EVIDENCE-BASED SENTENCING: RISKS AND OPPORTUNITIES

Hannes WESTERMANN[1]

ABSTRACT

When sentencing a defendant, an important consideration a judge has to take is the risk of recidivism of the defendant. Normally, this is evaluated in an unstructured, discretionary manner. Recently, new tools known as evidence-based sentencing (EBS) have appeared to make this assessment more empirical, by inferring risk based on statistics of previous defendants. These tools have been embraced as providing more accurate predictions, eliminating biases and harmonizing the decisions. However, there is a growing body of work that criticizes the methods for failing to live up to these promises and even for being an infringement on the fundamental rights of the defendant or increasing biases in society. This paper attempts to give an overview of the arguments in favor and against these new tools.

INTRODUCTION

After a defendant is proven to be guilty of having committed a crime, the judge has to decide which punishment is appropriate. In this decision, they have to incorporate many different factors, one of which is the risk of recidivism of an individual. Typically, risk is assessed in a clinical way, e.g. in a discretionary manner by the judge or other court officers. Recently, however, a new kind of statistical tool, known as evidence-based sentencing (EBS) or actuarial risk assessment has started to be used in many jurisdictions. These tools provide judges with a risk score for the defendant, based on empirical methods.

These methods face significant criticisms. One criticism is that the statistical reliance on averages based on factors such as age, income bracket and location go against the right of a fair trial, since an individual is judged based on factors related to his group belonging, rather than being judged as an individual. This might reinforce societal biases and discriminate against certain groups. Whether the algorithms work better

1 Cyberjustice Laboratory, Université de Montréal. hannes.westermann@umontreal.ca

than a judge assessing individuals has also been called into question. Another criticism is the opacity of the algorithms. These are often proprietary and protected as trade secrets, meaning that they are not disclosed to the defendant, making the risk scores impossible to defend against by challenging the underlying algorithms or data. This opacity further makes judges unable to properly assess the value of the risk scores, leading them to potentially play a too large role in the decision. Finally, the relevance of the question answered by the tools, namely risk for recidivism, for sentencing has also been called into question.

In this paper, I will first describe the sentencing decision of a judge, and how evidence-based sentencing is integrated into the process. Then, I will describe some of the potential advantages of the tools, and the criticisms they have faced. Finally, I will summarize my findings. While it is difficult to delve deeply into the many complex issues raised by EBS in this short space, I hope that the paper will provide the reader with an introduction to this fascinating and societally important debate.

72

1. EVIDENCE-BASED SENTENCING IN CRIMINAL JUSTICE

This section will briefly explain how sentencing is carried out traditionally, and how the new empirical tools are incorporated into the sentencing process. I will also explain the functioning of these tools and discuss a few high-profile examples.

1.1. SENTENCING

After an individual has been found guilty of a crime, the judge has to decide which punishment is appropriate for the person[2]. This is known as sentencing. Possible decisions include whether to incarcerate the defendant, and in this case for how long, or whether to use alternative punishments such as community service, as well as which rehabilitation measures should be implemented[3]. The judge has to weigh several different theories of criminal justice, such as retribution, rehabilitation, incapacitation, general deterrence and specific deterrence against each

2 see, for example, *Canadian Criminal Code, (R.S.C., 1985, c. C-46)* article 720.

3 Danielle Leah Kehl & Samuel Ari Kessler, "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing" (2017) Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, online: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041> at 13.

other to arrive at the decision[4]. In Canada, the fundamental purpose of sentencing is to "protect society" and to "contribute [...] to respect for the law and the maintenance of a just, peaceful and safe society"[5]. While the court is bound by several principles, it has a large degree of discretion in arriving at a decision[6].

To support them in this decision, judges are usually provided with a pre-sentence investigation report (PSI)[7]. This report is typically written by an officer at the court from a background of social work[8], and includes information such as the defendant's criminal record, other personal details and information from interviews with individuals who have been in contact with the defendant[9].

73

One variable that the judge is likely to consider in the sentencing of the individual is the risk of recidivism of the defendant[10]. If an individual is likely to commit another crime, the judge might decide to incapacitate them by ordering a prison sentence. This might not be necessary in cases where the risk for recidivism is low[11]. Which treatment an individual is given can also differ based on their risk for recidivism[12].

Typically, the estimation of the risk of an individual is taken in a discretionary way. An officer or the judge would use their personal experience and expertise to evaluate whether the defendant is likely to reoffend. This type of analysis is referred to as "unstructured" or "clinical assessment"[13].

4 *Ibid*; Kirk Heilbrun, "Risk Assessment in Evidence-Based Sentencing: Context and Promising Sues Symposium: Evidence-Based Sentencing: The New Frontier in Sentencing Policy and Practice: Risk and Needs Assessment" (2009) 1 Chap J Crim Just 127–142 at 128; *Canadian Criminal Code, (R.S.C., 1985, c. C-46)*, *supra* note 2 article 718.

5 *Canadian Criminal Code, (R.S.C., 1985, c. C-46), supra* note 2 article 718.

6 *Ibid* article 718.

7 Steven L Chanenson & Jordan Hyatt, "The use of risk assessment at sentencing: Implications for research and policy" (2016) Villanova Law/Public Policy Research Paper, online: <https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps> at 3; *Canadian Criminal Code, (R.S.C., 1985, c. C-46)*, *supra* note 2 article 721.

8 Chanenson & Hyatt, "The use of risk assessment at sentencing", *supra* note 7 at 4.

9 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 15.

10 *What We Know and Don't Know About Risk Assessment with Offenders of Indigenous Heritage*, by Leticia Gutierrez, L Maaike Helmus & R Karl Hanson, 2017-R009 (Public Safety Canada, 2017) at 3.

11 Chanenson & Hyatt, "The use of risk assessment at sentencing", *supra* note 7 at 3.

12 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 13; Heilbrun, "Risk Assessment in Evidence-Based Sentencing", *supra* note 4 at 128.

13 Shoba Sreenivasan et al, "Actuarial risk assessment models: A review of critical issues related to violence and sex-offender recidivism assessments" (2000) 28 Journal-American Academy of Psychiatry and the Law 438–448 at 2; JC Oleson et al, "Training to see risk: Measuring the accuracy of clinical and actuarial risk assessments among federal probation officers" (2011) 75 Federal Probation 52–56 at 52.

This type of assessment has the advantage that it allows all relevant factors to be taken into account. The decision is taken truly on a case-by-case basis. However, it also has disadvantages. The decision taken might be dependent on unconscious biases of the judge[14]. A study has shown that people with an African-American background on average receive a 20% longer prison sentence than white people in Maryland, for example[15]. Further, the decision might simply be inaccurate – humans are, after all, fallible[16]. This could mean that society is exposed to dangerous individuals that do not receive an appropriate prison sentence or treatment. Further, the liberty of individuals might be unduly affected if judges overestimate someone's risk to reoffend. Finally, the decisions might be inconsistent between different judges, causing a loss of trust for the legal system.

1.2. EVIDENCE-BASED SENTENCING

To combat these issues, there has been a strong movement towards evidence-based sentencing. In these systems, empirical and statistical methods are used to support judges in reaching their decisions regarding sentencing[17]. The systems work by empirically discovering factors related to whether a person will re-offend, and then building a model that weighs these factors to produce an estimation of risk for recidivism[18]. Typically, the systems incorporate a number of different factors to predict whether a person will re-offend, such as demographic and socioeconomic variables[19]. The systems often provide scores, showing the risk for a person re-offending, and might group defendants into low, medium and high-risk categories. This score is presented in the pre-sentence investigation report. The judge can then determine an appropriate punishment and treatment, taking into account the risk for recidivism as indicated by the score[20].

14 Oleson et al, "Training to see risk", *supra* note 13 at 52.

15 Shawn D Bushway & Anne Morrison Piehl, "Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing" (2001) 35:4 Law & Society Review 733–764.

16 Sreenivasan et al, "Actuarial risk assessment models", *supra* note 13 at 1.

17 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 7; Sonja B Starr, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination" (2014) 66 Stanford Law Review 803–872 at 815.

18 Christopher Slobogin, "Risk Assessment" in The Oxford Handbook of Sentencing and Corrections (Oxford University Press, 2012) 196 at 198.

19 Starr, *supra* note 17 at 811.

20 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 8.

The factors used in the prediction vary between different such systems. Typically, they are split into risk factors and protective factors. Risk factors are circumstances that raise the risk for recidivism, while protective factors lower the risk for recidivism[21]. Further, the factors can be split into static and dynamic factors. Static factors, such as age and prior criminal history, cannot be altered by human intervention[22]. Dynamic factors, such as substance use and family support, on the other hand, can be affected by interventions or the defendant themselves[23].

The algorithms used to arrive at a risk score are often kept secret. This is one of the main points of criticism, as discussed below. Starr claims that the systems are generally simple regression models, that develop a points system based on provided variables. For example, the effect of an offender's gender on risk for recidivism in historical cases might be statistically evaluated. If this analysis shows that men typically have a larger risk for recidivism, they will be assigned a higher risk score than women who otherwise have identical profiles[24].

Many of the tools are adapted from a pre-trial risk assessment context, where the purpose is to predict whether an individual will commit a crime before the trial begins, or skip the trial, which warrants detention of the individual[25]. Tools for sentencing are used in at least 20 States in the United States[26], and is also spreading to jurisdictions such as Canada and the UK[27].

1.3. EXAMPLES OF EBS TOOLS

One of the most notorious risk prediction tools is Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). It contains risk models for General Recidivism Risk and Violent Recidivism risk,[28] and is

21 Heilbrun, "Risk Assessment in Evidence-Based Sentencing", *supra* note 4 at 127.

22 Slobogin, *supra* note 18 at 196.

23 *Ibid*.

24 Starr, *supra* note 17 at 813.

25 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 10; Chanenson & Hyatt, "The use of risk assessment at sentencing", *supra* note 7 at 4.

26 Starr, *supra* note 17 at 809.

27 *Ibid* at 805.

28 *Practitioner's Guide to COMPAS Core* (equivant, 2019) at 1.

based on a questionnaire of 137 questions,[29] including questions about the current charges, criminal history, family criminality, peers, leisure activities and personality of the defendant.[30]

Level of Services Inventory-Revised (LSI-R) is another very popular risk prediction system.[31] It includes objective variables, but also subjective evaluations on performance at work, level of crime in their neighborhood and social involvement with companions.[32] A survey containing 54 items, including financial means, education, employment and personality traits, has to be filled out for the calculation of LSI-R.[33]

Public Safety Assessment (PSA) is a risk prediction tool for use in pre-trial bail decisions. Unlike the previous two, the PSA is public – anyone can see the way the algorithm is applied. It is based on 9 variables.[34]

76

2. ARGUMENTS IN FAVOR OF EVIDENCE-BASED SENTENCING

While many researchers are critical of evidence-based sentencing, there is also a large group of academics and institutions in favor of their use, arguing that the tools can help judges render more accurate, fair and consistent decisions. This section elaborates on these arguments in favor of evidence-based sentencing.

2.1. MORE ACCURATE DECISIONS

As mentioned above, accurate predictions for recidivism risk can be important in achieving the goals set out by the criminal justice system. Many authors believe that actuarial risk assessment is a necessary step towards improving the accuracy of these decisions.[35] Some meta-studies seem to indicate that statistical predictions are more accurate than clinical judgments in predicting risk for recidivism.[36]

---

29 Cynthia Rudin, Caroline Wang & Beau Coker, "The age of secrecy and unfairness in recidivism prediction" (2018) arXiv:181100731 [cs, stat], online: <http://arxiv.org/abs/1811.00731> at 25, arXiv: 1811.00731.

30 Julia Angwin, *COMPAS Risk Assessment Sample* (ProPublica).

31 Kelly Hannah-Moffat, "Actuarial Sentencing: An 'Unsettled' Proposition" (2012) 30 Justice Quarterly 1–27 at 4.

32 Starr, *supra* note 17 at 812.

33 Slobogin, *supra* note 18 at 199.

34 "Public Safety Assessment (PSA)", online: <https://psapretrial.org/>.

35 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 11–12.

36 Oleson et al, "Training to see risk", supra note 13 at 52; Roberta Cooper Ramo et al, Model Penal Code: *Sentencing - Proposed Final Draft* (The American Law Institute, 2017) at 67.

The accurate labeling of true negatives, e.g. people that will not recidivate, has the important function of enabling judges to choose alternative punishments, such as fines or house arrest, for individuals unlikely to reoffend. This is much less resource-intensive and can reduce the prison population.[37] Reducing incarceration rates in the United States is seen as one of the main motivations of introducing evidence-based sentencing practices.[38] Further, resources do not have to be spent rehabilitating individuals that are unlikely to recidivate.

The same goes for true positives, or individuals that will reoffend. In these cases, if the prediction is accurate, the courts are able to incapacitate the individuals, thereby protecting society. Further, proper treatments can be deployed to rehabilitate the individual and make them able to reenter society after their punishment.[39]

77

2.2. LESS BIAS IN DECISIONS

Another advantage is the potential decrease of bias in sentencing decisions.[40] As already mentioned, biases have been shown to play an important role in sentencing decisions. By introducing standardized metrics, the argument goes, the discretion of the judge can be somewhat limited, thereby eliminating some of this bias.[41] Of course, this argument depends on two assumptions – namely that the actuarial risk assessment tools themselves do not contain biases and that the judges end up following the decisions. These issues are addressed below.

2.3. HARMONIZATION OF SENTENCING DECISIONS

Judges have historically had a large amount of discretion in sentencing decisions. The introduction of risk assessment into the process can somewhat curtail this discretion.[42] The standardized nature of the tools

37 Ramo et al, *supra* note 36 at 176.

38 Starr, *supra* note 17 at 815–816.

39 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." (2018) 42:3 Law and Human Behavior 181–214.

40 Han-Wei Liu, Ching-Fu Lin & Yu-Jie Chen, "Beyond State v Loomis: artificial intelligence, government algorithmization and accountability" (2019) 27:2 Int J Law Info Tech 122–141 at 124.

41 compare Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 6.

42 Hannah-Moffat, "Actuarial Sentencing", *supra* note 31 at 3.

could further reduce discrepancies between different judges. Two individuals with the same background, risk profile and crime committed would receive similar punishments which could potentially increase trust for the legal system.

3. ARGUMENTS AGAINST EVIDENCE-BASED SENTENCING

The previous section details some of the arguments proponents of evidence-based sentencing systems use to argue for the implementation of such systems. However, the systems have also faced significant criticisms from different sources. This section gives an overview over some of these criticisms.

78

3.1. FOCUS ON GROUPS RATHER THAN INDIVIDUALS

One of the cornerstones of the criminal justice system is that everyone is considered as an individual, regardless of gender, ethnicity, socioeconomic background and other such factors.[43] Everyone has a right to an individualized sentence, where their defendant is considered as an individual, and not as part of a group. Some argue that the use of actuarial risk assessment tools goes against this right.

Instead of looking at the defendant as an individual, the tools consider individuals as being made up of different factors, which often include gender, education level or neighborhood.[44] Sonia Starr argues that, while the actuarial risk assessment tools might be accurate at determining a group average risk of recidivism, they are poor at determining the *individual* risk of recidivism. This is because there is a lot of variance within the group, much of which is unfeasible to predict using the variables used in the tools. The risk score presented by the tools is an average of the group the defendant is a member of but might not say much about the individual.[45]

43 Starr, *supra* note 17 at 804.

44 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 23.

45 Starr, *supra* note 17 at 842–843; compare Slobogin, *supra* note 18 at 208.

This distinction is, according to Starr, often lost on judges. A high risk-score means that an individual shares characteristic with people who reoffended – not necessarily that they themselves are likely to reoffend.[46]

This argument was tried in State v Loomis, one of the most significant decisions in the area of Evidence-Based Sentencing. In the case, Eric Loomis was accused of being involved in a drive-by shooting. He was sentenced to 6 years in prison without probation, supported by a risk analysis of the COMPAS tool, where he was scored as having a high risk of reoffending.[47] Loomis appealed the decision on multiple grounds and reached the Wisconsin Supreme Court. One of arguments was that the use of COMPAS deprived Loomis of the right to an individualized sentence.[48] The court conceded that the COMPAS results were used to detect high risk groups rather than individuals. If COMPAS was the determining factor, this would therefore infringe on the right to an individualized sentence. However, the courts using COMPAS were encouraged to disagree with the risk score and exercise their discretion in coming to a decision. The score was supposed to be used to provide the sentencing courts with more complete information, *not* as the sole determinant of the sentence. Therefore, the court ruled, Loomis had received an individualized sentence, although it did caution the courts in the use of COMPAS.[49]

3.2. DISCRIMINATION AND BIAS

The fact of considering individuals as members of a group rather than individuals can in practice have the consequence of discriminating against individuals based on their ethnicities, gender or socioeconomic background, which might infringe on human rights against discrimination. Racial disparities exist at many levels in society, and a large fear with the use of risk assessment tools is that these may reinforce or increase these.[50] Bias might be contained in the datasets used to create the algorithms at many levels. For example, while the algorithms aim to predict recidivism, they are created using data of re-*arrests*. Certain

46 Starr, *supra* note 17 at 848–850; see also Hannah-Moffat, "Actuarial Sentencing", *supra* note 31 at 12.

47 *State v Loomis 881 NW2d 749 (Wis 2016) 754 (US)* p. 8-22; Liu, Lin & Chen, "Beyond State v Loomis", *supra* note 40 at 126–127.

48 note 47 paragraph 34.

49 *Ibid* paragraph 67-74.

50 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 29.

crimes, such as possession of drugs, might be disproportionately pursued by the police in poor and minority neighborhoods.[51] The data might thus indicate that people living in these neighborhoods have a higher risk of committing these crimes, even though the real difference is a higher police presence in these areas. This could lead to unjust and discriminatory predictions.

Sonia Starr argues that the use of risk assessment scores based upon demographics and socioeconomic status is an embrace of overt discrimination.[52] Nobody would generally agree with defendants from, for example, a poorer background being punished more harshly than other defendants. Yet this is exactly what the use of the actuarial risk assessment tools enables, hidden behind a technocratic framing.[53] She believes the use of tools including demographic and socioeconomic variables to be unconstitutional in the United States.[54]

80

Of course, in none of the systems, ethnicity is explicitly used as a variable for the prediction of recidivism. However, this does not mean that it does not enter the final analysis. There might be variables that are captured by the questionnaires that are linked to ethnicity. For example, asking for the level of education might disparately impact people with certain ethnicities. Further, questions such as how often a person has been stopped by the police might also be linked to ethnicity, as certain groups are much more likely to be approached by the police.[55] This kind of discrimination might be illegal even if statistically motivated.[56]

In 2016, an article by ProPublica attacked the COMPAS system (here used for pre-trial decisions), claiming that it is biased against African Americans. In their analysis, ProPublica obtained data on risk scores, and showed that black defendants that did not reoffend were twice as likely to be classified as high risk compared to white defendants that did not

51 *Ibid* at 30.

52 Starr, *supra* note 17 at 806.

53 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 23–24.

54 Starr, *supra* note 17 at 823–836; compare Slobogin, *supra* note 18 at 204.

55 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 24.

56 *Ibid*.

reoffend.[57] The company itself and some other academics disagreed with the analysis.[58]

The disagreement seems to come down to different definitions of fairness between ProPublica and equivant, the company behind COMPAS.[59] This is a problem in general when analyzing evidence based sentencing tools. Most people agree that risk assessment algorithms should be fair, but few agree on what exactly fairness means in this context. One could argue that the more accurate an algorithm is, the fairer it is. However, this might imply that certain groups of people are treated differently than others. In order to remove this disparate treatment, the accuracy of the algorithm in predicting recidivism necessarily has to be decreased, which makes the algorithms more unfair from another perspective.[60] Further, since the data used to build the algorithm might have issues, even a seemingly accurate algorithm might in practice be unfair. Since there is not uniform definition of fairness, and creating such as definition might be difficult and differ between different contexts, objectively arguing about fairness in algorithms is very difficult.[61]

Beyond ethnicity, factors such as gender and socioeconomic status are routinely included in the risk analysis tools.[62] In the Loomis case, the defendant claimed that the use of gender as a variable was a violation of his right to due process. The court found that the inclusion of gender is allowed if it promotes accuracy of the algorithm and was therefore permitted. Further, it did not consider that Loomis had discharged his burden of proof in showing that the judge had actually relied on gender in his decision, since the judge did not mention gender in his rationale for the decision.[63] It can, of course, be questioned whether the factors a

57 Julia Angwin et al, "Machine Bias" *ProPublica* (23 May 2016), online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; Jeff Larson et al, "How We Analyzed the COMPAS Recidivism Algorithm", (23 May 2016), online: *ProPublica* <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

58 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" (2016) 80:2 Fed Probation 38–46; *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, by William Dieterich, Christina Mendoza & Tim Brennan (Nortpointe Inc, 2018).

59 Sam Corbett-Davies et al, "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.", *Washington Post* (17 October 2016), online: <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/>; Rudin, Wang & Coker, *supra* note 29 at 2.

60 Karen Hao & Jonathan Stray, "Can you make AI fairer than a judge? Play our courtroom algorithm game" *MIT Technology Review* (17 October 2019), online: <https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>.

61 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 30.

62 *Ibid* at 25–26.

63 *Ibid* at 19; note 47 paragraph 86.

judge considers are limited limited to what they mention in their rationale. The influence of risk scores on judges is discussed below.

The incorporation of certain static factors might also have a normative effect of increasing discrimination in society. By stigmatizing certain groups in EBS systems, they might be tainted in the eyes of the justice system or the public. This could lead to an increase of crimes committed by individuals in these groups.[64]

3.3. INACCURACY OF THE ALGORITHMS

Many of the proponents of EBS argue that it should be used due to being more accurate than clinical assessments. Several researchers have disputed that this is the case. Inaccuracies in the decisions of the algorithms can cause damage to society and unduly restrict the freedom of individuals.

82

Evaluating actuarial risk assessment tools is not trivial. One popular metric is the "area under the curve" (AUC). Here, each person who does reoffend is compared with a randomly selected person who does not. Ideally, in these cases, the person who does reoffend would have a higher risk score than the one who does not. If this is always the case, the AUC score is 1. Randomly assigning risk scores gives an AUC score of 0.5. Most actuarial risk assessment tools seem to score between 0.65 and 0.75.[65]

Dressel and Farid asked groups of randomly selected people on the internet to rate the risk of recidivism rate of individuals, based on a short description containing 7 variables, and compared this to the COMPAS score of the same defendants. Surprisingly, they found that the AUC score for the experiment participants was the same as for the COMPAS tool (which uses 137 variables).[66] They believe these results to cast "significant doubt on the entire effort of algorithmic recidivism prediction".[67]

Starr also questions the research aiming to show that EBS tools are more accurate than the traditional, clinical approach. While some studies do show an advantage for EBS tools, the performance seem to vary wildly

64 Slobogin, *supra* note 18 at 206.

65 Starr, *supra* note 17 at 846; Slobogin, *supra* note 18 at 201.

66 Julia Dressel & Hany Farid, "The accuracy, fairness, and limits of predicting recidivism" (2018) 4:1 Science Advances eaao5580 at 2.

67 *Ibid* at 1.

between the tools and what they aim to predict, as well as the group of individuals it is used for.[68] EBS tools are routinely used in other contexts than where they were developed, perhaps without testing the accuracy in the new context.[69] This application of tools developed in one context in another context was criticized in a decision by the Canadian Supreme Court. The court criticized the Correctional Services of Canada for using an EBS tool in relation to Indigenous offenders, without confirming its validity in this context.[70]

3.4. OPACITY OF THE METHODS USED TO ASSESS RISK

Another common point of criticism against EBS tools is their lack of transparency. There is a large difference in the level of transparency between different evidence-based sentencing tools. Some tools, such as the Public Safety Assessment, are developed in the open and available to the public for examination. Others, like the COMPAS tool, are proprietary. Information about how these algorithms work is sparse, as the companies developing them might claim trade secret protection on the algorithms.[71] In these cases, the defendant together with an officer of the court fills out a questionnaire, which is processed by the company, which then returns a calculated risk score. How the answers by the defendant are treated, and which weight they are assigned, is not clear to the defendant, the judge or the public.

Transparency of an algorithm is not binary, and it is not always clear which level of transparency would be "sufficient" in a particular case. One could consider the answers in the questionnaire together with the risk score of the algorithm as a form of transparency. However, this does not explain how the answers were processed, and why. A more in-depth explanation might therefore require disclosure of which weight the different answers are assigned to inside the algorithm and how exactly used to determine the risk score. This is, of course, commercially sensitive, as competitors would be able to reproduce the systems by knowing these weights. Even disclosure of the weights does not contain

68 Starr, *supra* note 17 at 850–855.

69 Slobogin, *supra* note 18 at 208.

70 *Ewert v Canada, 2018 SCC 30*, [2018] 2 SCR 165 (Supreme Court Judgments, 2018).

71 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 28.

an explanation of *why* the variables are taken into account in a certain way. For this, more in-depth explanations of the underlying statistics have to be provided. This opacity carries a number of problems for the different stakeholders. I will go through these one by one below.

3.4.1. THE DEFENDANT

The lack of explanation of the calculated score can be seen as an infringement on the right of due process of the defendant. In the United States, the Supreme court has recognized that the information provided in the pre-sentencing report has to satisfy the requirement of due process, and that the defendant must thus be provided with the opportunity to respond to the information the judge uses in arriving at a sentencing decision.[72]

84

Whether the opacity of the algorithm violates the requirement of due process was analyzed in the Loomis case. Loomis argued that the opacity of the COMPAS algorithm prevented him from challenging the scientific validity of the risk assessment, since he was not able to see how the variables were weighted.[73] The court disagreed with this argument. Since the risk scores were provided to Loomis, he had the opportunity to refute them, according to the court.[74] Further, since the COMPAS score was calculated based on answers to questions that the defendant provided, as well as public registers, he had the opportunity to correct the information.[75] Further, the court ruled, several studies have shown that COMPAS is reliable in many cases.[76]

This part of the Loomis decision has been criticized. The defendant is, of course, able to control and challenge the answer they give to the correctional officer. However, this does not enable them to challenge the risk score, since the function that is used to arrive at the risk score from these answer remains hidden.[77] The researchers therefore believe the algorithms used should be disclosed to the defendant or an expert

72 *Ibid* at 22.

73 note 47 paragraph 46, 52.

74 *Ibid* paragraph 53.

75 *Ibid* paragraph 55.

76 *Ibid* paragraph 57-65.

77 Liu, Lin & Chen, "Beyond State v Loomis", *supra* note 40 at 133.

committee.[78] They also discuss the issue of the technical black box, where in some instances the system remains inscrutable even for the creators of the algorithms.[79]

Defendants generally have the right to question witnesses on the other side and confront evidence against them. This is impossible if a proprietary algorithm is used. Some experts have argued that the use of algorithmic risk assessment is similar to an anonymous, un-questionable expert witness.[80] This inability of challenging data used by a judge in determining a sentence might be seen as a significant impediment of the right to due process.

3.4.2. THE PUBLIC

Beyond the defendant, the public has an interest in having transparent algorithms in sentencing decisions. Currently, researchers often have to guess the functioning of EBS tools due to the unavailability of the algorithms and underlying data. The ProPublica researchers, for example, did not have access to the COMPAS algorithm but had to rely on examples of the risk scores obtained via public record requests.[81] Only if algorithms are transparent can they be verified for biases and disparate impacts on minorities,[82] and the policy implications and effects they might have. Outsourcing an important part of the sentencing process to proprietary and secret algorithms causes democratic issues, as sentencing policy can potentially be changed without anyone being able to affect or even realize it. Rudin et al. argue that transparency is, in many ways, a precursor to fairness – only if algorithms are transparent can they be properly assessed and debated in the public.[83] They also argue for the replacement of complicated proprietary models, like COMPAS, for interpretable, transparent models, that often show similar performances. [84]The public development of algorithms could also give the public the

78 *Ibid* at 135.

79 *Ibid* at 136.

80 Leah Wisser, "Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments In Sentencing" 56 American Criminal Law Review 1811–1832 at 1824.

81 Larson et al, *supra* note 57.

82 Wisser, *supra* note 80 at 1813.

83 Rudin, Wang & Coker, *supra* note 29 at 4.

84 *Ibid* at 25.

possibility of adapting them to specific context and improving them over time.[85]

3.4.3. THE JUDGE

For the judge, it is similarly crucial to have an explanation and understanding of how the algorithms work. Only in these cases can the judge properly assess the value of the algorithm and determine the influence it should have on the sentencing decision. This question is dealt with in the next section.

3.5. THE JUDGE'S DIFFICULTY OF DETERMINING THE VALUE OF THE RISK-SCORE

86

One common defense of evidence based sentencing tools is that they are only providing information to the judge, which the judges are free to incorporate into their decision or ignore.[86] However, this is only possible if judges are able to properly determine the value of a risk score. This might not only be challenging due to the opacity of the algorithms, as discussed above, but also due to the potential difficulty for legally trained judges to assess the statistical validity of actuarial risk instruments.[87]

Starr does not believe the scores to be mere information. Unless the judges completely ignore the risk scores, the score will affect the sentence for some individuals.[88] Another defense is that judges already assess risk of recidivism in sentencing. What is the risk of replacing this implicit assessment with statistical, potentially more accurate tools? Starr does not buy this argument either. Risk prediction is hard, perhaps causing judges to use it carefully in sentencing decisions. However, presenting a calculated risk score might cause judges to assign a greater weight to risk of recidivism in their decisions.[89] Starr conducted an experiment with law students, which seems to suggest that the presentation of a risk score potentially does increase the weight of risk prediction in a sentencing decision.[90]

---

85 Wisser, *supra* note 80 at 1830.

86 Starr, *supra* note 17 at 862.

87 Hannah-Moffat, "Actuarial Sentencing", *supra* note 31 at 25; "State v. Loomis" (2017) 130 Harvard Law Review 1530–1537 at 1535.

88 Starr, *supra* note 17 at 863.

89 *Ibid* at 865–867.

90 *Ibid* at 867–870.

In Loomis, the court introduced a disclaimer outlining the shortcomings of systems such as COMPAS, to be provided together with the risk scores. This includes information such as the fact that COMPAS is based on group averages and has not been validated on local samples, and should therefore be treated with care.[91] Researchers have questioned whether this is enough. Even with the disclaimer, there is a significant possibility for an anchoring effect where judges are psychologically bound to the computer-generated number.[92] Further, society has no way of verifying to which extent the risk scores contributed to a certain decision.[93]

3.6. ARE WE ASKING THE RIGHT QUESTION?

87

Finally, concerns have been raised about the question the actuarial risk assessment tools are trying to answer. The systems used generally aim to predict the risk of recidivism of individuals. In pre-trial decisions, there is a direct link between this risk and an action a judge should take – if a person is likely to reoffend before trial, they should remain in jail.

This link is far weaker in sentencing decisions.[94] It is unclear how a judge should react to an individual having a high risk-score in deciding on a sentence. As mentioned before, the criminal justice system has several goals that have to be balanced by judges in determining a sentence. These include retribution, rehabilitation, deterrence and incapacitation. Some argue that retribution is the only valid justification for punishment, and that it is not legitimate to punish individuals based on future crimes they have not yet committed.[95] This conflict is not new, and discussing the merits of the different purposes and ways of sentencing are beyond the scope of this paper. However, the introduction of risk scores might alter the *balance* between the different purposes.[96]

The risk score is only relevant for some of the purposes of the criminal justice system. Both rehabilitation and incapacitation depend on future crimes committed by an individual and can thus be seen to have a clear link to risk of recidivism, while retribution and general deterrence might

91 note 47 paragraph 100.

92 Liu, Lin & Chen, "Beyond State v Loomis", *supra* note 40 at 130–131; note 87 at 1536.

93 Wisser, *supra* note 80 at 1816.

94 Slobogin, *supra* note 18 at 197–198; Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 13.

95 Starr, *supra* note 17 at 818.

96 *Ibid*; *ibid* at 865; Slobogin, *supra* note 18 at 205; Hannah-Moffat, "Actuarial Sentencing", *supra* note 31 at 26.

have weaker or no links at all to risk scores. The proliferation of risk scores is likely to increase the importance of risk management of defendants, strengthening some purposes of the criminal justice system at the expense of others.[97]

Starr believes that the Actuarial Risk Assessment tools are answering a question that is fundamentally wrong. Instead of answering whether an individual will recidivate or not, she believes the tools should focus on the question of how a certain punishment or treatment *affects* this risk. This information would help judges decide between the costs of a certain punishment to society and the individual versus the reduction in future crimes caused by the punishment.[98] While incarceration obviously reduces the possibility of individuals to commit crimes while they are in prison, it might increase the risk of reoffending after release. Further, individuals might be affected by an increased prison terms in different ways.[99] This is not considered by current risk assessment tools.

88

4. SUMMARY

In this paper, I have attempted to provide an overview over some of the arguments in favor of and against evidence based sentencing tools.

The tools generally show promise in producing more accurate decisions, that are less biased and more consistent between judges. However, the issues that have been raised are grave and make it unclear whether the use of risk scores is a good idea in practice.

First, the tools treat everyone as part of a group, rather than an individual. One of the pillars of the criminal justice system is that everyone has the right to be judged as an individual. By using EBS tools, judges can be said to assess thousands of previous defendants that are in some way similar to the current defendant, rather than the individual defendant. People are suddenly judged by statistics of their group – basing decisions to restrict individual liberty on this seems unfair.

Second, this focus on groups has the potential to increase discrimination and bias in society. Individuals can, hidden behind the veil of statistics, be

97 Kehl & Kessler, "Algorithms in the Criminal Justice System", *supra* note 3 at 27; Starr, *supra* note 17 at 818.

98 Starr, *supra* note 17 at 855–856.

99 *Ibid* at 857–858.

punished more harshly if they are poor or uneducated. The potential for reinforcement and increase of racial discrimination seems enormous.

Third, even the argument that risk predictions are more accurate than clinical assessments might not be fully justified. There does not seem to be a consensus of which method works better, and randomly selected internet participants are apparently equally accurate in predicting recidivism as some of the tools. This does not inspire confidence in the accuracy of EBS tools.

Fourth, the fact that the EBS tools are proprietary and their algorithms are secret makes it difficult to evaluate the previously discussed issues for the public. For the individual defendant, it makes it impossible to respond to the risk scores, since there is no way to criticize or even understand how they are created. This could be seen as a significant imposition on the defendants right of due process.

Fifth, this lack of transparency might lead judges to have difficulty assessing the value of the risk score. They might not understand the statistics the scores are based on, believing them to be more accurate than they are. The psychological anchoring effect might make diverging from the score arduous, meaning a large part of the sentencing decision might in practice be taken by potentially flawed algorithms.

Finally, the question answered by the tools – whether an individual will recidivate – can be criticized. It might not have direct relevance to sentencing, depending on the purpose of sentencing in a court's jurisdiction. The inclusion of a risk score might alter the balance between different purposes of the criminal justice system, causing more emphasis on the risk management aspect of the courts. Whether we agree with this direction or not, this seems like a policy decision that should not be determined by the adoption of certain software tools.

One thing is for certain. No matter what we think of the adoption of evidence based sentencing tools, they are coming. Courts all over the world are implementing actuarial methods into their sentencing process. It is now up to us, as citizens, to demand transparency into the inner workings of the tools and the process of their creation, or even to push for not using them in the sentencing process. Whatever our reaction might be, it is important to realize that the tools are probably no simple

addition of information to the criminal justice process, but are likely to have practical, policy and constitutional implications. Understanding these issues is the first step in being able to decide how to handle them. I hope that this paper can serve as a step for the reader to obtain an understanding of the issues and stakes involved.

90

Bibliography

Canadian Criminal Code, (R.S.C., 1985, c. C-46).

Angwin, Julia et al. "Machine Bias" *ProPublica* (23 May 2016), online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Bushway, Shawn D & Anne Morrison Piehl. "Judging Judicial Discretion: Legal Factors and Racial Discrimination in Sentencing" (2001) 35:4 Law & Society Review 733–764.

Chanenson, Steven L & Jordan Hyatt. "The use of risk assessment at sentencing: Implications for research and policy" (2016) Villanova Law/Public Policy Research Paper, online: <https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=1201&context=wps>.

Corbett-Davies, Sam et al. "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.", *Washington Post* (17 October 2016), online: <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/>.

Dressel, Julia & Hany Farid. "The accuracy, fairness, and limits of predicting recidivism" (2018) 4:1 Science Advances eaao5580.

Flores, Anthony W, 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" (2016) 80:2 Fed Probation 38–46.

Hannah-Moffat, Kelly. "Actuarial Sentencing: An 'Unsettled' Proposition" (2012) 30 Justice Quarterly 1–27.

Hao, Karen & Jonathan Stray. "Can you make AI fairer than a judge? Play our courtroom algorithm game" *MIT Technology Review* (17 October 2019), online: <https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>.

Heilbrun, Kirk. "Risk Assessment in Evidence-Based Sentencing: Context and Promising Sues Symposium: Evidence-Based Sentencing: The New Frontier in Sentencing Policy and Practice: Risk and Needs Assessment" (2009) 1 Chap J Crim Just 127–142.

Kehl, Danielle Leah & Samuel Ari Kessler. "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing" (2017) Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, online: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>.

Liu, Han-Wei, Ching-Fu Lin & Yu-Jie Chen. "Beyond State v Loomis: artificial intelligence, government algorithmization and accountability" (2019) 27:2 Int J Law Info Tech 122–141.

Oleson, JC et al. "Training to see risk: Measuring the accuracy of clinical and actuarial risk assessments among federal probation officers" (2011) 75 Federal Probation 52–56.

Rudin, Cynthia, Caroline Wang & Beau Coker. "The age of secrecy and unfairness in recidivism prediction" (2018) arXiv:181100731 [cs, stat], online: <http://arxiv.org/abs/1811.00731>.

Slobogin, Christopher. "Risk Assessment" *in The Oxford Handbook of Sentencing and Corrections* (Oxford University Press, 2012) 196.

Sreenivasan, Shoba et al. "Actuarial risk assessment models: A review of critical issues related to violence and sex-offender recidivism assessments" (2000) 28 Journal-American Academy of Psychiatry and the Law 438–448.

Starr, Sonja B. "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination" (2014) 66 Stanford Law Review 803–872.

Viljoen, Jodi L, Dana M Cochrane & Melissa R Jonnson. "Do risk assessment tools help manage and reduce risk of violence and reoffending? A systematic review." (2018) 42:3 Law and Human Behavior 181–214.

Wisser, Leah. “Pandora’s Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments In Sentencing” 56 American Criminal Law Review 1811–1832.

“State v. Loomis” (2017) 130 Harvard Law Review 1530–1537.

Angwin, Julia. COMPAS *Risk Assessment Sample* (ProPublica).

Dieterich, William, Christina Mendoza & Tim Brennan. *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, by William Dieterich, Christina Mendoza & Tim Brennan (Nortpointe Inc, 2018).

Gutierrez, Leticia, L Maaike Helmus & R Karl Hanson. *What We Know and Don’t Know About Risk Assessment with Offenders of Indigenous Heritage*, by Leticia Gutierrez, L Maaike Helmus & R Karl Hanson, 2017-R009 (Public Safety Canada, 2017).

Larson, Jeff et al. “How We Analyzed the COMPAS Recidivism Algorithm”, (23 May 2016), online: *ProPublica* <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

Ramo, Roberta Cooper et al. Model Penal Code: *Sentencing - Proposed Final Draft* (The American Law Institute, 2017).

Ewert v Canada, 2018 SCC 30, [2018] 2 SCR 165 (Supreme Court Judgments, 2018).

Practitioner’s Guide to COMPAS Core (equivant, 2019).

“Public Safety Assessment (PSA)”, online: <https://psapretrial.org/>.

State v Loomis 881 NW2d 749 (Wis 2016) 754 (US).