one takes into account socioeconomic and neighborhood-level predictors. According to a 2009 study, high school dropouts, for example, are forty-seven times as likely to be incarcerated as college graduates, and approximately twenty-two percent of young black male dropouts are incarcerated at any given time.[152] An ample literature documents these disparities and their effects on communities.[153]

The EBS instruments produce higher risk estimates, other things equal, for subgroups whose members 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 demographically concentrate mass incarceration's impact.

---

148. *See supra* note 44 and accompanying text.

149. *See* HEATHER C. WEST, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, PRISON INMATES AT MIDYEAR 2009—STATISTICAL TABLES 21 tbl.18 (2010), *available at* http://www.bjs.gov/content/pub/pdf/pim09st.pdf (estimating that 4749 black men per 100,000 were in custody in 2009 as compared with 91 white women per 100,000).

150. JENIFER WARREN, PEW CTR. ON THE STATES, ONE IN 100: BEHIND BARS IN AMERICA 2008, at 3 (2008), *available at* http://www.pewstates.org/uploadedFiles/PCS_Assets/2008/one%20in%20100.pdf.

151. THOMAS P. BONCZAR, BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, PREVALENCE OF IMPRISONMENT IN THE U.S. POPULATION, 1974-2001, at 1 (2003), *available at* http://www.bjs.gov/content/pub/pdf/piusp01.pdf.

152. Andrew Sum et al., *The Consequences of Dropping Out of High School: Joblessness and Jailing for High School Dropouts and the High Cost for Taxpayers* 10 (Ctr. for Labor Mkt. Studies Publ'ns, Paper No. 23, 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, at 20, 20 (discussing neighborhood effects).

153. *See generally* MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS (rev. ed. 2012); TODD R. CLEAR, IMPRISONING COMMUNITIES: HOW MASS INCARCERATION MAKES DISADVANTAGED NEIGHBORHOODS WORSE (2007); IMPRISONING AMERICA: THE SOCIAL EFFECTS OF MASS INCARCERATION (Mary Pattillo et al. eds., 2004).

Moreover, EBS is likely to further concentrate mass incarceration's *racial* impact. I have ignored race in my constitutional analysis because the instruments do not include it. But the socioeconomic and family variables that they do include are highly correlated with race, as is criminal history, so they are likely to have a racially disparate impact.[154] Given widespread de facto residential segregation and the concentration of crime in urban neighborhoods of color, the neighborhood crime rate variables found in some instruments are particularly disturbing.[155] Rather than requiring specific information on neighborhood crime rate, for example, Pennsylvania's new EBS instrument simply assigns extra risk points to Philadelphia County and Allegheny County (Pittsburgh) while treating "rural counties" as lowest risk and "smaller urban counties" as medium risk. The likely racial impact of this decision is obvious.[156] Although the courts have not recognized equal protection claims grounded in disparate impact alone,[157] 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, or on young people generally. Rather, it will principally affect a subgroup of young men—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

---

154. *See* U.S. CENSUS BUREAU, STATISTICAL ABSTRACT OF THE UNITED STATES: 2012, at 151 tbl.229 (2012), *available at* http://www.census.gov/compendia/statab/2012/tables/ 12s0229.pdf (showing that blacks and particularly Hispanics have lower high school and college graduation rates than whites); Harcourt, *supra* note 64 (manuscript at 2) ("[P]rior criminal history has become a proxy for race."); *Table A-2: Employment Status of the Civilian Population by Race, Sex, and Age*, BUREAU LAB. STAT. (Sept. 6, 2013), http://www.bls.gov/ news.release/empsit.t02.htm (showing unemployment rates approximately twice as high for blacks than for whites).

155. *See* Kyle Crowder et al., *Neighborhood Diversity, Metropolitan Constraints, and Household Migration*, 77 AM. SOC. REV. 325, 325-26 (2012) (discussing continued "moderate to high" levels of black-white neighborhood segregation in large metropolitan areas and "steady or increased" segregation of Asian and Latino populations from blacks and whites).

156. *See* Monahan & Skeem, *supra* note 11, at app. 2.

157. *See, e.g.*, Washington v. Davis, 426 U.S. 229, 238-39 (1976).

a substantial part of a generation of men being absent from their communities and compounding 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.[158]

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 about the matter. It involves the state explicitly telling judges 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.[159] Dressing up that generalization in scientific language may have succeeded in forestalling public criticism, but mostly because few Americans understand these instruments or are even 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.[160]

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.[161] Generally, however, states are

---

158. *See, e.g.*, CLEAR, *supra* note 153, at 97-98; William A. Darity, Jr. & Samuel L. Myers, Jr., *Family Structure and the Marginalization of Black Men: Policy Implications*, *in* THE DECLINE IN MARRIAGE AMONG AFRICAN AMERICANS: CAUSES, CONSEQUENCES, AND POLICY IMPLICATIONS 263, 286 (M. Belinda Tucker & Claudia Mitchell-Kernan eds., 1995); Elizabeth I. Johnson & Jane Waldfogel, *Children of Incarcerated Parents: Multiple Risks and Children's Living Arrangements*, *in* IMPRISONING AMERICA: THE SOCIAL EFFECTS OF MASS INCARCERATION, *supra* note 153, at 87, 98; James P. Lynch & William J. Sabol, *Effects of Incarceration on Informal Social Control in Communities*, *in* IMPRISONING AMERICA: THE SOCIAL EFFECTS OF MASS INCARCERATION, *supra* note 153, at 135, 135-60; Bruce Western et al., *Incarceration and the Bonds Between Parents in Fragile Families*, *in* IMPRISONING AMERICA: THE SOCIAL EFFECTS OF MASS INCARCERATION, *supra* note 153, at 21, 21-43.

159. For a recent, prominent reflection on the way such generalizations about black men have affected African American communities, see Remarks on the Verdict in *State of Florida v. George Zimmerman*, 2013 DAILY COMP. PRES. DOC. 509 (July 19, 2013).

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

161. *See, e.g.*, RICHARD COUZENS, EVIDENCE-BASED PRACTICES: REDUCING RECIDIVISM TO INCREASE PUBLIC SAFETY; A COOPERATIVE EFFORT BY COURTS AND PROBATION 10 (2011), *available at* http://www.courts.ca.gov/documents/EVIDENCE-BASED-PRACTICES-Summary-6-27-11.pdf; Etienne, *supra* note 38, at 50; *cf.* Kleiman, *supra* note

not actually using risk prediction instruments in a systematically one-sided manner; risk predictions are simply provided to sentencing judges and parole boards. There is no persuasive reason to believe access to risk predictions would only tend to reduce sentences rather than to also increase them in some 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.[162] This line of argument is belied, however, by the fact that 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.[163]

It is possible to try to *force* unidirectional use of risk assessments, via legislation mandating their use only for mitigation or diversion from incarceration, but it may be difficult to enforce that limitation. 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.[164] 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 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 many of the instruments are quite simple, one would expect the parties to calculate the scores themselves and plan accordingly rather than 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 specifically on constitutionally and normatively problematic demographic and socioeconomic characteristics (instead of past or present criminal conduct or other personal, behavioral assessments).

36, at 301 (explaining that Virginia's EBS program diverts "25 percent of nonviolent, prison-bound offenders into alternate sanctions").

162. *See, e.g.*, Marcus, *supra* note 44, at 751.

163. *See* Rachel E. Barkow, *Federalism and the Politics of Sentencing*, 105 COLUM. L. REV. 1276, 1278-81 (2005).

164. Analogously, limiting instructions given to juries—instructions to consider evidence for one purpose but not another—are "notoriously ineffective" and "may be counterproductive because they draw jurors' attention to the evidence that is supposed to be ignored." J.J. Prescott & Sonja Starr, *Improving Criminal Jury Decision Making After the* Blakely *Revolution*, 2006 U. ILL. L. REV. 301, 323 (citing studies).

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 twenty-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 she has no criminal history and no other risk factors and has committed a relatively minor offense.[165] Other instruments that consider gender and a wider variety of socioeconomic and family traits could be even more strongly driven by those factors.[166]

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.[167] 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 the bulk of the population another reason not to worry, even if those policies will have the immediate effect of somewhat restraining that growth. They should instead look for strategies to reduce excessive incarceration in ways that do not discriminate against the disadvantaged.

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 some form of heightened constitutional scrutiny. One must consider how strongly EBS advances competing state interests. In the next Part, then, I turn to the question of whether the studies support the optimism of EBS advocates.

---

165. *See* Wolff, *supra* note 30, at 112-13.

166. 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 he was denied a government benefit on the basis of some improper consideration. *See, e.g.*, Fisher v. Univ. of Tex. at Austin, 133 S. Ct. 2411 (2013) (reviewing a challenge based on the denial of state university admission).

167. *See* William J. Stuntz, *Unequal Justice*, 121 Harv. L. Rev. 1969, 1974 (2008); James Forman, Jr., *Why Care About Mass Incarceration?*, 108 Mich. L. Rev. 993, 1001 (2010) (book review).

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 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, with a focus on three main concerns. In Subpart A, I consider the limited ability of EBS instruments to precisely and accurately predict individual recidivism, as opposed to group re-cidivism rates. This individual-versus-group distinction has been raised by oth-ers, particularly in criminological literature, but it is often elided by EBS advo-cates. Because its legal implications are important (and previously unrecognized), I discuss it in some detail. Next, I turn to more novel criticisms. In Subpart B, I critically examine the underlying research and show that there is not yet any persuasive evidence even to support the basic claim that the actuar-ial instruments outperform judges' individual risk predictions, much less to show that the constitutionally problematic variables add substantial marginal predictive value to the instruments. In Subpart C, I argue that the instruments do not even address the right empirical question: the effect of sentencing deci-sions on the defendant's recidivism risk.

A.   *Precision, Group Averages, and Individual Predictions*

The instruments' first serious limitation is that they do not provide any-thing even approaching a precise prediction of an individual's recidivism risk. The models are designed to predict the *average* recidivism rate for all offenders who share with the defendant whichever characteristics are included as varia-bles in the model. If the model is well specified and based on an appropriate and large enough sample, then it might perform this task well. But because in-dividuals vary much more than groups do, even a relatively precisely estimated model will often not do well at predicting individual outcomes in particular cases. "It is a statistical truism that the mean of a distribution tells us about eve-ryone, yet no one."[168] 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 simply, let's start with an example that involves predicting a continuous outcome (height)—rather than a binary future event—using just one

---

168. David J. Cooke & Christine Michie, *Limits of Diagnostic Precision and Predictive Utility in the Individual Case: A Challenge for Forensic Practice*, 34 LAW & HUM. BEHAV. 259, 259 (2010).

explanatory variable (sex). The height distributions of the U.S. male and female populations look approximately like Figure 1, which is based on average heights of 70 inches for males and 65 inches for females, and normal distributions with standard deviations of 3 inches and 2.5 inches, respectively.[169]

FIGURE 1
Height Distributions by Gender
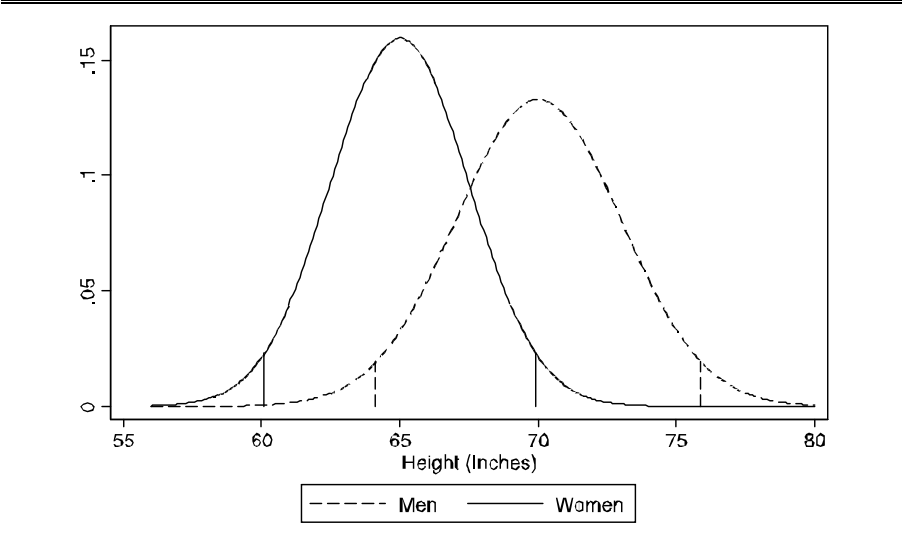
But suppose one did not know the true population distributions and had to estimate them using a random sample. With a large enough sample, it is easy to obtain quite precise estimates of the male and female averages and the difference between them. This point is illustrated in Table 1. I created simulated 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 of 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.

169. This is a rough approximation. For a more detailed height distribution segmented by age, see U.S. CENSUS BUREAU, *supra* note 154, at 137 tbl.209.

TABLE 1

Precision of Predicted Means Versus Individual Forecasts: An Illustration

	Male Height in Inches			Female Height in Inches		
N	Mean & Forecast	95% C.I. for Mean	95% Indiv. Forecast Interval	Mean & Forecast	95% C.I. for Mean	95% Indiv. Forecast Interval
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]

Note: Samples are drawn from a simulated "true population" with population means of 70.0 ($\sigma = 3.0$) for men and 65.0 ($\sigma = 2.5$) for women, normally distributed.

How close each sample comes to approximating the true population means involves chance: different random samples of the same size may have different means. But chance can be expected to play a smaller role as the sample gets larger. This expectation is captured in the 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% confidence in quite a precise estimate of the mean: for males, between 69.6 inches and 70.4 inches, and for females, between 64.5 inches and 65.3 inches.[170] If you keep drawing additional 400-person samples, they don't tend to differ very much; with that sample size, one can generally do quite a good job approximating the underlying population, which is why the confidence interval is narrow. Meanwhile, the 20-person sample produces much wider confidence intervals, spanning more than three inches.

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 would be the female mean from your sample, which is 64.9 inches. But you wouldn't be *nearly* as confident in that prediction as you would be in the prediction for the group mean. In fact, within the same 400-person sample used above, only 13.5% of women have heights that are between 64.5 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% individual forecast interval shown in Table 1. In other words, you don't know much at all about how tall to expect the next woman to be.[171]

---

170. To describe something as a 95% confidence interval for an estimated group mean is to express confidence that for 95% of random samples, the same estimation procedure will produce an interval containing the true group mean for the underlying population.

171. The estimated uncertainties in Table 1 are based on a regression of height on gender using standard Stata postestimation prediction commands. By construction, the uncer-

One could make the example much more complicated, but the individual forecast interval is always wider than the confidence interval for the mean—generally *much* wider.[172] Note that while the confidence intervals for the means get much narrower as the sample grows, the individual forecast interval does not. The underlying uncertainty that it reflects is not mainly sampling error; it's the variability in the underlying population that we saw in Figure 1.

The same basic intuition applies to models of binary outcomes, such as whether a defendant will recidivate. Some of the recidivism risk prediction instruments include confidence intervals for the probabilities they predict. Indeed, some scholars have urged that confidence intervals (rather than mere point estimates) should always be provided so that judges can get an idea of how precise the instruments are.[173] But given that judges are using the instruments to predict *individual* recidivism risk, providing them with a confidence interval for the *group* recidivism rate might misleadingly represent the instrument's precision. 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.[174]

With binary outcomes, though, while the confidence interval for the mean may be misleading, the individual forecast interval is not a very useful alternative, because it tells you nothing that was not made clear by the point estimate itself. Unless the predicted probability is extremely low or extremely high, a 95% individual prediction interval will always run from 0 to 1, meaning that the only prediction that can be made with 95% confidence is that the individual will either recidivate or not.[175] This fact does not reflect poorly on the design of the prediction instruments or on the underlying research. It reflects the inherent uncertainty of this predictive task and the binary nature of the outcome.

---

tainties are the same for men and women. Another way to estimate a 95% prediction interval for an individual woman is to just ignore the men and give the range within which the middle 95% of the women in your sample fall. This range is marked on Figure 1.

172. *See* Cooke & Michie, *supra* note 168, 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).

173. *See, e.g.*, McGarraugh, *supra* note 67, at 1095-96, 1113.

174. My objection here is not that the models cannot establish "individual-level causation." *Id.* at 1101-02. The models 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.

175. *See* R. Karl Hanson & Philip D. Howard, *Individual Confidence Intervals Do Not Inform Decision-Makers About the Accuracy of Risk Assessment Evaluations*, 34 LAW & HUM. BEHAV. 275, 278-79 (2010).

In order to assess how well a model predicts recidivism risk for individuals, some other metric is necessary.[176] There is no single agreed-upon method for assessing the individual predictive accuracy of a binary model. One common metric used in the recidivism prediction literature is called the "area under the curve" (AUC) approach.[177] 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, giving it an AUC score of 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, unlike simple point systems based on objective factors.[178] Many studies have reported AUC scores closer to 0.65.[179] By comparison, a prominent meta-analysis of studies of psychologists' clinical predictions of violence found a mean AUC score of 0.73, which the author characterized as a "modest, better-than-chance level of accuracy."[180] As another point of comparison, if one turns

---

176. *See id.* at 276. Stephen D. Hart et al., *Precision of Actuarial Risk Assessment Instruments*, 190 Brit. J. Psychiatry s60 (2007), offers an alternative way of calculating the individual forecast interval. Hart et al. use a traditional method for estimating the confidence interval for a probability prediction given a point estimate for the probability and a sample size, *see* Edwin B. Wilson, *Probable Inference, the Law of Succession, and Statistical Inference*, 22 J. Am. Stat. Ass'n 209 (1927), and calculate it for each risk-level category in two common violence prediction instruments, using a sample size of 1, *see* Hart et al., *supra*, at s61. The resulting intervals do not run from 0 to 1, but they are always very wide, ranging between 79 and 89 percentage points in width. *Id.* at s63. The authors conclude that it is "impossible to make accurate predictions about individuals using these tests." *Id.* at s64.

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." *Id.* at s62 (internal quotation marks omitted). 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. *See id.* at s63. The actual observed rate for a sample of one individual must always be 0 or 1, however, so I agree with Hanson & Howard, *supra* note 175, at 278-79, that the forecast interval for all but the extreme cases should be from 0 to 1 (rather than, say, from 0.10 to 0.94). But either way, it is wide.

177. *See* Douglas Mossman, *Assessing Predictions of Violence: Being Accurate About Accuracy*, 62 J. Consulting & Clinical Psychol. 783, 785-90 (1994) (describing the method as well as competing approaches).

178. *See* Office of Prob. & Pretrial Servs., Admin. Office of the U.S. Courts, An Overview of the Federal Post Conviction Risk Assessment 9 (2011), *available at* http://www.uscourts.gov/uscourts/FederalCourts/PPS/PCRA_Sep_2011.pdf; M. Dolan & M. Doyle, *Violence Risk Prediction: Clinical and Actuarial Measures and the Role of the Psychopathy Checklist*, 177 Brit. J. Psychiatry 303, 305-07 (2000).

179. Dolan & Doyle, *supra* note 178, at 307.

180. Mossman, *supra* note 177, at 788, 790.

height into a binary variable called "tall" (above-average height), our basic, one-variable model does much better at predicting who will be tall than any actuarial model does at predicting who will recidivate; it has an AUC score of 0.825.[181] This is despite the fact that, as we saw, the model gives rather wide bounds for individual predictions of height; gender is actually quite a strong predictor of height, but it still leaves considerable individual variation unexplained.[182]

Another simple measure of prediction accuracy is the linear correlation between predicted probabilities and actual outcomes for offenders. This measure will be 0 if the instrument explains nothing more than chance and 1 if it predicts perfectly.[183] In 1994, a prominent meta-analysis of studies comparing several actuarial recidivism prediction instruments found that the LSI-R had the highest reported correlation with outcomes, at 0.35.[184] 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 recidivism risk 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 underlying 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 mental health treatment purposes, 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, the accuracy measures discussed here assess the *total* predictive power of each model. The marginal predictive power added by just the constitutionally problematic variables is considerably less, as discussed in the next Subpart.

The basic difference between individual and group predictions has been pointed out by some scholars in the empirical literature surrounding risk predic-

---

181. This is estimated in the same 400-person sample used above.

182. Note that a 95% confidence interval for an individual for "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."

183. The square of this correlation coefficient is one variant on the "pseudo R-squared" statistic, a "fit" measure. 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 *FAQ: What Are Pseudo R-Squareds?*, INST. DIGITAL RES. & EDUC., http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm (last updated Oct. 20, 2011).

184. *See* Paul Gendreau et al., *A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!*, 34 CRIMINOLOGY 575, 587 tbl.4 (1996).

tion instruments.[185] 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, 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 being trained in the use and interpretation of risk tools, practitioners tended to struggle with the meaning of probability scores . . . . Instead of understanding that an individual with a high risk score *shares characteristics* with an aggregate group of high-risk offenders, practitioners are likely to perceive the individual *as* a high-risk offender. In practical terms, *correlation becomes causation* and potential risk is translated into administrative certainty. When used at the pre-sentence stage, the courts may assume that a "high-risk" offender poses a greater danger to society and sentence him/her accordingly.[186]

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 Reidy 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 subdiscipline of violence prediction, draws its fundamental scientific character from its willingness to draw insights from data collected on groups and apply them to individuals.[187] 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?'"[188] Jennifer Skeem and John Monahan similarly argue "that group data theoretically can be, and in many areas empirically are, highly informative

---

185. *See, e.g.*, Cooke & Michie, *supra* note 168, at 259; Hart et al., *supra* note 176, at s61-62.

186. Hannah-Moffat, *supra* note 66, at 278.

187. Mark D. Cunningham & Thomas J. Reidy, *Violence Risk Assessment at Federal Capital Sentencing: Individualization, Generalization, Relevance, and Scientific Standards*, 29 CRIM. JUST. & BEHAV. 512, 517 (2002); *see also* Jessica M. Tanner, *"Continuing Threat" to Whom?: Risk Assessment in Virginia Capital Sentencing Hearings*, 17 CAP. DEF. J. 381, 402-05 (2005) (similarly arguing that "*all* scientifically derived expertise . . . is derived from collective data" and that probabilities based on groups can be very useful in formulating individual predictions).

188. Redding, *supra* note 34, at 12 n.52 (first and second alterations in original) (quoting PAUL E. MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION: A THEORETICAL ANALYSIS AND A REVIEW OF THE EVIDENCE 138 (1954)).

when making decisions about individual cases."[189] They ask us to consider the following analogy, first posed by William Grove and Paul Meehl:

> Two revolvers are put 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"?[190]

These responses strike me as off base. I do not argue, nor could anybody, that group averages have nothing to do with individual behavior. But that does not always mean that the group average tells us *much* about what to expect for any given individual. The question is how much individual variation there is in a given population and how much of that variation the variables in the 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.[191]

One could defend the instruments on the ground that the precision of individual predictions does not matter from an efficiency perspective. If the group-based estimates are good, then the model will, averaged across cases, improve judges' predictions of recidivism, leading to 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. 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.[192] 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 on which 0.18% of the drivers were drunk, I would choose the latter every time. But just that disparity did not suffice, in *Craig v. Boren*, to justify a gender-discriminatory alcohol law.[193] When demographic and socioeconomic characteristics are used to justify the

---

189. Jennifer L. Skeem & John Monahan, *Current Directions in Violence Risk Assessment*, 20 CURRENT DIRECTIONS PSYCHOL. SCI. 38, 40 (2011).

190. *Id.* (quoting William M. Grove & Paul E. Meehl, *Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy*, 2 PSYCHOL. PUB. POL'Y & L. 293, 305-06 (1996)).

191. In the Russian roulette hypothetical, the decisionmaker is given the only variable that matters: the number of bullets. The recidivism models are not in the same ballpark.

192. *See supra* notes 93-107, 126 and accompanying text.

193. 429 U.S. 190, 201 (1976).

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" predictions of risk—might actually be more efficient because it captures more of that variation. This point is explored further in the next Subpart.

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

The *Bearden* test requires assessment of whether other available and non-discriminatory (or less discriminatory) alternatives could accomplish the state's penological objectives. Here, I consider two such alternatives: (1) actuarial methods that *do not* rely on constitutionally troubling variables; and (2) 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 it seeks to replace, it does not substantially advance the state's penological interests and is also undesirable on policy grounds.

Advocates of EBS have concluded that it is superior to available alternatives, but they have had to stretch the existing evidence quite far to support this claim. Oleson, for instance, argues that even inclusion of race would be constitutionally permissible and concludes that it is "straightforward" to show that no less restrictive means is available.[194] 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."[195] 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 that was 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.

---

194. *See* Oleson, *supra* note 9, at 1385-86; *see also id.* at 1387 ("Once the constitutional door is open to race, all other sentencing factors can pass through: gender, age, marital status, education, class, and so forth.").

195. *See id.* at 1386 (citing Petersilia & Turner, *supra* note 139, at 173).

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 provided by correlates of race (and which no sentencing scheme ignores).[196] 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.[197] This is presumably because past 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 "[t]he average accuracy of predictions based on past behavior is higher" than those based on either mental health professionals' clinical judgments or actuarial instruments.[198]

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 magnitude of the *coefficients* on those variables).[199] But Petersilia 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? In the literature on evidence-based criminal justice practices, it is vir-

---

196. Petersilia & Turner, *supra* note 139, at 171 fig.1 (showing, in the table for "[a]ll convicted felons," that 57% of outcomes could be accurately predicted by chance, 60% when nonracially correlated factors were added, 67% when crime characteristics were added, 70% when criminal history variables were added, and still 70% when demographic and "[o]ther" variables were added).

197. *Id.* at 169-70 (describing the "demographic[]" and "other" variables).

198. Mossman, *supra* note 177, at 789-90.

199. One exception is a 2009 North Carolina report that found that eliminating three potentially problematic variables (race, gender, and prior arrests not resulting in conviction) from the state's risk prediction instrument would have little effect on its predictive accuracy. N.C. SENTENCING & POLICY ADVISORY COMM'N, RESEARCH FINDINGS AND POLICY RECOMMENDATIONS FROM THE CORRECTIONAL PROGRAM EVALUATIONS, 2000-2008, at 13 (2009), *available at* http://www.nccourts.org/Courts/CRS/Councils/spac/Documents/correctionalevaluation_0209.pdf.

tually gospel that they do.[200] 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 between 1945 and 1994 and that met certain quality criteria; just five criminal recidivism studies made the cut (compared to 131 studies from other fields).[201] Overall, actuarial prediction performed on average about 10% better, but the authors warned: "[O]ur 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."[202]

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 some states actually use. Take, for instance, a study by Grant Harris, Marnie Rice, and Catherine Cormier, which has been cited by EBS advocates, testing an instrument called the Violence Risk Appraisal Guide (VRAG).[203] The VRAG consists of twelve variables, "[t]he first and most heavily weighted" of which "is itself a 20-item scale involving the systematic assessment of such behaviors as 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."[204] Assessing these factors requires an elaborate psychological profile, which was carried out in the study by groups of mental health clinicians who "knew the patients well."[205] Nothing like this is typically involved in EBS. Even in the case of sentencing instruments that try to use fairly 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

---

200. *See* sources cited *supra* note 43.

201. William M. Grove et al., *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 12 PSYCHOL. ASSESSMENT 19, 22-24 (2000) (listing studies).

202. *Id.* at 25.

203. 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); Wolff, *supra* note 35, at 1406 & n.73 (citing Harris et al., *supra*).

204. Harris et al., *supra* note 203, at 378.

205. *Id.* at 379.

comparability of the populations is also dubious; Harris et al., for example, involved Canadian psychiatric patients.[206]

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, his ultimate predictions will be better informed if he then uses an actuarial model to tell him 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 between actuarial and clinical weighting of variables; it is between actuarial weighting of a few variables and clinical weighting of a much wider range of variables.[207] It is possible that the actuarial instruments would win in that comparison, but we cannot reach such a conclusion based on existing research.

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.[208] Two of the five studies found no appreciable advantage for actuarial prediction.[209] One of two studies that found a substantial advantage involved an archaic prediction instrument in which the most strongly predictive variable was the offender's (clinically assessed) "social development pattern," defined by the following categories: "respected citizen," "inadequate," "fairly conventional," "ne'er-do-well," "floater," "dissipated," and "socially maladjusted."[210] It also involved very few clinical decisionmakers (four psychiatrists and four sociologists who worked in a parole

---

206. *See id.* at 381.

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

208. *See* Grove et al., *supra* note 201, at 22 tbl.1 (listing studies evaluated by meta-analysis).

209. Terrill R. Holland et al., *Comparison and Combination of Clinical and Statistical Predictions of Recidivism Among Adult Offenders*, 68 J. APPLIED PSYCHOL. 203, 207 (1983) (finding that individual decisionmakers better predict violent recidivism, but actuarial prediction instruments better predict some measures of overall recidivism); James Smith & Richard I. Lanyon, *Prediction of Juvenile Probation Violators*, 32 J. CONSULTING & CLINICAL PSYCHOL. 54, 56 (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).

210. Daniel Glaser, *The Efficacy of Alternative Approaches to Parole Prediction*, 20 AM. SOC. REV. 283, 285-86 (1955) (capitalization altered).

system in the 1940s),[211] so one possible explanation for the results is that a couple of these people might not have been terribly good at their jobs.[212] A study by Stephen Wormith and Colin Goldstone evaluated an instrument with more objective criteria and also found that it predicted recidivism better than did the parole board's actual (clinical) decisions,[213] 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."[214] The authors also warned that their measures of clinical and actuarial judgment were not really fairly comparable; the "clinical prediction" was not actually a risk prediction at all (instead, it was a binary parole decision), whereas the actuarial prediction was.[215] Finally, a study by Howard Sacks included a brief analysis of the clinical-versus-actuarial comparison, but the comparison it drew 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.[216]

Nor are more recently published studies more compelling. Oleson et al. purports to compare the accuracy of clinical and actuarial judgment in federal probation officers' assessment of a probationer's recidivism risk.[217] The study included over a thousand decisionmakers (but only one case vignette) and used a modern instrument, recently developed by the Administrative Office of the U.S. Courts (AOUSC), called the Federal Post Conviction Risk Assessment (PCRA).[218] The researchers asked officers to watch a video about an individual

---

211. *See id.* at 285.

212. Problems like this recur in other actuarial-versus-clinical studies as well. These studies 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 decisionmakers involved in the study, meaning that standard errors should instead be calculated with clustering on the decisionmaker.

213. J. Stephen Wormith & Colin S. Goldstone, *The Clinical and Statistical Prediction of Recidivism*, 11 CRIM. JUST. & BEHAV. 3, 30 (1984).

214. *Id.* at 29.

215. *See id.* at 17-21. 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 and Goldstone explain that it is unsurprising that the parole decision does not predict recidivism as well as an actuarial prediction, because the parole decision might be affected by factors unrelated to risk prediction and by the desire to err on the side of caution. *See id.* at 20.

216. Howard R. Sacks, *Promises, Performances, and Principles: An Empirical Study of Parole Decisionmaking in Connecticut*, 9 CONN. L. REV. 347, 402-03 (1977).

217. J.C. Oleson et al., *Training to See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessments Among Federal Probation Officers*, FED. PROBATION, Sept. 2011, at 52, 54-55.

218. *See id.* at 53-54. This instrument includes qualitative and dynamic factors plus objective factors such as age and education. OFFICE OF PROB. & PRETRIAL SERVS., ADMIN.

and predict his risk, and then to redo the exercise after being given training in the PCRA method and the individual's PCRA score.[219] The researchers concluded that the officers were "more accurate" when they had the PCRA.[220] But their only evidence for that claim was that officers' risk scores were *more consistent with the PCRA* after the officers were given the PCRA and instructed on its implementation. 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.[221]

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. At least so far, there is little evidence that the recidivism risk prediction instruments offer any discernible 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 has the highest risk of recidivism. They do not tell us whose risk of recidivism will be *reduced the most* by incarceration. The two questions are not the same, and only the latter directly pertains to the state's penological interests. In this Subpart, I first set out this argument and then explore the possibility of instruments that *do* address the right question, examining the relevant existing empirical research and pointing to ways in which that research would have to be enriched.

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 policy goals 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

---

OFFICE OF THE U.S. COURTS, *supra* note 178, at 10. It is used for planning probation supervision and treatment interventions, not sentencing. *Id.* at 1.

219. Oleson et al., *supra* note 217, at 53-54.

220. *Id.* at 54-55.

221. *See id.* The AOUSC's other validation study for the PCRA did not compare its effectiveness to clinical prediction and did not find anything close to *perfect* accuracy. OFFICE OF PROB. & PRETRIAL SERVS., ADMIN. OFFICE OF THE U.S. COURTS, *supra* note 178, at 9 (finding AUC scores ranging from 0.709 to 0.783).

great harm on society, as does excessive incarceration. Striking an appropriate balance between these concerns is an enormous and vital challenge.[222]

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,[223] the question is whether the additional two years of 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 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.[224]

This point has been ignored by the EBS literature. Harcourt makes a similar point, however, about the general-deterrence consequences of police profiling and sentencing enhancements based on criminal history.[225] Harcourt observes that a group's higher crime rate does not mean that it is more readily deterred by policing, and thus concentrating police resources on high-risk groups does not always maximize deterrence.[226] 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 communi-

---

222. 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 incarceration of one defendant as trading off with incarceration of another. Instead, they think about whether the particular sentence in question is worth its costs.

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

224. 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, three years of incarceration and the other group's rate is the average after five. 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.

225. HARCOURT, *supra* note 6, at 122-36.

226. *Id.* at 123.

ties. The relevant issue, Harcourt argues, is not rate of crime commission; it is "elasticity" to policing.[227]

Harcourt's argument focuses on general-deterrence effects on community crime rates, but a similar problem arises when one specifically considers the effects of marginal changes in incarceration 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.[228] If the incapacitation effect were the *only* effect that incarceration had 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 *postrelease* 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 (interference with subsequent employability, establishment of 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 net risk less than would locking up some low-risk offenders.

---

227. *Id.*; *see also* Bernard E. Harcourt, *A Reader's Companion to* Against Prediction*: A Reply to Ariela Gross, Yoram Margalioth, and Yoav Sapir on Economic Modeling, Selective Incapacitation, Governmentality, and Race*, 33 LAW & SOC. INQUIRY 265, 269 (2008) (explaining that traditional actuarial methods do not measure comparative elasticities of different groups to policing).

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

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 incarceration sentence for felonies in state courts was seventeen months.[229] Moreover, advocates of EBS often emphasize its value in determining whether a person should be incarcerated at all versus being given 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.

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 crime commission by *other individuals*, there are other factors to consider that have no intuitive connection to recidivism risk scores, such as general deterrence and the impact of expressive effects on social norms.

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 the simple fact that persons released from prison recidivate at higher rates than probationers.[230] 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 extensive prior criminal histories. 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 net 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.

---

229. SEAN ROSENMERKEL ET AL., BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, FELONY SENTENCES IN STATE COURTS, 2006—STATISTICAL TABLES 6 tbl.1.3 (2009), *available at* http://www.bjs.gov/content/pub/pdf/fssc06st.pdf.

230. *E.g.*, 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'G REP. 320, 320 (2008); McGarraugh, *supra* note 67, at 1107.

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 such as crime type and criminal history.[231] 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 decisionmaker'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.[232] 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 sentence length on subsequent recidi-

---

231. *See, e.g.*, OR. DEP'T OF CORR., THE EFFECTIVENESS OF COMMUNITY-BASED SANCTIONS IN REDUCING RECIDIVISM 18-19 (2002), *available at* http://www.oregon.gov/doc/CC/docs/pdf/effectiveness_of_sanctions_version2.pdf.

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

vism. Other studies take advantage of legal reforms that introduce sentencing variation.[233]

These studies have fairly consistently found that, on average, increased sentence length reduces subsequent offending, although the effect seems to be nonlinear: the marginal effect of increasing sentence length declines and eventually disappears as the total sentence length gets longer.[234] Thus, the specific-deterrence effect of increasing sentence length on average cuts in the same direction as incapacitation effects do.[235] Reported incapacitation effects typically appear larger,[236] but the results of the two types of studies are hard to compare. Incapacitation studies generally estimate the number of crimes avoided during each person-year of incarceration,[237] 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 remaining lifetime). Instead, such studies mostly have quite short follow-up periods and generally measure not number of crimes committed but rather binary recidivism rates (or recidivism "survival," that is, the length of time before recidivism) within that short period.[238] Moreover, incapacitation studies sometimes use reported crime as their measure,[239] whereas recidivism studies use the more underinclusive measures of rearrest or reconviction.[240]

Regardless, what the existing research on causal effects has *not* done is estimate the relationship between either specific deterrence or incapacitation elasticity and 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 Avinash Bhati, did estimate incapacitation elasticities that are gender-, race-, and state-specific, but not specific-deterrence elasticities, and the estimated incapacitation elasticities were not broken down by socioeconomic status. The study found no major differences in

---

233. *See, e.g.*, 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.

234. *See* Abrams, *supra* note 232, at 936.

235. *See id.* at 936-39 (reviewing incapacitation studies).

236. *See id.* at 929-39 (reviewing specific-deterrence and incapacitation studies); *id.* at 962 tbl.5, 964 tbl.6 (applying these studies' findings to calculate the expected specific-deterrence and incapacitation effects of possible changes in incarceration policy).

237. *See, e.g.*, Rucker Johnson & Steven Raphael, *How Much Crime Reduction Does the Marginal Prisoner Buy?*, 55 J.L. & ECON. 275, 297-98 (2012).

238. *See, e.g.*, Kuziemko, *supra* note 233, at 30 (estimating recidivism rates during a three-year follow-up period).

239. *See, e.g.*, Johnson & Raphael, *supra* note 237, at 293.

240. *See, e.g.*, Kuziemko, *supra* note 233, at 12 & n.18.

the total number of crimes averted when broken out by gender and race.[241] Notably, variations by state were far more dramatic, suggesting the need to worry about another problem with 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 found that sentence length increases have a "much stronger [deterrent] effect for older offenders than younger ones, for whom time served actually increases recidivism (though weakly)."[242] That is, young age—one of the most heavily weighted predictors of increased recidivism risk in the current instruments—actually appears to correspond with lower effectiveness of sentence length increases in deterring postrelease 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) use of valid causal identification methods, for example, exploiting random assignment of judges;

(2) application of those methods to obtain estimates of incarceration's effects as a function of the variables that the state seeks to include in the instrument;

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

(4) long enough follow-up periods to allow researchers to meaningfully approximate the change in an individual's lifetime recidivism risk;[243]

---

241. AVINASH SINGH BHATI, URBAN INST., AN INFORMATION THEORETIC METHOD FOR ESTIMATING THE NUMBER OF CRIMES AVERTED BY INCAPACITATION 24 tbl.4.2 (2007), *available at* http://www.urban.org/uploadedpdf/411478_crimes_averted.pdf (showing estimated male elasticities that were slightly greater than female elasticities in most, but not all, states, 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.

242. Kuziemko, *supra* note 233, at 23.

243. Collecting data over the course of 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 six or seven years (after release, if they were incarcerated) have quite low subsequent recidivism rates. *E.g.*, Megan C. Kurlychek et al., *Scarlet Letters and Recidivism: Does an Old Criminal Record Predict Future Offending?*, 5 CRIMINOLOGY & PUB. POL'Y 483, 499 (2006). Thus, to study the effect of the 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.

(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 and on a representative sample; and

(7) evidence of the *marginal* explanatory power added by each constitutionally problematic variable 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 if it is, the problems discussed above 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 objections by arguing that its use will not change very much at all. This "defense" comes in two forms. First, EBS advocates observe that the risk prediction instruments don't directly determine the sentence—they merely provide information to judges. Second, advocates point out 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 Subparts A and B, respectively.

A. *Do the Instruments Merely Provide Information?*

One response to disparity concerns is to defend the instruments as innocuous insofar as they only provide information rather than completely control the sentence.[244] 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

---

244. *See, e.g.*, Malenchik v. State, 928 N.E.2d 564, 573 (Ind. 2010); Kleiman, *supra* note 36, at 301, 302; David E. Patton, *Guns, Crime Control, and a Systemic Approach to Federal Sentencing*, 32 CARDOZO L. REV. 1427, 1456 (2011).

the risk prediction instrument does not itself amount to disparate "treatment" at all. Rather, it merely provides social scientific information to a government decisionmaker, 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,[245] and the instrument tells the judge how to calculate that risk. Unless judges completely ignore the instruments (rendering them pointless), some defendants will inevitably 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, the information provision itself is arguably 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."[246] 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 ten-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 if a court's decisionmaking is unconstitutional in substance, it

---

245. *See, e.g.*, 18 U.S.C. § 3553(a)(2)(C) (2012) (instructing federal courts to consider "protect[ing] the public from further crimes of the defendant").

246. Bearden v. Georgia, 461 U.S. 660, 671 (1983).

surely cannot become constitutional through obscurity of reasoning. In any event, the use of the discriminatory factor here is not obscure, even if its specific consequence for any given defendant is not transparent. A defendant subjected to an unconstitutional decisionmaking process should be entitled to resentencing.[247] 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 characterized it in dissent, as a "factor of a factor of a factor of a factor" that very likely was not the reason the plaintiff in the case was denied admission.[248]

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.[249] Accordingly, the fact that the risk prediction instruments do not completely displace all other sentencing factors is a point in their favor when assessing narrow tailoring, but it is hardly dispositive, as *Fisher* suggests. One must also consider the extent to which the instruments 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.[250] 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 used by the instruments are not so closely linked to those interests.

---

247. *See* Chapman v. California, 386 U.S. 18, 23-24, 26 (1967).

248. 133 S. Ct. 2411, 2434 (2013) (Ginsburg, J., dissenting) (quoting Fisher v. Univ. of Tex. at Austin, 631 F.3d 213, 262 (5th Cir. 2011) (Garza, J., specially concurring)) (internal quotation marks omitted).

249. *Id.* at 2416, 2418 (majority opinion).

250. *Id.* at 2418 (quoting Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 315 (1978) (opinion of Powell, J.)) (internal quotation mark omitted).

B. *Does Evidence-Based Sentencing 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. 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 merely allow them to do so more accurately.[251] 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.[252] These possibilities have 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 have already 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 decisionmaking 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.[253] If it does, it will almost surely increase judicial focus on the individual demographic and socioeconomic characteristics used in the EBS instruments, and those characteristics are not relevant to retributive motivations for punishment (or may even cut in 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

---

251. *See, e.g.*, MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a at 53, 55 (Tentative Draft No. 2, 2011); Branham, *supra* note 34, at 169; Kern & Bergstrom, *supra* note 35, at 185; Marcus, *supra* note 44, at 757; Denise Meyerson, *Risks, Rights, Statistics and Compulsory Measures*, 31 SYDNEY L. REV. 507, 513-15 (2009); Oleson, *supra* note 9, at 1373; Patton, *supra* note 244, at 1456; Jennifer Skeem, *Risk Technology in Sentencing: Testing the Promises and Perils*, 30 JUST. Q. 297, 298 (2013).

252. 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.

253. *See, e.g.*, Hyatt et al., *supra* note 34, at 266.

discount this factor. If such a judge is presented with a quantified risk assessment that is framed as scientifically established, they may well give it more weight.[254] In many legal, policy, and other contexts, scholars have observed that judges and other decisionmakers often defer both to scientific models that they do not really understand and to "expert" viewpoints.[255] Moreover, sentencing is high-stakes, complex decisionmaking that many judges describe as weighing heavily on their emotions,[256] rendering the use of a simple, seemingly objective algorithm potentially appealing.[257] For elected judges, research has shown that political considerations influence sentencing,[258] and reliance on risk predictions might provide political cover for release decisions while making it politically difficult to release offenders rated as high risk.[259] Prosecutors might

---

254. *See* Hannah-Moffat, *supra* note 66, at 277 ("Risk scores impart a sense of moral certainty and legitimacy into the classifications they produce, allowing people to accept them as normative obligations and therefore scripts for action." (internal quotation marks omitted)); Harcourt, *supra* note 227, at 273 (describing "the pull of prediction" (internal quotation marks omitted)).

255. *E.g.*, Kathryn M. Campbell, *Expert Evidence from "Social" Scientists: The Importance of Context and the Impact on Miscarriages of Justice*, 16 CAN. CRIM. L. REV. 13, 14 (2011); Robert L. Kane, *Creating Practice Guidelines: The Dangers of Over-Reliance on Expert Judgment*, 23 J.L. MED. & ETHICS 62, 63 (1995); 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 (2012) (discussing the problem of judicial overreliance on expert claims of causation); Janine Pierson, Comment, *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 about dangerousness in civil commitment hearings); *see also* Charles D. Case, *Problems in Judicial Review Arising from the Use of Computer Models and Other Quantitative Methodologies in Environmental Decisionmaking*, 10 B.C. ENVTL. AFF. L. REV. 251, 256 (1982) (describing courts' and policymakers' tendency to overrely on models and perceived expertise in the environmental context); Robert E. Schween & Steven P. Larson, *Groundwater Modeling: Capabilities and Limitations, Use and Abuse*, 32 ROCKY MTN. MIN. L. INST. 22-1 (1986) (same).

256. *See* 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."); D. Brock Hornby, *Speaking in Sentences*, 14 GREEN BAG 2d 147, 157 (2011); Oleson, *supra* note 9, at 1330 & n.2 (citing sources); Robert Pratt, *The Implications of* Padilla v. Kentucky *on Practice in United States District Courts*, 31 ST. LOUIS U. PUB. L. REV. 169, 169 (2011) ("Sentencing is unquestionably the most difficult job of any district court judge.").

257. This point may help to explain the continuing heavy weight federal judges give to the Federal Sentencing Guidelines, even though they are no longer required to follow them.

258. 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 a positive correlation between sentence length and the approach of elections).

259. Hannah-Moffat, *supra* note 66, at 290 ("The use of risk instruments can be particularly appealing to elected judges and prosecutors who must defend their decisions to an electorate concerned with security.").

similarly feel politically pressured to push for harsh sentences for offenders rated high risk but free to offer better deals to those rated low risk.[260]

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 his professional diagnostic skills than judges are in their ability to foresee a defendant's future.[261] Even if a particular judge does not really trust the instrument, its prediction might influence her thinking through anchoring.[262] 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 eighty-three criminal 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 conviction for underage drinking). 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 twenty-one-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—

---

260. *See id.*

261. *See, e.g.*, ATUL GAWANDE, THE CHECKLIST MANIFESTO: HOW TO GET THINGS RIGHT 158-62 (2009).

262. *See* Prescott & Starr, *supra* note 164, at 325-30 (reviewing anchoring research); Cass R. Sunstein et al., *Predictably Incoherent Judgments*, 54 STAN. L. REV. 1153, 1170-71 (2002).

whereas those inclined to rely on a defendant's characteristics to assess future dangerousness would likely give William a longer sentence.[263] Subjects were given a wide statutory sentencing range (0 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 would score -1 ("below average").[264] Subjects considered the scenarios in a prescribed, randomized order.

TABLE 2

An Experiment: Risk Prediction Instruments and Relative Sentence Outcomes

	Probability Low-Risk Def. Gets Higher Sentence[†]	Probability High-Risk Def. Gets Higher Sentence[†]	Rank-Order of Sentences[††]	Sentence Length in Years[†††]
RRPI Score Given	-0.603* (0.305)	0.710* (0.284)	0.662** (0.257)	-0.871 (0.733)
High-Risk				-0.711 (0.473)
High-Risk and RRPI Score Given				1.67* (0.61)

† Probit regressions of indicators for giving the "low-risk" or "high-risk" defendant, respectively, a higher sentence.

†† 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.

††† OLS regression with sentence length in years as the outcome. An indicator for which case the subjects considered first was also included. Standard errors in parentheses.

* Significant at 5% level.

** Significant at 1% level.

263. Students' comments after completing the exercise supported this interpretation.
264. *See* Wolff, *supra* note 30, at 112-13.

The results in Table 2 suggest that the RRPI score sharply affected the relative sentences some subjects gave to Robert and William. Among the forty-three students who were not given the RRPI score, seventeen gave Robert (the "low-risk" defendant) the higher sentence, thirteen gave them the same sentence, and thirteen gave William the higher sentence. Among the forty students who received the RRPI score, only eight gave Robert the higher sentence, nine gave them the same sentence, and twenty-three gave William the higher sentence.

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 (column 1) or that William would (column 2). These two are not just mirror image inquiries, given that 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 (column 3). Next, I used the recommended sentence, in years, as the outcome variable, an approach that estimates the magnitude and not just the direction of the sentencing distinctions between the defendants and the RRPI's effect on those distinctions (column 4). Although probit and ordered probit coefficients do not have a very intuitive interpretation, 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 when the RRPI score was provided than when it was not; the average sentence given to Robert was about 0.9 years lower when the score was provided.[265]

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 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. Moreover, although much experimental research on decisionmaking 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—the evidence is far more detailed, and

---

265. 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 or the effect of the RRPI.

the stakes are far higher. Moreover, judges are not law students—their experience and expertise (with EBS and with sentencing generally) may make them less suggestible. Still, it cannot be assumed that judges are wholly resistant to attempts to influence their sentencing decisionmaking. After all, judges still tend to defer to the nonbinding Federal Sentencing Guidelines, and research from other legal settings suggests that courts tend to defer to scientific expertise.[266] 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 the use of actuarial instruments in parole decisions, which is now routine in thirty states, including almost all of those that have not abolished discretionary parole.[267] This practice has been given little attention by legal scholars or the public[268] and has rarely been challenged in court, perhaps because of the absence of counsel in parole proceedings or because parole decisionmaking is not very transparent. Many prisoners may not even know

---

266. *See supra* note 255 and accompanying text.

267. HARCOURT, *supra* note 6, at 78-80.

268. Scholarly criticism has focused on procedural concerns—mainly the prisoner's lack of counsel at parole hearings. For this reason, the commentary to the revised 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). MODEL PENAL CODE: SENTENCING § 6B.09 cmt. a at 54 (Tentative Draft No. 2, 2011); *see also* McGarraugh, *supra* note 67 (advocating the elimination of the use of risk prediction instruments in the context of parole and the increased use of these same instruments in the context of sentencing).

that risk prediction instruments exist, much less understand how they work or their constitutional infirmities.[269] But while risk prediction unquestionably is properly central to the parole decision,[270] the use of demographic and socioeconomic variables to predict risk raises the same disparate treatment concerns that EBS does.[271] Moreover, the parole context may offer additional 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 the assignment of prisoners, probationers, and parolees to correctional and reentry programming (e.g., job training) and in the shaping of conditions of supervised release (e.g., drug tests).[272] In this context, risk assessment is often combined with instruments assessing criminological needs and predicting responsivity to various interventions. The empirical merits of such instruments are beyond this Article'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 the *Griffin* equality principle to socioeconomic classifications and other traits that are not treated as suspect outside the criminal justice

---

269. 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* note 67, at 1079 & n.5.

270. 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 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. *Cf.* Christopher Slobogin, *Prevention as the Primary Goal of Sentencing: The Modern Case for Indeterminate Dispositions in Criminal Cases*, 48 SAN DIEGO L. REV. 1127, 1128-30 (2011).

271. 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) (Stevens, J., dissenting) (internal quotation marks omitted). States have no obligation to provide a system of parole, but once they do, its operation is constrained by the Constitution. Bd. of Pardons v. Allen, 482 U.S. 369, 377 n.8, 378 n.10 (1987); Morrissey v. Brewer, 408 U.S. 471, 482 (1972).

272. *See* CASEY ET AL., *supra* note 36, at 16-20; Melissa Aubin, *The District of Oregon Reentry Court: An Evidence-Based Model*, 22 FED. SENT'G REP. 39, 42 (2009) (discussing evidence-based practice in federal reentry courts); Warren, *supra* note 13, at 155-56.

context. Distributing access to correctional programming based on risk, needs, and responsivity assessments is not particularly different from distributing access to noncorrectional 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. Although EBS and similar parole practices have so far escaped serious constitutional challenge, that could easily change, and it should. Defendants whose gender or socioeconomic status subject them to an adverse score that is taken into account in sentencing or parole decisions should consider challenging those decisions as the product of an unconstitutional process. Such challenges would be well grounded in Supreme Court doctrine.

Risk prediction is here to stay as part of sentencing, and perhaps actuarial instruments can play a legitimate role. But they should not include demographic and socioeconomic variables, which do not in any event offer very 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.