

# RISKY BUSINESS: CRITIQUING PENNSYLVANIA’S ACTUARIAL RISK ASSESSMENT IN SENTENCING

Jessica Corey\*

*The use of statistical risk assessment tools is growing in many different areas of the criminal justice field. This Note critiques this growing trend towards using risk assessment tools during the sentencing phase. Part II of this Note will survey the differing types of risk assessment tools used across the United States to track the development of risk under the assessment protocol standards of the criminal justice systems. Part III will review the current criticisms on actuarial risk assessment. Part IV will outline the proposed risk assessment tool Pennsylvania is currently developing. Finally, Part V will weigh Pennsylvania’s model against the criticisms of actuarial risk assessment discussed in Part II.*

I.	INTRODUCTION.....	151
II.	RISK ASSESSMENT TOOLS.....	153
	A. First-Generation Risk Assessment Tools .....	153
	B. Second-Generation Risk Assessment Tools .....	154
	C. Third- and Fourth-Generation Risk Assessment Tools .....	156
III.	OVERVIEW OF CRITICISMS .....	157
	A. Ethical Concerns with Punishing Future Conduct .....	157
	B. Invidious Punishment Criteria and Distributive Concerns.....	160
	C. Statistical Reliability.....	165
	D. Judges Do Not Understand Risk Assessment....	169
	E. Crime Rates and Elasticity to Incarceration .....	170

---

\* J.D. Candidate 2017, Columbia Law School; B.A. 2013, Furman University. The author would like to thank Jennifer Laurin, who provided tireless feedback and helped make this Note possible. The author would also like to thank Jonathan Allen and Sarah Freeman for their helpful comments on previous versions of this Note.

IV.	PENNSYLVANIA'S MODEL OF RISK ASSESSMENT IN SENTENCING .....	172
V.	CRITICISM OF PENNSYLVANIA'S SYSTEM .....	176
A.	Ethical Concerns with Punishing Future Conduct .....	176
B.	Invidious Punishment Criteria and Distributive Concerns.....	177
C.	Statistical Reliability.....	181
D.	Judges Do Not Understand Risk Assessment....	185
E.	Crime Rates and Elasticity to Incarceration .....	188
VI.	CONCLUSION .....	189

## I. INTRODUCTION

Mass incarceration has wreaked havoc in the United States, leaving many states with budget deficits and overcrowded prisons.<sup>1</sup> Now, as demands for reform to eliminate institutional racism within our criminal justice system increase, the wide disparity in treatment between minority and non-minority individuals in the criminal justice system cannot continue to be ignored by lawmakers.<sup>2</sup> Current practices are unsustainable, and as a result, states are searching for ways to reduce the prison population and the disparate treatment of defendants without sacrificing public safety. Enter evidence-based sentencing, which uses actuarial risk assessment tools during the sentencing phase to predict the sentence based on a defendant's risk of recidivism. Actuarial risk assessment tools use statistics to determine a defendant's risk of recidivism based on varying characteristics such as age, gender, and criminal history. Proponents say evidence-based sentencing will help aid

---

<sup>1</sup> See Bruce Western & Christopher Wildeman, *Punishment, Inequality, and the Future of Mass Incarceration*, 57 U. KAN. L. REV. 851, 857–66 (2009); Robert J. Sampson & Charles Loeffler, *Punishment's Place: The Local Concentration of Mass Incarceration*, 139 DAEDALUS 20 (2010).

<sup>2</sup> Jacqueline Johnson, *Mass Incarceration: A Contemporary Mechanism of Racialization in the United States*, 47 GONZ. L. REV. 301 (2012).

judicial decision-making, reduce costs, and help reduce prison populations.<sup>3</sup>

Risk assessment tools are not new; they have been utilized in different areas of the criminal justice system since their inception in the early 20th century, with actuarial risk assessment cropping up in the late 1970s.<sup>4</sup> Currently, roughly twenty states use some form of actuarial risk assessment in areas such as pretrial detention, bail setting, and probation or parole.<sup>5</sup> Most states, however, have resisted using risk assessment tools during sentencing. Nevertheless, with the current trend of budget shortages and lack of space in prisons, states are increasingly considering evidence-based sentencing as a practical option. Indeed, influential legal institutions, such as the American Law Institute and the National Center for State Courts, have endorsed the use of these tools.<sup>6</sup> The American Law Institute's draft revisions to the Model Penal Code ("MPC") would *require* actuarial risk assessment tools to be used in sentencing, which creates an important moment to examine these tools and how they will be used.<sup>7</sup>

This Note will focus specifically on Pennsylvania's exploration of evidence-based sentencing, as it is currently the only state in the process of implementing a fully-automated<sup>8</sup> actuarial risk assessment tool for all defendants at sentencing.<sup>9</sup> Part II will survey the various types of risk assessment tools used across the United States to track the

---

<sup>3</sup> Douglas A. Berman, *Re-Balancing Fitness, Fairness, and Finality for Sentences*, 4 WAKE FOREST J.L. & POL'Y 151, 158 (2014).

<sup>4</sup> Kelly Hannah-Moffat, *Actuarial Sentencing: An "Unsettled" Proposition*, 30 JUST. Q. 270, 274 (2013).

<sup>5</sup> Dawinder S. Sidhu, *Moneyball Sentencing*, 56 B.C. L. REV. 671, 674 (2015).

<sup>6</sup> *Id.*

<sup>7</sup> *Id.*

<sup>8</sup> "Fully automated" means that there will be little to no involvement of clinical professionals in creating the risk score. That is, the information will be entered into the computer, which will automatically assign a risk profile based on the input information.

<sup>9</sup> While Virginia has evidence-based sentencing, it is not fully automated, nor is it used across all defenders.

development of risk under the assessment protocol standards of the criminal justice system. Part III will review the current criticisms on actuarial risk assessment. Part IV will outline the proposed risk assessment tool Pennsylvania is currently developing. Finally, Part V will weigh Pennsylvania's model against the criticisms of actuarial risk assessment discussed in Part II. Although Pennsylvania has taken steps to avoid problems typical of evidence-based sentencing, this Note argues that actuarial risk assessment tools have no place in sentencing, given their current unreliability and the systemic racial biases they entrench. As a result, evidence-based sentencing is not the "new way forward" as many have suggested, and Pennsylvania should abandon its potential move to evidence-based sentencing.

## II. RISK ASSESSMENT TOOLS

There are currently more than sixty different risk assessment tools across the United States.<sup>10</sup> These risk assessment tools vary widely in their approach and accuracy. In general, risk assessment tools are understood to be part of one of four different "generations" of risk assessment,<sup>11</sup> which this section will briefly outline to provide context for understanding the type of tool Pennsylvania will implement.

### A. First-Generation Risk Assessment Tools

First-generation risk assessment tools, unlike all subsequent generations, do not utilize statistical data; instead, they are based on clinical evaluations.<sup>12</sup> Until recently, the primary way criminal justice actors calculated a defendant's likelihood of recidivism was through these clinical

---

<sup>10</sup> Anna Maria Barry-Jester et al., *The New Science of Sentencing: Should Prison Sentences be Based on Crimes That Haven't Been Committed Yet?*, THE MARSHALL PROJECT (Aug. 4, 2015, 7:15 A.M.), <https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing#.abQZxa461> [<https://perma.cc/2VF6-RNXD>].

<sup>11</sup> JAMES BONTA & D.A. ANDREWS, RISK-NEED-RESPONSIVITY MODEL FOR OFFENDER ASSESSMENT AND REHABILITATION 3–4 (2007).

<sup>12</sup> Hannah-Moffat, *supra* note 4.

risk assessment tools.<sup>13</sup> With a clinical model, mental health professionals evaluate an individual based on a wide range of criteria, analyzing factors such as family history, education, and criminal history. These mental health professionals then make a judgment based on all the available information to determine the defendant's risk of recidivism, and communicate their findings to the court.<sup>14</sup>

First generation tools, or clinical risk assessment tools, are divided into either unstructured or structured assessments.<sup>15</sup> Unstructured clinical assessments are general interviews with no preset criteria, whereas structured clinical assessments have standardized questions that are based on empirical research.<sup>16</sup> Those who prefer clinical risk assessments value them for their focus on the individual, as each assessment is a comprehensive analysis of the defendant and her background.<sup>17</sup> Although clinical risk assessment tools are still widely used, they have been criticized for a lack of accuracy and for high degrees of subjectivity, generally attributed to the wide variance of results depending on the individual conducting the assessment.<sup>18</sup>

## B. Second-Generation Risk Assessment Tools

Second-generation risk assessment tools (the type Pennsylvania will implement) are actuarial.<sup>19</sup> They use static factors (factors that cannot be changed by the defendant) linked to recidivism (e.g., a history of substance abuse), as well as other factors, in order to make statistical predictions about

---

<sup>13</sup> *Id.* at 271.

<sup>14</sup> BONTA & ANDREWS, *supra* note 11, at 3.

<sup>15</sup> Patricia M. Harris, *What Community Supervision Officers Need to Know About Actuarial Risk Assessment and Clinical Judgment*, 70 FED. PROB. 8, 8 (2006).

<sup>16</sup> *Id.*

<sup>17</sup> Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 853 (2014).

<sup>18</sup> Melissa Hamilton, *Back to the Future: The Influence of Criminal History on Risk Assessments*, 20 BERKELEY J. CRIM. L. 75, 92 (2015).

<sup>19</sup> BONTA & ANDREWS, *supra* note 11.

the defendant's likelihood of recidivism.<sup>20</sup> The presence of a "risk factor," in any category, such as being male, produces a score of one or more depending on the variable and instrument; its absence receives a score of zero.<sup>21</sup> The points derived from the presence of these factors are added together to generate a percentage that is supposed to predict the chance of recidivism based on data previously gathered about *other* defendants with the same factors as the current defendant.<sup>22</sup> The data used in actuarial risk assessment tools is created using different criminal databases and then tested to find relevant indicators of recidivism.<sup>23</sup> One reason actuarial risk assessments are purported to be more reliable than clinical judgments is their uniformity in application—any individual conducting the assessment should get the same result.<sup>24</sup> For example, if you are male, twenty years of age, with two prior arrests, you will receive the same risk assessment score as every other male who is twenty years of age and has two prior arrests, regardless of who conducts the assessment.

Evidence has consistently shown that actuarial risk assessments perform better than *unstructured* clinical risk assessments in terms of accurately predicting recidivism, which seems to support the notion that the statistical method is better than the older interview type methods.<sup>25</sup> This evidence has led many to decry clinical risk assessment tools, and move towards actuarial tools.<sup>26</sup> But the accuracy benefits of second generation tools and beyond may only exist when comparing actuarial tools to unstructured clinical assessments. Studies have found that the structured clinical

---

<sup>20</sup> *Id.*

<sup>21</sup> *Id.*

<sup>22</sup> Hannah-Moffat, *supra* note 4.

<sup>23</sup> *Id.*

<sup>24</sup> Sidhu, *supra* note 5, at 691; Michael Marcus, *MPC—The Root of the Problem: Just Deserts and Risk Assessment*, 61 FLA. L. REV. 751, 769 (2009).

<sup>25</sup> Harris, *supra* note 15, at 9–10.

<sup>26</sup> Dawinder S. Sindhu, *Moneyball Sentencing*, 56 B.C. L. REV. 671, 674 (2015).

assessments can produce results as accurate or even more accurate than the second-generation actuarial tools, depending on the population being assessed and the tool being used.<sup>27</sup> Overall, there is a slight edge in accuracy for actuarial prediction tools over structured clinical assessments due to their ease of use and uniformity in application.<sup>28</sup> However, critics have also questioned the de-individualization of second-generation risk assessment tools.<sup>29</sup> That is, unlike clinical risk assessment, second-generation risk assessment tools fail to account for the individual's particular circumstances. Instead, they view the individual narrowly under a set of basic characteristics and ignore nuances.

### C. Third- and Fourth-Generation Risk Assessment Tools

These criticisms led to the development of third- and fourth-generation risk assessment tools, which combine successful clinical elements with actuarial tools to predict recidivism.<sup>30</sup> Proponents of these tools claim that they are more effective than either method alone, because they utilize best practices from each of the two prior generations of tools to create a more individualized, yet still scientific, risk assessment tool.<sup>31</sup> Significantly, these tools—dubbed “risk-needs tools”—include dynamic risk factors that an individual can change over time and are focused on identifying criminogenic needs (traits that are linked to an individual's likelihood of recidivism).<sup>32</sup> These tools are predicated on the idea that criminal behavior is predictable and that the areas of risk that are likely to lead an individual to re-offend are traceable.<sup>33</sup> Those who support the third-generation of risk-

---

<sup>27</sup> *Id.*

<sup>28</sup> *Id.* at 10–11; Hannah-Moffat, *supra* note 4, at 276.

<sup>29</sup> Hannah-Moffat, *supra* note 4, at 276.

<sup>30</sup> *Id.* at 275–76.

<sup>31</sup> 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, 603 (2009).

<sup>32</sup> Hannah-Moffat, *supra* note 4, at 276–80.

<sup>33</sup> *Id.* at 276.

need assessment tools assert “third-generation risk-need instruments offer a way of monitoring the effectiveness, or ineffectiveness, of programs and supervision strategies,” which is a major drawback of second-generation tools.<sup>34</sup>

Fourth-generation risk assessment tools build upon third generation tools, as they are also risk-need tools, but they refine their assessments of risk to align more directly with treatment possibilities.<sup>35</sup> Thus, not only do fourth-generation tools identify areas of need, they also identify treatment target goals and assess progress of the offender moving forward.<sup>36</sup> These tools, however, are relatively new and are not widely used.<sup>37</sup> As such, these tools have not been tested as often as other generations of tools have, and their reliability remains in question.

### III. OVERVIEW OF CRITICISMS

This section will explore the criticisms of actuarial risk assessment in more depth, and group them into five primary categories: 1) ethical concerns with punishing future conduct; 2) static and dynamic factors; 3) statistical reliability; 4) judicial understanding; and 5) impact on incarceration.

#### A. Ethical Concerns with Punishing Future Conduct

One of the more overarching concerns with risk assessment tools being used during sentencing is the concern of punishing future conduct.<sup>38</sup> Historically, the United States has followed a conception of justice that punishes defendants

---

<sup>34</sup> BONTA & ANDREWS, *supra* note 11, at 4.

<sup>35</sup> Hannah-Moffat, *supra* note 4, at 279–84.

<sup>36</sup> BONTA & ANDREWS, *supra* note 11, at 4–5.

<sup>37</sup> *Id.*

<sup>38</sup> Jordan M. Hyatt et al., *Reform in Motion: The Promise and Perils of Incorporating Risk Assessments and Cost-Benefit Analysis into Pennsylvania Sentencing*, 49 DUQ. L. REV. 707, 732 (2011); John Monahan, *A Jurisprudence of Risk Assessment: Forecasting Harm Among Prisoners, Predators, and Patients*, 92 VA. L. REV. 391, 393 (2006); Sidhu, *supra* note 5, at 686.

for the crimes they have committed—a theory of justice that focuses on just outcomes.<sup>39</sup> The United States has emphasized liberty as one of the highest tenets of its society. Indeed, the Declaration of Independence proclaims that the right to liberty is unalienable.<sup>40</sup> In light of our emphasis on personal liberty, we afford a defendant guaranteed rights designed to protect liberty until there is proof beyond a reasonable doubt that he has committed a crime.<sup>41</sup> Evidence-based sentencing, however, shifts the paradigm by attempting to predict the likelihood that a defendant will commit another crime and reenter the criminal justice system. A system that gives longer sentences to defendants based on the outcome of a risk assessment tool is at direct odds with our fundamental values of liberty and innocence until proven guilty.

Some advocates of actuarial risk assessment do not believe it punishes future conduct and therefore avoid confronting this problem.<sup>42</sup> Michael Marcus, an Oregon state judge, holds this view, stating that, “Considering real disparity in risk does not mean that we are ‘punishing offenders for future crimes,’ but rather that we are managing the risk these offenders represent by fashioning a sentence in response to a present conviction for a past crime.”<sup>43</sup> This position is untenable. If a judge’s sentence is based on a statistical model which believes the defendant has a high-risk of committing a crime in the future, and the defendant is sentenced more harshly because of that high-risk designation, then the sentence length is only increased as a result of the prediction that the defendant will commit a crime in the future. It is difficult to conceive how sentencing based on a “high” risk of recidivism could not punish the individual for future conduct when it is built precisely to consider future conduct in sentencing determinations.

---

<sup>39</sup> Marcus, *supra* note 24, at 751.

<sup>40</sup> THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

<sup>41</sup> U.S. CONST. amend. V, § 1, cl. 3.

<sup>42</sup> Marcus, *supra* note 24.

<sup>43</sup> *Id.* at 754.

Most proponents accept that risk assessment involves future-oriented sentencing; however, they do not view this sentencing form as problematic because judges, they argue, already implicitly use their own arbitrary assessment of a defendant's risk during sentencing and may impose a longer sentence if they believe he will commit more crimes in the future.<sup>44</sup> Actuarial tools, proponents claim, will only help to increase the accuracy of the judge's predictions.<sup>45</sup> The first problem with this reasoning is that utilizing evidence-based sentencing will likely only increase the role future-oriented risk assessment plays,<sup>46</sup> perhaps overshadowing more traditional assessments of moral desert. That is, even though most proponents say that the risk assessment tool will only be "one piece of the sentencing puzzle," it is still likely to have a significant impact if judges rely on this model to make sentencing decisions.

If judges do not rely on these tools, and they do not have a significant impact on judicial decision-making, then pouring limited resources into the creation of these tools will waste resources that could have been better used in other areas of criminal justice reform. If judges do choose to rely on these tools and sentence an individual based on statistics, then this increases the emphasis on future-oriented sentencing, which is directly at odds with the notion of innocent until proven guilty and fundamental liberty. It seems that the more likely of the two scenarios is that judges will rely on these risk assessment tools, especially elected judges who have their record scrutinized. Actuarial risk assessment could help remove accountability—the judge can look towards a defendant's risk score in the event of an incorrect decision to avoid any accompanying political backlash.

---

<sup>44</sup> 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, 185 (2013); David A. Soulé & Stacy S. Najaka, *Paving a Path to Informed Sentencing Decisions*, 25 FED. SENT'G REP. 181, 182 (2013).

<sup>45</sup> Sidhu, *supra* note 5, at 686.

<sup>46</sup> Sonja B. Starr, *supra* note 17 at 808.

## B. Invidious Punishment Criteria and Distributive Concerns

As demonstrated earlier, not all risk assessment tools are created equal. Second-generation risk assessment tools only focus on static characteristics, such as age, gender, or criminal history.<sup>47</sup> A defendant cannot change these factors, meaning he cannot improve his risk score over time. Third- and fourth-generation risk assessment tools include these static factors, but they also utilize dynamic factors, which can change over time, such as employment history or education.<sup>48</sup> While risk assessment tools have steered clear of explicitly using race as a factor because of the clear constitutional bar, they still consider other static immutable characteristics, which is directly at odds with current federal sentencing practices.<sup>49</sup> As University of Virginia School of Law professor John Monahan notes:

With the single exception of criminal history—which the Guidelines state ‘is relevant in determining the appropriate sentence’—virtually all of the variables that potentially could be used as scientifically valid risk factors for violence under a forward-looking consequentialist ‘crime control’ theory of punishment are *explicitly excluded from consideration in federal sentencing procedures*.<sup>50</sup>

In other words, federal judges are prohibited by law under the Sentencing Reform Act of 1984<sup>51</sup> from considering any of the many factors, such as age and gender, which risk assessment tools explicitly use to forecast future risk.

---

<sup>47</sup> BONTA & ANDREWS, *supra* note 11, at 3.

<sup>48</sup> *Id.*

<sup>49</sup> U.S. SENTENCING COMMISSION, GUIDELINES MANUAL, §§ 5H1.10, 5H1.12, at 425 (NOV. 2004).

<sup>50</sup> Monahan, *supra* note 38, at 397–98 (emphasis added).

<sup>51</sup> Pub. L. 98-473, 98 Stat. 1837 (1984).

One might not take issue with the inclusion of these factors. If we know that males, for instance, commit more crimes than females do, using static factors such as gender could allow us to be “smart on crime.” But it is unethical to sentence defendants using static characteristics that they can neither control nor change. Not only are defendants potentially being sentenced more harshly for factors outside of their control, but also defendants would be unable to reduce their risk assessment score in the future. For example, if a young, unemployed male sold narcotics once and received a risk score of fifty percent, and the next year enrolled in college courses and found meaningful employment, his risk assessment score would not change. Since the time of his conviction, he has drastically changed his position in life, and it seems patently unfair to deny acknowledgement of the positive changes he has made. Additionally, evidence bears the idea that work and education reduce recidivism; studies show that released prisoners who participate in transitional work programs have a recidivism reduction of eighteen percent.<sup>52</sup> One study has found that formerly incarcerated individuals without some type of employment are three times more likely to recommit a crime.<sup>53</sup>

Yet, second-generation tools do not account for this variance in any way. Some scholars note that sentencing based on immutable static factors goes against fundamental notions of fairness.<sup>54</sup> As University of New Mexico School of Law professor Dawinder Sidhu opined, by measuring risk in this way, “risk-assessment tools sever the link between punishment and individual conduct, and between punishment and individual control.”<sup>55</sup> Our modern criminal justice system is built on punishing individuals based on their choices and

---

<sup>52</sup> Bruce Western, *From Prison to Work: A Proposal for a National Prisoner Reentry Program*, 2008 BROOKINGS INST. 1, 2–5.

<sup>53</sup> Steven D. Bell, Note, *The Long Shadow: Decreasing Barriers to Employment, Housing, and Civic Participation for People with Criminal Records Will Improve Public Safety and Strengthen the Economy*, 42 W. ST. L. REV. 1, 10 (2014).

<sup>54</sup> *Id.*

<sup>55</sup> Sidhu, *supra* note 5, at 675.

acts, and not on immutable characteristics. Inability to improve a risk score is one reason why second-generation risk assessment tools are criticized more harshly than third- and fourth-generation risk assessment tools. These later generations include dynamic factors and clinical assessment elements, meaning an individual's risk score has the potential to improve over time.<sup>56</sup>

However, the inclusion of dynamic factors introduces a new host of issues that make them problematic. For example, unemployment is often considered a risk factor, and defendants without jobs will be given a higher risk score than those with jobs.<sup>57</sup> While factoring in employment may seem beneficial, it will likely disproportionately affect people of color because minority unemployment rate is higher than the white unemployment rate,<sup>58</sup> especially post-incarceration. Dynamic tools also often factor in education levels, which is likely to negatively affect minority individuals at a disproportionate rate due to the disparate educational opportunities and graduation rates.<sup>59</sup> This draws into question the fairness of assessing risk using dynamic factors; these factors could cause more sentencing errors because they do not account for the systemic racism that pervades our system. A recent study done by ProPublica in Broward County, Florida confirms this idea.

Broward County may begin using a private company, Northpointe, along with the 137-question dynamic risk assessment tool designed by Northpointe.<sup>60</sup> However, an analysis of the tool's efficacy found that black individuals were

---

<sup>56</sup> Hannah-Moffat, *supra* note 4, at 275.

<sup>57</sup> BONTA & ANDREWS, *supra* note 11, at 4–5.

<sup>58</sup> Western, *supra* note 52, at 6.

<sup>59</sup> State High School Graduation Rates By Race, Ethnicity, GOVERNING, <http://www.governing.com/gov-data/education-data/state-high-school-graduation-rates-by-race-ethnicity.html> [https://perma.cc/QFQ5-PZ64].

<sup>60</sup> Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-riskassessments-in-criminal-sentencing> [https://perma.cc/PY2R-SDMP].

almost twice as likely as white individuals to be labeled a higher risk but not actually re-offend.<sup>61</sup> It had the opposite effect among white individuals: white individuals were much more likely than black individuals to be labeled lower risk and recidivate.<sup>62</sup> While no clear research has been done to examine the error rates between black and white defendants for static risk assessment tools, the factors used in static risk assessment tools are likely still subject to many of the same pitfalls because of the criteria they use.

The utilization of criminal history, which is seen by many as the only valid criterion for actuarial risk assessment,<sup>63</sup> is also problematic. As one leading scholar put it, “prison populations are not random, they are the products of past sentencing policies and patterns and they disproportionately represent black people and other socially disadvantaged groups.”<sup>64</sup> Indeed, a black man is more likely to be incarcerated than a white woman,<sup>65</sup> and minority communities have considerably higher rates of police presence and surveillance than white communities do.<sup>66</sup> Thus, those who commit crime are considerably more likely to be caught in neighborhoods with added police surveillance (largely minority communities) than they would in communities with lower police presence (largely white communities).<sup>67</sup>

---

<sup>61</sup> *Id.*

<sup>62</sup> *Id.*

<sup>63</sup> Monahan, *supra* note 38, at 397–98.

<sup>64</sup> Hannah-Moffat, *supra* note 4, at 279–84.

<sup>65</sup> HEATHER C. WEST, U.S. DEP’T OF JUSTICE, BUREAU OF JUSTICE STATISTICS, PRISON INMATES AT MIDYEAR 2009—STATISTICAL TABLES 21 tbl.18 (2010), <http://www.bjs.gov/content/pub/pdf/pim09st.pdf> [<https://perma.cc/8PAS-36RZ>].

<sup>66</sup> 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 (2008). *See also* Sarah Childress, *The Problem with “Broken Windows” Policing*, FRONTLINE (June 28, 2016), <http://www.pbs.org/wgbh/frontline/article/the-problem-with-broken-windows-policing/> [<https://perma.cc/DCC6-WNVH>].

<sup>67</sup> *Id.*

This concept is borne out especially concerning drug offenses. Though white and black people use and sell drugs at roughly the same rate,<sup>68</sup> black people are 3.6 times more likely to be arrested for selling drugs and 2.5 times more likely to be arrested for possessing them.<sup>69</sup> Minorities who are arrested at higher rates are then more likely to serve longer sentences when sentenced in accordance with an actuarial risk assessment tool because they will be deemed a higher risk for having been arrested for more crimes.<sup>70</sup> But this higher arrest rate for minorities exists, at least in part, because the police are paying more attention to them and their communities.<sup>71</sup> The resulting effect is what some have termed a “ratchet” or a “feedback” effect. Bernard Harcourt, a leading scholar on actuarial risk assessment in criminal justice, has argued that the feedback effect will:

[H]ave significant detrimental consequences on the employment, educational, familial, and social outcomes of the profiled populations—including, in the case of racial profiling, the devastating effects associated with the notion of [b]lack criminality that pervades the public imagination . . . The high costs associated with any ratchet effect should temper our embrace of the actuarial.<sup>72</sup>

Compounding the issue that minority communities are more heavily policed, many risk instruments assign points multiple times for a single prior criminal event, particularly those that maintain numerous and overlapping criminal offending items

---

<sup>68</sup> Some studies have suggested that white people use and sell drugs at even higher rates than black people. Jonathan Rothwell, *How the War on Drugs Damages Black Social Mobility*, BROOKINGS INSTIT. (Sept. 30, 2014), <http://www.brookings.edu/blogs/social-mobility-memos/posts/2014/09/30/how-the-war-on-drugs-damages-black-social-mobility> [https://perma.cc/9KX3-UJ6B].

<sup>69</sup> *Id.*

<sup>70</sup> BONTA & ANDREWS, *supra* note 11, at 3.

<sup>71</sup> Rothwell, *supra* note 68.

<sup>72</sup> Harcourt, *supra* note 66, at 271.

in their scoring sheets.<sup>73</sup> This results in a double counting effect. For example, if an individual has committed a juvenile drug crime, depending on the risk assessment tool, they could have the same event counted three times—once under prior offense, again for its nature as a drug crime, and a third time for it being a juvenile crime—corresponding to a much higher risk assessment score for a single crime. Sonia Starr summed it up well:

Although we do not know whether EBS [evidence-based sentencing] 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.<sup>74</sup>

In other words, while the morality issue is still up for debate, *even if* it were morally acceptable to use these tools punishing future conduct, the data on which these tools rely is largely grounded in structural racism. Using these types of tools will only perpetuate the structural racism that has plagued the criminal justice system.

### C. Statistical Reliability

Not only is punishing defendants based on factors outside of their control problematic, it is also near impossible to treat individuals fairly if they are treated as monolithic groups, unshaped by the particular contexts of their background. This is one of the biggest problems facing actuarial risk assessment tools: group statistics cannot adequately determine individual conduct. From a moral standpoint, the courts have argued that each individual must be sentenced as such. For instance, the Supreme Court's case

---

<sup>73</sup> Hamilton, *supra* note 18, at 98.

<sup>74</sup> Starr, *supra* note 17, at 837.

law on indigent defendants has consistently held that disparate treatment cannot be justified based on statistical group generalizations.<sup>75</sup>

These generalizations are not just morally questionable; they are also scientifically problematic. That is, knowledge of group tendencies, even when precise, cannot accurately forecast individuals.<sup>76</sup> Actuarial risk assessment tools identify risk levels based on a defendant's membership in certain groups, and assign the *same* risk profile to every defendant who contains those same characteristics. But we know that individuals vary widely—something for which actuarial risk assessments do not account. One thirty-year-old female with a criminal record can be viewed differently from a thirty-year-old female without a criminal record. Moreover, if sixty percent of males under twenty-one recidivate, this is inadequate evidence that a random twenty-one-year-old male will recidivate. But second-generation actuarial risk assessment tools do not consider these variances. Quite opposite, one of the main supporting arguments for evidence-based sentencing is that it will reduce the variance of judicial decision-making.<sup>77</sup> However, an overreliance on risk assessment encourages the judge to reduce the individual to a risk number, which values consistency over fairness and justice.

All of the aforementioned helps to account for one of the most fundamental misconceptions about risk assessment: that risk assessment tools predict risk in some absolute sense. In reality, risk assessment tools cannot say, for example, that an individual is sixty percent likely to recidivate—they can only say that sixty percent of the people with those same group characteristics did recidivate. Importantly, the tool

---

<sup>75</sup> *Id.* at 807.

<sup>76</sup> Timothy P. Cadigan & Christopher T. Lowenkamp, *Implementing Risk Assessment in the Federal Pretrial Services System*, 75 *FED. PROB.* 30, 31 (2011); David J. Cooke & Christine Michie, *Limitations of Diagnostic Precision and Predictive Utility in the Individual Case: A Challenge for Forensic Practice*, 34 *LAW & HUM. BEHAV.* 259 (2010).

<sup>77</sup> Cadigan & Lowenkamp, *supra* note 72.

cannot distinguish between whether the defendant is in the sixty percent of people who did recidivate or the forty percent of people who did not. As might be expected, this leads to high false positive rates (predicting that the defendant would recidivate when they did not) because the group statistics just cannot predict the future.<sup>78</sup> Indeed, the commentary to the MPC revision on risk assessment finds similarly, stating 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.”<sup>79</sup>

To better explain why actuarial risk assessment error rates are so high, Sonia Starr, a leading critic of actuarial risk assessment tools, provides an analysis of group to individual statistics. Her analysis uses a sample of 400 women and their height averages. She demonstrates how relatively easy it is to take the average height of this group of women, and then predict the average height of another group of women based off of the original sample—there is a high statistical likelihood that you would be correct in this scenario.<sup>80</sup> She continues on to ask, if, instead of trying to ascertain the average height of another group of women, we wanted to simply guess the height of the next woman we encountered based on our sample. She states:

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

---

<sup>78</sup> Brian Netter, *Using Group Statistics to Sentence Individual Criminals: An Ethical and Statistical Critique of the Virginia Risk Assessment Program*, 97 J. CRIM. L. & CRIMINOLOGY 699, 712 (2007).

<sup>79</sup> MODEL PENAL CODE: SENTENCING § 6B.09, at 56–57 (Tentative Draft No. 2, 2011).

<sup>80</sup> Starr, *supra* note 17, at 844.

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 . . . In other words, you don't know much at all about how tall to expect the next woman to be.<sup>81</sup>

Though recidivism is somewhat different because it is a bivariate statistic, which means there are only two options: the individual will recidivate or the individual will not recidivate. Starr's example is illustrative of how difficult it is to use group statistics for a particular characteristic to predict the occurrence of that characteristic in an individual.

Many risk assessment models use the Area Under the Curve ("AUC") to validate and determine the statistical significance of their tool—that is, to gauge the tool's accuracy in predicting recidivism. AUC 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 was given a higher risk score.<sup>82</sup> AUC is measured from zero to one. A perfect model would receive a score of one, meaning it would always predict correctly. In contrast, a coin flip would have an AUC of .5, signifying a fifty percent chance of either outcome occurring.<sup>83</sup> The best published scores for risk assessment tools are around 0.75.<sup>84</sup> Tools with this level of predictive capacity are generally the third and fourth-generation models, which include dynamic risk factors, such as a psychological profile.<sup>85</sup> "Many studies have reported AUC scores closer to 0.65."<sup>86</sup> Applying this model to height,

---

<sup>81</sup> *Id.*

<sup>82</sup> *Id.* at 846.

<sup>83</sup> *Id.*

<sup>84</sup> *Id.* at 846–47.

<sup>85</sup> *Id.*

<sup>86</sup> Starr, *supra* note 17, at 846.

Professor Starr notes that if height is changed into a binary model (tall or not tall), then height has an AUC score of 0.825.<sup>87</sup> This is despite the fact that the model was not very good at individual predictions. This example demonstrates how even a high AUC does not necessarily mean that the predictive power of recidivism functions on an individual level.

#### D. Judges Do Not Understand Risk Assessment

Researchers who investigated how judges used risk information in decision-making found that judges know little about risk technologies and thus often use information that already fits their perception of risk.<sup>88</sup> In a comprehensive study, researchers found that although court actors received training on risk assessment tools and their interpretation:

[P]ractitioners tended to struggle with the meaning of the risk score and the importance of the items contained in the assessment tools . . . (Rather than understanding that an individual who obtains a high risk score *shares characteristics* of an aggregate group of high-risk offenders, the individual is likely to become *known* as a high-risk offender.)<sup>89</sup>

Unless actors within the criminal justice system are made fully aware of the drawbacks of the tools they are implementing, they have the potential to misuse them. James Bonta, a staunch advocate for risk assessment, notes that:

It is one thing for scientists to demonstrate that a risk instrument or a treatment program can work but it is a very different matter to make it

---

<sup>87</sup> *Id.* at 846–47.

<sup>88</sup> Hannah-Moffat, *supra* note 4, at 288.

<sup>89</sup> KELLY HANNAH-MOFFAT, *THE SAGE HANDBOOK OF PUNISHMENT AND SOCIETY: PUNISHMENT AND RISK* 12 (Jonathan Simon & Richard Sparks eds., 2013).

work in correctional agencies with a diverse work force in terms of education, values and experience, conflicting criminal justice policies and management practices that are not conducive to selecting and training of staff in effective assessment techniques.<sup>90</sup>

The effectiveness and fairness of these tools relies in large part on how the actors within the criminal justice system choose to utilize them. If the same tool is used differently across different jurisdictions and even across different judges, the claim that these tools decrease subjectivity holds considerably less weight. Furthermore, because there are already fairness and accuracy concerns with the tools themselves, adding human error will likely decrease the efficacy of these tools.

#### E. Crime Rates and Elasticity to Incarceration

In addition to punishing individuals for crimes they have not yet committed, actuarial tools also have high false positive rates that rely on data skewed against racial minorities. Risk assessment tools are supposed to reduce incarceration rates and prevent more crime. However, as demonstrated, the tools can only say whom has the highest risk of recidivism based on other defendants with the same characteristics who have recidivated. This is not reliable. This does not divulge the most important information necessary to prevent crime and reduce recidivism: the elasticity to incarceration or the extent at which incarceration reduces crime. Without this information, risk assessment tools cannot adequately serve the public safety component that they purport to serve. That is, if it is unclear whether incarcerating certain groups of individuals for longer will actually reduce their recidivism rates, then we have no basis for incarcerating individuals for longer just because they are deemed “higher risk.” In fact, studies almost unequivocally

---

<sup>90</sup> BONTA & ANDREWS, *supra* note 11, at 15.

show that a longer sentence does not decrease offender recidivism.<sup>91</sup>

One commonly advanced line of reasoning in support of longer sentences is that individuals in prison can no longer commit new crimes. This logic is flawed. As one University of Houston Law Center professor notes, there is insufficient confirmatory data that increasing sentence length or using imprisonment to deter and incapacitate offenders predicted to be high-risk are effective at reducing recidivism or crime rates.<sup>92</sup> In fact, studies have shown that when they are released, inmates are more likely to commit future crimes than those not incarcerated are or than those individuals incarcerated for shorter periods.<sup>93</sup> Advocates for risk assessment tools ignore an overarching point: a longer sentence for individuals may increase the chances of recidivism in some individuals.

Failing to study the elasticity to incarceration is likely to have detrimental effects on communities of color. These tools have not been adequately developed or tested for minority populations.<sup>94</sup> Much of the science has been based on white offender populations. As a result, there has been “a lack of attention to the racialized nature of offending and imprisoned populations and to the specific needs of nonwhite [sic] offenders.”<sup>95</sup> Therefore, not only does actuarial risk assessment lack information on elasticity to incarceration limit its effectiveness, it especially lacks data regarding the effect these tools will have on minority populations.

Increasing sentence length will only perpetuate the current cycle of racial injustice. Studies have shown that, because high incarceration rates disproportionately affect

---

<sup>91</sup> Warren, *supra* note 31, at 594.

<sup>92</sup> Anne R. Traum, *Mass Incarceration at Sentencing*, 64 HASTINGS L.J. 423, 434 (2013).

<sup>93</sup> Warren, *supra* note 31, at 594.

<sup>94</sup> Hannah-Moffat, *supra* note 4, at 279–84.

<sup>95</sup> *Id.* at 280.

minority communities, “those communities are harmed when prisoners are incarcerated and when they return with diminished political, economic, and social status.”<sup>96</sup> Incarceration has an “enormously harmful effect on the life prospects of those imprisoned. Released prisoners suffer a thirty to forty percent loss of income, their domestic partnerships are often ruptured, and their marriage, job, and housing prospects reduced.”<sup>97</sup> Minority populations experience these effects more keenly. For example, post-release, sixty percent of black individuals suffered consequences in the job search as opposed to only thirty percent of white individuals. In other words, a criminal history is considerably more damaging to black individuals. When combined with the statistics on the reduction of recidivism in those defendants who have secured employment, racial minorities face disparate impact. These factors contribute directly to the likelihood of recidivism. Yet evidence-based sentencing fails to account for this basic premise of the detriment to people of color.

#### IV. PENNSYLVANIA’S MODEL OF RISK ASSESSMENT IN SENTENCING

In November 2015, Pennsylvania had approximately 50,000 people in state custody—about 2,000 more individuals than for which it has permanent beds.<sup>98</sup> To accommodate this overflow, Pennsylvania has increased the percentage of its budget devoted to incarceration.<sup>99</sup> Thirty years ago, Pennsylvania spent less than two percent of its total budget on incarceration—it now spends more than seven percent, or two billion USD, each year on its corrections system.<sup>100</sup> Unsurprisingly, greater incarceration rates has not “fixed” these individuals; between twenty and twenty-nine percent of

---

<sup>96</sup> Traum, *supra* note 92, at 434.

<sup>97</sup> Lynn Adelman, *What the Sentencing Commission Ought to be Doing: Reducing Mass Incarceration*, 18 MICH. J. RACE & L. 295, 309 (2013).

<sup>98</sup> Barry-Jester et al., *supra* note 10.

<sup>99</sup> *Id.*

<sup>100</sup> *Id.*

individuals return to jail within one-year of release.<sup>101</sup> Thus, Pennsylvania is spending a significant portion of its budget on a program that has yet to address the severe problems it faces. As a result, the Pennsylvania legislature has been looking for ways to reduce the incarceration rate, as well as the amount spent on incarceration. In 2010, the legislature and governor passed Act 95, which mandated that the Pennsylvania Commission on Sentencing develop a risk assessment tool to assist judges at the sentencing phase.<sup>102</sup> Specifically, according to the Pennsylvania Commission on Sentencing report, Act 95 mandated that the Commission undertake the following:

- Adopt a risk assessment instrument to be used at sentencing;
- Consider the risk of re-offense and threat to public safety;
- Help determine if offender is a candidate for alternative sentencing programs [CIP, SIP, Recidivism Risk Reduction Incentive Program, Department of Correction's Boot Camp]; and
- Develop an empirically based worksheet using factors predicting recidivism.<sup>103</sup>

In response to Act 95 of 2010, the Commission has undertaken a "Risk Assessment Project," which is aimed at developing a fully automatic second-generation risk assessment tool for all defendants at the sentencing phase.<sup>104</sup> Phase I of the project, which was conducted between

---

<sup>101</sup> NICOLETTE BELL ET AL., PA. DEP'T OF CORR., RECIDIVISM REPORT 2013, 11 (2013), <http://www.nationalcia.org/wp-content/uploads/2013-PA-DOC-Recidivism-Report.pdf> [<https://perma.cc/X8ZZ-RFFU>].

<sup>102</sup> PA. COMM'N ON SENTENCING, RISK/NEEDS ASSESSMENT PROJECT, PHASE II INTERIM REPORT 1: DEVELOPMENT OF A RISK ASSESSMENT SCALE BY OFFENSE GRAVITY SCORE FOR ALL OFFENDERS (2015), <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-ii-reports/Interim-Rpt-1-Phase-2/view> [<https://perma.cc/5VMX-VUJM>] [hereinafter "INTERIM REPORT 1"].

<sup>103</sup> *Id.*

<sup>104</sup> *Id.* at 1.

2010–2014, studied different types of risk assessment tools and resulted in the development and validation of an initial risk assessment tool for defendants sentenced under levels three and four of the sentencing guidelines.<sup>105</sup> After analyzing the results, the Commission decided on eight factors that correlated with recidivism: age; gender; location; total number of prior arrests; prior property arrests; prior drug arrests; property offender; and Offense Gravity Score (“OGS”), which has fourteen levels that correspond to the seriousness of the crime.<sup>106</sup>

Phase II of the project is currently ongoing.<sup>107</sup> For Phase II, the Commission has expanded the scope of its risk assessment tool and developed a scale for defendants sentenced under all five levels of the guidelines. The Commission is in the process of validating the expanded Phase II tool.<sup>108</sup> For this phase, the Commission used the total sample of offenders from 1998–2000, consisting of 112,702 offenders, to generate its data on recidivism rates and factors.<sup>109</sup> This sample was then split at random into two sub-samples: one for the development sample and the other for the validation sample.<sup>110</sup> The Commission used another sample of offenders from 2004–2006 as a second validation sample.<sup>111</sup> The final development sample had 51,131 defendants after some cases were removed for missing data.<sup>112</sup>

The Phase II tool, in addition to covering all five levels of the guideline, also built a scale for each different OGS,

---

<sup>105</sup> *Id.*

<sup>106</sup> PA. COMM’N ON SENTENCING, RISK/NEEDS ASSESSMENT PROJECT, INTERIM REPORT 7: VALIDATION OF RISK SCALE (2013), <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-7-validation-of-risk-scale/view> [https://perma.cc/G4MN-ZCSL] [hereinafter “INTERIM REPORT 7”].

<sup>107</sup> INTERIM REPORT 1, *supra* note 102, at 3.

<sup>108</sup> *Id.*

<sup>109</sup> *Id.* at 3–4.

<sup>110</sup> *Id.*

<sup>111</sup> *Id.*

<sup>112</sup> INTERIM REPORT 1, *supra* note 102, at 3–4.

which tracks the seriousness of the crime.<sup>113</sup> For example, criminal mischief with damage under \$500 has an OGS of one, while murder of a law enforcement official has an OGS of fifteen.<sup>114</sup> This was a change from the Phase I tool, which had one tool for OGS's different levels and used the OGS level as a factor in the risk assessment scale.<sup>115</sup>

The Commission chose a set of nine factors for Phase II, four of which (age, gender, county, and number of prior arrests) were a part of the Phase I study.<sup>116</sup> The nine factors are: age; gender; county; current offense type; multiple current convictions; number of prior arrests; prior offense type; prior record score; and prior juvenile adjudication.<sup>117</sup> The factors that were dropped or modified for Phase II were prior property arrests, prior drug arrests, property offender, and OGS. In Phase II, the Commission combined prior property arrests and prior drug arrests into prior offense type, changed property offender to type of current offense, and dropped OGS as a factor in favor of having a differing scale for each OGS as explained above.<sup>118</sup>

According to the Commission's website, Phase III will include an expansion of the study to determine the impact of the inclusion of more dynamic factors, such as drug use, employment status, and education, on the risk of recidivism and the validity of the tool.<sup>119</sup> However, Carol Zeiss, the Commission's Director of Research, stated that Phase III may not occur given the difficulty in obtaining data.<sup>120</sup> She

---

<sup>113</sup> *Id.* at 7.

<sup>114</sup> PA. CODE § 303.15 (2015).

<sup>115</sup> INTERIM REPORT 1, *supra* note 102, at 7.

<sup>116</sup> *Id.* at 1.

<sup>117</sup> *Id.* However, the Commission made the decision to remove county as a factor from subsequent tools.

<sup>118</sup> *Id.* at 6.

<sup>119</sup> *Risk Assessment Project*, PA. COMM'N ON SENTENCING, <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment> [https://perma.cc/4PHU-TG2V].

<sup>120</sup> Telephone Interview with Carol Zeiss, Manager of Data, Pa. Sentencing Comm'n (Nov. 11, 2016) [hereinafter Interview with Carol

explained that for most cases, data simply was not available, and a reliable tool could not be developed using the small sample of defendants for whom information was available.<sup>121</sup> Ms. Zeiss said that if the Commission were to incorporate dynamic factors into their risk assessment tool, it would take place years down the road, after the successful implementation of the Phase II risk assessment tool in the field.<sup>122</sup> In fact, she stated that before adding dynamic factors would be reconsidered, it is likely that the Commission will focus on creating a separate tool for sex offenders and violent crime.<sup>123</sup>

## V. CRITICISM OF PENNSYLVANIA'S SYSTEM

### A. Ethical Concerns with Punishing Future Conduct

Pennsylvania is not be the first state to implement an actuarial risk assessment tool at sentencing; Virginia has used an actuarial risk assessment tool in sentencing for more than a decade.<sup>124</sup> Nevertheless, there are important differences between the Pennsylvania and Virginia risk assessment models that necessitate examining Pennsylvania's tool more closely. Pennsylvania will be the first state to utilize a risk assessment tool for *all defendants*, whereas Virginia's risk assessment tool only focuses on low-risk offenders.<sup>125</sup> The Virginia risk assessment model *requires* that twenty-five percent of all low-risk offenders be diverted

---

Zeiss]. Ms. Zeiss's statements are subject to change, pending additional guidance from the Pennsylvania Sentencing Commission.

<sup>121</sup> *Id.*

<sup>122</sup> *Id.*

<sup>123</sup> *Id.*

<sup>124</sup> BRIAN J. STROM ET AL., VA. CRIMINAL SENTENCING COMM'N, OFFENDER RISK ASSESSMENT IN VIRGINIA: A THREE-STAGE EVALUATION 1, 9 (2002), [http://www.vcsc.virginia.gov/risk\\_off\\_rpt.pdf](http://www.vcsc.virginia.gov/risk_off_rpt.pdf) [https://perma.cc/377P-T2ER]

<sup>125</sup> BRIAN J. STROM ET AL., VA. CRIMINAL SENTENCING COMM'N, OFFENDER RISK ASSESSMENT IN VIRGINIA: A THREE-STAGE EVALUATION 5 (2002), [http://www.vcsc.virginia.gov/risk\\_off\\_rpt.pdf](http://www.vcsc.virginia.gov/risk_off_rpt.pdf) [https://perma.cc/Z6HH-LN37].

into alternative incarceration programs, while Pennsylvania's model has no such mandate.<sup>126</sup> If Pennsylvania uses actuarial risk assessment tools to increase sentence length, as its model suggests, then the state is incarcerating its citizenry for longer based on the potential that they will commit future crimes using a statistical model that is not always accurate. This issue is minimized when risk assessment tools are used only to divert individuals from prison, as in Virginia, but it is a very real ethical problem that the Pennsylvania system faces.

Pennsylvania's utilization of actuarial risk assessment tools is problematic. Many advocates for evidence-based sentencing suggest that actuarial risk assessment tools should only be used to mitigate sentences, as they understand that risk assessment tools are challenging for a number of reasons.<sup>127</sup> Even the commentary to the MPC revision asserts that "section 6B.09 takes an attitude of skepticism and restraint concerning the use of high-risk predictions as a basis of elongated prison terms, while advocating the use of low-risk predictions as grounds for diverting otherwise prison-bound offenders to less onerous penalties."<sup>128</sup>

## B. Invidious Punishment Criteria and Distributive Concerns

As previously stated, the Pennsylvania risk assessment tool is a second-generation tool, meaning it judges individuals on a variety of static characteristics, such as age, gender, and prior criminal history. Choosing a second-generation risk assessment as opposed to a third- or fourth-generation risk assessment tool limits the tool's utility. Second-generation tools are criticized for assessing

---

<sup>126</sup> *Id.*

<sup>127</sup> RICHARD COUZENS, EVIDENCE-BASED PRACTICES: REDUCING RECIDIVISM TO INCREASE PUBLIC SAFETY: A COOPERATIVE EFFORT BY COURTS AND PROBATION 10 (2011), <http://www.courts.ca.gov/documents/EVIDENCE-BASED-PRACTICES-Summary-6-27-11.pdf> [<https://perma.cc/AR2T-QEKJ>].

<sup>128</sup> MODEL PENAL CODE: SENTENCING § 6B.09 cmt. at 54 (Tentative Draft No. 2, 2011).

defendants based on immutable characteristics. The Commission responded to criticisms of unfairness by issuing a special report “to examine, from a research perspective, the impact of” including static factors such as age, gender, and location in the tool.<sup>129</sup> This study tested the validity of the tool without each factor, and concluded that including each factor increased the reliability of the tool. Therefore, they asserted that these factors would remain in subsequent versions of the tool. However, Carol Zeiss, the Commission’s Director of Research, states that the Commission dropped location as a factor as they move forward with testing and validation of the tool due to its racial implications.<sup>130</sup> However, the Commission will continue to utilize age and gender. As a result, young males will have considerably higher risk assessment scores, which will likely impact young men of color on an even greater scale.

The Pennsylvania model will suffer from the “feedback” effect where defendants are caught in a cycle that increasingly punishes them for factors outside of their control. Minority defendants will be more likely to have a higher risk factor because their neighborhoods tend to be more heavily policed, and they are therefore more likely to have a prior criminal record.<sup>131</sup> Even more problematic, however, is that a single crime could result in several different points towards the risk assessment score, drastically increasing the risk the tool calculates—a double-dipping effect. For example, Pennsylvania’s tool counts not only the defendant’s number of prior arrests, but also prior offense type, prior record score and prior juvenile adjudication, each of which will punish the offender for their criminal history.<sup>132</sup> This means that

---

<sup>129</sup> PA. COMM’N ON SENTENCING, RISK/NEEDS ASSESSMENT PROJECT, SPECIAL REPORT: IMPACT OF REMOVING DEMOGRAPHIC FACTORS (2015), <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-ii-reports/special-report-impact-of-removing-demographic-factors/view> [https://perma.cc/2FAP-GP7P] [hereinafter “SPECIAL REPORT”].

<sup>130</sup> Interview with Carol Zeiss, *supra* note 120.

<sup>131</sup> Harcourt, *supra* note 66.

<sup>132</sup> INTERIM REPORT 1, *supra* note 102, at 1.

Pennsylvania's tool will be particularly susceptible to an effect where one juvenile drug offense would count as a risk factor under prior arrests, prior offense type, prior record score and prior juvenile adjudication, which results in four or more additional points for the exact same crime.

This system disproportionately impacts people of color in each factor. For example, when considering prior juvenile adjudication, it is necessary to recognize that black individuals are disproportionately more likely to be arrested as juveniles. "In 2010, African Americans comprised seventeen percent of all juveniles, but thirty-one percent of all arrests."<sup>133</sup> Therefore, black defendants are more likely to have an increased risk score in the juvenile category than white defendants. The prior offense type is also likely to disproportionately affect people of color, as the only two types of prior offenses that count as a risk factor under Pennsylvania's model are drug and property crimes. As previously stated, black individuals are far more likely than white individuals to be arrested and convicted for drug crimes despite using and selling at the same rate.

Under the Pennsylvania model, in five out of nine OGS levels, a defendant will get a one-point increase in their score if they have a prior drug crime.<sup>134</sup> "Number of prior arrests" will also disproportionately affect communities of color, because, as discussed, these communities have higher rates of police presence and are therefore more likely to be arrested for the same crimes than those in communities with a lower police presence. The statistics on drug crimes provide an illustrative example. The National Association for the Advancement of Colored People ("NAACP") reports that nationally, black individuals represent twelve percent of the

---

<sup>133</sup> *Disproportionate Minority Contact in the Juvenile Justice System*, THE SENTENCING PROJECT, 1 (May 2014), <http://www.sentencingproject.org/wp-content/uploads/2015/11/Disproportionate-Minority-Contact-in-the-Juvenile-Justice-System.pdf> [<https://perma.cc/N5FZ-VAK7>].

<sup>134</sup> INTERIM REPORT 1, *supra* note 102, at 1.

total population of drug users, but thirty-eight percent of those arrested for drug offenses.<sup>135</sup> In Philadelphia, the American Civil Liberties Union (“ACLU”) found that black individuals represent eighty-two<sup>136</sup> percent of those arrested for marijuana possession, despite only constituting a small fraction of the total population.<sup>137</sup> These statistics illuminates how the tool treats race differently, with the majority of the double-dipping disparity affecting black defendants.

In an interim report, the Pennsylvania Commission on Sentencing questioned: “is it appropriate to consider factors that are closely linked to those already considered by the guidelines? . . . The issue for utilizing prior arrests is primarily whether counting both prior arrests and prior convictions would be considered [‘double-dipping’] and potentially punishing an offender twice for the same conduct (i.e., a prior arrest and prior conviction for the same crime).”<sup>138</sup> Nonetheless, there is no readily available evidence that the Commission has further considered, much less resolved, the issue of double-dipping.

In general, responses to the criticism that actuarial risk assessment tools condemn individuals based on group characteristics that they cannot change have been quite unsatisfactory. Mark Bergstrom, the Director of the Pennsylvania Commission on Sentencing, states that “this is also true of risk assessments as practiced in the field of medicine, yet the treatment regimens guided by such

---

<sup>135</sup> *Criminal Justice Fact Sheet*, NAACP, <http://www.naACP.org/pages/criminal-justice-fact-sheet> [https://perma.cc/6U3A-2QD6].

<sup>136</sup> *The War on Marijuana in Black and White*, ACLU, 106 (Jun. 2013), <https://assets.documentcloud.org/documents/708721/aclu.pdf> [https://perma.cc/3ZQN-RL8J].

<sup>137</sup> *Id.* at 17–20.

<sup>138</sup> PA. COMM’N ON SENTENCING, FACTORS THAT PREDICT RECIDIVISM FOR VARIOUS TYPES OF OFFENDERS, INTERIM REPORT 3, at 12 (2011), <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-3-factors-that-predict-recidivism-for-various-types-of-offenders/view> [https://perma.cc/B2EK-TQJX] [hereinafter “INTERIM REPORT 3”].

individual comparison with an aggregate group lead to treatment regimens that result in better overall patient outcomes than if such information were ignored.”<sup>139</sup>

This is an unpalatable comparison. First, actuarial risk assessment tools in the medical field have a considerably higher accuracy rate. Generally, these actuarial risk assessment tools must have an AUC above a .8 to be considered even useful, and above a .9 to be considered truly reliable.<sup>140</sup> For comparison, the AUC for Pennsylvania’s tool was between .66-.71, which rates between the “poor” category of predictive value and just barely into the “fair” category for one scale with a .71 AUC.<sup>141</sup> Much more importantly, in the medical field, individuals are not punished for sharing characteristics with the larger group with which they are compared, nor are they punished for their inability to change their characteristics. At best, Mr. Bergstrom’s response goes to the accuracy of these tools and is silent on the ethics of fairness. While the medical field may use static characteristics to treat patients, it is tenuous to assert that it is also acceptable for the criminal justice system to use such characteristics to incarcerate certain individuals for longer periods.

### C. Statistical Reliability

As discussed above, Pennsylvania has worked on developing a risk scale for each OGS. The AUC of Pennsylvania’s tool varies depending on which OGS scale being looked at, which ranges from .66 to .73.<sup>142</sup> The first OGS level has an AUC of .66, which falls in line with the studies that show that many actuarial risk assessment tools have an

---

<sup>139</sup> Kern & Bergstrom, *supra* note 44, at 187.

<sup>140</sup> Thomas G. Tape, *The Area Under an ROC Curve*, UNIV. NEB. MED. CTR., <http://gim.unmc.edu/dxtests/roc3.htm> [<https://perma.cc/2S8P-2CBZ>].

<sup>141</sup> *Id.*

<sup>142</sup> SPECIAL REPORT, *supra* note 129, at 20–29.

AUC around .65.<sup>143</sup> Flipping a coin has an AUC of .5, which shows that for many of the OGS levels, the statistical reliability is not that much greater than a completely random selection. On the other end, OGS level eight has an AUC of .73, which seems to indicate it is at least decently reliable, but this is only for one category of offenses. Yet even a high AUC does not necessarily mean the tool is reliable, because again, even aside from the concerns of fairness, error rates matter. Error rates are very instructive as to how good the predictive power is on the individual scale. The higher the error rate, the lower the predictive power of the tool for individuals. For example, Pennsylvania reported an error rate of sixteen people wrongly predicted to recidivate (false positive) for every one person falsely predicted not to recidivate (false negative) in its first validation sample findings.<sup>144</sup> It is concerning that the false positive rate is considerably higher than the false negative rate. This suggests that the Commission made an implicit value judgment that it was better to falsely incarcerate more individuals than it was to incarcerate fewer.

Pennsylvania studied the false positive/negative rate when determining a low-risk and high-risk cutoff point, as the cutoff point directly relates to the error rate. For example, Pennsylvania's risk assessment tool is scaled from zero to fourteen, so if the cutoff rate for low risk was between zero and ten, there would likely be an over prediction of low-risk individuals as opposed to if the cutoff rate was between zero and three. In the seventh report of the validation study, the cutoff rate for low-risk was between zero and four points and high-risk was between five and fourteen, which means that if a defendant scored four or fewer risk points they would be designated low-risk, and above four, high-risk.<sup>145</sup> With these cutoffs, the false positive rate was around forty-four percent.<sup>146</sup> This is an incredibly high error rate. The Commission deemed it is a better policy "to err on the side of

---

<sup>143</sup> Starr, *supra* note 17, at 846.

<sup>144</sup> INTERIM REPORT 7, *supra* note 106, at 1.

<sup>145</sup> *Id.* at 24.

<sup>146</sup> *Id.*

over-predicting recidivism [which results in more offenders incorrectly predicted to recidivate] than to err on the side of under-predicting recidivism [which results in more offenders incorrectly predicted to not recidivate].”<sup>147</sup>

In an earlier development report, the Committee decided to test a risk assessment tool that divided defendants into low- (zero to four), medium- (five to ten), and high-risk (eleven to fourteen), instead of just high- and low-risk.<sup>148</sup> With three categories of risk, the error rate was much lower, as it hovered around fifteen percent, depending on the sample.<sup>149</sup> It is confusing, to say the least, why the later validation sample would only include high-risk and low-risk samples when the earlier report showed a considerably lower false positive rate when the tool included the medium-risk category.

Unfortunately, it is difficult to say why the Commission would choose the two-category risk division instead of the three-category tool. One explanation might be that the percent accuracy for truly high-risk individuals was higher in the two-category risk assessment. More specifically, in the three-category risk study, the percentage of defendants who were accurately assessed to be high-risk was around seventy-six percent.<sup>150</sup> Eighty-six percent of those defendants who were actually high-risk (i.e., they recidivated within a three-year period) were correctly forecasted.<sup>151</sup> In contrast, for the high- and low-risk validation study, while only fifty-six percent of individuals who were predicted to be high-risk actually were, of those who were actually high risk, ninety-

---

<sup>147</sup> *Id.* at 19.

<sup>148</sup> *Id.* at 4. See also PA. COMM’N ON SENTENCING, DEVELOPING CATEGORIES OF RISK, INTERIM REPORT 5, at 7 (2012), <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-5-developing-categories-of-risk/view> [https://perma.cc/Q484-4FR8] [hereinafter “INTERIM REPORT 5”].

<sup>149</sup> *Id.* at 12.

<sup>150</sup> *Id.* at 11.

<sup>151</sup> *Id.*

five percent of them were predicted to be high-risk.<sup>152</sup> This suggests that the Commission made a value judgment to focus on getting and keeping truly high-risk individuals behind bars rather than with ensuring that only those who are actually high-risk are designated high risk (i.e., a lower false positive rate). While it is unclear whether this assessment is correct, this assessment does fall in line with the Commission's determination that it would rather over-predict than under-predict. Regardless, an error rate of forty-three percent falsely predicted to recidivate seems untenable. Such a high error rate should have deterred the Commission from moving forward with the binary risk category model. Leading statisticians are skeptical whether an error rate above thirty percent could ever be defensible.<sup>153</sup> However, even if it was assumed that the Commission will use the high-, medium-, and low-risk model, an error rate of twenty-seven percent is still quite high.

In addition, there is much to be concerned about in regards to a potential error rate of around thirty percent, or at the worse end, around forty-five percent (depending on whether the Commission uses a high-, medium-, and low-risk model or simply high- and low-risk. If the Commission has decided that over-prediction is better than under-prediction, it will lead to more individuals being incarcerated for longer than they otherwise would have been with a different predictive model, and they will be incarcerated based on a faulty tool that cannot predict for them individually. If Pennsylvania wants to reduce incarceration rates, implementing a tool that not only has a high false positive rate, but also drastically favors false positives over false negatives, will not achieve that goal. The problem with false positive rates can be summed up by a quote from William Blackstone that "it is better that ten guilty persons escape, than that one innocent suffer."<sup>154</sup> The false positive rate is

---

<sup>152</sup> *Id.*

<sup>153</sup> Netter, *supra* note 78, at 712.

<sup>154</sup> 4 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND: 1765-1769 (9th ed. 1966).

even more troubling when considering future punishment, because, as one scholar puts it, “offenders 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.”<sup>155</sup>

Pennsylvania has not published any studies on the error rates of these tools for communities of color.<sup>156</sup> Such research is crucial to understanding how these tools impact minority individuals, and to ensuring that these tools treat racial groups equally. One might expect that, because of the issues with double-dipping criminal history and with the development of the tool based primarily on white criminology,<sup>157</sup> actuarial risk assessment will disproportionately affect people of color. If this is the case, error rates are likely to be greater for minority defendants as the tool will have rated them as higher risk than they actually are. Therefore, it is critical to a fair system to ensure that the error rates do not increase with a minority defendant as opposed to with a white one.

#### D. Judges Do Not Understand Risk Assessment

To Pennsylvania’s credit, the Commission tested the statistical understanding of the criminal justice actors who would use the risk assessment tool. Interestingly, the Commission found “significant differences” among the numeracy levels (understanding of numbers) of the actors within different counties, with some counties having considerably higher numeracy levels than others.<sup>158</sup>

---

<sup>155</sup> Netter, *supra* note 78, at 712 (emphasis added).

<sup>156</sup> *Risk Assessment Project*, *supra* note 119.

<sup>157</sup> Hannah-Moffat, *supra* note 4, at 280–82.

<sup>158</sup> PA. COMM’N ON SENTENCING, COMMUNICATING RISKS AT SENTENCING, INTERIM REPORT 8, at 1 (2014), <http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment/phase-i-reports/interim-report-8-communicating-risk-at-sentencing/view> [<https://perma.cc/MGT3-S5UP>] [hereafter “INTERIM REPORT 8”].

Numeracy levels were measured by testing the degree to which the individuals felt comfortable with numerical information, using a nine-question survey about statistics.<sup>159</sup>

Even more intriguing is that the counties measuring the lowest numeracy tended to show a different pattern of risk assessment than the other counties. In some cases, the lowest numeracy county went in the opposite direction recommended by the tool from the higher numeracy counties.<sup>160</sup> When the risk assessment tool designated the offender as high- or low-risk, the low numeracy counties chose the opposite designation, which does not bode well for one of the supposed benefits of risk assessment in decreasing variance in judicial decision-making. The Commission did not indicate which counties had lower numeracy levels.

All users received full information regarding the risk assessment tool during at least one point in the study, and despite this information, there was still a variance in understanding among county actors.<sup>161</sup> If the lower numeracy counties demonstrated a lower understanding of the risk assessment tools (as evidenced by choosing the opposite of the higher numeracy counties), there is a concern that defendants who reside in the lower numeracy counties will be disadvantaged in relation to their counterparts in higher numeracy counties. If Pennsylvania is going to implement a risk assessment tool, the Commission should take steps to equalize the numeracy level of all actors who will be using the tool. Clearly, this numeracy training cannot simply be limited to providing full information about the risk assessment tool, as participants in this study had such information. Unsurprisingly, across the board, high numeracy individuals were significantly more likely to report that the information was easy to understand than low numeracy individuals.<sup>162</sup>

---

<sup>159</sup> *Id.*

<sup>160</sup> *Id.*

<sup>161</sup> *Id.*

<sup>162</sup> *Id.*

Providing risk score information only had a modest effect on actors' assessment of an offender's recidivism risk. The study found that the risk assessment tool significantly affected risk judgments "in four of the six cases, but only ten percent of all decisions and, overall, the decisions of only forty-four percent of the sample."<sup>163</sup> The Commission hypothesized reasons why this might be, considering that if the participants knew the study was a simulation they may not have taken the study as seriously as they would in sentencing real cases. Additionally, based on their numeracy studies, they claimed, "it is possible that the participants did not understand the risk information and therefore were not influenced by it."<sup>164</sup> They concluded that "this possibility is unlikely, however, in that the average movement was in the direction toward the actuarial information."<sup>165</sup> This explanation is nonsensical when paired with the Commission's studies on numeracy. The Commission noted that the lowest numeracy counties assessed risk in the opposite direction of the high numeracy counties. Even though the average score moved in the direction of the risk assessment, it does not mean that the actors necessarily understood the risk assessment information.

If the reason is that judges did not take the tool as seriously as they might in a real situation, the Commission should be concerned about the overreliance on this risk assessment tool. The Director of the Commission specifically stated that the tool should only be one instrument in an array of sentencing information.<sup>166</sup> However, there is the possibility that elected judges are more likely to be persuaded to rely on the risk assessment tool for political protection, as they can cite to the risk assessment tool as the scientific basis for their decisions.<sup>167</sup> Most of the judges in Pennsylvania are elected

---

<sup>163</sup> *Id.*

<sup>164</sup> INTERIM REPORT 8, *supra* note 158.

<sup>165</sup> *Id.*

<sup>166</sup> Kern & Bergstrom, *supra* note 44, at 187–88.

<sup>167</sup> In Virginia, for example, there is ample evidence that judges do rely on these types of tools quite substantially. *Id.* at 183–84.

and not appointed,<sup>168</sup> which increases the chances of this scenario. One prominent University of Wisconsin Law School professor, Cecelia Klingele, shares this concern. Ms. Klingele states that “unless criminal justice system actors are made fully aware of the limits of the tools they are being asked to implement, they are likely to misuse them.”<sup>169</sup>

### E. Crime Rates and Elasticity to Incarceration

If Pennsylvania is going to increase sentences based on group averages to reduce crime, then it is necessary to analyze how incarcerating its members reduces crime. This is an important question left unaddressed by second-generation risk assessment tools. As Sonia Starr puts it, “[t]here is no particular reason to believe that groups that recidivate at higher rates are also more responsive to incarceration. [Evidence-based sentencing] 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.”<sup>170</sup> When you consider how high the Pennsylvania model’s error rate is, incarcerating those that the tool deems the riskiest would not reduce crime for forty-three percent of the “high-risk” population because the tool falsely predicted that they would recidivate.

Certain types of so-called “higher-risk” defendants might be less responsive to enhanced penalties and therefore might be more vulnerable to its criminogenic effects.<sup>171</sup> If this is the case, lengthening high-risk defendants’ sentences might

---

<sup>168</sup> See How Judges Are Elected, The United Judicial System of Pennsylvania, <http://www.pacourts.us/learn/how-judges-are-elected> [https://perma.cc/B8P5-V9XJ].

<sup>169</sup> Cecelia Klingele, *The Promise and Perils of Evidence-based Correction*, U. WIS. LAW SCH., LEGAL STUDIES RESEARCH PAPER SERIES PAPER NO. 1368, (Nov. 30, 2015) <http://www.wisconsinappeals.net/wp-content/uploads/2015/12/Klingele-article.pdf> [https://perma.cc/C9BB-BLG9].

<sup>170</sup> Starr, *supra* note 17, at 857.

<sup>171</sup> Hamilton, *supra* note 18, at 102.

increase the risk they pose after they are released, as some scholars have suggested.<sup>172</sup> This is one of the major drawbacks of using a second-generation risk assessment tool that Pennsylvania has chosen. As discussed earlier, third- and fourth-generation tools are augmented with information about a defendant's criminogenic needs—that is, the deficiencies most strongly correlated with risk of future criminality. This allows actors within the criminal justice system to tailor sentencing conditions to target for intervention people most likely to benefit from correctional programming.<sup>173</sup> Without information on the elasticity to incarceration, second-generation risk assessment tools have limited utility in achieving their stated goal of reducing incarceration, even setting aside their problems with accuracy and the entrenchment of racial bias.

## VI. CONCLUSION

Reducing incarceration rates is a laudable goal. But evidence-based sentencing simply is not the answer to that problem. Advocates of risk assessment should use caution when proclaiming actuarial risk assessment is the “new way forward.” When examining the statistical reliability of the tools, proponents of risk assessment “have sometimes gone too far in describing the power of evidence-based practices to revolutionize the criminal justice system” as they currently

---

<sup>172</sup> Michael Neminski, *The Professionalization of Crime; How Prisons Create More Criminals*, <http://www.bu.edu/av/core/journal/xxiii/Neminski.pdf> [https://perma.cc/AY25-YC5K]; See Francis T. Cullen et al., *Prisons Do Not Reduce Recidivism: The High Cost of Ignoring Science*, 91 PRISON J. 48S, 48S (Supp. 2011) (reviewing studies, concluding “the use of custodial sanctions may have the unanticipated consequence of making society less safe”); Robert DeFina & Lance Hannon, *For Incapacitation, There is No Time Like the Present: The Lagged Effects of Prisoner Reentry on Property and Violent Crime Rates*, 39 SOC. SCI. RES. 1004, 1013 (2010) (concluding “any crime-reducing benefits of increased incarceration are completely wiped out by the crime-promoting effects associated with the increasing prevalence of ex-inmates”).

<sup>173</sup> Hannah-Moffat, *supra* note 4, at 279–84.

stand.<sup>174</sup> Pennsylvania has made valiant efforts to be thoughtful about the issues actuarial risk assessment presents; however, despite these efforts, considerable issues remain. We should be wary of promoting punishment based on future actions, especially when those estimates of future action are not as reliable as claimed. Even if actuarial tools increase accuracy in predicting recidivism, the concerns outlined here outweigh the utility of the slight increase in accuracy. In short, the criminal justice system needs to be conscious of systemic racism and take steps to dismantle the racial bias that plagues our system. Given the current state of actuarial risk assessment, evidence-based sentencing will only further contribute to the disparate treatment defendants of color receive.

---

<sup>174</sup> Klingele, *supra* note 169.