Marquette Law Review

Volume 103
Issue 3 *Symposium 2020*

Article 6

2020

Violence Risk Assessment: Current Status And Contemporary Issues

Sarah L. Desmarais

Samantha A. Zottola

VIOLENCE RISK ASSESSMENT: CURRENT STATUS AND CONTEMPORARY ISSUES

SARAH L. DESMARAIS* & SAMANTHA A. ZOTTOLA**

Despite the increased use of violence risk assessment instruments in the criminal justice system, they remain the topic of continued discussion and debate. This Article will discuss the state of science and practice as it relates to assessing risk for violent recidivism, highlighting current issues of concern and empirical investigation. We first provide an overview of violence risk assessment and describe the role of violence risk assessments instruments in this process. We then discuss their current status in science and practice, including the accuracy with which violence risk assessment instruments forecast violent recidivism, their impact on criminal justice decisions, and their effectiveness as a strategy to reduce violent recidivism. Finally, we turn our attention to contemporary issues in violence risk assessment, including the notion of fairness and the potential benefits, as well as concerns related to the application of technological and statistical advances in violence risk assessment—most notably, artificial intelligence. We conclude that the use of violence risk assessment instruments represents the state-of-the-art approach, but that there remain critical avenues for continued research and discussion.

* Sarah L. Desmarais, Ph.D., is a Professor of Psychology and Director of the Center for Family and Community Engagement at North Carolina State University.

** Samantha A. Zottola, M.A., is a PhD Candidate in Applied Social and Community Psychology at North Carolina State University. She will be completing her degree summer 2020.

I. INTRODUCTION

Violence risk assessment is an integral component of criminal justice decision-making. Every day, those who work in the criminal justice system engage a process of assessing the threat to public safety posed by the individuals before them. To do so, criminal justice professionals may consider information regarding the characteristics of a person, their past, their current case, or other circumstances to inform a decision regarding threat to public safety. The process may be implicit and informal in nature, involving little structure or even conscious consideration, or it may be explicit and formal in nature with strict guidelines, procedures, or even the application of a violence risk assessment instrument.

Briefly, violence risk assessment instruments are designed to increase structure, consistency, and accuracy in the evaluation of the likelihood of violent recidivism through consideration of items associated with violence recidivism. In this way, violence risk assessment instruments may help (1) identify and differentiate between people who pose lesser and greater risk of violent recidivism, (2) support criminal justice decision-making, and (3) inform risk management and interventions strategies. The process of assessing risk of violence and violence risk assessment instruments are not one and the same. While violence risk assessment instruments may be used in the process of violence risk assessment, they do not supplant or replace criminal justice decision-making. *State v. Loomis* asserts that scores produced by violence risk assessment instruments may not be the decisive factor nor the only piece of information considered in decisions of release.[1] The process of violence risk assessment will occur with or without the use of a violence risk assessment instrument; however, their use may contribute to more consistent, transparent, and accurate decisions. To that end, the use of violent risk assessment instruments to support these decisions has come to be recognized as a key component of evidence-based criminal justice policy, practice, and reform.

II. VIOLENCE RISK ASSESSMENT INSTRUMENTS

More than a hundred risk assessment instruments have been designed to forecast risk of future violent behavior. These instruments typically use one of two general approaches used to estimate violence risk: (1) algorithms or (2) structured professional judgment. Both are empirically based approaches that combine information about a person and their social environments or circumstances. These approaches differ in the strategies through which this

1. 2016 WI 68, ¶ 9, 371 Wis.2d 235, 881 N.W.2d 749; Laurel Eckhouse, Kristian Lum, Cynthia Conti-Cook, & Julie Ciccolini, *Layers of Bias: A Unified Approach for Understanding Problems with Risk Assessment*, 46 CRIM. JUST. & BEHAV. 185, 199–200 (2019).

information is used to inform estimates of risk for future violence. Instruments that use algorithms represent a more mechanical or automated approach to violence risk assessment. Typically, these instruments include items that showed the strongest statistical associations with violence in the development samples.[2] Item ratings are combined and weighted based upon statistically derived models (i.e., algorithms) to create total scores.[3] These total scores are cross-referenced (by hand or via a computer program) with actuarial tables that describe probabilities or rates of violence seen in development or norming samples.[4] In contrast, violence risk assessment instruments that use a structured professional judgment approach guide assessors to consider and rate items that have been shown in the research literature broadly, rather than in the instrument's development sample, to be associated with future violence. Though items are scored, assessors use the item ratings to inform a judgment of risk for future violence based on their professional judgment rather than using total scores.[5] While developers of algorithmic and structured professional judgment instruments have debated their relative merits, research reviews show that they estimate the likelihood of violence with comparable reliability (i.e., consistency between assessors) and predictive validity (i.e., accuracy in forecasting future violence).[6] We return to predictive validity later in this Article.

Items that are included in violence risk assessment instruments typically represent characteristics of a person, their past, their social environment, or their current circumstances that are associated with increases in the likelihood of violent recidivism—typically referred to as "risk factors." Some violence risk assessment instruments also include characteristics that, when present, mitigate or buffer these risks to reduce the likelihood of violent recidivism—typically referred to as "protective factors."[7] Even though few violence risk

---

2. N. Zoe Hilton, Grant T. Harris, & Marnie E. Rice, *Sixty-Six Years of Research on the Clinical Versus Actuarial Prediction of Violence*, 34 COUNSELING PSYCHOLOGIST 400, 401 (2016).

3. *Id.* at 402, 405.

4. *Id.* at 402.

5. Laura S. Guy, Ira K. Packer, & William Warnken, *Assessing Risk of Violence Using Structured Professional Judgment Guidelines*, 12 J. FORENSIC PSYCHOL. PRAC. 270, 271 (2012).

6. Jay P. Singh, Martin Grann, & Seena Fazel, *A Comparative Study of Violence Risk Assessment Tools: A Systematic Review and Metaregression Analysis of 68 Studies Involving 25,980 Participants*, 31 CLINICAL PSYCHOL. REV. 499, 501 (2011); Mary Ann Campbell, Sheila French, & Paul Gendreau, *The Prediction of Violence in Adult Offenders: A Meta-Analytic Comparison of Instruments and Methods of Assessment*, 36 CRIM. JUST. & BEHAV. 567, 583 (2009).

7. Corine de Ruiter & Tonia L. Nicholls, *Protective Factors in Forensic Mental Health: A New Frontier*, 10 INT'L J. FORENSIC MENTAL HEALTH 160, 161 (2011); John Monahan & Jennifer L.

assessment instruments include protective factors, research demonstrates that they contribute unique information that improves the predictive validity of violence risk assessments.[8] Both risk and protective factors can be either static or dynamic in nature. Static factors are historical or otherwise unchangeable characteristics, such as history of violent behavior or age at first arrest, whereas dynamic factors are characteristics that may change over time and/or when targeted in treatment, such as substance abuse.[9] Both algorithmic and structured professional judgment instruments can include risk and protective factors that are static or dynamic in nature. However, algorithmic instruments tend to rely more heavily on static and historical risk factors than do structured professional judgment instruments.

The manner through which information is gathered to inform item ratings and the sources of this information also differ across violence risk assessment instruments. Some violence risk assessment instruments may be filled out exclusively based upon official records. Some are filled out by the individuals themselves in a questionnaire format. Others, still, require structured interviews and observations. Some violence risk assessments instruments are computerized and automated, while others are paper-based and completed by hand. Regardless of the specific approach, format, or even contents, violence risk assessment instruments are the accepted state of science and practice when it comes to forecasting violence risk.[10]

Many violence risk assessment instruments were first developed to speak to dangerousness in the context of civil commitment hearings for individuals with serious mental health problems. Indeed, much of modern violence risk assessment has its foundations in involuntary commitment laws that shifted in focus from a person's need for mental health treatment to their dangerousness

---

Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. CLINICAL PSYCHOL. 489, 498 (2016) [hereinafter *Risk Assessment in Criminal Sentencing*].

8. Sarah L. Desmarais, Tonia L. Nicholls, Catherine M. Wilson, & Johann Brink, *Using Dynamic Risk and Protective Factors to Predict Inpatient Aggression: Reliability and Validity of START Assessments*, 24 PSYCHOL. ASSESSMENT 685, 686 (2012) [hereinafter *Using Dynamic Risk and Protective Factors*]; Michiel de Vries Robbé, Vivienne de Vogel, & Kevin S. Douglas, *Risk Factors and Protective Factors: A Two-Sided Dynamic Approach to Violence Risk Assessment*, 24 J. FORENSIC PSYCHIATRY & PSYCHOL. 440, 452 (2013).

9. Kevin S. Douglas & Jennifer L. Skeem, *Violence Risk Assessment: Getting Specific About Being Dynamic*, 11 PSYCHOL., PUB. POL'Y, & L. 347, 349–50 (2005).

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

to self or others in the 1960s.[11] These instruments were designed to predict violence, which may or may not qualify as violent *recidivism*; that is, they were designed to forecast risk for behavior that is violent but does not necessarily meet standards to be prosecuted as a crime. For example, the Violence Risk Appraisal Guide-Revised (VRAG-R)[12] uses an actuarial approach to estimate violence risk years into the future, while the Historical Clinical Risk Management-20, Version 3 (HCR-20V3)[13] and the Structured Assessment of Violence Risk in Youth (SAVRY),[14] use the structured professional judgement approach to estimate future violence over several months to a years. As one more example, the Classification of Violence Risk (COVR) uses a decision tree approach to guide the assessor though a chart review and brief interview to estimate a psychiatric patient's violence risk after discharge into the community.[15] Many violence risk assessment instruments, including the VRAG and HCR-20, have been validated for the prediction of violent recidivism, specifically, among justice-involved persons with and without mental health problems.[16]

There also are many risk assessment instruments designed to predict *general* recidivism risk, including committing a new crime and violating conditions of probation or parole. A review of studies published prior to 2012 identified more than sixty such instruments being used in criminal justice

11. Alec Buchanan, Renee Binder, Michael Norko, & Marvin Swartz, *Psychiatric Violence Risk Assessment*, 169 AM. J. PSYCHIATRY 340, 340 (2012).

12. VERNON L. QUINSEY, GRANT T. HARRIS, MARNIE E. RICE, & CATHERINE A. CORMIER, VIOLENT OFFENDERS: APPRAISING AND MANAGING RISK 341 (2nd ed. 2013).

13. *See generally* KEVIN S. DOUGLAS, STEPHEN DAVID HART, CHRISTOPHER D. WEBSTER, & HENRIK BELFRAGE, HCR-20V3: ASSESSING RISK OF VIOLENCE: USER GUIDE (2013).

14. RANDY BORUM, PATRICK BARTEL, & ADELLE FORTH, STRUCTURED ASSESSMENT OF VIOLENCE RISK IN YOUTH (SAVRY) 3–5 (2006).

15. John Monahan, Henry J. Steadman, Paul S. Appelbaum, Thomas Grisso, Edward P. Mulvey, Loren H. Roth, Pamela Clark Robbins, Stephen Banks, & Eric Silver, *The Classification of Violence Risk*, 24 BEHAV. SCI. & L. 721, 723 (2006).

16. KEVIN S. DOUGLAS, CATHERINE SHAFFER, ADAM J. E. BLANCHARD, LAURA S. GUY, KIM A. REEVES, & JOHN WEIR, HCR-20 VIOLENCE RISK ASSESSMENT SCHEME: OVERVIEW AND ANNOTATED BIBLIOGRAPHY 14 (2014), http://hcr-20.com/hcr/wp-content/uploads/2013/03/HCR-20-Annotated-Bibliography-Version-12-January-2014.pdf [https://perma.cc/EH2F-XUKQ]; Anthony J. J. Glover, Frances P. Churcher, Andrew L. Gray, Jeremy F. Mills, & Diane E. Nicholson, *A Cross-Validation of the Violence Risk Appraisal Guide—Revised (VRAG–R) Within a Correctional Sample*, 41 L. & HUM. BEHAV. 507, 508 (2017); Ana Cristina Neves, Rui Abrunhosa Gonçalves, & José Manuel Palma-Oliveira, *Assessing Risk for Violent and General Recidivism: A Study of the HCR–20 and the PCL–R with a Non-Clinical Sample of Portuguese Offenders*, 10 INT'L J. FORENSIC MENTAL HEALTH 137, 138 (2011).

settings in the United States.[17] Some—but not all—of these instruments produce separate estimates for risk of general recidivism and risk of violent recidivism, such as the Correctional Offender Management Profile for Alternative Sanctions (COMPAS)[18] and the Static Risk and Offender Needs Guide for Recidivism (STRONG-R).[19] Others do not include specific estimates for violent recidivism, but have been used and validated in that way, such as the Level of Service Inventory-Revised (LSI-R),[20] the Youth Level of Service/Case Management Inventory (YLS/CMI),[21] and the Federal Post Conviction Risk Assessment (PCRA).[22]

Other violence risk assessment instruments estimate risk of specific forms of violent offending, including sexual violence and domestic violence, such as the Static-99[23] for predicting sexual recidivism and the Ontario Domestic Assault Risk Assessment (ODARA)[24] for predicting domestic violence recidivism. Other instruments, still, are designed to estimate risk of violence and other outcomes during specific timeframes, such as during the pretrial period (i.e., between the time of arrest and the start of the trial). Such pretrial risk assessment instruments are typically designed to forecast risk for failure to appear in court or perpetration of a new crime.[25] Some of the pretrial risk

---

17. Sarah L. Desmarais, Kiersten L. Johnson, & Jay P. Singh, *Performance of Recidivism Risk Assessment Instruments in U.S. Correctional Settings*, 13 PSYCHOL. SERV. 206, 208 (2016) [hereinafter *Performance of Recidivism Risk Assessment*].

18. Tim Brennan, William Dieterich, & Beate Ehret, *Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System*, 36 CRIM. JUST. & BEHAV. 21, 34 (2009).

19. Zachary Hamilton, Alex Kigerl, Michael Campagna, Robert Barnoski, Stephen Lee, Jacqueline Van Wormer, & Lauren Block, *Designed to Fit: The Development and Validation of the STRONG-R Recidivism Risk Assessment*, 43 CRIM. JUST. & BEHAV. 230, 238–39 (2016).

20. D. A. ANDREWS & JAMES L. BONTA, LEVEL OF SERVICE INVENTORY–REVISED (LSI-R): USER'S MANUAL (2001).

21. R. D. HOGE & D. A. ANDREWS, YOUTH LEVEL OF SERVICE/CASE MANAGEMENT INVENTORY (YLS/CMI) (2006).

22. James L. Johnson, Christopher T. Lowenkamp, Scott W. VanBenschoten, & Charles R. Robinson, *The Construction and Validation of the Federal Post Conviction Risk Assessment (PCRA)*, 75 FED. PROB. 16 (2011).

23. Amy Phenix & Douglas L. Epperson, *Overview of the Development, Reliability, Validity, Scoring, and Uses of the Static-99, Static-99R, Static-2002, and Static-2002R*, *in* SEXUAL OFFENDING: PREDISPOSING ANTECEDENTS, ASSESSMENTS AND MANAGEMENT 437 (Amy Phenix & Harry M. Hoberman eds., 2016).

24. N. ZOE HILTON, GRANT T. HARRIS, & MARNIE E. RICE, RISK ASSESSMENT FOR DOMESTICALLY VIOLENT MEN: TOOLS FOR CRIMINAL JUSTICE, OFFENDER INTERVENTION, AND VICTIM SERVICES 3, 6 (2010).

25. SARAH L. DESMARAIS & EVAN M. LOWDER, PRETRIAL RISK ASSESSMENT TOOLS: A PRIMER FOR JUDGES, PROSECUTORS, AND DEFENSE ATTORNEYS 5 (2019) [hereinafter PRETRIAL RISK ASSESSMENT TOOLS].

assessment instruments, such as the Public Safety Assessment (PSA) and the COMPAS Pretrial, produce specific estimates of risk for new violent crime during the pretrial period.[26] In this way, these instruments may be used to forecast risk for violent recidivism, but in the pretrial context. Similarly, some risk assessment instruments produce estimates of risk for multiple adverse outcomes, including but not limited to violent recidivism. The Short-Term Assessment of Risk and Treatability (START), for example, is designed to estimate risk of short-term violence (i.e., over a period of several weeks to a few months), as well as six other public health and public safety outcomes.[27]

Lastly, there are some instruments that are used in the context of violence risk assessment that are not risk assessment instruments at all, but instead assess specific attitudes, beliefs, behaviors, or functioning. These may include clinical inventories, such as the Beck Depression Inventory[28] or Novaco Anger Scale,[29] personality assessment tools, such as the Psychopathy Checklist-Revised[30] or the Personality Assessment Inventory,[31] or criminal thinking scales, such as the TCU Criminal Thinking Scales[32] or the Psychological Inventory of Criminal Thinking.[33] Although these instruments do not assess risk for recidivism or violence, they frequently are used for that purpose in correctional settings in the United States and elsewhere.[34] However, research demonstrates that while their

---

26. MARIE VANNOSTRAND & CHRISTOPHER T. LOWENKAMP, ASSESSING PRETRIAL RISK WITHOUT A DEFENDANT INTERVIEW 5 (2013); PRETRIAL RISK ASSESSMENT TOOLS, *supra* note 25, at 3.

27. CHRISTOPHER D. WEBSTER, MARY-LOU MARTIN, JOHANN BRINK, TONIA L. NICHOLLS, & SARAH L. DESMARAIS, SHORT-TERM ASSESSMENT OF RISK AND TREATABILITY (START) MANUAL (1.1 ed. 2009).

28. Aaron T. Beck, Robert A. Steer, & Margery G. Garbin, *Psychometric Properties of the Beck Depression Inventory: Twenty-Five Years of Evaluation*, 8 CLINICAL PSYCHOL. REV. 77, 78 (1988).

29. Raymond W. Novaco, *Anger as a Risk Factor for Violence Among the Mentally Disordered*, *in* VIOLENCE AND MENTAL DISORDER: DEVELOPMENTS IN RISK ASSESSMENT 21, 34 (John Monahan & Henry J. Steadman eds., 1994).

30. ROBERT D HARE, THE HARE PSYCHOPATHY CHECKLIST-REVISED (2nd ed. 2003).

31. LESLIE C. MOREY, THE PERSONALITY ASSESSMENT INVENTORY PROFESSIONAL MANUAL (2nd ed. 2007).

32. Kevin Knight, Bryan R. Garner, D. Dwayne Simpson, Janis T. Morey, & Patrick M. Flynn, *An Assessment for Criminal Thinking*, 52 CRIME & DELINQ. 159, 163 (2006).

33. Glenn D. Walters, *The Psychological Inventory of Criminal Thinking Styles: Part I: Reliability and Preliminary Validity*, 22 CRIM. JUST. & BEHAV. 307, 307 (1995).

34. Jay P. Singh, Sarah L. Desmarais, Cristina Hurducas, Karin Arbach-Lucioni, Carolina Condemarin, Kimberlie Dean, Michael Doyle, Jorge O. Folino, Verónica Godoy-Cervera, Martin Grann, Robyn Mei Yee Ho, Matthew M. Large, Louise Hjort Nielsen, Thierry H. Pham, Maria Francisca Rebocho, Kim A. Reeves, Martin Rettenberger, Corine de Ruiter, Katharina Seewald, & Randy K. Otto, *International Perspectives on the Practical Application of Violence Risk Assessment: A Global Survey of 44 Countries*, 13 INT'L J. FORENSIC MENTAL HEALTH (2014); *Performance of*

results may be associated with violent recidivism (and recidivism more generally), they generally have less predictive capacity than violence risk assessment instruments.[35] In other words, these assessments can account for some variation in risk for future violent behavior but less than would be accounted for by those produced using violence risk assessment instruments. They simply do not provide the full picture.

III. CURRENT STATUS

A comprehensive review of the content, methods, strengths, and limitations of each violence risk assessment instrument is beyond the scope of this Article. Readers are referred to the systematic reviews and meta-analyses referenced earlier for more detailed discussion.[36] Instead, in the Sections that follow, we consider the current status of the violence risk assessment instruments, summarizing findings of the extant research literature regarding their validity and impact, as well as highlighting areas of continued debate and investigation.

A. *Predictive Validity*

Predictive validity speaks to the accuracy with which risk assessment instruments produce scores or estimates that can—and do—forecast violent recidivism. It is well established in the research literature that structured, empirically based assessments produce more accurate estimates of future behavior than do evaluations made by clinicians, judges, or others based upon their professional training and experience in the absence of structured checklists, protocols, or guidelines.[37] For more than half a century, researchers have been examining and comparing the accuracy of unstructured and structured predictions of human behavior, from violence to suicide to other behaviors of interest or concern. Research efforts related to violence risk assessment, specifically, were spurred on by Ennis and Litwack's 1974

Recidivism Risk Assessment, *supra* note 17, at 206; Jodi L. Viljoen, Kaitlyn McLachlan, & Gina M. Vincent, *Assessing Violence Risk and Psychopathy in Juvenile and Adult Offenders: A Survey of Clinical Practices*, 17 ASSESSMENT 377, 377 (2010).

35. *See, e.g.*, Neves, Gonçalves, & Palma-Oliveira, *supra* note 16, at 137, 139–40; Mark E. Hastings, Shilpa Krishnan, June P. Tangney, & Jeffrey Stuewig, *Predictive and Incremental Validity of the Violence Risk Appraisal Guide Scores with Male and Female Jail Inmates*, 23 PSYCHOL. ASSESSMENT 174, 182 (2011); *Using Dynamic Risk and Protective Factors*, *supra* note 8, at 695.

36. *See supra* notes 2–6 and accompanying text.

37. *See* Hilton, Harris, & Rice, *supra* note 2, at 400 ("[S]tatistical prediction was about 10% more accurate than clinical prediction and was consistently superior across date and source of publication, type of judge . . . general or task-relevant experience, type of data . . . and amount of data available.").

monograph in which they compared psychiatrists' predictions of dangerousness to "[f]lipping coins in the courtroom" because they were so frequently biased and inaccurate.[38] Since then, there have been dozens of investigations and comparisons of unstructured professional judgment and structured, empirically based assessments. In its totality, decades of empirical investigation demonstrate the superiority of structured, empirically based assessments over unstructured ones in terms of their accuracy in forecasting future behavior. Moreover, the benefits of structure seem to be particularly strong for predictions of violence and other criminal outcomes.[39] We summarize some key findings below.

As early as the 1950s and 1960s, reviews of unstructured versus empirically based predictions of human behavior concluded that the latter often produced more accurate predictions of future behavior than the former.[40] The first meta-analysis of the research (i.e., quantitative synthesis of findings across research)[41] examined studies conducted between 1966 and 1988.[42] Results showed that empirically based predictions often demonstrated superiority over unstructured professional judgements of future health and behavior, regardless of the task, judge, amount of experience, or types of information. Another meta-analysis conducted about ten years later showed even more compelling results: empirically based predictions were more accurate than unstructured judgments overall, with increases of 17% seen for predictions of violence and other criminal outcomes, specifically.[43] This meta-analysis showed that out of 1,000 predictions, empirically based predictions accurately identified 90 more individuals who went on to be violent than did unstructured professional judgments.[44] Another meta-analysis examining predictions of sexual

38. Bruce J. Ennis & Thomas R. Litwack, *Psychiatry and the Presumption of Expertise: Flipping Coins in the Courtroom*, 62 CALIF. L. REV. 693, 719, 728, 737 (1974).

39. Stefanía Ægisdóttir, Michael J. White, Paul M. Spengler, Alan S. Maugherman, Linda A. Anderson, Robert S. Cook, Cassandra N. Nichols, Georgios K. Lampropoulos, Blain S. Walker, Genna Cohen, & Jeffrey D. Rush, *The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction*, 34 COUNSELING PSYCHOLOGIST 341, 368 (2006).

40. PAUL E. MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION: A THEORETICAL ANALYSIS AND REVIEW OF THE EVIDENCE iii (1996); Jack Sawyer, *Measurement and Prediction, Clinical and Statistical*, 66 PSYCHOL. BULL. 178, 192 (1966).

41. Anna-Bettina Haidich, *Meta-Analysis in Medical Research*, 14 HIPPOKRATIA 29, 29 (2010).

42. William M. Grove, David H. Zald, Boyd S. Lebow, Beth E. Snitz, & Chad Nelson, *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 12 PSYCHOL. ASSESSMENT 19, 20 (2000).

43. Ægisdóttir, White, Spengler, Maugherman, Anderson, Cook, Nichols, Lampropoulos, Walker, Cohen, & Rush, *supra* note 39, at 356, 360, 367–68.

44. *Id.* at 368; Hilton, Harris, & Rice, *supra* note 2, at 400.

recidivism found even greater advantage conferred by the use of risk assessment instruments;[45] specifically, the effect size was almost 90% larger for empirically based assessments compared to unstructured professional judgments.[46] In short, there is both substantial and consistent evidence from meta-analyses—considered the highest level of scientific evidence[47]—regarding the increases in accuracy associated with the use of violence risk assessment instruments.

There have now been more than a dozen peer-reviewed, meta-analytic studies of the predictive validity of violence risk assessments produced by specific instruments.[48] These meta-analyses, representing a cumulative analysis of hundreds of thousands of risk assessments, have all shown statistically significant associations between the results of violence risk

---

45. *See generally* R. KARL HANSON & KELLY MORTON-BOURGON, PREDICTORS OF SEXUAL RECIDIVISM: AN UPDATED META-ANALYSIS (2004).

46. Hilton, Harris, & Rice, *supra* note 2, at 402–03.

47. M. Hassan Murad, Noor Asi, Mouaz Alsawas, & Fares Alahdab, *New Evidence Pyramid*, 21 EVIDENCE-BASED MED. 125, 126 (2016).

48. These studies include, but are not limited to, Campbell, French, & Gendreau, *supra* note 6, at 568; Seena Fazel, Jay P. Singh, Helen Doll, & Martin Grann, *Use of Risk Assessment Instruments to Predict Violence and Antisocial Behaviour in 73 Samples Involving 24,827 People: Systematic Review and Meta-Analysis*, 345 BRIT. MED. J. 1, 2 (2012); Rachael Lofthouse, Laura Golding, Vasiliki Totsika, Richard Hastings, & William Lindsay, *How Effective are Risk Assessments/Measures for Predicting Future Aggressive Behaviour in Adults with Intellectual Disabilities (ID): A Systematic Review and Meta-Analysis*, 58 CLINICAL PSYCHOL. REV. 76, 78 (2017); Laura E. O'Shea & Geoffrey L. Dickens, *Performance of Protective Factors Assessment in Risk Prediction for Adults: Systematic Review and Meta-Analysis*, 23 CLINICAL PSYCHOL.: SCI. & PRAC. 126, 128 (2016); Laura E. O'Shea, Amy E. Mitchell, Marco M. Piccioni, & Geoffrey L. Dickens, *Moderators of the Predictive Efficacy of the Historical, Clinical and Risk Management-20 for Aggression in Psychiatric Facilities: Systematic Review and Meta-Analysis*, 18 AGGRESSION & VIOLENT BEHAV. 255, 258 (2013); Mark E. Olver, Keira C. Stockdale, & J. Stephen Wormith, *Risk Assessment with Young Offenders: A Meta-Analysis of Three Assessment Measures*, 36 CRIM. JUST. & BEHAV. 329, 333 (2009) [hereinafter *Risk Assessment with Young Offenders*]; Mark E. Olver, Keira C. Stockdale, & J. Stephen Wormith, *Thirty Years of Research on the Level of Service Scales: A Meta-Analytic Examination of Predictive Accuracy and Sources of Variability*, 26 PSYCHOL. ASSESSMENT 156, 159 (2014); Singh, Grann, & Fazel, *supra* note 6, at 502; Leslie-Maaike Helmus & David Thornton, *Stability and Predictive and Incremental Accuracy of the Individual Items of Static-99R and Static-2002R in Predicting Sexual Recidivism: A Meta-Analysis*, 42 CRIM. JUST. & BEHAV. 917, 921 (2015); Min Yang, Stephen C. P. Wong, & Jeremy Coid, *The Efficacy of Violence Prediction: A Meta-Analytic Comparison of Nine Risk Assessment Tools*, 136 PSYCHOL. BULL. 740, 745 (2010); Jodi L. Viljoen, Sarah Mordell, & Jennifer L. Beneteau, *Prediction of Adolescent Sexual Reoffending: A Meta-Analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99*, 36 L. & HUM. BEHAV. 423, 425 (2012); Craig S. Schwalbe, *A Meta-Analysis of Juvenile Justice Risk Assessment Instruments: Predictive Validity by Gender*, 35 CRIM. JUST. & BEHAV. 1367, 1371 (2008); Craig S. Schwalbe, *Risk Assessment for Juvenile Justice: A Meta-Analysis*, 31 L. & HUM. BEHAV. 449, 452 (2007) [hereinafter *Risk Assessment for Juvenile Justice*].

assessment instruments and future violent behavior.[49] Across these meta-analyses, there are some differences in predictive validity as a function of various factors, such as the follow-up periods, outcomes, and populations. However, experts agree there is no single violence risk assessment instrument that produces the most accurate assessments but, instead, that many instruments produce estimates that forecast future violence with roughly comparable rates of accuracy, when they are implemented in practice with fidelity (i.e., following the guidelines provided in the instrument manual, coding protocol, etc.).[50]

At the same time, there is some research evidence that supports the need for specificity in the application of risk assessment instruments. In particular, risk assessment instruments often produce the most accurate predictions when used to assess risk for the intended outcome and in the intended population. Although there are some exceptions, research demonstrates that violence risk assessment instruments generally produce more accurate predictions of violent recidivism than do instruments designed to estimate risk of sexual violence or non-violent recidivism.[51] And, sexual violence risk assessment instruments typically are better at forecasting sexual recidivism than are (general) violence risk assessment instruments.[52] Similar results have been found with respect the predictive validity of risk assessments designed for different age groups; to demonstrate, risk assessment instruments designed for adults produce assessments with greater predictive validity when applied to adults than to youth.[53] This is not to say, however, that risk assessment instruments cannot predict other outcomes in other populations, but rather that their generalizability must be established via validation studies.

Predictive validity is not a property of a violence risk assessment instrument itself; instead, the assessments completed using a given instrument can have

---

49. *Risk Assessment for Juvenile Justice*, *supra* note 48, at 458.

50. *Current Directions in Violence Risk Assessment*, *supra* note 10, at 40–41.

51. Joel K. Cartwright, Sarah L. Desmarais, Justin Hazel, Travis Griffith, & Allen Azizian, *Predictive Validity of HCR-20, START, and Static-99R Assessments in Predicting Institutional Aggression Among Sexual Offenders*, 42 L. & HUM. BEHAV. 13, 22 (2018); Ægisdóttir, White, Spengler, Maugherman, Anderson, Cook, Nichols, Lampropoulos, Walker, Cohen, & Rush, *supra* note 39, at 344–45; J. Stephen Wormith, Sarah Hogg, & Lina Guzzo, *The Predictive Validity of a General Risk/Needs Assessment Inventory on Sexual Offender Recidivism and an Exploration of the Professional Override*, 39 CRIM. JUST. & BEHAV. 1511, 1529–30 (2012).

52. HANSON & MORTON-BOURGON, *supra* note 45, at 11.

53. Kevin M. Williams, J. Stephen Wormith, James Bonta, & Gill Sitarenios, *The Use of Meta-Analysis to Compare and Select Offender Risk Instruments: A Commentary on Singh, Grann, and Fazel (2011)*, 16 INT'L J. FORENSIC MENTAL HEALTH 1, 4, 10–11 (2017); *see* Gina M. Vincent, Dara Drawbridge, & Maryann Davis, *The Validity of Risk Assessment Instruments for Transition-Age Youth*, 87 J. CONSULTING & CLINICAL PSYCHOL. 171, 173–74, 182 (2019).

predictive validity.[54] This is a subtle but important distinction. A well-validated violence risk assessment instrument may produce assessments with poor accuracy in forecasting violent recidivism for many different reasons including, but not limited to, the accuracy and availability of information required to complete the assessments, the attitudes, training, and knowledge of individuals completing the assessments, and the base rate of violent recidivism in that jurisdiction.[55] These should be key considerations in the selection of a violence risk assessment instrument.[56] Absent the necessary information and time, implementation with fidelity is simply not possible, even with highly motivated, knowledgeable, and well-trained staff. Further, there may be some jurisdictional differences, including variations in penal codes and base rates, which necessitate modifying the instrument.[57] However, if such modifications are required, the instrument developers should be consulted and the modified version should be shared widely and subject to local evaluation.[58] Even in the absence of such modifications, a pilot implementation and evaluation should be conducted prior to full implementation to establish jurisdiction-specific base rates and predictive validity.[59]

B. *Impact*

For all the research that has been done on predictive validity, there has been much less investigation of the impact of violence risk assessment instruments. A recent systematic review and meta-analysis examined whether the use of risk assessment instruments (not specific to violence) decreased restrictive placements, including pretrial detention, post-conviction placements, and

---

54. AM. EDUC. RESEARCH ASS'N, AM. PSYCHOLOGICAL ASS'N & NAT'L COUNCIL ON MEASUREMENT IN EDUC., STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (2014).

55. Jeremy F. Mills, Michael N. Jones, & Daryl G. Kroner, *An Examination of the Generalizability of LSI-R and VRAG Probability Bins*, 32 CRIM. JUST. & BEHAV 565, 580–82 (2005); Gina M. Vincent, Melissa L. Paiva-Salisbury, Nathan E. Cook, Laura S. Guy, & Rachael T. Perrault, *Impact of Risk/Needs Assessment on Juvenile Probation Officers' Decision Making: Importance of Implementation*, 18 PSYCHOL., PUB. POL'Y, & L. 549, 550 (2012); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2118, 2236–37, 2243–50, 2284–86 (2019).

56. *Performance of Recidivism Risk Assessment*, *supra* note 17, at 217.

57. Vivienne de Vogel & Michiel de Vries Robbé, *Adapting Risk Assessment Tools to New Jurisdictions*, *in* INTERNATIONAL PERSPECTIVES ON VIOLENCE RISK ASSESSMENT 26, 27, 30 (Jay P. Singh, Stål Bjørkly, & Seena Fazel eds., 2016).

58. *Id.* at 31.

59. GINA M. VINCENT, LAURA S. GUY, & THOMAS GRISSO, RISK ASSESSMENT IN JUVENILE JUSTICE: A GUIDEBOOK FOR IMPLEMENTATION 9 (2012).

release, for adult and adolescent defendants or offenders.[60] Overall, results showed that the use of risk assessment instruments was associated with decreases in restrictive placements in more than two-thirds of the twenty-two included studies.[61] Broken down across stages of case processing, 64% of studies showed decreases in pretrial detention, 60% of studies showed decreases in post-conviction placements, and 100% of studies showed increased release from custody.[62] However, many of these studies—thirteen out of twenty-two—were rated at serious risk of study bias (e.g., confounding variables, selection biases, missing data).[63] When the authors reran their analyses without these thirteen potentially biased studies, just over half of the remaining studies showed reductions in restrictive placements associated with the use of risk assessment instruments.[64]

These findings highlight the importance of stakeholder buy-in during implementation. Insufficient training, lack of time, a preference for unstructured approaches, and skepticism about whether violence risk assessment instruments really work can impact their uptake in practice.[65] Interviews with judges underscore their preference for unstructured risk assessment and skepticism about whether violence risk assessment instruments forecast violence with any accuracy.[66] Surveys of judges also emphasize concerns regarding the availability of community resources to manage even low levels of risk. For instance, when judges were asked in a recent survey about availability of local treatment resources that would allow for alternative sentencing, seven out of ten rated the current resources as "less than adequate" and 5% rated local resources as "virtually non-existent."[67] A follow-up study

60. Jodi L. Viljoen, Melissa R. Jonnson, Dana M. Cochrane, Lee M. Vargen & Gina M. Vincent, *Impact of Risk Assessment Instruments on Rates of Pretrial Detention, Postconviction Placements, and Release: A Systematic Review and Meta-Analysis*, 43 L. & HUM. BEHAV. 397, 399, 401 (2019) [hereinafter *Impact of Risk Assessment Instruments on Rates*].

61. *Id.* at 402.

62. *Id.*

63. *Id.* at 401, 402.

64. *Id.* at 402.

65. *See, e.g.*, Tamara L. F. De Beuf, Vivienne de Vogel, & Corine de Ruiter, *Implementing the START:AV in a Dutch Residential Youth Facility: Outcomes of Success*, 5 TRANSLATIONAL ISSUES PSYCHOL. SCI. 193, 196 (2019); Impact of Risk Assessment Instruments on Rates, supra note 60, at 383, 386.

66. Anne Metz, John Monahan, Brandon Garrett, & Luke Siebert, *Risk and Resources: A Qualitative Perspective on Low-Level Sentencing in Virginia*, J. COMMUNITY PSYCHOL. 1476, 1483, 1486 (2019).

67. John Monahan, Anne L. Metz, & Brandon Garrett, *Judicial Appraisals of Risk Assessment in Sentencing*, 36 BEHAV. SCI. & LAW 565, 569 (2018).

showed that the percent of offenders who scored low on a risk assessment instrument and actually received the recommended alternative sentence varied from 22% to 67% between circuit court judges in Virginia.[68] This study also showed that the rate of alternative sentences increased as did the availability of treatment in the community.[69] So, there are many reasons that violence risk assessment instruments are not having a widespread impact on criminal justice practices. Some of these issues, such as judges' knowledge of violence risk assessment instruments, could be addressed during implementation, while others, such as the availability of community-based treatment options, may be more challenging to overcome.

There has been even less rigorous, scientific study of whether violence risk assessment instruments contribute to reductions in violent recidivism. A recent systematic review identified twelve studies that examined whether using risk assessment instruments reduced violence and general offending.[70] These studies included two randomized controlled trials (RCTs) that showed reductions in violence associated with the use of the Brøset Violence Checklist among psychiatric hospital patients in Switzerland and the Netherlands.[71] A third RCT examining the START failed to find an impact on violence in outpatient forensic psychiatric patients in the Netherlands.[72] Two pre-post studies found reductions in violence or reoffending following the implementation of the PSA and HCR-20, respectively, while six other pre-post studies of various risk assessment instruments did not find such effects.[73] One study of SAVRY assessments completed for youth on probation found fewer

68. Brandon L. Garrett, Alexander Jakubow, & John Monahan, *Judicial Reliance on Risk Assessment in Sentencing Drug and Property Offenders: A Test of the Treatment Resource Hypothesis*, 46 CRIM. JUST. & BEHAV. 799, 807–08 (2019).

69. *Id.* at 808.

70. Jodi L. Viljoen, Dana M. Cochrane, & Melissa R. Jonnson, *Do Risk Assessment Tools Help Manage and Reduce Risk of Violence and Reoffending? A Systematic Review*, 42 L. & HUM. BEHAV. 181, 184, 198 (2018) [hereinafter *Do Risk Assessment Tools Help Manage and Reduce Risk*].

71. Christoph Abderhalden, Ian Needham, Theo Dassen, Ruud Halfens, Hans-Joachim Haug & Joachim E. Fischer, *Structured Risk Assessment and Violence in Acute Psychiatric Wards: Randomised Controlled Trial*, 193 BRIT. J. PSYCHIATRY 44, 48 (2008); R. van de Sande, H. L. I. Nijman, E. O. Noorthoorn, A. I. Wierdsma, E. Hellendoorn, C. van der Staak & C. L. Mulder, *Aggression and Seclusion on Acute Psychiatric Wards: Effect of Short-Term Risk Assessment*, 199 BRIT. J. PSYCHIATRY 473, 476 (2011).

72. N. A. C. Troquete, R. H. S. van den Brink, H. Beintema, T. Mulder, T. W. D. P. van Os, R. A. Schoevers, & D. Wiersma, *Risk Assessment and Shared Care Planning in Out-Patient Forensic Psychiatry: Cluster Randomised Controlled Trial*, 202 BRIT. J. PSYCHIATRY 365, 366, 368 (2013).

73. *Do Risk Assessment Tools Help Manage and Reduce Risk*, *supra* note 70, at 200–02.

violent offenses but no differences in general offenses or probation violations.[74] These mixed findings across studies may be attributable to variations in in the settings, populations, and instruments tested, as well as the research methods. In particular, only three studies employed an RCT design, which is the gold standard research design for testing intervention effects.[75] However, even the RCTs suffered methodological problems, including limited use of the violence risk assessment instrument under investigation.

Taken together, findings of the extant research suggest that violence risk assessment instruments can have a positive impact on criminal justice practices and violent recidivism. However, they also speak to the importance of implementation. Indeed, evaluations of the impact of violence risk assessment instruments may produce erroneous conclusions due to implementation issues, as noted above. Another unpublished effort to study the impact of the HCR-20 on violence among psychiatric patients failed to show significant differences between the group assessed using the HCR-20 and a comparison group assessed using a different instrument.[76] Further examination of the records for the HCR-20 revealed that very few HCR-20 assessments were actually completed.[77] Thus, apparent lack of effectiveness of the HRC-20 actually reflected unsuccessful implementation. On the whole, there is a critical need for largescale research studies—and RCTs, in particular—to establish the impact of violence risk assessment on criminal justice practices and their effectiveness in reducing violent recidivism.

IV. CONTEMPORARY ISSUES

Thus far, we have provided an overview of violence risk assessment instruments and a summary of current scientific knowledge regarding their validity in forecasting violent recidivism and impact on criminal justice outcomes. We now delve into two contemporary issues in violence risk assessment: (1) fairness and (2) artificial intelligence.

74. LAURA S. GUY, GINA M. VINCENT, THOMAS GRISSO, & RACHAEL PERRAULT, ADVANCING USE OF RISK ASSESSMENT IN JUVENILE PROBATION 22–23, 39, 78, 85 (2015).

75. Murad, Asi, Alsawas, & Alahdab, *supra* note 47, at 125.

76. *See* Adrian Cree, *Perceived Barriers to the Implementation of Violence Risk Assessment Tools*, *in* INTERNATIONAL PERSPECTIVES ON VIOLENCE RISK ASSESSMENT 166 (Jay P. Singh, Stål Bjørkly & Seena Fazel eds., 2016).

77. *Id.* at 166–67.

A. *Fairness*

Much of the recent debate regarding the use of risk assessment instruments in the criminal justice system centers on fairness (or the lack thereof) and, more specifically, racial bias. "Generally [speaking], a process or decision is considered fair if it does not discriminate against people on the basis of their membership in a protected group."[78] Fairness in the context of risk assessment has been defined in many different ways. Some have discussed fairness in risk assessment as reflecting three criteria; specifically, to be considered fair, risk assessment instruments must produce: (1) risk scores that have similar meanings for all groups (i.e., similar likelihood of recidivism), (2) similar scores across non-recidivists of all groups, and (3) similar scores across recidivists of all groups.[79] Others have described fairness in risk assessment as comprising two separate issues: (1) *predictive bias,* also known as differential prediction, which is found when a violence risk assessment instruments demonstrates different levels of predictive validity across groups, and (2) *disparate impact*, which is seen when violence risk assessment results are applied in ways that are unequal or unfair across groups.[80]

Discussion of predictive bias must address three separate but related issues: first, whether certain groups have higher rates of violent recidivism than others; second, whether certain groups of people receive higher risk scores than other groups; and, third, whether certain groups are overclassified at higher risk levels and underclassified at lower risk levels relative to their actual rates of violent recidivism and relative to other groups.[81] Individuals with higher rates of true violent recidivism (i.e., actual behavior) should receive higher risk scores and those with lower rates, lower risk scores. Concerns have been raised that structured, empirically based risk assessment approaches, especially algorithmic methods, overestimate risk of violent recidivism among persons of color and underestimate risk of violent recidivism among white individuals.[82]

---

78. Songül Tolan, Marius Miron, Emilia Gómez, & Carlos Castillo, *Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia*, 17 INT'L CONF. ON ARTIFICIAL INTELLIGENCE & L. 83, 83 (2019).

79. Jon Kleinberg, Sendhil Mullainathan, & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, 8 INNOVATIONS THEORETICAL COMPUTER SCI. CONF. 43:1, 43:5 (2017).

80. Jennifer L. Skeem & Christopher T. Lowenkamp, *Risk, Race, and Recidivism: Predictive Bias and Disparate Impact*, 54 CRIMINOLOGY 680, 685 (2016); Evan M. Lowder, Megan M. Morrison, Daryl G. Kroner, & Sarah L. Desmarais, *Racial Bias and LSI-R Assessments in Probation Sentencing and Outcomes*, 46 CRIM. JUST. & BEHAV. 210, 213–14 (2019).

81. Lowder, Morrison, Kroner, & Desmarais, *supra* note 80, at 215.

82. *Id.* at 212.

Factors that contribute to these over and underclassifications, and thus, the potential for predictive bias, may include differences in base rates and problems inherent in the data used to complete and to design the instruments. We consider each of these in turn.

Information on rates of violent crime will—and should—be used in some way to inform estimates of risk for violent recidivism.[83] Meta-analytic research demonstrates both higher base rates of violence as well as greater risk of violent recidivism among persons of color compared to white persons.[84] Whether higher rates of violence reflect true differences in behavior or, more likely, the result of systemic factors, such as unfair police practices and discriminatory prosecutorial decisions, is the subject of much examination and debate.[85] Indeed, much has been written about the extent to which systemic factors have contributed to the overrepresentation of people of color in the criminal justice system.[86] However, individual behavior is influenced by a complex and nested system of individual, social, community, and societal level factors.[87] Consequently, we may see higher rates of violent recidivism in certain groups for many reasons.[88] Regardless, differences in base rates of violent behavior and recidivism may result in statistical models that over or underestimate risk of certain groups of people. Ideally, "risk assessments should provide similar ability to discriminate between risk classifications for different racial groups, regardless of the base rate of offending in each group."[89] A laudable goal that is challenging in practice.

Other information used in the violence risk assessment process also may contribute to predictive bias. Some argue that items that measure criminal history, employment, education level, debt, or housing stability may serve as proxies for race because they reflect racial marginalization[90] and that their

83. JOHN MONAHAN, THE CLINICAL PREDICTION OF VIOLENT BEHAVIOR 108 (1981).

84. Alex R. Piquero, Wesley G. Jennings, Brie Diamond, & Jennifer M. Reingle, *A Systematic Review of Age, Sex, Ethnicity, and Race as Predictors of Violent Recidivism*, 59 INT'L J. OFFENDER THERAPY & COMP. CRIMINOLOGY 5, 10, 11, 17 (2015).

85. *See* ELIJAH ANDERSON, CODE OF THE STREET: DECENCY, VIOLENCE, AND THE MORAL LIFE OF THE INNER CITY 9 (1999).

86. *See* MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS 6–8 (rev. ed. 2012).

87. Etienne G. Krug, James A. Mercy, Linda L. Dahlberg, & Anthony B. Zwi, *The World Report on Violence and Health*, 360 LANCET 1083, 1085 (2002).

88. Piquero, Jennings, Diamond, & Reingle, *supra* note 84, at 6.

89. Lowder, Morrison, Kroner, & Desmarais, *supra* note 80, at 214.

90. Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 805–06 (2014); Bernard E. Harcourt, *Risk as a Proxy for Race: The Dangers of Risk Assessment*, 27 FED. SENT'G REP. 237, 237 (2015).

inclusion in the process of violence risk assessment may disadvantage people of color due to structural inequalities.[91] Others, however, contend that these items do not serve as proxies for race because they are stronger predictors of violent recidivism than race, and thus, cannot serve as proxies for something with lesser predictive capacity.[92] Rather, they argue that these variables may overlap or interact with race but are not being used to conceal the use of race as a predictor of violent recidivism.[93] Moreover, there is evidence that criminal history mediates the relationship between race and future violent arrest rather than serving as a proxy for race.[94] In other words, criminal history helps explain some but not all of the statistical association between race and recidivism. Strongly opposing views of whether criminal history is a proxy for race likely reflect the use of different definitions of the term "proxy."[95] Nevertheless, to the extent that there is bias in the information used to complete the assessments, there may be bias in their predictions, too.

What does the scientific literature tell us regarding predictive bias against racial and ethnic minorities in violence risk assessment? While the empirical literature is relatively small, in general, meta-analyses and rigorous, largescale studies fail to find lower rates of predictive validity for racial and ethnic minorities compared to white individuals or other evidence of predictive bias. As one example, a recent empirical investigation of more than 30,000 PCRA assessments completed on federal offenders across the United States found minimal mean differences in PCRA scores as a function of race and strong validity in predicting re-arrest for violent crime among black and white offenders.[96] This study also found that any given PCRA score corresponded to the same probability of violent recidivism for both black and white offenders.[97] Systematic reviews and meta-analyses also have found limited evidence of predictive bias in diverse racial and ethnic groups across various risk assessment instruments, jurisdictions, and populations.[98]

---

91. Lowder, Morrison, Kroner & Desmarais, *supra* note 80, at 226, 228.

92. *Risk Assessment in Criminal Sentencing*, *supra* note 7, at 499.

93. *Id.*

94. Skeem & Lowenkamp, *supra* note 80, at 700.

95. Mayson, *supra* note 55, at 2232–33.

96. Skeem & Lowenkamp, *supra* note 80, at 691.

97. *Id.* at 692.

98. *Performance of Recidivism Risk Assessment*, *supra* note 17, at 216; *Risk Assessment with Young Offenders*, *supra* note 48, at 343; Jay P. Singh & Seena Fazel, *Forensic Risk Assessment: A Metareview*, 37 CRIM. JUST. & BEHAV. 965, 984 (2010); Holly A. Wilson & Leticia Gutierrez, *Does One Size Fit All? A Meta-Analysis Examining the Predictive Ability of the Level of Service Inventory (LSI) with Aboriginal Offenders*, 41 CRIM. JUST. & BEHAV. 196, 214 (2014).

This body of research is in stark contrast with the assertions of civil rights advocates, some scholars, and, notably, the American Bail Coalition, who are leading efforts to eliminate the use of risk assessment instruments in the criminal justice system.[99] The ProPublica article by Angwin and colleagues is widely cited in support of these efforts as providing *clear evidence* that there is racial bias in risk assessment instruments.[100] Briefly, Angwin and colleagues obtained COMPAS risk scores for more than 7,000 people arrested in Broward County, Florida in 2013 and 2014 and compared these ratings against the rates of new crimes over the next two years.[101] Angwin and colleagues report low rates of predictive accuracy, with only 20% of people predicted to commit violent crimes actually engaging in violent recidivism.[102] Further, they report higher rates of false positives vis-à-vis violent recidivism for black than white defendants.[103] Consequently, the story was published with the headline: "There's software used across the country to predict future criminals. And it's biased against blacks."[104] This claim fueled calls to abolish the use risk assessment instruments in the criminal justice system.

In the scientific literature, however, the ProPublica study has been widely criticized. Critiques include its misapplication of COMPAS assessments to a population for which they were not intended, faulty assumptions regarding the structure of their data, misinterpretation of their statistical results, and failure to use established standards for testing for bias and determining significance.[105]

99. *See, e.g.*, Chelsea Barabas, Karthik Dinakar, & Colin Doyle, *The Problems with Risk Assessment Tools*, N.Y. TIMES (July 17, 2019), https://www.nytimes.com/2019/07/17/opinion/pretrial-ai.html [https://perma.cc/LZ6N-H25E]; *see also* Sarah Desmarais, Brandon Garrett, & Cynthia Rudin, *Risk Assessment Tools Are Not a Failed Minority Report*, LAW 360 (July 19, 2019), https://www.law360.com/articles/1180373 [https://perma.cc/8UNB-R65F].

100. Julia Angwin, Jeff Larson, Surya Mattu, & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/L292-FFG]; *see also* DAVID G. ROBINSON & LOGAN KOEPKE, CIVIL RIGHTS AND PRETRIAL RISK ASSESSMENT INSTRUMENTS 4 n.7 (2019); Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES 1, 5 n.2 (2018); Eckhouse, Lum, Conti-Cook, & Ciccolini, *supra* note 1, at 206; Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 328–29 n.161 (2018).

101. Angwin, Larson, Mattu, & Kirchner, *supra* note 100.

102. *Id.*

103. *Id.*

104. *Id.*

105. Anthony W. Flores, Kristin Bechtel & Christopher T. Lowenkamp, *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks."*, 80 FED. PROB. 38, 39–40 (2016); Cynthia Rudin, Caroline Wang & Beau Coker, *The Age of Secrecy and Unfairness in*

At least two groups of leading scholars have re-analyzed the data shared by ProPublica and have found that the COMPAS predicts recidivism in very similar ways for both black and white defendants.[106] These re-analyses show the COMPAS assessments to be good predictors of violent recidivism in both white and black defendants.[107] These re-analyses also have failed to find significant interaction effects between race and COMPAS scores nor have they found that race by COMPAS score interactions add any predictive power to their analytic models, which are the two criteria needed to establish predictive bias in the context of assessment.[108] These findings suggest that Angwin and colleagues' assertion regarding racial bias in the COMPAS was erroneous. As Rudin and colleagues noted, "faulty assumptions about a proprietary algorithm lead to faulty conclusions that go unchecked."[109]

Moving on to disparate impact, there has been less research examining whether the use of violence risk assessment instruments contribute to increased racial disparities in criminal justice practices. In the meta-analysis of the impact of risk assessment instruments on restrictive placements discussed earlier, only six of twenty-two studies examined racial disparities.[110] While these studies focused on pretrial risk assessment instruments rather than violence risk assessment instruments specifically, their results suggest that using risk assessments may improve outcomes for people of color. Specifically, all but one found study that the absolute rates of restrictive placements decreased anywhere from 6% to 57% across these studies.[111] Comparison of detention rates for people of color and their white counterparts showed that rates of restrictive placements decreased *more* for people of color in three studies and decreased at the same rate in one study.[112] Two studies found mixed results.[113] However, four of these six studies were rated at high risk of study bias.[114]

Fairness in violence risk assessment remains a critical avenue of discussion and investigation. Herein we focused on fairness conceptualized as reflecting the related but distinct constructs of predictive bias and disparate impact. We

---

Recidivism Prediction, HARV. DATA SCI. REV., 2020, at 1, 3 [hereinafter *The Age of Secrecy and Unfairness*].

106. Flores, Bechtel & Lowenkamp, *supra* note 105, at 40–41, 44.

107. Id. at 44.

108. AM. EDUC. RESEARCH ASS'N, *supra* note 54.

109. *The Age of Secrecy and Unfairness*, *supra* note 105, at 1.

110. *Impact of Risk Assessment Instruments on Rates*, *supra* note 60, at 408, 410.

111. *Id.* at 408.

112. *Id.*

113. *Id.*

114. *Id.* at 406.

failed to find compelling evidence to support the existence of either in the context of violence risk assessment. In fact, on the whole, research to date suggests that the use of violence risk assessments instruments can—and often does—contribute to reductions in rates of restrictive placements for people of color at rates equivalent to or even greater than the reductions seen in their white counterparts. That said, there have been only a handful of studies on racial disparity; more rigorous and unbiased scientific study is needed. Nonetheless, though risk assessment instruments do not appear to exacerbate racial disparities, they also may not produce racial equity in criminal justice practices. Thus, while "risk assessment tools may not achieve a defined notion of fairness, but rather be comparatively better than the status quo."[115]

B. *Artificial Intelligence*

Most current violence risk assessment instruments use rather unsophisticated, classical methods to weight and combine information to forecast likelihood of violent recidivism. However, there is considerable interest—and perhaps even more debate—regarding the application of artificial intelligence (AI) via machine learning into this process. Briefly, AI refers the ability of machines to carry out tasks in smart or human-like ways that can lead to decision making without human intervention. Machine learning represents the application of AI, such that machines access data and learn from that data to inform decisions.[116] So, while traditional violence risk assessment instruments are designed to inform but not replace criminal justice decision-making, AI approaches may in fact be designed with the intention of replacing judicial decisions.

Machine learning algorithms can be created to detect patterns in data and to then develop predictions about future actions based on those patterns. In this way, data that are already collected during the course of police investigations or jail intake, for example, could be used to train models that would predict a person's risk of engaging in violence in the future.[117] On the one hand, machine learning may have many potential benefits. For instance, machine learning

115. PARTNERSHIP ON AI, REPORT ON ALGORITHMIC RISK ASSESSMENT TOOLS IN THE U.S. CRIMINAL JUSTICE SYSTEM 10, https://www.partnershiponai.org/artificial-intelligence-research-and-ethics-community-calls-for-standards-in-criminal-justice-risk-assessment-tools/ [https://perma.cc/EGW4-DULF].

116. Christopher Rigano, *Using Aritifical Intelligence to Address Criminal Justice Needs*, NAT'L INST. J., Jan. 2019, at 1, 2.

117. *See, e.g.*, Vincent Menger, Marco Spruit, Roel van Est, Eline Nap, & Floor Scheepers, *Machine Learning Approach to Inpatient Violence Risk Assessment Using Routinely Collected Clinical Notes in Electronic Health Records*, 2 JAMA NETWORK OPEN 2, 4 (2019).

might help overcome human error, and it can process information with increased speed, quality, and accuracy.[118] Machine learning also can incorporate a vast amount of data into its calculations, far more data than any human could process, and use it to make decisions.[119] Violence risk assessments completed via machine learning may happen more quickly than manual risk assessment tools. This may be especially true if the program is set up to extract data automatically, as part of an agency's natural workflow. Consequently, individuals do not have to dedicate time to completing the assessment; instead, they can simply pull the results produced automatically via machine learning from the computer program. This may be one of the greatest potential benefits of machine learning, as finding the time to complete the violence risk assessment instruments is one of the most commonly cited barriers to using them.[120]

On the other hand, there are many concerns regarding the application of machine learning in the context of violence risk assessment. One important limitation of machine learning is that the accuracy of the model is contingent upon the accuracy of the data from which it learns.[121] If models are based on incomplete or biased data, then the predictions may be unfounded or biased. This is, of course, a primary concern in the context of violence risk assessment, given the systemic bias that is reflected in criminal justice records.[122] The perfect scenario would be to train a model using data that captured whether a person had *actually* committed a violent crime rather than relying on measures such as arrests or charges. Relying on such measures can contribute to bias because some crimes are underreported (e.g., rape, human trafficking),[123] and as mentioned earlier, people of color are more likely to be arrested for a given behavior than are their white counterparts.[124] Further, any given machine learning algorithm is limited to a particular context and for a particular task; applying an algorithm to a different context or task renders the results useless.[125] For example, if an algorithm is created using data from defendants in the

118. Rigano, *supra* note 116, at 2–3.

119. *Id.* at 5.

120. Menger, Spruit, van Est, Nap, & Scheepers, *supra* note 117, at 2; *see also* Sara K. Levin, Per Nilsen, Preben Bendtsen, & Per Bulow, *Structured Risk Assessment Instruments: A Systematic Review of Implementation Determinants*, 23 PSYCHIATRY, PSYCHOL. & L. 602 (2015).

121. PARTNERSHIP ON AI, *supra* note 115, at 3.

122. *See* Angwin, *supra* note 100 (commenting on the systemic bias reflected in criminal justice records).

123. *See id.* at 16 n.16; PARTNERSHIP ON AI, *supra* note 115, at 16–17.

124. *See supra* notes 100–04 and accompanying text.

125. PARTNERSHIP ON AI, *supra* note 115, at 14.

pretrial period to predict their likelihood of arrest for a new crime in that period, it *cannot* and *should not* be applied to determine risk of violent recidivism post-conviction.

Another potential concern with machine learning models is the fact that they can operate without human guidance or oversight.[126] This is especially true for models that do not make their contents and algorithms available to the public; doing so precludes scientific scrutiny via peer review and replication studies and consideration of the relevance of the content to a particular case. A related concern is that such black box models conceal not only what is, but also what is not being used in the model; there may be factors relevant to a specific case that are not being considered.[127] Finally, machine learning models are sometimes marketed as providing forecasts of whether or not an individual will engage in violent recidivism rather than producing risk scores that are associated with different likelihoods of future criminal activity.[128] This is another subtle but important distinction. The former suggests greater certainty about a given individual and their future behavior whereas the latter more appropriately reflects the uncertainty involved in violence risk assessment, even when the most sophisticated statistical methods are applied.

More complex statistical models are not necessarily more accurate. When working with structured data that has meaningful features, there is not typically a significant difference in the performance of simple classifiers (e.g., logistic regression) compared to complex classifiers (e.g., deep neural networks).[129] Therefore, complicated algorithms may not be necessary for accurate prediction of violent recidivism. One recent study compared violence predictions produced via machine learning to the violence risk estimates produced using the SAVRY.[130] Results showed that the slight increases in predictive accuracy conferred by machine learning were offset by problems of group fairness.[131] Across three metrics of fairness, SAVRY assessments outperformed machine learning.[132] Another study compared data mining, machine learning, and modern statistical models with classical statistical methods—namely, logistic

126. Michael E. Donohue, *A Replacement for Justitia's Scales?: Machine Learning's Role in Sentencing*, 32 HARV J. L. & TECH. 657, 659, 666 (2019).

127. *Id.* at 664–65, 671–72.

128. *Id.* at 661.

129. Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACHINE INTELLIGENCE 206, 207 (2019) [hereinafter *Stop Explaining Black Box Machine Learning Models for High Stakes*].

130. Tolan, Miron, Gómez, & Castillo, *supra* note 78, at 84.

131. *Id.*

132. *Id.*

regression and linear discriminant analysis (LDA).[133] Results showed only slight differences in the accuracy of predictions of general and violent recidivism between the three more complex techniques (i.e., data mining, machine learning, and modern statistical models) and logistic regression or LDA.[134]

Together, these findings suggest that there is no real advantage to using complex statistical models that are challenging for the layperson to understand. Instead, the extant research literature supports the continued use of violence risk assessment instruments that rely on structured professional judgment or simple algorithms that weight variables and include them in a logistic regression or LDA model.[135] An alternative strategy that still takes advantage of AI advances would be to create simple and interpretable machine learning models.[136] Such simple machine learning models afford easier detection of predictive bias or other measures of unfairness and greater transparency and interpretability. Such models also would be easier to monitor, debate, and adjust if problems are detected. We must balance a desire for accuracy and precision in forecasting future violence with the need to clearly understand the process by which these forecasts are made, so that they can be refuted, if necessary.[137]

V. CONCLUSION

Violence risk assessment instruments represent the current state-of-the-art approach to forecasting the likelihood of violent recidivism. Our review of the scientific evidence supports their continued use to inform criminal justice decision-making and failed to find substantial benefits associated with the application of new technologies, such as machine learning. Further, and in contrast with much of the current narrative surrounding risk assessment, we found relatively limited evidence of predictive bias and disparate impact, instead finding more evidence of predictive parity and, even, reductions in racial disparities in rates of restrictive placements. However, continued discussion and research is needed to clarify points of debate, including the definitions of fairness and proxies for race, and, ultimately, to establish whether

---

133. N. Tollenaar & P. G. M. van der Heijden, *Which Method Predicts Recidivism Best?: A Comparison of Statistical, Machine Learning and Data Mining Predictive Models*, 176 J. ROYAL STAT. SOC'Y 565, 582 (2013).

134. *Id.* at 574–75.

135. *Id.* at 582.

136. *Stop Explaining Black Box Machine Learning Models for High Stakes*, *supra* note 129, at 1, 15.

137. Richard Berk & Jordan Hyatt, *Machine Learning Forecasts of Risk to Inform Sentencing Decisions*, 27 FED. SENT'G REP. 222, 222 (2015).

the use of violence risk assessment instruments reduce or exacerbate racial disparities in the criminal justice system. In the end, the implementation of a violence risk assessment instrument will not improve criminal justice outcomes in and of itself. Their results must be used in meaningful ways to inform criminal justice practices.