Journal of Criminal Law and Criminology

Volume 97
Issue 3 *Spring*

Article 1

Spring 2007

Using Groups Statistics to Sentence Individual Criminals: An Ethical and Statistical Critique of the Virginia Risk Assessment Program

Brian Netter

0091-4169/07/9703-0699
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY
Vol. 97, No. 3

Printed in U.S.A.

CRIMINAL LAW

USING GROUP STATISTICS TO SENTENCE INDIVIDUAL CRIMINALS: AN ETHICAL AND STATISTICAL CRITIQUE OF THE VIRGINIA RISK ASSESSMENT PROGRAM

BRIAN NETTER*

Virginia's program of nonviolent offender risk assessment uses predictions of recidivism to recommend which felons should be incarcerated. Unlike many sentencing schemes that rely upon the severity of the offense and the offender's criminal history, Virginia's depends on a statistical study commissioned by the legislature that purports to match offender characteristics with future behavior. New offenders are given recidivism "scores" that depend on gender, employment status, marital status, and age—all factors seemingly unrelated to the criminal conduct itself. This Essay criticizes the Virginia approach as ethically suspect and mathematically unsound and calls for greater public discourse as to the hidden assumptions underlying the sentencing apparatus.

I. INTRODUCTION

In an ideal crime-fighting world, we would know every convict's criminal proclivities. An offender could be detained for precisely the right amount of time as we effortlessly balanced the many competing interests served by our system of criminal justice. Of course, such a system would, of necessity, require invasions of civil liberties that even the toughest on crime might deem unreasonable.[1] Thus, our myriad schemes for criminal

* J.D., Yale Law School; B.S.E., M.S.E., University of Michigan. Thanks to Michael Maltz, Richard Kern, Nancy Gertner, Stephen Pollock, Doug Kysar, Kate Stith, Jed Rubenfeld, Wendy Netter Epstein, and Karen Dunn.

[1] *See* Norval Morris & Marc Miller, *Predictions of Dangerousness*, 6 CRIME & JUST. 1, 1 (1985) ("The use of predictions of dangerousness requires a political judgment balancing the risk and harm to society with the intrusion on the liberty of each member of a preventatively

sentencing rely upon proxies to estimate how best to achieve the goals of "deterring crime, incapacitating the offender, providing just punishment, and rehabilitating the offender."[2]

To enhance these proxies, considerable efforts have been devoted to developing tools to predict criminal behavior and criminological trends.[3] But these efforts fall far short of predictive accuracy. In light of these failings, there is a threshold ethical dilemma that we seldom consider: how good must predictive efforts be to justify using them to take restrictive actions that implicate the liberties of others?

The Commonwealth of Virginia sits at the forefront of predictive techniques. Anticipating that prison space would become scarce after implementing a truth-in-sentencing initiative to extend sentences for violent crime, Virginia has spent a decade studying how to use its prison space efficiently to reduce future crime. The goal has been to divide nonviolent criminal offenders into groups—those deemed most likely to recidivate are imprisoned while those assessed to pose less of a threat to society are given alternative sanctions.[4] On its face, such a program seems pedestrian and

---

detained group."). To state this predicament in more mathematical terms, "[a] defendant's right not to be a false positive must be balanced against the public's right not to be set on by a false negative." Don M. Gottfredson, *Prediction and Classification in Criminal Justice Decision Making*, 9 CRIME & JUST. 1, 13 (1987). In pop culture, this theme served as the basis for the blockbuster MINORITY REPORT (Cruise/Wagner Prods. et. al. 2002), which explored the consequences of burying false negatives in pursuit of a world without murder. A direct connection was drawn between Virginia's risk assessment program and *Minority Report* in Emily Bazelon, *Sentencing by the Numbers*, N.Y. TIMES, Jan. 2, 2005, § 6 (Magazine), at 18. *See also infra* Part III.A.2.

[2] U.S. SENT'G COMM'N, UNITED STATES SENTENCING COMMISSION GUIDELINES MANUAL § 1A1.1 (2006), *available at* http://www.ussc.gov/2006guid/gl2006.pdf.

[3] For example, the City of Chicago has recently launched a network of citywide surveillance cameras to alert the police to suspicious activities on the streets. *See* Andrew Buchanan, *On Chicago Streets, Cameras Are Watching*, CHRISTIAN SCI. MONITOR, July 30, 2003, at 1. Also, behavioral geneticists have worked to determine what biological indicators signal a predisposition to criminality. *See, e.g.*, Erik Stokstad, *Psychology: Violent Effects of Abuse Tied to Gene*, 297 SCI. 752 (2002) (summarizing the results of research that a certain genetic characteristic leads men to be more violent if they were abused as children); Avshalom Caspi et al., *Role of Genotype in the Cycle of Violence in Maltreated Children*, 297 SCI. 851 (2002) (finding that the likelihood a maltreated child will develop antisocial behavior depends on gene characteristics).

[4] In Virginia, alternative sanctions include release with intensive supervision, home electronic and telephonic monitoring, day reporting, boot camp, diversion centers, detention centers, and additional local sanctions. BRIAN J. OSTROM ET AL., NAT'L CTR. FOR STATE COURTS, OFFENDER RISK ASSESSMENT IN VIRGINIA 20-22 (2002). The vast majority of diverted offenders are sentenced to supervised probation. VA. CRIM. SENT'G COMM'N, 2003 ANNUAL REPORT 69 fig.46 (2003) [hereinafter 2003 ANNUAL REPORT] (showing sentence of supervised probation for 80.6% of diverted offenders).

unremarkable. After all, judges routinely make determinations about offender riskiness during sentencing hearings.[5] But Virginia chose not to rely on the expertise of judges for this initiative; instead, the Commonwealth commissioned a risk-assessment tool, based on a statistical study of recidivism, to make an initial division between the dangerous destined for prison and those others who could be welcomed back into society without visiting the state penitentiary. The system relies upon simple worksheets that tally demerits for past crimes with additional penalties for demographic characteristics found to be correlated with the commission of crime. Thus, a young, unemployed, never-married man is considerably more likely to face jail time than an older, divorced woman who held a job prior to committing an identical crime.

The Virginia risk assessment program presents serious issues regarding the ethical propriety of predictive techniques in criminal sentencing. An evaluation of the program forces us to ask what classes of information we are comfortable considering when penalizing criminal misdeeds. Our answers to these questions are influenced—at least in part—by the effectiveness of predictive techniques and the hidden assumptions that govern their conclusions.[6]

This Essay evaluates Virginia's risk assessment program as a means to probe our (dis)comfort with the use of predictive group statistics on individual criminals and to determine whether the statistical techniques adopted in Virginia are suitable for this important task. Part II gives the history of risk assessment in Virginia. Part III discusses the ethics of predictive sentencing of this sort. Part IV critiques the methodology of Virginia's approach and the errors that are introduced therein. Part V concludes the analysis.

---

[5] *See, e.g.*, Jurek v. Texas, 428 U.S. 262, 275 (1976) ("[A]ny sentencing authority must predict a convicted person's probable future conduct when it engages in the process of determining what punishment to impose."). This judicial competency can be seen in related contexts as well. In determining whether mental health sufferers should be incapacitated after criminal sentences have lapsed, the Supreme Court has found that the expected dangerousness of the detainee must be considered. Jones v. United States, 463 U.S. 354, 363-66, 368-69 (1983); *see also* Mark H. Moore, *Purblind Justice: Normative Issues in the Use of Prediction in the Criminal Justice System*, *in* 2 CRIMINAL CAREERS AND "CAREER CRIMINALS" 314, 314 (Alfred Blumstein et al. eds., 1986) (noting that judges must frequently confront risk in the context of bail hearings); Leslie T. Wilkins, *The Politics of Prediction*, *in* PREDICTION IN CRIMINOLOGY 34, 35 (David P. Farrington & Roger Tarling eds., 1985) ("It may seem absurd to ask whether human behavior should be predicted, because all persons predict the behavior of others all the time.").

[6] As Leslie Wilkins aptly observed, "[i]n the development of prediction instruments it is necessary, but not sufficient, to seek the most efficient set of equations." Wilkins, *supra* note 5, at 49.

II. THE PATH TO RISK ASSESSMENT IN VIRGINIA

Virginia's experimentation with actuarial risk assessment in criminal justice has been marked by incremental change and paced implementation. The project began in earnest in 1994, when the General Assembly created the Virginia Criminal Sentencing Commission (VCSC or Commission) and instructed it to study the feasibility of placing twenty-five percent of nonviolent felons into alternative arrangements by determining who among the newly convicted posed the smallest risk to society.[7]

This initiative was a direct consequence of Virginia's truth-in-sentencing reforms. After George Allen won the 1993 gubernatorial race largely by promising to eliminate parole and to increase penalties for violent crime, a commission recommended the widespread redrafting of Virginia's criminal statutes.[8] Unsurprisingly, with violent criminals serving as little as twenty-nine percent of their sentences for crimes as serious as first-degree murder,[9] there was pressure from both the public and Congress[10] to keep violent criminals behind bars for longer terms. Under Virginia's truth-in-sentencing program, violent convicts must serve at least eighty-five percent of their sentences, and in the first decade of the program, the average criminal served fully ninety percent.[11] Over that same decade, from 1994 to 2003, Virginia lowered its incidence of murder by twenty-eight percent.[12]

While implementing truth-in-sentencing, Virginia was understandably concerned that increased prison sentences would lead to widespread prison overcrowding. In light of their instincts "[t]o reserve the most expensive resources for the most dangerous offenders, reformers underscored the importance of making the most efficient use of the state's remaining correctional resources to punish nonviolent offenders."[13] The newly-

---

[7] OSTROM ET AL., *supra* note 4, at 9 n.1; *see also* VA. CODE ANN. § 17-235(5) (1995), *amended by* VA. CODE ANN. § 17.1-803(5) (2006).

[8] *See* BRIAN J. OSTROM ET AL., NAT'L CTR. FOR STATE COURTS, TRUTH-IN-SENTENCING IN VIRGINIA 4 (1999).

[9] *Id.* at 28 (reporting statistics for 1988-1992).

[10] Under the Violent Crime Control and Law Enforcement Act of 1994, Pub. L. No. 103-322, tit. II, 108 Stat. 1796 (1994) (codified at 42 U.S.C. § 13703 (2000)), truth-in-sentencing incentive grants were made available for states in which violent offenders would serve at least eighty-five percent of their sentences. *See* OSTROM ET AL., *supra* note 8, at 5.

[11] VA. CRIM. SENT'G COMM'N, A DECADE OF TRUTH-IN-SENTENCING IN VIRGINIA, No. 1 (2004), *available at* http://www.vcsc.state.va.us/ReptCdPDFfinal.pdf.

[12] *Id.* at No. 8. Of course, this statistic does not establish causation.

[13] VA. CRIM. SENT'G COMM'N, 2004 ANNUAL REPORT 44-45 (2004) [hereinafter 2004 ANNUAL REPORT].

created VCSC's charge, then, was to find ways to keep unlikely recidivists out of the system.

It bears mentioning, however, that this utilitarian-driven initiative conflicts with the underpinnings of truth-in-sentencing. The latter—also adopted by the federal government and several other states[14]—springs from a notion of "just deserts."[15] Violent sentences were increased and judicial discretion was largely curtailed under the theory that imprisonment is a "deserved punishment rather than . . . a means for rehabilitation and treatment."[16] Conversely, the risk assessment pilot was grounded in a different, utilitarian theory of criminal law. It focused neither on punishing the guilty nor on rehabilitating the rehabilitatable. Rather, the question presented to the VCSC was the following: "Which nonviolent felony offenders are at a low risk of re-offending and can thus be safely placed in alternative sanction programs?"[17]—i.e., how can sentencing procedures be designed efficiently to protect the public-at-large?

In developing its approach, Virginia confronted a number of initial decisions; namely, whether to conduct risk assessment or needs assessment and whether to rely upon clinical risk assessment or statistical risk assessment.

In deciding between risk assessment and needs assessment, Virginia selected the former. The differences are meaningful. Risk assessment determines sentencing based only on the probability that a criminal will recidivate.[18] Although there is considerable disunity as to the definition of recidivism,[19] this approach focuses on the public and its instrumental interest in avoiding crime. Needs assessment, on the other hand, emphasizes an individual offender's need for services, so the potential effectiveness of treatment, counseling, or structured confinement can be incorporated into penal decisionmaking.[20] The VCSC, interpreting its mandate as limited to the "risk to public safety," chose to conduct only risk assessment.[21] Thus, the tool ultimately developed by the VCSC makes only a binary determination as to whether an offender should be sentenced to confinement in the penitentiary or whether some unspecified form of

---

[14] *See* PAULA M. DITTON & DORIS JAMES WILSON, BUREAU OF JUST. STATS., TRUTH IN SENTENCING IN STATE PRISONS (1999).

[15] OSTROM ET AL., *supra* note 8, at 5; *see also infra* Part III.

[16] *Id.*; *see also* ANDREW VON HIRSCH, DOING JUSTICE: THE CHOICE OF PUNISHMENTS (1976).

[17] OSTROM ET AL., *supra* note 4, at 10.

[18] *Id.* at 43-44.

[19] *See infra* note 112 and accompanying text.

[20] OSTROM ET AL., *supra* note 4, at 43-44.

[21] *Id.* at 43.

alternative sanction is objectively warranted.[22] The subjective determination of how best to accommodate eligible offenders' needs is made by judges and/or probation officers without any reference to statistics.[23]

Despite the preference for clinical judgment in deciding *between* alternative sanctions, Virginia chose to implement statistical risk assessment to determine *whether* alternative sanctions would be recommended. Clinical models rely upon intuition and discretion often left in the hands of judges or medical professionals.[24] Conversely, statistical classification relies upon actuarial techniques to objectify the risks of future misconduct.[25] Inevitably, actuarial models cause individuals to suffer the consequences of the groups to which they belong. Nonetheless, various studies have found that actuarial methods can achieve better results than their clinical counterparts.[26] The Virginia legislature mandated a statistical approach[27] and the VCSC followed course.

In pursuit of the legislated goal, the VCSC studied a random sample of roughly 1500 fraud, larceny and drug offenders who had each been incarcerated and released between July 1, 1991, and December 31, 1992.[28] The Commission tracked these subjects and tallied who among them was reconvicted for another felony within three years of initial release by mining data in a Pre/Post-Sentence Investigation Database.[29] Using multivariate logistic regression to study the correlation between personal characteristics and recidivism among those eligible for risk assessment,[30]

---

[22] *Id.* at 3-4, 36-38, 42.

[23] *Id.*; *see also* 2003 ANNUAL REPORT, *supra* note 4, at 42, 66.

[24] OSTROM ET AL., *supra* note 4, at 28.

[25] *Id.* Actuarial techniques are commonplace in society; they form "the basis of all insurance and of a great many of our efforts to share and shift risk in the community." Morris & Miller, *supra* note 1, at 14. On the differences between clinical and statistical predictions generally, see THOMAS GABOR, THE PREDICTION OF CRIMINAL BEHAVIOUR: STATISTICAL APPROACHES (1986).

[26] *See* Stephen D. Gottfredson & Don M. Gottfredson, *Accuracy of Prediction Models*, *in* 2 CRIMINAL CAREERS AND "CAREER CRIMINALS," *supra* note 5, at 247 ("[I]n virtually every decision-making situation for which the issue has been studied, it has been found that statistically developed predictive devices outperform human judgment."); Morris & Miller, *supra* note 1, at 1. Note that this objective correlation does not speak to the ethicality of this guilt by association.

[27] *See supra* note 7 and accompanying text.

[28] 2003 ANNUAL REPORT, *supra* note 4, at 65.

[29] *Id.*

[30] Eligibility for risk assessment was limited to fraud and larceny offenders and drug offenders except those convicted for selling an ounce or more of cocaine. *See* OSTROM ET AL., *supra* note 4, at 15 & n.4.

the Commission developed a worksheet incorporating eleven factors and their expected contributions to recidivist tendencies.

Figure 1

Factors Used To Predict Recidivism[31]

Current Offense Information:
1. Offender gender
2. Offender age
3. Offender marital status
4. Offender employment status
Current Offense Information:
5. Whether the offender acted alone when committing the crime
6. Whether there were additional offenses at conviction
Prior Adult Criminal Record:
7. Whether the offender had been arrested or confined within the past 12 months
8. Offender's prior criminal record
9. Whether the offender had prior drug felony convictions
10. Whether the offender had been incarcerated as an adult
Prior Juvenile Record:
11. Whether the offender had been incarcerated as a juvenile

The Commission's study revealed that race was statistically significant but chose to omit this variable from the set of predictors because "the Commission viewed race as a proxy for social and economic disadvantage."[32]

The risk assessment worksheet is a mandatory component of sentencing for nonviolent offenders. Judges, however, have discretion to deviate from the recommendations by "stat[ing] in writing the reason for the departure."[33] In 2004, eighty-one percent of sentences conformed to the worksheet recommendation, with the deviations roughly split between

[31] *Id.* at 27; *see also, e.g.*, *id.* at 123 (providing the risk assessment worksheet for fraud). Note that these factors were amended after a validation study. *See infra* Part IV.B.2.

[32] 2003 ANNUAL REPORT, *supra* note 4, at 65-66.

[33] *See* 2004 ANNUAL REPORT, *supra* note 13, at 46.

upward and downward adjustments.[34] The program began as a voluntary pilot in scattered judicial circuits across Virginia; after judges were interviewed, the Commission conducted a validation study to test and refine the risk assessment instrument in 2001.[35] The revised instrument was then implemented statewide in 2002, with those offenders scoring thirty-five points or fewer on the worksheet receiving a favorable recommendation.[36] The threshold point level for favorable recommendations was raised to thirty-eight in July 2004.[37]

III. ETHICAL AND CONSTITUTIONAL CONCERNS

> Classification and prediction methods must rest on a basis in science, but they must be applied in a world of values.[38]

Virginia's novel venture into risk-based sentencing lies at the intersection of a variety of thorny ethical issues. This nexus is unsurprising, as "[p]eople have been arguing about the justifications for punishment and for the criminal law for thousands of years."[39] These issues remain unresolved and this Essay will not present a definitive solution. However, when assessing a legislative mandate such as this, it is prudent to consider the relevant reference points in the philosophical debate.

Part III examines four of the significant ethical issues raised by offender risk assessment. Section A looks at the appropriateness of punishing on the basis of hypothesized future crime. Section B considers the significance of saddling individuals with the burdens of generalizations on the groups to which they belong. Section C addresses the feedback effects that differential sentencing could have on criminal deterrence, leading to moral hazard. Section D questions whether the fundamentality of the prison/no-prison decision affects our conclusions.

A. PUNISHING FOR FUTURE CRIME

In our criminal justice system, defendants are presumed innocent until proven guilty and have a fundamental right to hold the state to its high

---

[34] *See id.* at 47.

[35] 2003 ANNUAL REPORT, *supra* note 4, at 66-67; *see also* VA. CRIM. SENT'G COMM'N, 2001 ANNUAL REPORT 55-61 (2001) [hereinafter 2001 ANNUAL REPORT] (announcing the results of the validation study).

[36] *See, e.g.*, Bazelon, *supra* note 1.

[37] *Id.*; *see also* 2003 ANNUAL REPORT, *supra* note 4, at 74 (recommending the change).

[38] Gottfredson, *supra* note 1, at 13.

[39] Michael Tonry, *Prediction and Classification: Legal and Ethical Issues*, 9 CRIME & JUST. 367, 368 (1987).

burden of proof.[40] Under Virginia's program of offender risk assessment, conclusions must be drawn about *future* behavior. These trends are speculative at best, so proof beyond a reasonable doubt is, in this context, unattainable.[41] The question of whether it is proper to attach penalties to such unprovable conduct goes to the most foundational postulates of criminal law and to the distinction between utilitarian and deontological considerations.[42]

The risk assessment program can be justified only if base utilitarianism is appropriate in the sentencing context. To further address the issue, this section is divided into four subsections. Subsection 1 presumes, counterfactually, that our predictive capabilities are perfect and questions whether predictive techniques are prima facie defensible. Subsection 2 introduces the inevitability of error and examines the consequences. Subsection 3 looks at whether the retributive underpinnings of Virginia's truth-in-sentencing can be reconciled with the utilitarian nature of statistical risk assessment. Subsection 4 questions whether it is problematic that the selection biases that drive criminal law enforcement carry through to sentencing.

1. Theories of Criminal Punishment

The goals of our correctional system are many. These goals relate to offenders, society, and the constraints of the correctional institutions.[43] Most criminal sanctions reflect a need to balance these constituencies,[44] but in the case of prediction, such balancing is tricky as two key theories come into conflict. Do we punish because criminals ought to receive their "just deserts" or because they are cogs in a system designed to minimize dangers to society? Virginia's approach is defensible only under the latter.[45]

---

[40] *See, e.g.*, Blakely v. Washington, 542 U.S. 296, 301-02 (2004); *In re* Winship, 397 U.S. 357, 361-62 (1970).

[41] *See* Tonry, *supra* note 39, at 391-97.

[42] *See generally* Moore, *supra* note 5; Tonry, *supra* note 39.

[43] MICHAEL D. MALTZ, RECIDIVISM 8-19 (1984).

[44] *See, e.g.*, 18 U.S.C. § 3553(a) (2000) (setting out the "[f]actors to be considered in imposing a sentence," notably including both "the need for the sentence imposed . . . to reflect the seriousness of the offense" and "the need . . . to protect the public from further crimes of the defendant").

[45] *See* David P. Farrington, *Predicting Individual Crime Rates*, 9 CRIME & JUST. 53, 88-90 (1987).

a. Retributivism

Under a retributivist, or "just deserts," theory of punishment, the criminal law acts to enforce moral norms; any deterrence is incidental.[46] The strictest versions of retributivism, attributable to Kant and Hegel, would allow nothing more than the details of the instant crime to be used in determining the instant sentence.[47] Once an offender repaid his debt to society, he would regain equality with non-offending citizens.

Retributivist theory has a number of powerful virtues. Punishments qualify as both horizontally equitable—similar crimes are punished similarly—and vertically equitable—more serious crimes are punished more harshly.[48] In addition, retributivist punishments are predictable, so we can see ourselves merely as doling out the consequences that informed criminals knew would attach to their wrongful acts.[49] This makes deterrence easier to calibrate.

Notwithstanding these attributes, few scholars find strict retributivism workable. Even the leading scholar of "just deserts" punishment, Andrew Von Hirsch, agrees that prior bad acts are relevant because this use of past data still represents "a deserved penalty based on the seriousness of his past criminal conduct."[50] Von Hirsch and others would draw a line between past bad acts and predicted future misconduct.[51] But this distinction does not survive a quick test for philosophical consistency. If a misdeed is supposed to carry a certain penalty, why should it have any force if an ex-convict recidivates after the penalty has been paid? Some scholars contend that the principal reason for incorporating past acts is because repeat offense is a powerful predictor of career criminality.[52] This, no doubt, is true, but making this predictive judgment seriously weakens the distinction between morals-based and consequentialist punishment.

---

[46] Stephen D. Gottfredson & Don M. Gottfredson, *Selective Incapacitation?*, 478 ANNALS AM. ACAD. POL. & SOC. SCI. 135, 137 (1985). In fact, retributivism may be the only theory of punishment for which prediction is not relevant.

[47] Wilkins, *supra* note 5, at 45; *see also* 1 CRIMINAL CAREERS AND "CAREER CRIMINALS," *supra* note 5, at 8.

[48] *See* 1 CRIMINAL CAREERS AND "CAREER CRIMINALS," *supra* note 5, at 77-83.

[49] *See* Moore, *supra* note 5, at 319.

[50] Von Hirsch, *supra* note 16, at 98-100.

[51] *See, e.g.*, PETER W. GREENWOOD WITH ALLAN ABRAHAMSE, SELECTIVE INCAPACITATION ix (1982) ("One could argue that 'deserts' is the predominant basis for current sentencing decisions.").

[52] *See* Wilkins, *supra* note 5, at 45.

b. Utilitarianism

Utilitarian theory takes a pragmatic approach to punishment. A criminal should be punished consistent with the extent to which future crime will be prevented. At the extreme, a strict utilitarian would find no justification for punishing a criminal at all if that punishment would not prevent subsequent crimes.[53] In the context of Virginia's risk assessment program, utilitarianists would have no problem designing sentences with reference to future impacts of correctional decisions.

In the context of prediction of future behavior, there are two utilitarian approaches that merit discussion: selective incapacitation and selective deinstitutionalization. Using either theory to assess Virginia's action results in the same critique but for different reasons.

"Selective incapacitation" refers to "sentencing policies that attempt to distinguish between higher-rate and lower-rate criminal offenders in determining who will be incarcerated and for how long."[54] Proponents of this approach presuppose that the goal of criminal justice is to maximize "incapacitation effects"—defined as "those crimes prevented while offenders are incarcerated"—subject to capacity constraints.[55] Choosing which subjects to incarcerate and which to set free based on a consideration of all statistically relevant factors allows for better results. Advocates dismiss retributivist theory as quaint but impractical because we have always made such judgments: "[M]en serve longer terms than women . . . ; defendants with prior records are more likely to be incarcerated than those without. . . . We are simply offering a more rational method for distinguishing among offenders."[56] Thus, as Reagan Administration

---

[53] Tonry, *supra* note 39, at 386. Of course, it would be a rare circumstance in which a serious crime would be best left unpunished. Even if a criminal is certain never to commit another crime, there are second-order deterrence effects on bystanders who may be considering similar offenses. However, "[t]he lack of relationship between crime seriousness and risk of future crime has ominous significance for corrections policymakers. . . . It is possible to optimize either just deserts or crime control, but not both simultaneously." Todd Clear, *Statistical Prediction in Corrections*, 1 RES. IN CORRECTIONS 1, 6 (1988).

[54] GREENWOOD WITH ABRAHAMSE, *supra* note 51, at 27.

[55] *Id.* at x, 29.

[56] *Id.* at 29-30. James Q. Wilson has been similarly skeptical of this retributivist argument:

> It is not enough to say, in opposition to selective capacitation, that it involves predicting behavior, as if that were intolerable and never done. The entire criminal justice system is shot through at every stage (bail, probation, sentencing, and parole) with efforts at prediction, and necessarily so; if we did not try to predict, we would release on bail or on probation either many more or many fewer persons, and make some sentences either much longer or much shorter.

JAMES Q. WILSON, CRIME AND PUBLIC POLICY 279 (1983).

National Institute of Justice Chief James Stewart maintained, "'[t]he most effective way to treat these criminals[] . . . is to put them in prison for the likely length of their criminal careers.'"[57] Critics complain that this approach can never be sufficiently accurate[58] and that punishment should not depend on characteristics that are unrelated to the blameworthiness of the offender.[59]

Whereas selective incapacitation deals with the initial decision of whether to incarcerate, selective deinstitutionalization does not. The latter approach, advanced by Gottfredson and Gottfredson, supports the use of risk-based analysis only when, for reasons of overcrowding or emergency, offenders who are already serving their sentences must be released.[60] The authors find a "fundamental difference" between Greenwood's selective incapacitation approach and their own method.[61] To be sure, selective deinstitutionalization never results in the heightening of sentences beyond their original deserts-based ideal; this approach claims merely to be a concession to necessity.

Other utilitarian scholars have found a middle ground—supporting the use of risk assessment to mitigate but not to aggravate sentences.[62] Perhaps this approach can be seen as allowing a deserts theory to make the first estimate of the proper sentence, but then authorizing downward adjustments in the name of mercy.[63]

In any case, the distinction between selective incapacitation and selective deinstitutionalization is not altogether meaningful in evaluating Virginia's sentencing apparatus. By setting an initial twenty-five percent threshold, Virginia created a zero-sum game. One offender's reprieve can come only at the expense of another inmate's incarceration. By setting this fixed target, alternative sanctions are less a prize and more a competition invoking all applicable principles of equity and liberty.

---

[57] *Key to Criminals' Futures: Their Pasts*, U.S. NEWS & WORLD REP., Oct. 18, 1982, at 17.

[58] *See infra* Part III.A.2.

[59] *See* Gottfredson & Gottfredson, *supra* note 46, at 142.

[60] *Id.* at 147-49.

[61] *Id.* at 149.

[62] Morris & Miller, *supra* note 1, at 6.

[63] *Id.* at 4 ("A merciful and just system of punishment presupposes leniency toward those who least threaten social injury; and this, in turn, inexorably involves predictions of dangerousness."); *see also* Jackson v. State, 329 N.W.2d 66 (Minn. 1983); State v. Magnan, 327 N.W.2d 147 (Minn. 1983) (proscribing the use of predictions of future dangerousness to impose lengthier prison sentences than Minnesota's sentencing guidelines would otherwise prescribe).

2. The Problem of Error

Even if predictive techniques were perfect, there would be moral questions about their propriety. Yet the reality of risk-based assessment is that it involves considerable error.[64] The error comes in two varieties: false positives and false negatives. In this context, a false positive is a person predicted to recidivate who is therefore incarcerated but, in reality, would have stayed crime-free. The false positive, then, spends time in prison that may be considered undeserved or at least inefficient. False negatives, conversely, are offenders predicted not to recidivate who do re-offend.[65] Society suffers harm for a false negative, whereas the burden of false positives falls squarely on the offender. Interestingly, the VCSC reports that "[a]n error that results in diverting an offender who then reoffends (false negative) is considered the more serious because it can endanger public safety."[66] However, "[t]his common practice results in unfair constraints being placed on low-risk offenders and wasted departmental resources."[67]

The problem of error brings to mind the famous assertion of William Blackstone that "it is better that ten guilty persons escape, than that one innocent suffer."[68] Nobody here is innocent in a strict sense, but there are parallels to those who are innocent of future crimes and therefore undeserving of punishment stemming therefrom.[69] Blackstone would likely choose a different ratio than his ten-to-one baseline, but if wrongful punishment for past crimes uncommitted is so serious as to justify leaving wrongdoers on the streets to endanger society, there must be a threshold across which "wrongful" punishment for mis-predicted future crimes becomes unpalatable, even if establishing this threshold means releasing some wrongdoers-to-be onto the streets. Predictably, some scholars have riposted that a moral line divides those who have been found guilty and those who have not.[70] Otherwise, why would we not adopt pre-criminal justice and pre-penalize our most dangerous?[71] We surely would be

---

[64] *See infra* notes 72-75 and accompanying text.

[65] *See* MINORITY REPORT, *supra* note 1 (a pop-culture commentary on the perils of hidden false negatives).

[66] OSTROM ET AL., *supra* note 4, at 30.

[67] Alex M. Holsinger et al., *Practitioners' Guide to Understanding the Basis of Assessing Offender Risk*, 65 FED. PROBATION 46, 47 (2001).

[68] WILLIAM BLACKSTONE, 4 COMMENTARIES *352.

[69] *See* Alfred Blumstein et al., *Delinquency Careers: Innocents, Desisters, and Persisters*, 6 CRIME & JUST. 187, 217 (1985); Tonry, *supra* note 39.

[70] *See* Wilkins, *supra* note 5, at 35.

[71] *See supra* note 1.

uncomfortable using this trigger metaphor to impose lifetime imprisonment on a first-time petty thief.

The inevitable errors of predictive punishment are quite substantial. Even the pioneers of risk-based recidivism analysis conceded that while their approaches were theoretically defensible, implementation was unjustifiable using current statistical methods. As Schmidt and Witte noted:

> Our levels of false positive and false negative results are high enough that it would be very hard to justify differential treatments on the basis of our predictions at least for more than a very small percentage of the sample. . . . [W]e believe that the point is basically moot until models are developed that predict more accurately than is currently possible.[72]

Schmidt and Witte were skeptical that any more than thirty percent of the error would *ever* be explainable.[73] Other studies have been similarly skeptical about error rates.[74] In the National Center for State Courts' ("NCSC") review of the VCSC's approach, it found a 24.3% recidivism rate (versus a 31.5% correct detection rate) among those scoring below the threshold for recommended alternative sanctions.[75] The peril is that false positives are empirically incalculable because potential false positives are in jail. But given that the false negative rate is consistent with Greenwood and Abrahamse's rate, there is no reason to believe that the false positive rate would not be similar, which would suggest an overall rate of error approaching one-half.[76]

Some scholars find this error even more troubling when it reflects estimations about the future rather than misjudgments about the past, because "[o]ffenders incorrectly predicted to commit crimes in the future would be exposed to criminal liabilities that are doubly undeserved: once because they were based on predictions rather than past deeds, and twice because the predictions were inaccurate."[77] If these errors really do

---

[72] PETER SCHMIDT & ANN DRYDEN WITTE, PREDICTING RECIDIVISM USING SURVIVAL MODELS 5-6, 120 (1988) (finding a false positive rate of forty-seven percent and false negative rate of twenty-eight percent).

[73] *Id.* at 14. In fact, "statistical risk assessment devices rarely explain more than twenty percent of the variance (i.e., differences among offenders) in criterion measures." Holsinger et al., *supra* note 67, at 47.

[74] *See, e.g.*, GREENWOOD WITH ABRAHAMSE, *supra* note 51, at xix-xx (finding a false positive rate of fifty-six percent and a false-negative rate of forty-six percent for their selective incapacitation study: "As long as our ability to discriminate between high- and low-rate offenders is imprecise, there will be legitimate concern about those who are improperly classified."); *see also* Morris & Miller, *supra* note 1, at 5 ("[T]hat our prediction capacities are poor[] is a regrettable truth.").

[75] OSTROM ET AL., *supra* note 4, at 72 fig.5.13 (measuring recidivism by rearrest).

[76] *See supra* note 74.

[77] Moore, *supra* note 5, at 315.

approach fifty percent, then this systemic dysfunctionality lends additional credence to the retributive theories of punishment.

3. The Impact of Truth-in-Sentencing

Comparing the relative benefits of retributivism and utilitarianism would be an academic exercise but for Virginia's existing fealty to retributivism. The truth-in-sentencing reforms that directly charted the course for risk assessment in Virginia are grounded in deserts theory.[78] It makes little sense for violent crimes to be governed by one moral baseline and nonviolent crimes to be regulated by another.

By piloting risk-based sentencing at the same time that the governor was preaching the importance of serving time for wrongful acts, the Virginia legislature sent a mixed message. To be sure, there are obvious meaningful distinctions between violent and nonviolent crimes. Also, of course, no state has a truly consistent doctrinal formulation for its criminal sentencing policy. However, it requires tortured logic to reconcile the two programs. Perhaps we could characterize the whole system as utilitarian with the added assumption that all violent acts produce criminals likely to recidivate. Yet the consistency that this empirical assertion would require is implausible.

The biggest problem with this conflict is that many criminals will start their careers with non-violent offenses and will perhaps get the wrong impression about the justice system's purpose. Instead of emphasizing liberty and punishment for wrongful deeds, the state can appear manipulative if the machinery of justice is seen merely as ends-driven.

4. Enforcement Effects and Circularity

A final problem with assessing current punishment based on expected future conduct is that detection of future conduct is strongly dependent on how we enforce the criminal code, and criminal enforcement is far from uniform. Criminals are more likely to be caught in neighborhoods with additional surveillance. These neighborhoods are disproportionately populated by the economically disadvantaged and racial minorities.

We tolerate this differential enforcement because all those who are convicted have, in fact, committed a crime, even if others have committed the same crime undetected. However, Virginia adds a layer of complexity

---

[78] *See* OSTROM ET AL., *supra* note 8, at 5 ("[Truth in sentencing] in deeply rooted in the determinate sentencing philosophy that dominated the 1980s. . . . The determinate model is based on a 'just deserts' philosophy . . . ."); *see also* 2004 ANNUAL REPORT, *supra* note 13, at 44-45; *supra* text accompanying notes 15-16.

by questioning the frequency of recidivism, which is defined as rearrest or reconviction for any misdemeanor or felony.[79] Since these later events are also functions of enforcement, there is a feedback effect. Those who are more likely to be caught are more likely to serve longer sentences, not just because they are *actually* more likely to commit crimes, but in part because the police are paying more attention to them. This phenomenon can falsely make certain demographic characteristics—like age or personal economics—appear to be more strongly correlated with crime commission than they actually are, because younger people and poorer people are more likely to live in areas of heavy police presence.

B. PUNISHING INDIVIDUALS FOR GROUP CHARACTERISTICS

An additional complication in any risk-based sentencing program is the determination of factors to include in the consideration. As set out in Figure 1, Virginia found such factors as gender, age, and past criminal history to be relevant, but left out characteristics such as race and annual income,[80] which may in reality be correlated with recidivism. As more distinctive characteristics are added to the model, the model becomes theoretically more accurate, but practically more difficult to populate because the number of people in each subgroup obviously decreases as the number of groups increases.

The statistical approaches employed by the VCSC are not intended to make predictions about individuals. Statistical predictions speak to group tendencies, not individual proclivities.[81] To be sure, we regularly ignore this point in the context of insurance, where individuals voluntarily assume a standardized risk corresponding to a group statistic. But forcing convicts to be saddled by the actions of others who are similarly aged, similarly employed, or similarly unmarried invokes principles of equity and justice that do not appear in the insurance context.

This Section addresses two issues. Subsection 1 considers the general flaws of applying group-based analysis to individual members. Subsection 2 looks to specific factors that raise particular concerns because of their lack of a relationship to the offense committed and to social norms.

---

[79] *See* 2001 ANNUAL REPORT, *supra* note 35, at 49, 53.

[80] *See supra* Figure 1.

[81] Morris & Miller, *supra* note 1, at 18; James Rowland, *A Review from the Practitioner's Perspective*, 1 RES. IN CORRECTIONS 47, 49 (1988).

1. When Individuals Don't Fit the Mold

A key reason why the error rates are so large in risk assessment is that key variables are always omitted and it is virtually impossible to distinguish between mere correlation and true causation. Variable omission is troubling in these contexts. This phenomenon is perhaps best explained by example. Suppose that "amount of time spent outside the house" ("time out" for short) is a key predictor of recidivism. Those who spend more time at bars and clubs are more likely to become associated with a criminal element and are more likely to commit crimes themselves. Suppose further that the "time out" variable is difficult to measure and that nobody thought to include it in the model. Does this mean that Virginia has merely lost some predictive value by not incorporating the time-out variable? In reality, the damage is considerably more extensive.

One of the variables in the Virginia model is marital status—specifically, whether an offender has ever been married.[82] It is probably safe to assume that those who are married spend less time at bars and clubs, so if time out is the true, but omitted, predictor, and marital status is correlated with that variable, then marital status will show up as a significant factor. However, it may well be that one's marital status, considered independently, has nothing to do with criminality. There just happen to be more single people who are out on the town. But by using the correlated variable (marital status) instead of the true variable (time out), those single people who do not frequent bars are encumbered with a burden they do not deserve.

Ideally, we could establish strong relationships between predictive variables and future criminality. Causation may be a misleading term in this circumstance (does a previous conviction really *cause* you to commit crime, or is this relationship merely a proxy for career criminality?), but the broader issue is whether key variables have been omitted and whether this omission raises ethical issues when employing a system where omitted variable bias remains a problem.

Even without statistical risk assessment, however, we regularly rely on heuristics and biases—accurate or not—to make judgments. So it may be unfair to critique statistical risk assessment on these grounds simply because this methodology formalizes an approach that judges apply informally. Nonetheless, this concern signals the importance of the variables chosen and of the thoroughness required to develop the correct set of predictors.

---

[82] *See supra* Figure 1.

2. Which Characteristics Are Off Limits?

Even if an ideal set of predictors could be developed to accurately assess an offender's risk of recidivism, there are some types of characteristics that we should be wary to include. There is almost unanimous consensus that past crimes are acceptable considerations in punishment because of their obvious relation to criminality, culpability, and dangerousness. So, too, do we consider the nature of the instant offense and any showings of remorse. Just as in the Virginia model, certain crimes are penalized more extensively when part of a conspiracy, as this connection may signal career criminality. If a model could be crafted based only on these criminologically-based variables, few would complain.

However, to state a broad countervailing principle, "[m]any people believe it unjust to base punishment decisions on factors over which the offender has no control."[83] Thus, even though violent crime is considerably more prevalent among males[84] and some racial minorities, "[u]nalterable characteristics such as race and sex are considered inadmissible as predictors, and would be even if they were found to be correlated with the frequency of serious offending."[85] Even reports advocating the use of prediction have maintained that "[a] fair tool does not discriminate against offenders on the basis of enduring personal traits (e.g., race, gender, or age) and permits offenders equal access to services and treatment."[86]

Any attempts to justify punishment on the basis of gender or race must be compelling,[87] because the racial and sexual dynamic of our country is complex. It would be entirely unsatisfying to decree that male crimes were somehow worse or that men "deserved" more time in jail than women on the basis of gender alone. Even using the instrumental tack—that men should be imprisoned longer because they are simply more likely to recidivate—is painfully uncomfortable because men and women are hardly monolithic groups. Yet in Virginia, gender is one of the most potent determinants of prison time for nonviolent offenders.[88]

---

[83] Tonry, *supra* note 39, at 397; *accord* H.L.A. HART, LAW, LIBERTY AND MORALITY 174 (1968); Moore, *supra* note 5, at 324.

[84] Gabor, *supra* note 25, at 28 ("Cross-national evidence indicates that men are far more likely to engage in criminal activity than are women and that this imbalance becomes more pronounced with the increased gravity of criminal conduct.").

[85] 1 CRIMINAL CAREERS AND "CAREER CRIMINALS," *supra* note 5, at 166.

[86] Holsinger et al., *supra* note 67, at 47.

[87] Tonry, *supra* note 39, at 372.

[88] *See* 2001 ANNUAL REPORT, *supra* note 35, at 60 fig.53; Nonviolent Risk Assessment Worksheet (2006), *available at* http://www.vcsc.state.va.us/worksheet_2004/DrugsSch.IIIWkst.pdf (adding eight points to risk scores for male offenders).

One way to assess the appropriateness of a predictor is to examine its "relationship to the blameworthiness of the offender and the empirical and logical relationship of the predictor to the behavior being predicted."[89] Under this standard, immutable characteristics like race, religion, and gender are ethically suspect.

Even age may be an inappropriate characteristic to consider. On the one hand, we subject juveniles to a different system of criminal justice. However, this distinction is justifiable on the grounds that juveniles lack the understanding of the consequences of their actions and are, optimistically thinking, more susceptible to rehabilitation. Once criminals reach the age of maturity, is there a valid justification for punishing younger offenders for longer? Under the VCSC results, younger offenders are more likely to recidivate within a short timeframe.[90] Plus, age may be indicative of omitted variables. Whereas *aggregate* age-specific crime commission rates are greatest for teenagers and fall sharply thereafter, some scholars have noted that each *individual*'s distribution is flat, so age is merely an inaccurate proxy for where in a criminal career an individual may be.[91]

Most interesting is the issue of race. Race was the only factor explicitly considered and rejected by the VCSC.[92] As the NCSC reported, "[a]lthough race was strongly significant in the analysis, the Sentencing Commission viewed race 'as a proxy for social and economic disadvantage,' and decided to exclude it from the risk assessment worksheet."[93] The VCSC saw race as "'standing in'" for other factors, including "economic deprivation, inadequate educational facilities, family instability, and limited employment opportunities, many of which disproportionately apply to the African-American population."[94]

---

[89] 1 CRIMINAL CAREERS AND "CAREER CRIMINALS," *supra* note 5, at 8.

[90] *See* 2001 ANNUAL REPORT, *supra* note 35, at 60 fig.53; Nonviolent Risk Assessment Worksheet, *supra* note 88, at 4 (adding thirteen points for offenders under thirty years old).

[91] *See* Chul W. Ahn et al., *Estimation of Arrest Careers Using Hierarchical Stochastic Models*, 6 J. QUANTITATIVE CRIMINOLOGY 131, 133 (1990); *see also* MALTZ, *supra* note 43, at 128 ("A covariate analysis may show, for instance, that a higher probability of recidivism is associated with individuals who are younger at first arrest. But it cannot predict what particular *individuals* will do; it can only predict, to some extent, what *cohorts* with different mixes of individuals might do. Applying such findings to individuals can be improper, as exemplified by Greenwood's (1982) reanalysis of the Chaikens' (1982) data. He showed that those scoring higher on a crime predictor scale did, on average, have higher crime rates; however, the percentage of cases incorrectly predicted was so high as to render the scale useless as a prediction device.").

[92] *See* OSTROM ET AL., *supra* note 4, at 26-28.

[93] *Id. at* 12; *accord* 2001 ANNUAL REPORT, *supra* note 35, at 49.

[94] OSTROM ET AL., *supra* note 4, at 27-28.

This assertion raises two important questions. First, what is it about social and economic disadvantage that makes for a uniquely unacceptable criterion where, apparently, race alone is not *per se* objectionable? Second, what steps must be taken to fully "exclude" a factor?

The VCSC has provided no further explanation about its appeal to "social and economic disadvantage." Surely, this cannot be the true reason for excluding race. Employment status is clearly correlated with economic disadvantage, but it was included as a factor. Perhaps the VCSC was merely acknowledging indirectly that race is off limits. But without explaining its rationale more coherently, it is difficult to guess how the VCSC distinguished between race and sex, ultimately including only the latter. Notably, sex is correlated with social and economic disadvantage, at least in the workplace.[95]

In order to truly exclude a factor from multivariate regression, it must be included in the initial analysis and then ignored when determining the worksheet.[96] According to the NCSC, the VCSC basically took this approach.[97] If a variable were included and ignored, then there would be serious omitted variable problems within the results because even facially neutral factors like education, housing, and employment vary across races.[98]

But is race the only demographic variable that affects, for example, employment prospects? Characteristics such as ethnicity and religion have both permissible and impermissible covariates. They deserve the same treatment as race.

C. THE POSSIBILITY OF MORAL HAZARD

An additional concern with moving away from sentencing based solely on the severity of crimes is that doing so changes the deterrent effect on potential first-time criminals. If older women are less likely to face serious criminal sanctions, then standard rationality suggests that more older women will commit crimes eligible for risk assessment.

In other words, punishment affects crime not only in the context of recidivism, but also in the first instance of criminality. Virginia does not appear to have studied whether this concern is legitimate. Moreover, this circularity problem was not incorporated into the model. An assumption of the model is that potential re-offenders make their criminal decisions independent of the consequences. This inference simply is not true.

---

[95] *See, e.g.*, Tamar Lewin, *Women Losing Ground to Men in Widening Income Difference*, N.Y. TIMES, Sept. 15, 1997, at A1.

[96] *See, e.g.*, SCHMIDT & WITTE, *supra* note 72, at 6.

[97] *See* OSTROM ET AL., *supra* note 4, at 27-28 & n.10.

[98] 1 CRIMINAL CAREERS AND "CAREER CRIMINALS," *supra* note 5, at 167.

A plausible counterargument to this concern is the voluntary nature of the risk assessment program. If an offender has callously disregarded the law because the spreadsheet recommends no jail time, then a judge would be well-served to imprison the offender in spite of the risk-based recommendation. But detecting such a calculation is easier said than done. The system is designed to create different levels of punishment for different types of people, and *Homo economicus* petty criminals will respond in kind by skewing the criminality of these groups because of the moral hazard.

D. ACCOUNTING FOR THE STIGMA OF INCARCERATION

A final ethical consideration is that Virginia's risk assessment program is being used not just to scale punishments but also to distinguish between penalties that are different in kind. Incarceration sends a very powerful message to society and to the inmate.[99] By and large, incarceration forever brands an ex-convict. It suggests to potential future employers, neighbors, and friends that the crime committed was serious enough to warrant physical incapacitation. On the other hand, sanctions of probation are considerably less damaging.

Is the Virginia public capable of sorting through the risk assessment program to properly stigmatize past offenders? Or, to put it differently, society will inevitably treat differently those who have spent their year in the state penitentiary and those who have not. Is it appropriate if the basis for this distinction is a prediction of future dangerousness based on group characteristics?

The point here is that the ramifications of imprisonment are not easily contained. While Virginia may be attempting to isolate recidivism and minimize crime with limited resources, it is simultaneously sponsoring a social response that diverges along the same parameters.[100] It is worth considering whether the difference-in-kind between prison and alternative sanctions sufficiently militates against error-prone predictive judgments that do not bear upon the significance of an offender's past crime.

---

[99] VON HIRSCH, *supra* note 16, at 110 ("The symbolism of being incarcerated compounds its pains: it is not pleasant to live where the very walls are a reminder that one has been singled out as a miscreant.").

[100] *Cf.* Brian Netter, *Avoiding the Shameful Backlash: Social Repercussions for the Increased Use of Alternative Sanctions*, 96 J. CRIM. L. & CRIMINOLOGY 187, 190 (2005) (arguing that when alternative sanctions attempt to capitalize on the social norms that stigmatize criminal offenders, "the act of punishing . . . carries the capacity to reshape those norms and to alter our social dynamics").

E. ETHICAL CONCLUSIONS

The above discussions highlight the ethical concerns raised by predictive and group-based sentencing. Of course, there are also benefits. The goal of reducing crime is substantial. But in light of the ethical concerns, any jurisdiction seeking to impose either risk-based prediction or demographic-based sentencing must find a considerable crime-fighting benefit. It is difficult to quantify the costs of violating these ethical principles, but only compelling countervailing benefits would warrant even considering a strictly utilitarian approach.

IV. STATISTICAL AND METHODOLOGICAL CONCERNS

Virginia's risk assessment program is problematic even if statistically precise, but additional questions about the methodology are equally troubling. Given the obvious importance of accuracy, a comparison between Virginia's approach and the statistical literature highlights some surprising shortcomings. At the very least, these tactical blunders render the tool less predictive; at worst, they render the program inequitable.

This Part will discuss the major methodological choices made by the VCSC and the consequences that result therefrom. Section A revisits the debate between risk assessment and needs assessment in the context of the quantitative objective of the Virginia approach. Section B examines the statistical model used to achieve this objective in light of the realities of criminal enforcement.

A. CHOOSING THE OBJECTIVE FUNCTION

The stated objective of Virginia's risk assessment program is to "predict the relative risk that a felon will become a threat to public safety" and sentence appropriately.[101] The reference to "risk" led the VCSC to choose "risk assessment"—where the public's safety is preeminent—over "needs assessment"—where the offender's interests are invoked.[102] However, these approaches need not be mutually exclusive, and the VCSC exposes Virginians to additional risk by refusing to account for the needs of individual offenders.

Consider Virginia's approach for developing the risk worksheet. The VCSC developed the original model based on a study of 2013 fraud, drug, and larceny offenders *as they were released from incarceration*.[103] The

---

[101] *See* VA. CODE ANN. § 17-235(5) (1995), *amended by* VA. CODE ANN. § 17.1-803(5) (2006).

[102] *See supra* notes 17-23 and accompanying text.

[103] OSTROM ET AL., *supra* note 4, at 25.

study examined which of these offenders was reconvicted of a felony within a period of roughly three years.[104] By formulating the study in this manner, the results suggested only the probability that an *incarcerated* felon recidivates but gave no data on the rate of recidivism for offenders diverted to alternative sanctions.

The VCSC's approach, then, was to incarcerate those felons most likely to recidivate after their release from incarceration. This formulation cannot be construed as effectively minimizing the rate of recidivism unless we assume that an offender's future conduct is unaffected by the present choice of punishment. Although "[r]adical proponents of [selective incapacitation] regard prisons as little more than warehouses from which inmates emerge in roughly the same shape they entered,"[105] the existence of a wide variety of alternative sanctions[106] and Virginia's commitment to using clinical judgment to best match a redirected offender to an appropriate alternative sanction[107] indicate that Virginia believes that choice of programs *does* matter. A meta-analytic study comparing risk- and need-based punishments agrees.[108]

To be consistent with the goal of reducing risk for the population as a whole, Virginia would need to conduct a study that examined recidivism as a function both of personal characteristics and subsequent punishment. This difference is not trivial. Take, for example, the case of a drug offender eligible for risk assessment. Under the nine-point pilot, prior felony drug convictions counted against the offender, presumably reflecting the frequency with which drug addicts are reconvicted.[109] Yet thinking qualitatively, drug offenders would appear to be among the best candidates for alternative sanctions: if effective drug treatment programs replaced jail

---

[104] *Id.*

[105] GABOR, *supra* note 25, at 5; *see also* GREENWOOD WITH ABRAHAMSE, *supra* note 51, at vii-viii ("The most generally accepted view now is that the likelihood of an offender's recidivism is not a function of the type of sentence he receives.").

[106] *See supra* note 4 and accompanying text.

[107] *See supra* note 24 and accompanying text.

[108] *See* Craig Dowden & D.A. Andrews, *Effective Correctional Treatment and Violent Reoffending: A Meta-Analysis*, 2000 CAN. J. CRIMINOLOGY 449, 459-60 ("[B]ehavioral/social learning programs were associated with substantially larger treatment effects than those produced by non-behavioral approaches. . . . The results from this study highlight the importance of incorporating cognitive-behavioral and social learning strategies into correctional treatment programming for violent offenders. . . . Clearly, multimodal treatment approaches should ensure that criminogenic needs are the primary focus of the intervention to maximize their therapeutic potential.").

[109] 2001 ANNUAL REPORT, *supra* note 35, at 58.

time, then they could contribute significantly to a drop in crime.[110] These concerns are especially potent given the established criminogenic nature of imprisonment.[111]

Another initial problem with the objective function is the definition of recidivism. Although recidivism is susceptible to upwards of thirteen separate definitions,[112] the binary indicator variable chosen by the VCSC—reconviction for a felony—may not be consistent with the stated goal of protecting the public. No felonies are "good," of course, but some are certainly worse than others. Assessing the *extent* of subsequent damage as well as the probability thereof would be appropriate if the goal were truly to minimize social damage caused by releasing these nonviolent offenders.

B. SELECTING THE MODEL

Having discussed the initial matter of the objective, we can now address the more fundamental questions of modeling. Both the predictive capabilities and ethical propriety of a mathematical model depend strongly on how accurately the model reflects the underlying process. The options for modeling are broad. Sometimes simplistic models can accurately represent complex phenomena; however, incorporating advanced knowledge often requires sophisticated methodology.

The range of approaches that have been applied to recidivism is outlined in Subsection 1. Subsection 2 looks to how the VCSC's choice of a simple logistic regression seems inadequate, especially in light of the NCSC's marginally more sophisticated approach that calls into question some of Virginia's results.

1. The Science of Recidivism

A large literature exists discussing various models for recidivism. It has been noted, however, that "the choice of technique often appears to be a

---

[110] Incidentally, in its 2001 Annual Report, the VCSC noted that when judges exercising their ultimate discretion changed a recommendation from alternative sanction to incarceration, the reason was the "defendant's immersion in the drug culture" six percent of the time. *Id.* at 52.

[111] *See* Clear, *supra* note 53, at 28; Christina DeJong, *Survival Analysis and Specific Deterrence: Integrating Theoretical and Empirical Models of Recidivism*, 35 CRIMINOLOGY 561 (1997).

[112] MALTZ, *supra* note 43, at 63 tbl.6-1 (listing recidivism definitions used in recent studies, including recorded police contact, new offense, severity of offense, arrest, parole suspension, parole revocation, technical violation, absconding, probation violation, court appearance, reconviction, sentencing, and return to prison).

matter of convenience rather than the result of careful thought and analysis."[113]

Sometimes models are robust and even the failure of some assumptions will not render the whole approach meaningless. However, there are considerable differences between blind statistical guesswork and logical mathematical modeling that contextualizes the numbers. In many respects, the story of the father of mathematical modeling, Siméon-Denis Poisson, is telling. In the early 1830s, Poisson and Adolphe Quetelet studied French conviction rate statistics to evaluate Quetelet's theory that man had an average penchant for crime that had steadily decreased through the late 1820s.[114] Poisson, however, recognized that conviction rates were a function of criminality, enforcement, and jury behavior—which was random.[115] Thus, Poisson found that Quetelet's data were principally explainable by stochasticity, not progress.[116] Much can be learned from the work of Poisson, but a key point is that mere variation does not imply explanatory capacity.

A number of key questions must be asked in the context of recidivism. First, does time matter, or is the key question simply whether or not a criminal recidivates? Second, to what extent should parameters concerning the shape of known distributions be incorporated into a model? Third, does everyone eventually recidivate? Examining these questions in turn allows us to highlight some of the approaches that have been introduced through the academic literature.

a. Does Time Matter?

Virginia collected data spanning over three years that reflected not only whether an offender recidivated but how long this process took. The availability of this information provides a threshold question of modeling: is it significant whether the offender is rearrested on the first day out as opposed to after three years on the street?

If the answer to this question is "no," then the simplest regression techniques may be appropriate. Multivariate regression is used to determine how a set of predictors controls an end result. When the output

---

[113] SCHMIDT & WITTE, *supra* note 72, at 2.

[114] Michael D. Maltz, *From Poisson to the Present: Applying Operations Research to Problems of Crime and Justice*, 12 J. QUANTITATIVE CRIMINOLOGY 3, 7-10 (1996).

[115] *Id.*

[116] *Id.* at 9.

data are binary, as here, logistic regression is used. To be sure, regression is a blunt tool and its assumptions are often violated.[117]

In this context, time *does* matter. Not every crime is caught; adopting Virginia's objective of minimizing the number of crimes committed, the maximum likelihood estimate of the number of crimes an offender would have committed while unincarcerated is the inverse of the time until capture.[118] Thus, special emphasis must be given to those who recidivate rapidly. When timing data are available, basic regression ignores useful information.

b. To Parameterize or Not To Parameterize?

A variety of distributions satisfy the requirement of accounting for time. These approaches typically employ a technique called censoring that accounts for the possibility that an event *would have* happened but for the early termination of the study.[119] This adjustment allows those times that have been collected to be utilized.

When additional information is known about the functioning of a random phenomenon, the phenomenon can be modeled with a particular distribution and the parameters of that distribution can be estimated. For example, if we knew that the time since an offender's last offense had no bearing on the likelihood that he would offend now, we could estimate the parameters of the exponential distribution, which shares this "memoryless" property.[120]

An approach known as Cox proportional hazards modeling obviates this concern and has been applied on occasion to predicting recidivism.[121] The key assumption of a Cox proportional hazards model is that everyone has a base hazard function—the likelihood of recidivating at any point in

---

[117] *See generally* WILLIAM D. BERRY, UNDERSTANDING REGRESSION ASSUMPTIONS (1993).

[118] Moreover, suppose offenders commit crimes with rate λ_k, where k represents the individual offender. Let π be the probability a crime is detected and let τ_i be the delay time between the $(i-1)^{st}$ and i^{th} crime. Then, if $E[N(T)]$ is defined as the expected number of crimes committed by time T, then conditioning on the first crime gives $E[N(T)] = \Pr\{\tau_1 < T\}(1 + (1-\pi)E[N(T-\tau_1)])$. So the probability that $\tau_1 < T$ is merely one multiplicative component of the function to be minimized. On quantitative modeling of crime data more generally, see Benjamin Avi-Itzhak & Reuel Shinnar, *Quantitative Models in Crime Control*, 1 J. CRIM. JUST. 185 (1973).

[119] *See, e.g.*, Ann D. Witte & Peter Schmidt, *An Analysis of Recidivism, Using the Truncated Lognormal Distribution*, 26 APPLIED STAT. 302, 303 (1977).

[120] *See, e.g.*, ERIC V. DENARDO, THE SCIENCE OF DECISION MAKING 362-63 (2002).

[121] *See, e.g.*, Brian D. Bunday & Victor A. Kiri, *Analysis of Censored Recidivism Data Using a Proportional Hazards-Type Model*, 41 STATISTICIAN 85 (1992).

time given that the person has not yet recidivated—and that personal characteristics merely transform this curve by a multiplicative constant.[122] A corollary to this assumption is that hazard rates can never cross; that is, among two people, the person who at one time is more likely to commit a crime is *always* more likely to commit a crime. This approach is akin to postulating a base rate of "stable criminality" just as Quetelet hypothesized and Poisson disproved.[123] The problem with Cox regression is twofold. First, if information about the distribution is known, this information must be ignored.[124] Second, the multiplicative relationship requirement is both illogical and inconsistent with some existing data.[125] These problems have led some scholars to conclude that "nonparametric methods like the proportional hazards model are best regarded as useful in preliminary work designed to select an appropriate parametric method."[126]

c. Does Everybody Recidivate?

Even among parametric survival models, there are options. Standard survival models were developed as part of reliability theory.[127] They assess the probability that a process will not have failed by a certain time. (In this case, the process would be living in society and failure would be recidivism.) The problem is that survival techniques were developed to study electronic components. All electronics eventually fail. Is it proper to extend this assumption to a study of recidivism? Do all criminals eventually err again? Michael Maltz has argued that "to assume that everyone released from a correctional program (whose purpose is to *prevent* failure) will eventually fail is, at worst, unconscionable, and at best, unduly

---

[122] *See, e.g., id.* at 86-87 (1992).

[123] *See* Michael D. Maltz, *Operations Research in Studying Crime and Justice: Its History and Accomplishments*, *in* OPERATIONS RESEARCH AND THE PUBLIC SECTOR 201, 228-29 (S.M. Pollock et al. eds., 1994).

[124] *See* MALTZ, *supra* note 43, at 136.

[125] Maltz, *supra* note 123, at 228-29; *see also* Christy A. Visher & R.L. Linster, *A Survival Model of Pretrial Failure*, 6 J. QUANTITATIVE CRIMINOLOGY 153 (1990). *But see* OSTROM ET AL., *supra* note 4, at 83 n.45 (finding insufficient evidence to reject the hypothesis of proportionality). Of course, the failure to reject the assumption does not prove its validity.

[126] SCHMIDT & WITTE, *supra* note 72, at 18; *see also* MALTZ, *supra* note 43, at 72 ("Without [the multiplicative constant] assumption the results are of little utility.").

[127] *See generally* RICHARD E. BARLOW & FRANK PROSCHAN, MATHEMATICAL THEORY OF RELIABILITY (1965).

pessimistic."[128] The truth is that some offenders will successfully desist from future wrongful behavior.[129]

Desistance can be incorporated by adding an additional parameter representing the binary outcome of whether or not an individual will recidivate. These models, called "split" or "incomplete" distributions, have been applied successfully to recidivism data.[130] Recognizing the Poisson distribution for crimes committed by released criminals who do recidivate, a split distribution can be modeled as exponential (where recidivism is defined by arrest) or lognormal (where recidivism is defined by conviction).[131] One study comparing split models to basic models (like linear or exponential regression) found that "the *worst* of [the] split models (the split exponential) was about as good as the *best* of the simple models (the LaGuerre)."[132]

2. Models Used for Virginia

With the well-developed literature on survival techniques and incomplete distributions, it is a bit surprising that the VCSC settled on logistic regression to model crime in Virginia, both in the original 1995-96 study and the 2001 validation.[133] Survival models, apparently with inevitable recidivism assumed, were considered and rejected both times.[134]

In the interim, the NCSC's evaluation study employed Kaplan-Meier survival analysis and Cox proportionality regression.[135] Both models account for the time of an offense—unlike the VCSC—but both approaches are nonparametric. Still, for the reasons discussed above, a survival approach is more meaningful to the process of recidivism than blind snapshot regression.[136]

It would be helpful to reevaluate the VCSC and NCSC data using *incomplete* parametric distributions, but the data are not publicly available.

---

[128] MALTZ, *supra* note 43, at 89.

[129] *See generally* Blumstein et al., *supra* note 69 (studying whether juveniles can properly be categorized as "innocents" who will not commit a crime, "desisters" who are likely to commit only one or two crimes, and "persisters" who are highly likely to recidivate).

[130] *See, e.g.*, JOHN E. ROLPH ET AL., METHODS FOR ESTIMATING CRIME RATES OF INDIVIDUALS (1981); Maltz, *supra* note 123, at 226-28.

[131] To be precise, Rolph et al. settled on a logit-lognormal distribution, which is logistic on the binary variable and lognormal on the others. *See* ROLPH ET AL., *supra* note 130, at 82.

[132] *Id.* at 81.

[133] 2001 ANNUAL REPORT, *supra* note 35, at 55 fig.49.

[134] *Id.* at 57 & fig.51.

[135] *See* OSTROM ET AL., *supra* note 4, at 65.

[136] *See supra* Part IV.B.1.a.

Still, comparing the VCSC and NCSC conclusions bears on the ethical issues discussed earlier.[137] Notably, the NCSC remarked that:

> Both types of survival analysis (Kaplan-Meier and Cox regression), for both measures of recidivism [(rearrest and reconviction)], produced very consistent results about which factors were "good" predictors of recidivism. The results indicated that only gender and factors related to prior record were useful for predicting recidivism. . . . None of the other factors on the worksheet were significant predictors of risk. This included the other demographic factors (employment status, marital status, and age of offender) and contemporaneous factors (offender alone and additional offenses at time of offense).[138]

After the NCSC report, the VCSC used different data to reevaluate its own use of demographic variables. Returning to its logistic regression models, the VCSC ran two models: one including age, marital status, and employment (Model 1); the other without (Model 2).[139] The commission found that Model 2 actually produced nonrecidivists with greater accuracy but that Model 1 produced a lower recidivism rate for those recommended for alternative sanctions.[140]

But this result was inevitable. If the models were designed to minimize recidivism, then Model 1 could perform no worse than Model 2 because Model 2's variables comprise a subset of Model 1's. The more important question is the extent of the significance here. Given the obvious ethical concerns about using demographic data, can Virginia possibly satisfy the threshold of substantiality? The answer is a resounding "no." There are sizable social costs associated with sentencing offenders on the basis of demographic characteristics. Double-penalizing the unemployed for their plight is heartless, and imprisoning offenders because they lack a spouse to come home to bears little relationship to the purposes of incarceration.

---

[137] *See supra* Part III.

[138] OSTROM ET AL., *supra* note 4, at 91.

[139] 2001 ANNUAL REPORT, *supra* note 35, at 58 fig.52.

[140] *Id.* at 57 fig.51. This result is nonsensical and violates Bayes's Theorem. As the statistics were given, 75.7% of nonrecidivists were accurately predicted by Model 1 and 76.6% by Model 2. This outcome can only refer to the probability of nonrecidivism given a point score in the lowest quartile (i.e., the grouping recommended for diversion). The only alternative would be the probability that a nonrecidivist were selected for diversion, the complement of false negatives. But the data render such an option impossible, because with overall recidivism hovering around one-third, identifying three-quarters of the nonrecidivists would necessitate identifying far more than one-quarter of the offenders. If this figure represents the probability of recidivism given diversion, then the second datum, the recidivism rate for offenders recommended for alternative sanctions, should be complementary. But the VCSC claimed that 12.4% of those diverted from Model 1 recidivate versus 14.5% in Model 2. *Id.* The dichotomous inconsistency is impossible if both models fixed a twenty-five point threshold.

Supporters of statistical risk assessment would question whether a judge, acting alone, could do any better. There are two responses. First, it is socially acceptable to analytically study the future behavior of offenders who commit certain crimes, who have certain criminal histories, or whose crimes included certain characteristics. However, marginal improvements that can be gained by adding demographic considerations must be balanced against the sizable equitable costs of imposing such a regime. There is a risk in detaching punishment from the punishable act. There is a risk in segmenting the population into these predictive groups. And there is the risk of false positives. To be sure, there will always be false positives, and jail time will frequently be served by those who pose little or no threat to society. But that penalty is justifiable only if it can be tied to the initial act and state retribution. Paying a penalty justified only by an immutable personal characteristic runs counter to nationwide trends in equity and imposes serious societal costs that Virginia has simply neglected.

Given the very small improvement produced by incorporating the four minor demographic factors in the validation study, Virginia would be well-served to revisit the model and omit them. Raw efficiency is not the only objective served by our correctional system, and the VCSC's current approach damages the integrity of criminal justice for Virginians.

V. CONCLUSIONS

Although effective criminal sentencing is a noble goal, we must be mindful that our tactics have collateral consequences. The Virginia risk assessment program may not appear dramatic, but its assumptions reflect a monumental change in how criminals are sentenced yet the program has been implemented with perilously little public discussion. While statistical methodology may not be appropriate fodder for public discourse, the fundamental basis for correctional judgments is not insubstantial. The equity concerns raised by predictive judgments that utilize demographic characteristics of dubious significance are worthy of discussions in both academic and political circles.

A review of the reports and the Virginia data should leave us unconvinced that Virginia has adopted an appropriate procedure for sentencing its nonviolent criminals. The Commonwealth has not met its burden to justify this new approach. In the end, perhaps sentencing extends the baseball analogy adopted by the Chief Justice at his confirmation hearings: judges should call the balls and the strikes.[141] In baseball, purists have long fought using cameras to review umpire's decisions, with no

[141] *See Quotation of the Day*, N.Y. TIMES, Sept. 13, 2005, at A2.

better reasoning than tradition. Here, there are strong reasons of substance to support vesting sentencing discretion in judges. Even if their sentencing is ultimately less efficient, it can, at the very least, limit our considerations to factors related to the crime we deem deserving of punishment.