Artificial Intelligence and Recidivism:

The Use of Risk Assessment as Prediction Instruments in Criminal Sentencings

By Paloma Loudermilk Bhatia
palomalb@uchicago.edu

A thesis submitted for partial fulfillment of the requirements for the degree of Bachelor of Arts in Public Policy Studies

Preceptor: Rubina Hundal
Second Reader: Jens Ludwig

Abstract
The use of recidivism risk assessment instruments throughout different stages of the criminal justice process continues to grow. Much of the existing literature fails to focus on the algorithm's consistency in decision-making and implementation. This thesis analyzes the effectiveness and uniformity of COMPAS recidivism risk assessment instruments in the sentencing process. I use quantitative data from the Wisconsin Department of Corrections and qualitative interviews with experts to better analyze the efficacy and consistency of the algorithms from a practical standpoint. I find that, while the COMPAS risk assessment instrument validly determines individuals' risk of recidivism on a general level, COMPAS risk level may not correspond to one's sentence length. This finding indicates that judges deviate from COMPAS risk levels during the sentencing process. This discrepancy points to a need for standardized training for state-operated risk assessment instruments, rigorous external validation studies, and protocols encouraging judges to articulate their decision-making processes. These measures aim to enhance the instrument's practical efficacy and fairness in sentencing decisions.

Acknowledgements
I would like to thank the countless individuals who supported me and contributed to this thesis. I begin by thanking those at the Wisconsin Department of Corrections who took time out of their busy schedules to help me. I also thank the interview participants for their honestly and vulnerability. They so openly shared their thoughts and proved to be an integral part of my research. I would like to thank my preceptor, Rubina Hundal, for her continuous guidance and support through every step of the thesis process. Her constant encouragement was integral and motivational. I would also like to thank Jens Ludwig for being my second reader and providing advice and inspiration. Thank you to Anthony Fowler for his input and assistance at the BA Thesis Symposium and throughout the writing process. Finally, I would like to thank my parents for inspiring my curiosity.

Table of Contents

Introduction

In 2013, Eric Loomis was charged with five criminal counts related to a drive by shooting (State v. Loomis, 2016). The court ordered a pre-sentence investigation that included a COMPAS risk assessment. COMPAS, or Correctional Offender Management Profiling for Alternative Sanctions, uses criminal record information and responses to a questionnaire to measure recidivism at different stages of the legal process and inform judicial decisions (Reiling, 2021, Gacutan & Selvadurai, 2020, Chohlas-Wood, 2020). These assessments have increasingly been used by criminal justice agencies on local, state, and federal levels (Bureau of Justice Assistance). Through the quantification of an individual's risk of reoffending, recidivism risk assessment instruments produce scores that are used in the "classification, management, and treatment of justice-involved populations" (Bureau of Justice Assistance). This paper employs the use of quantitative data analysis and qualitative interpretations of interviews to analyze the efficacy of risk assessment instruments, specifically COMPAS, in accurately predicting an individuals' risk of reincarceration in relation to sentencing time. I explore whether the use of risk assessment instruments as predictive tools in sentencing effectively predicts an individual's recidivism risk uniformly across different groups and is accurately reflected in one's sentence length.

Loomis's COMPAS risk scores indicated a high risk of reoffending, resulting in Loomis becoming ineligible for probation (State v. Loomis, 2016). Moreover, his risk scores were used as a factor in his sentencing process. In early April of 2016, Eric Loomis filed a case against the State of Wisconsin with the Wisconsin Supreme Court (State v. Loomis, 2016). Loomis argued that the circuit court's use of the COMPAS risk assessment instrument in sentencing violated his right to due process (State v. Loomis, 2016). The Wisconsin Supreme Court, however, ultimately disagreed with Loomis's argument and held that the that both the circuit court and Loomis had access to the same information (State v. Loomis, 2016). They emphasized that if risk assessment instruments are considered properly with an understanding of the "limitations and cautions," their use does not violate one's right to due process (State v. Loomis, 2016).

While the Wisconsin Supreme Court determined that the use of the COMPAS risk assessment instrument did not infringe upon Loomis's right to due process, the case of *State of Wisconsin v. Eric Loomis* brings to light questions regarding both the influence of the instruments on individual's criminal sentences through risk levels and the efficacy of the algorithm in predicting one's risk of reoffending. The algorithms within risk assessment instruments can bring consistency and accuracy to the court room, potentially minimizing discrimination and bias (Chohlas-Wood, 2020). They also, however, have the potential to perpetuate bias and cloud judicial decisions with a lack of transparency. Risk assessment instruments have faced claims of potential for discrimination (Chohlas-Wood, 2020).

This paper explores the variation in outcomes between actuarial sentencings and an individual's COMPAS risk assessment score. In addition, it seeks to investigate the efficacy of COMPAS risk assessment scores in predicting an individual's recidivism risk. I begin by examining existing literature and research related to the use of recidivism risk assessment instruments in criminal justice in the United States. I explore existing literature on machine learning and risk assessment in relation to governance, the relationship between artificial intelligence and human judgement, and algorithmic liability.

This paper then investigates the variation in criminal sentencing as a result of recidivism risk assessment instruments using collection and analysis of booking and release data from the Wisconsin Department of Corrections. This data analysis includes regression analysis of individual's COMPAS risk assessment level and the resulting criminal sentences, both expected and actual. I use the difference in the actual and expected sentences, in comparison to an individual's risk level, to understand the extent to which COMPAS risk level accurately reflects reincarceration risk and whether this risk is then accurately accounted for in sentence length. Moreover, I also analyze the relationship between the COMPAS risk score, sentence term length, and actual reincarceration. I compare COMPAS risk scores across different demographic groups to explore potential biases. I then use interviews with experts, risk assessment instruments, and those involved in the criminal justice field to contextualize the current and future use of these instruments. This multidimensional approach integrating quantitative data with qualitative offers a nuanced view of the role COMPAS plays in sentencing and highlights the complexities of using AI in judicial decisions.

From this study, I found a positive relationship between risk level and reincarceration. On average, a high-risk level person is more likely to be reincarcerated than a medium or low risk level person. This finding emphasizes the validity of the algorithm, meaning that the algorithm effectively predicts recidivism risk. It, however, did not provide consistent results across various demographic groups when applied to sentencing. On average, a group facing a higher recidivism risk and actual reincarceration rate did not necessarily experience longer time spent in prison. A low recidivism risk level, rather than a high recidivism risk level, is associated with a longer period in prison. Moreover, I found that there is no relationship between an increase in recidivism risk level and sentence length on average. This prompts the conclusion that, during the sentencing process, judges deviate from COMPAS risk level recommendations. This deviation can either be accounted for by judges balancing other factors beyond minimizing reincarceration in their sentencing process or judges making mistakes and failing to use sentencing as a tool to mitigate reincarceration.

Background Context

Recidivism risk assessment instruments are actuarial tools used to determine the likelihood an individual will reoffend using population scale data (United States Department of Justice). This allows the justice system to funnel their efforts, such as offering release and providing different levels of supervision and interventions, to perpetuate best outcomes (United States Department of Justice). There are four generations of risk assessment instruments that reflect the changing role of risk assessment instruments in the criminal justice system. At first, staff at prisons used their personal training and experience to assess one's recidivism risk without a structured system (United States Department of Justice). Due to the risk of human mistakes and cognitive biases, a second generation of risk assessment instruments emerged that focused on "numeric predictions derived" from analyzing unchanging risk factors (United States Department of Justice). As second-generation risk assessment instruments failed to reflect changes over time to a criminal's risk factors, third generation risk assessments emerged to capture the changes in the factors' used to determine individuals' recidivism risk over time (United States Department of Justice). The fourth generation of risk assessment instruments employs the use of machine learning to monitor inmates over time and "to maximize treatment and supervision benefits" (United States Department of Justice).

The use of risk assessment instruments in sentencing arose in a few state-level Supreme Court cases that reaffirmed their use. In the aforementioned 2016 case *State of Wisconsin v. Loomis*, the Wisconsin Supreme Court held that the defendant's due process rights were upheld with the use of the algorithmic risk assessment COMPAS in sentencing (State v. Loomis, 2016, Chohlas-Wood, 2020). The court's use of risk assessment instruments was once again upheld in the 2018 case *State of Iowa v. Guise* (State v. Guise, 2018). The court affirmed that the use of the Iowa Risk Revised assessment tool in sentencing did not violate the defendant's due process rights.

Risk assessment instruments have also come under media scrutiny for the potential of being discriminatory. A 2016 article by ProPublica argues that the use of COMPAS in Broward County, Florida was biased because it resulted in a higher false positive rate for Black individuals compared to white individuals (Chohlas-Wood, 2020). While the ProPublica analysis might include some statistical flaws, it emphasizes the importance of analyzing potential discrimination and sources of bias in these instruments.

Literature Review

Machine Learning, Risk Assessment, and Governance

Most cases a judge hears are a simple assessment with a relatively predictable outcome, meaning that the case could be partly automated using artificial intelligence (Reiling, 2020). Courts across the United States and globally have done just that with the use of risk assessment tools. These tools have been used in the United States criminal justice system since the 1920s (Nishi, 2019). The use of these automation tools in the criminal justice systems allows for increased pattern recognition, advising, and predicting, providing potential advantages of using a machine-decision maker compared to using a human-decision maker (Reiling, 2020). The increased development of machine learning and data accumulation has allowed for algorithms to mediate the flow of information, increasing the use of risk assessment tools (Meserole, 2022). Modern risk assessment tools use algorithms that automate human processes, such as assessing one's recidivism risk, and influence a judge's decision making (Henman, 2021, Nishi, 2019). In addition, these algorithms can handle big data and derive connections in data beyond human capability (Fagan & Levmore 2019). The use of these automation tools in the criminal justice systems allows for increased pattern recognition, advising, and predicting, providing potential benefits for the use of a machine-decision maker when compared to the use of a human-decision maker (Reiling, 2020).

In the context of courts, the automation of governance decision-making means embedding existing laws, procedures, and policies into the code (Henman, 2021). However, if an algorithm inaccurately reflects policy and law, a systematic error becomes magnified (Henman, 2021). Whether the decisions are presented without human involvement or provided as additional information for human-made decisions is a critical consideration when implementing various machine learning algorithms in court systems (Henman, 2021, Nishi, 2019). These risk assessment instruments are often used in varying manners by different judges, presenting a degree of unpredictability. Due to the vulnerable nature of the pre-trial and sentencing process, risk assessment's potential to magnify systemic biases and the risk of varying implementations of the instruments speak to potential downsides and the importance of carefully using the tool. Fagan & Levmore (2019) argue that artificial intelligence algorithms are most effective when used in conjunction with a human judge and recommend that judges evaluate the responses from the machines, rather than blinding accepting the result. However, there are unobserved variables playing into a judges' decision (Kleinberg et al., 2018). These unobserved variables may impact outcomes and may lack transparency.

Courts have implemented the use of risk assessment instruments to predict recidivism and advise courts on bail, parole, and sentencing decisions. With the increased incorporation of predictive technology, states have developed legislation requiring the use of various risk assessment algorithms, such as COMPAS, by judges in bail, parole, and sentencing decision-making (Nishi 2019). COMPAS, or Correctional Offender Management Profiling for Alternative Sanctions, uses criminal record information and responses to a 137-question questionnaire to measure recidivism risk; the outcome is then used to influence decisions regarding pre-trial detention, sentencing, and early release in some states and jurisdictions (Reiling, 2021, Gacutan & Selvadurai, 2020). The fourth generation of COMPAS employs machine learning to predict a forecast for the offender's likelihood of reoffending (Donohue, 2019). This forecast provides a prediction of whether the offender will reincarcerated. As demonstrated by COMPAS, risk assessment algorithms are continuously being used to categorize risk profiles for individuals (Henman, 2021). An individual's risk is assumed by comparing shared characteristics and assuming a particular outcome. As one can imagine, this presents concerns about algorithm bias.

Artificial intelligence algorithms have been characterized as "black boxes" because they determine a set of outputs from inputs through a process that lacks transparency and is not clearly explained (Nishi, 2019, Deeks, 2019). Black box algorithms use machine-learning to make predictions, but the explanation for the algorithm's decision remains unknown (Price, 2018).[1] This presents some potential negatives of the use of risk assessment instruments due to the potential for biases to be inserted into the algorithm. Moreover, the algorithm relies on data to establish connections between an input and an output. Ultimately, if the data used for developing risk assessment algorithms contains a "false sense of objectivity," the predictions will be inherently inaccurate (Nishi, 2019, Mazur, 2021, Fagan and Levmore, 2019). In addition, if there is a lack of data, artificial intelligence algorithms fall victim to omitted variable bias (Fagan

[1] These algorithms are characterized as black boxes because of the lack of transparency regarding what factors play into their decision making. The user does not understand or know what is occurring inside of the model.

& Levmore, 2019).[2] This issue can be ameliorated through the collection of more data. Nishi (2019) characterizes the development of a risk assessment algorithm as an opportunity for outside influences to impact the output. Since algorithms are developed by private companies, this presents a confounding factor as these private companies may not have to disclose what information was used in the algorithm.

Artificial Intelligence and Human Judgement

Despite the characterization of artificial intelligence algorithms as removed from human thought processes, some claim that human judgement occurs naturally in artificial intelligence because human judgement is exercised in the development of the algorithm (Spaulding, 2021). Others argue that algorithms remove human frailties in decision making, allowing for outcomes to potentially be determined in a more objective and impartial manner (Henman, 2021). Some scholars claim that artificial intelligence ameliorates the impacts of the hungry judge effect.[3]

Due to reliance on data collection and human development of the algorithm, artificial intelligence itself can contain some biases. Moreover, when risk assessment instruments, such as COMPAS, are used in sentencing, they present quantitative inputs to the judge. As a result, this quantitative measure may unprecedentedly override qualitative measures, such as a defendant's individual circumstances, that may be impacting a judge's decision (Donohue, 2019). Moreover, unobserved and unknown factors may be influencing a judge's decision-making process (Kleinberg et al., 2018). COMPAS, like other software, incurs criticism due to the potential for

[2] The algorithm draws conclusion from two variables, even if those two variables are not causally related (Fagan & Levmore, 2019).

[3] In the hungry judge effect, as the day progresses, judges are increasingly impacted by human emotions, such as hunger or exhaustion, which impacts the outcome of their judgements. Essentially, human judgements are inherently impacted by biases (Chatziathanasiou, 2022).

bias, lack of transparency, and the interaction between the automated decision-making and human judges (Chatziathanasiou, 2022).

While artificial intelligence appears to remove human judgement, these two concepts remained tied up in one another. Spaulding (2021) notes that, specifically in the domain of law, human judgement is impacted by a dependence on artificial intelligence.[4] This is seen by judge's use of risk assessment instruments in sentencing. Using a case study of the Social Security Administration Disability Program as an example, artificial intelligence, when combined with human staff, can mitigate the issues caused by human's varying capabilities, knowledge, and performance (Glaze et al., 2022). This emphasizes the importance of combining automated decision-making tools with human decision-makers.

Liability in the Age of Artificial Intelligence

Private companies are often contracted for the development of risk assessment algorithms due to their expertise (Nishi, 2019). Since the development of the algorithm is outsourced, the reasoning behind one's risk assessment remains opaque even to the judge. This means that there are two layers of opacity in the sentencing process: the lack of transparency in the recidivism risk level and a lack of clarity in the variables used in the judge's sentencing decision. So, who remains liable for ensuring that the risk assessment algorithms remain constitutionally compliant? Using experiments exploring how humans allocated blame about the harm caused by AI software, it is determined that people blame the developers of the algorithm and then the companies under most conditions (Sullivan et al., 2022).

---

[4] Algorithms have social, political, and aesthetic dimensions that both reflect and influence human judgement. (Spaulding, 2021).

Judges remain unable to understand the operations of risk assessment algorithms due to legal protections of the algorithms and their lack of expertise regarding the technology (Nishi, 2019). The opaque nature of risk assessment algorithms prevents judicial review since judges are not able to objectively review the decisions made within the algorithm (Nishi, 2019, Gacutan & Selvadurai, 2020). This provides room for the possibility that risk assessment tools are being used with "meaningful technical flaws in the software" (Donohue, 2019). The opaqueness of the risk assessment algorithms potentially prevents these instruments from perpetuating fairness and justness.

As mentioned before, in *State of Wisconsin v. Loomis*, a defendant questioned the use of privately developed risk assessment tool in determining his prison sentence, but ultimately the court held that the defendant and the sentencing judge were given access to the same information (Nishi, 2019). Neither the judge nor the defendant had a full understanding how the risk assessment algorithm operated or what information was used by the algorithm to reach the determination. The court cautioned that judges should not rely solely on risk assessment instruments when making sentencing decisions (Donohue, 2019). One, however, does not know when and how judges will choose to employ automated risk assessment tools (Gacutan & Selvadurai, 2020). Moreover, due to trade secret protections, the private company who developed the risk assessment tool were not made to disclose the algorithms operations (Nishi, 2019, Mazur, 2021).[5]

With the increased use of both the digital space and artificial intelligence, there have been discussions regarding the development of digital rights, specifically regarding traditional privacy

[5] Crawford (2019) argues that private developers of artificial systems are like other private across performing government functions. As a result, these developers should serve as state actors due to the need for constitutional liability. Under state action doctrine, as private actors wield the power of the state, the court holds them accountable as if they were the states themselves (Crawford, 2019).

and data protection. Sullivan et al. (2022) argues that, given a new motivation to blame artificial intelligence, policymakers should encourage further research in AI accountability and liability. While total transparency of artificial intelligence algorithms in risk assessment instruments is unnecessary, rights to explanation should be implemented to successfully reap the benefits of the algorithm while also protecting personal rights and interests (Gacutan & Selvadurai, 2020).

The place for risk-assessment instruments in the courts remains unclear as these instruments' present positives, such as increased pattern recognition and lack of human bias, and negatives, such as algorithmic bias and lack of transparency. Current research exists that seeks to determine the validity of various recidivism risk assessment instruments. Austin et al. conducted a validation study of the Level of Service Inventory-Revised that intended to determine if the algorithm effectively sorted incarcerated individuals into risk categories that predicted likelihood of recidivism. Overall, they found that "noise" present in the instrument prevented it from being reliable in predicting risk of reoffending (Austin et al. 2009). However, a validation study of the COMPAS risk assessment instrument in 2010 determined that, on average, the risk scores were indicative of the likelihood that one would recidivate, especially for high-risk individuals (Blomberg et al. 2010). Hyatt and Chanenson's research provides a qualitative analysis of the judicial attitudes regarding the implementation of risk assessment instruments (2016). This paper does not focus on the use of recidivism risk assessment instruments in the State of Wisconsin more specifically.

Other researchers, however, do not approach the subject from both a qualitative and quantitative angle for an analysis of the COMPAS risk assessment instrument and the State of Wisconsin more generally. I intend to determine whether the automation of decision-making using risk-assessment instruments, such as COMPAS, result in variation in sentencing compared to the results determined by human judgements. Moreover, I intend to conduct an exploration to better understand the efficacy of recidivism risk assessment as a tool for prediction and as a tool for providing better outcomes. While navigating the lack of transparency in risk-assessment instruments, I seek to explore what role artificial intelligence should play in sentencing. Should artificial intelligence be used alone or in conjunction with human judgement to determine sentencing outcomes? Or does artificial intelligence cause more harm than good in the court rooms and should not be allowed?

Methodology and Data

Quantitative Methods

To explore the variation in criminal sentencing between actuarial sentencings, individual's COMPAS risk assessment score, and recidivism risk, I decided to conduct a primarily quantitative study with qualitative interviews as a supplement to help contextualize my findings. To begin my quantitative analysis, I obtained a private research dataset from the Wisconsin Department of Corrections. I decided to acquire my data and focus on Wisconsin because Wisconsin is one of the few states that uses machine-learning based risk assessment instruments during the sentencing process. Moreover, Wisconsin provides a unique case study due to the *Wisconsin v. Loomis* case that was mentioned before. Through this case, the Wisconsin Supreme Court has determined that the use of recidivism risk assessment instruments as a tool in the sentencing process does not infringe upon due process rights, and thus the use of these instruments is deeply implemented into the Wisconsin sentencing process. In addition, the Wisconsin Department of Corrections provided guidance and assistance by walking me through the available data, allowing me to truly understand the data and variables that were included in the dataset.

This study uses individual inmate records as the unit of measure and pulls all data related to physical admissions to the Wisconsin Department of Corrections adult prisons in 2019. I chose to focus my study on admissions in the year 2019 for a few reasons. To begin, my discussions with experts in the field revealed that the COVID-19 pandemic changed the landscape of the sentencing process. My interview with a pro-bono lawyer in Wisconsin emphasized that the COVID-19 pandemic resulted in a turnover from older judges to newer judges who relied heavily on recidivism risk assessment instruments in their sentencing decisions. As a result, the data from recently after the COVID-19 pandemic may be biased. This would not allow us to accurately understand both the efficacy of the recidivism risk assessment instruments in predicting recidivism risk and the judges' use of the instruments during the sentencing process. Instead, it might reflect other institutional changes occurring. In addition, I felt it was important to use data a few years after the *State of Wisconsin v. Loomis* Wisconsin Supreme Court case occurred since this case reaffirmed the states' use of the risk assessment tools in sentencing. Time between this case and the year of my analysis would allow for the use of the risk assessment tools in sentencing to be more standardized and better understood by both those conducting the assessments and those using the risk level scores.

I collected data through the research request team at the Wisconsin Department of Corrections. While there were other ways to obtain the necessary data, such as through various county jail departments, I found that the private research dataset acquired from the Wisconsin DOC was the most comprehensive and had the largest scope. In addition, the Wisconsin DOC research team walked me through the available variables and provided insights into which variables they believed would be most informative. This dataset included information related to all individuals who were admitted to the Wisconsin Department of Corrections adult prison during 2019. It includes admission date, expected release data, COMPAS Risk level at admission and at the most recent taking, and demographic information. The Wisconsin DOC also provided a codebook alongside the dataset that provided a key explaining what each variable meant once again.

Once I obtained this dataset, I cleaned and created a final dataset that I would then use for the regression analysis. I conducted my analysis using Stata MP version 18.0. Before I began my analysis, I wanted to clean this dataset to clearly include variables indicating demographic categories, one's COMPAS risk level, actual time spent in prison, expected time spent in prison, and whether they were reincarcerated. To achieve this, I began by assigning numbers to correspond to the various categories under each included demographic variable to allow for easier analysis. I then coded individuals who were determined to be high risk from the COMPAS risk assessment both at admission into the prison and their most recent assessment as 3. Those who were determined to be medium risk were coded as a 2 and those who were determined to be low risk were coded as 1. I also generated a variable to account for actual length of time spent in prison from admission to release. I then calculated the actual length of time in prison by looking at the time between the admission date and the actual release date. The Wisconsin Department of Corrections included actual release dates up until mid-October of 2023, which is when I obtained the dataset. I also generated a variable to account for the expected time in prison. This is calculated from analyzing the time between the expected release date and the admission date.

I then generated three different indicator variables indicating if an individual is high risk, medium risk, or low risk for COMPAS scores provided. For example, when creating an indicator variable for high risk level, those who were determined have a high COMPAS risk level were coded as a 1, while those who were medium or low COMPAS risk level were coded as a 0. A similar line of reasoning was employed when creating indicator variables for medium and low risk levels. I also created an indicator variable that demonstrates whether an individual was reincarcerated anytime from 2019 to June 30, 2023. This indicator variable uses a 1 to indicate all individuals who were reincarcerated; a 0 indicates that the individual was not reincarcerated within the designated time frame.

While the dataset I used is a comprehensive dataset from the Wisconsin Department of Corrections, it still presents some limitations that are important to consider. First, an individual may be counted more than once in a year if they were admitted multiple times. In addition, this dataset does not include individuals under Division of Community Corrections supervision who were admitted on a temporary hold. Moreover, this dataset only includes physical admissions to a Wisconsin Department of Corrections adult prison. This does not include juvenile inmates or those who were not admitted to the adult prison. It also does not include individuals who were admitted to county jails or were moved to an alternative treatment center. I intentionally made the choice to use data from the adult prisons as this provided more comprehensive data related to the use of COMPAS risk assessment instruments. It is also important to consider that individuals may be serving sentences related to crimes committed in Wisconsin during 2019 but may be serving these sentences in other states. As a result, the data related to their prison admission and release would not include in this dataset. Moreover, this dataset is limited to 2019. While it includes individuals' release up until mid-October 2023, the dataset only includes those who were admitted in 2019. As a result, information related to individuals actual release may not be reflected in the dataset if they are still held in the prison after mid-October 2023, even if information related to their expected release is included. These limitations have been accounted for in my study and in my analysis of my findings. Overall, this dataset provides a strong micro-level data set that was compiled by a reliable department of corrections.

Using this compiled dataset that includes a myriad of coded indicator variables, I then ran various regressions to explore the variation in sentences and reincarceration related COMPAS risk levels. I denote these risk levels as high risk, medium risk, and low risk. I conduct two levels of analysis. The first analysis intended to explore the variation in sentence term length and COMPAS risk assessment levels. The second analysis explored the relationship between reincarceration and COMPAS risk assessment levels. In both levels of analysis, COMPAS risk assessment levels constitute the independent variable. In the first exploration, the length in time in the prison represents the dependent variable as the COMPAS risk level is a factor in determining time spent in prison (Table 1). In the second exploration, whether or not an individual was reincarcerated represented the dependent variable (Table 2). If COMPAS risk levels accurately reflect one's recidivism risk, then one's likelihood of reincarceration could be predicted by their risk level. I then explore the relationships between age, race, ethnicity, and marital status at time of admissions and COMPAS risk level. After speaking with experts, these were determined to be some of the most important demographic variables when determining the likelihood of an individual committing a crime. Therefore, it is important to control for these covariates in the regression. Moreover, given the potential that the length of sentence could influence one's reincarceration likelihood either by keeping someone incarcerated or through the effects of incarceration on future behavior, I also controlled for length of sentence and when the inmates were released from prison in my regression.

Table 1. Variable Description for Analysis of COMPAS Risk Scores and Length of Sentence

Variable	**Measuring**
Dependent	Length of Sentence

Independent	COMPAS Risk Score
Control	Age of Admission
Control	Race
Control	Ethnicity
Control	Marital Status at Admission

Table 2. Variable Description for Analysis of COMPAS Risk Scores and Incident of Recidivism

Variable	**Measuring**
Dependent	Reincarceration (before 6/30/2023)
Independent	COMPAS Risk Score
Control	Age of Admission
Control	Race
Control	Ethnicity
Control	Marital Status at Admission
Control	Length of Sentence
Control	Date of Release

I then used a regression to better understand the relationship between the independent and dependent variables. I first ran a regression to determine the average impact of COMPAS risk score on actual and estimated sentence lengths. I am looking to better understand how, as COMPAS risk level increases, time spent in prison changes. I first ran two regressions that explored how actual and expected sentences changed as risk level changes from low to medium to high-risk level. I used the following regression:

$$E[Y_t] = \beta_0 + \beta_1(COMPAS\ Risk\ Level) + Controls + \varepsilon$$

- $E[Y_t]$ is the length of sentence, whether actual or estimated, in days.
- β_0 is the average prison sentence length in the absence of high, medium, and low-risk level.

- β_1 is the change in the individual's sentence on average as COMPAS risk level moves up a risk level.[6]

I also ran the following regression accounting for the presence of medium-risk level and low-risk level using the indicator variables, as well as for both estimated and actuarial sentences, compared to high-risk level individuals:

$$E[Y_t] = \beta_0 + \beta_1(COMPAS\ Indicator) + \beta_2(COMPAS\ Indicator) + Controls + \varepsilon$$

- $E[Y_t]$ is the length of sentence, whether actual or estimated, in days.
- β_0 is the average prison sentence length for an individual in the absence of risk level.
- β_1 is the change in the individual's sentence on average because of being low-risk level from COMPAS.[7]
- β_2 is the change in the individual's sentence on average because of being medium-risk level from COMPAS.

For my second level of analysis, I also wanted to see how, as COMPAS risk level increases, the likelihood of being reincarcerated changes. I began this analysis by running a regression that looked at how the likelihood of reincarceration changes as risk-level increases. I used the following regression:

$$E[Y_t] = \beta_0 + \beta_1(COMPAS\ Risk\ Level) + Controls + \varepsilon$$

- $E[Y_t]$ is whether an individual was reincarcerated between the time they were released from prison to June 30, 2023.
- β_0 is the average likelihood an individual was reincarcerated in the absence of high, medium, and low-risk level.

[6] For example, as risk level moves from low risk level to medium risk level. Or, as risk level moves from medium risk level to high risk level.

[7] The full equation is:

$$E[Y_t] = \beta_0 + \beta_1(Low\ Risk\ Level\ Indicator) + \beta_2(Medium\ Risk\ Level\ Indicator) + Controls + \varepsilon$$

- β_1 is the change in the individual's likelihood of reincarceration on average as a result of an increase in COMPAS risk level.

I ran another regression using indicator variables comparing the inmates who were medium and low risk to high risk. The following regression intended to determine the relationship between COMPAS risk level and reincarceration:

$$E[Y_t] = \beta_0 + \beta_1(COMPAS\ Indicator) + \beta_2(COMPAS\ \ Indicator) + Controls + \varepsilon$$

- $E[Y_t]$ is whether an individual was reincarcerated between the time they were released from prison to June 30, 2023.
- β_0 is the average likelihood an individual was reincarcerated in absence of COMPAS risk level.
- β_1 is the change in the individual's likelihood of reincarceration due to being low risk.
- β_2 is the change in the individual's likelihood of reincarceration due to being medium risk.

While this regression methodology provides insights into the relationships between COMPAS risk scores, reincarceration, and sentence lengths, it is important to also note some limitations to this methodology. To begin, the controls were limited to the data that was available by the Wisconsin Department of Corrections. Some demographic data was taken and self-reported by the inmates at intake when they were admitted to the prisons. Some covariates, such as education level, income level, and substance abuse, could not be accounted for because this data was not reported or available at the Wisconsin Department of Corrections. These factors, however, are likely accounted for in the COMPAS risk assessment scores. In addition, the length of sentences is calculated from the time of admission to the time of release (actual or expected). While I am analyzing both the actual and expected criminal sentencings to best account for potential differences in the two, this study does not account for the implications of being released early or later than expected, or the actuarial sentence lengths are given. This is due to a limitation in the data available from the Wisconsin Department of Corrections. This study instead measures the time served in prison from admission to prison to the actual or expected release date. In addition, the COVID-19 pandemic resulted in some prison populations being released early.

Moreover, this analysis and data is limited to Wisconsin and specifically the use of COMPAS risk assessment instruments in sentencing. Wisconsin's average demographics, however, potentially represents United States' average demographics, allowing us to draw broader conclusions from the state-level results. In 2021, the median household income in Wisconsin was $67,080 compared to the national average of $69,021 (US Census Bureau). 10.7% of the Wisconsin population is in poverty compared to 11.5% of the US population living in poverty (US Census Bureau). Education levels in Wisconsin are also similar to education levels in the United States. 92.9% of individuals in the US have a high school graduate degree or higher compared to the 88.9% of individuals in Wisconsin (US Census Bureau). As of September 2023, the Wisconsin unemployment rate rested at 3.1%, compared to a national unemployment rate of 3.8% (Bureau of Labor Statistics). It is important to account for similarities between Wisconsin and the United States, on average, when drawing conclusions from this study.

While this study focuses on Wisconsin specifically, I believe aspects of my findings are generalizable to other regions within the United States that employ the use of recidivism risk assessment instruments specifically in the sentencing process. I do not believe these findings would be applicable to areas that use recidivism risk assessment instruments that are not employed in the sentencing process.[8] I believe these my findings could likely be applied to the State of Iowa and the Iowa Risk Revised assessment tool since the use of the tool during the sentencing process was affirmed during the 2018 case of *State of Iowa v. Guise.* Given the similar underlying similarities between Iowa and Wisconsin and their affirmed use of recidivism risk assessment instruments in sentencing, I believe much of this analysis could be applicable to the State of Iowa. In addition, Pennsylvania implemented the Sentence Risk Assessment Instrument in 2020. Much like COMPAS, this tool is used to help in establishing an appropriate sentence within guidelines (Pennsylvania Sentencing Commission, 2020). I do, however, believe that the findings from this study cannot be generalized to include the Sentence Risk Assessment Instrument in Pennsylvania. Since the instrument was implemented in July of 2020, this is following the start of the COVID-19 pandemic. Given the judge turnover and crucial changes that occurred during this time, the results of this study may not provide a genuine effect in Pennsylvania. The Virginia Criminal Sentencing Commission also developed a risk assessment instrument that is used during the sentencing process. Unlike COMPAS, however, this tool is used in identifying those who have a low probability of reincarceration and providing alternative treatment (Ostrom et al., 2002). Thus, I do not believe these findings can be expanded specifically to the State of Virginia.

Qualitative Analysis

In addition to my analysis of the individual level prison admission data from the Wisconsin Department of Corrections, I also conducted six interviews with individuals who had experience working with recidivism risk assessment instruments. All 6 of these interviews lasted

[8] Many recidivism risk assessment instruments, such as Public Safety Assessment and Ohio Risk Assessment System, are not used in the sentencing process.

between 15 to 30 minutes. These interviews were categorized into three categories based on individual's experience with the instruments: (1) Algorithm Development, (2) Legal, and (3) Field Experts.

I had the opportunity to speak with two individuals working for different companies involved in the development and implementation of specific risk assessment instruments. Over this remote Zoom call, one individual presented their own company's presentation regarding the process of development. This provided important perspective on the implementation of the algorithm and steps that the developer had taken to ensure proper implementation. I then engaged in another Zoom call with another individual involved in the development and training related to a different recidivism risk assessment instrument. I also spoke with an individual working at a state Department of Justice who had experience working directly with risk assessment instruments. This conversation took the form of a phone call. In addition, I interviewed an individual who worked for a legal aid company dedicated to providing free legal assistance to those in need. He provided a unique perspective on the impacts risk assessment instruments have on cases from an actuarial standpoint over a Zoom call.

I then spoke to two experts who dedicated time and energy towards researching the use of risk assessment instruments, as well as the long-term impact these instruments have on communities. Through remote video conferences, these individuals helped to provide insights on existing research related to the field. Each of these individuals were happy to share their perspective on how they believe risk assessments should be used, their hesitations, and the benefits they see to using the algorithm. This qualitative analysis was employed to provide contextualization and guidance for my analysis of the dataset. During these interviews, I took extensive notes that helped to paint a picture of each individual's perspective and to reference throughout the research process.

Table 3. Qualitative Interview Summary and Categorization

Category	**Interview**
Algorithm Development	Higher education organization
Algorithm Development	Non-profit dedicated to alternative justice
Legal Career	Working at Wisconsin Department of Justice
Legal Career	Working for a pro-bono legal clinic in Wisconsin
Field Experts	Expert researching use of tools generally
Field Experts	Expert exploring validity and efficacy of tools

These qualitative interviews had limitations to them. To begin, individuals self-selected by agreeing to participate in the interviews. To meet the field experts, I began by reading related research and reaching out to experts who seemed like they would be able to provide a unique perspective. These individuals then chose to participate in the conversation at a time and over a video conferencing platform of their choosing. Moreover, this also relied on snowball sampling, as some conversations resulted in experts suggesting I speak to someone else. To reach out to developers of risk assessment instruments, I first compiled a list of all implemented instruments and the companies in charge of developing these algorithms. I then reached out to these developers via email. After receiving responses, I was then put in contact with specific representatives. Not all companies responded or were willing to have a conversation beyond the information available on the website. Moreover, I also reached out to pro bono law clinics and was put in touch with individuals who were believed to have extensive experience regarding these instruments or a strong opinion on the instruments. The self-selection and process by which individuals became involved in the interview process presents some potential forms of biases.

Results/Findings

This findings section seeks to frame the results from the regression analysis and interviews outlined above. The following findings are divided into three sections: Risk Level and Sentence Term Length, Risk Level and Reincarceration Rates, and COMPAS Risk Assessment Instruments and Bias. Through these sections, I intend to examine the efficacy of COMPAS risk assessment tools as prediction mechanism of reincarceration during the sentencing process while using time in prison as the mechanism. While the existing literature indicates there has been past research into the use of risk assessment instruments in various steps throughout the criminal justice process and the potential biases the use of these instruments introduces, I specifically wanted to explore the efficacy of these machine learning tools in predicting risk of reincarceration through their impact on individuals' sentence term length.

In addition, I was particularly interested in the effectiveness of the COMPAS tool since it employs machine learning to predict individuals' risk levels and is used, specifically in Wisconsin, as an instrument in determining individual's sentence length. If COMPAS risk tools are not accurately predicting risk level or being used effectively in the sentencing process, this presents issues with real life implications; the use of COMPAS risk assessment instruments creates a potential for algorithmic bias, which could result in unfair outcomes due to biased data being used to train the algorithm. I explore how the use of the instrument influences varying demographic groups differently in assessing reincarceration risk and in sentencing. It is important to understand the efficacy of the tools in predicting reincarceration and the actuarial use in the sentencing process. My findings reveal a complicated landscape surrounding the efficacy of COMPAS tools as a tool to predict reincarceration rates and the use of the COMPAS risk scores in the sentencing process.

First, I broke down the population of inmates who were admitted to the Wisconsin Department of Corrections in 2019 by COMPAS risk level. As seen in Table 1, the majority of the population of those who were admitted and received a COMPAS risk assessment on admission were determined to be of medium risk. Low risk individuals comprised the smallest percent of the population who were admitted and received a COMPAS risk assessment on admission.

Table 4. Percent of Admitted Inmates in Different COMPAS Risk Level

Risk Level	**Percent of Population**
High Risk	37.82%
Medium Risk	40.72%
Low Risk	21.46%

Risk Level and Sentence Term Length

COMPAS risk assessment instruments are used in determining individuals' sentence length by employing one's risk level to determine one's risk of reoffending. My conversation with an individual who works for an organization that helped develop the Public Safety Assessment emphasized that they do not recommend the recidivism risk assessment instruments for sentencing because they believed that sentencing needs to be individualized.[9] They argued that individual analysis is necessary to provide the best treatment for the person, rather than using instruments that are intended for punishment. In their opinion, risk assessments need to be used when determining how to help a person, rather than how to punish someone through sentencing. However, in Wisconsin, COMPAS risk levels is used as a piece of information during the sentencing process. It is important to understand the relationship between COMPAS

[9] Public Safety Assessment is a risk assessment instrument that uses actuarial data to estimate failure to appear pretrial in court and new criminal arrests during pretrial release. It provides information to judges when considering if a defendant should be released pretrial.

risk level and sentence lengths to better understand how these tools are used during the sentencing process.

In order to analyze the relationship between time in prison and COMPAS risk level, I first explored the average sentence length of inmates broken down by their COMPAS risk level, as seen as Table 5.

Table 5. Average Length of Sentence by COMPAS Risk Level

Risk Level	Average Sentence (Days)	Average Sentence (Years)
High Risk	622.76	1.71
Medium Risk	615.72	1.69
Low Risk	668.32	1.83

This table indicates that, on average, those who are determined to have a low COMPAS risk level have the longest sentence (668.32 days). Those who are determined to have a high-risk level have a longer sentence length, on average, than those who are indicated to be of medium risk level but a shorter average sentence length than those who were determined to be of low risk. There is, however, a marginal average difference between the sentence length for those who are determined to be high risk and medium risk. Those who were determined to be low risk served a longer time in prison, on average, than those who were determined to be medium or high risk.

Figure 1. Average Length of Sentence (Days) by COMPAS Risk Level

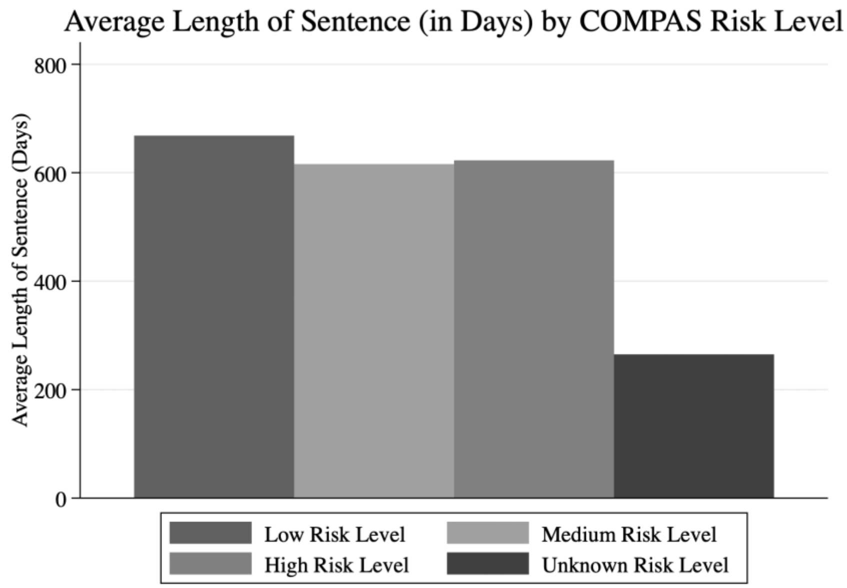


Figure 1 visually depicts that, on average, low risk level corresponds with a higher average sentence length than medium and high risk levels. One would expect that sentence length would increase as COMPAS risk level increased; this, however, is not the case. It is important to note, however, that not every inmate receives a COMPAS risk assessment at admission or throughout their prison sentence. This is accounted for in Figure 1 by the Unknown Risk Level category, which on average, has the lowest sentence length. In addition, not all individuals who are admitted into the prison are eligible for a risk assessment. My conversation with the individual working at the Department of Justice emphasized that risk assessment instruments taken at admission are part of a pre-sentence investigation. The determination of whether to undergo a presentence investigation is determined on a county-to-county basis; as a result, there may be variation in the data, which presents a source of bias.

I then ran various regressions that intended to uncover the relationship between a higher COMPAS risk level and length of sentence. Table 6 emphasizes that, on average, as COMPAS risk level increases from a low risk level to a high risk level, the actual length of sentence decreases on average when accounting for COMPAS risk levels taken at admission. When accounting for only the values that are determined to statistically significant, a higher COMPAS risk level corresponds to a lower actual time incarcerated on average (Table 6). As the COMPAS risk level increases by one point, the actual sentence length decreases by around 133 days on average.[10]

Table 6. COMPAS Risk Level and Length of Sentence Regression Table

	Sentence Length (Days)	
	Actual	**Expected**
COMPAS Risk at Admission	-133.454*	56.945
Age at Admission	-0.747*	6.550*
Race	-33.532*	-170.286*
Ethnicity	30.010*	138.32*
Marital Status	-1.242	34.955

**Value is statistically significant and has a p-value of 0.05 or lower.*

I also wanted to explore the relationship between the length of sentence under different indicator variables for high, medium, and low COMPAS risk levels at admission and when most recently assessed. This will allow me to better understand the relationship between COMPAS risk level and sentence length. When accounting only for statistically significant values, I found that, on average, a low COMPAS risk level was associated with a higher sentence length. For COMPAS risk levels taken most recently, it was found that an individual who was determined to be of low risk was likely to have a longer actual sentence length than a high risk individual (Table 7). A lower COMPAS risk level was associated with more time spent in prison than those who were determined to have a higher COMPAS risk level. On average, those who received a low COMPAS risk assessment level at admission had a sentence length 239 days longer than

[10] Increases by one point indicates a change from low to medium risk level or a change from medium to high risk level.

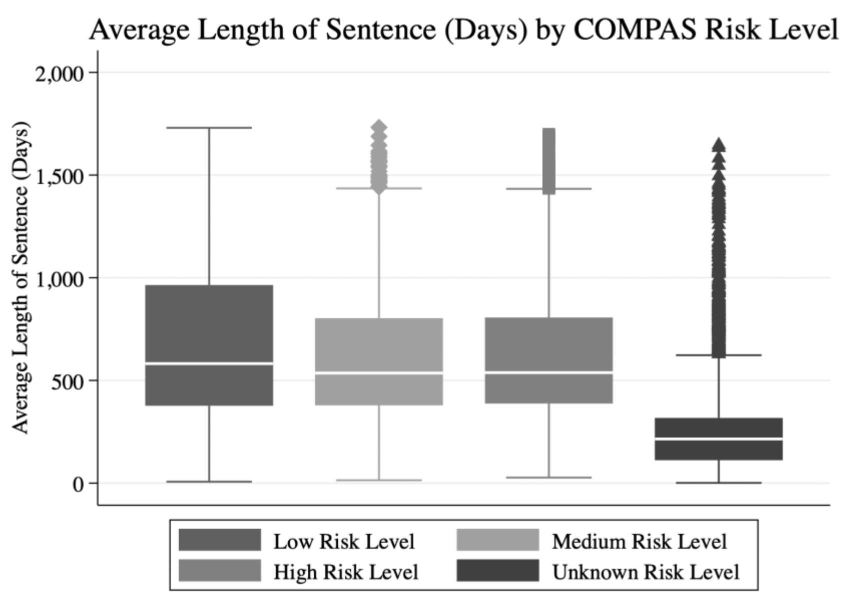
those who were high risk (Table 7). On average, those with a low COMPAS risk score taken most recently had a sentence that was 130 days longer than those who were high risk. Overall, across instances when the risk level is taken both at admission and at a more recent time, it was found that, on average, low risk individuals had longer actual and expected sentence lengths than those who were medium or high risk (Table 7). Figure 2 reiterates that the category of low COMPAS risk level corresponds with the highest average and individual-level sentence length.

Table 7. Low and Medium COMPAS Risk Level at Admission Compared to High COMPAS Risk Level and Length of Sentence

	Sentence Length (Days)			
	Actual		**Expected**	
COMPAS Low Risk Admission	239.14*		50.85	
COMPAS Medium Risk Admission	186.80*		-100.04*	
COMPAS Low Risk Recent		130.86*		506.78*
COMPAS Medium Risk Recent		33.64*		99.61*
Controls	Yes	Yes	Yes	Yes

**Value is statistically significant and has a p-value of 0.05 or lower.*

Figure 2. Average Length of Sentence (in Days) by COMPAS Risk Level



From these results, I find that judges deviate from risk level recommendations when determining sentence length. They do not necessarily give longer sentences to people who are determined to be higher risk. In sentencing, reducing recidivism is one of many goals that influences the sentencing process. From these findings, I determine that either judges are accounting for more factors than just recidivism risk or, if they are accounting for solely recidivism risk, they are making errors in employing the algorithm as a tool for reducing reincarceration. The other factors judges may be accounting for in the sentencing process, such as ability to be rehabilitated, personal circumstance, and external factors, may be inversely related to their COMPAS risk level. For example, if someone is a single parent attempting to support their family, a judge may be more inclined to seek a shorter sentence even if the person is marked as high risk. It is critical to mention the importance of alternative interventions; the possibility of providing alternative interventions may change a judge's sentencing decision.

I had the opportunity to speak with a professor at a law school who focused their research on risk assessment instruments, and they emphasized that there is a correlation between the presence of mental health or substance abuse treatment resources and the use of alternative, non-jail sentences. They also highlighted that these tools are used to differentiate people with a high likelihood of reoffending. This professor noted, however, that risk assessment instruments should be used "for instances that are not serious crimes. [Their use] should not be the case for instances that involve murder." Moreover, I do not believe these results can be tied to parole or early release from prison due to Wisconsin Department of Corrections policy regarding parole and early release. Only offenders in the Wisconsin Department of Corrections who were convicted of crimes before December 31, 1999 are eligible for parole interviews (Wisconsin Department of Corrections).[11]

My conversation with the lawyer at the pro-bono clinic in Wisconsin provided insights into the use of the risk assessment instruments as tools to predict recidivism risk. From their experience, they found that risk assessment instruments provide more equitable outcomes than a "bad" judge but provided less perception than a "good" judge. They argued that the risk assessment instrument "doesn't pick up on factors like being a single parent" and focuses on other elements. These additional factors could help to explain the judge's variation from the COMPAS sentencing recommendations. Moreover, the lawyer emphasized that there are real-life implications for being designated "high-risk" that judges may not realize.

Risk Level and Reincarceration

My interview with an individual working at the Wisconsin Department of Justice emphasized that court rulings such as *State v. Loomis* provide indication of Wisconsin's movement toward more evidence-based solutions at the county level through the use of risk assessment instruments. Due to the real life implications of COMPAS risk levels, it is important to analyze the validity of the tools in predicting reincarceration risk.

To explore this, I conducted an independent validation study to analyze whether the use of COMPAS risk assessment instruments as a tool for predicting an individual's risk of reoffending accurately reflected actuarial results. I intended to determine whether individuals who were determined to be high risk either at admission or at their most recent assessment were more likely to be reincarcerated between the time of their release and June 30, 2023, on average.

[11] According to Wisconsin's Truth-in-Sentencing laws, "any person who commits a felony offense on or after December 31, 1999 and is sentenced to at least one year in prison" is not eligible for parole (Wisconsin Department of Corrections).

I began by breaking down the population of inmates who were reincarcerated by COMPAS risk level. I found that, on average, individuals who were high risk were more likely to be reincarcerated than individuals who were demonstrated to be medium and low risk (Table 8 and Table 9). Those who were high risk comprised the largest percent of individuals who were reincarcerated. This finding held true when I looked at whether the COMPAS risk level being analyzed was taken at admission or during the most recent assessment, in addition to whether I was analyzing just the population of prisoners who were reincarcerated or all individuals who were admitted to the prison system.

Figure 3. Average Reincarceration Rates by COMPAS Risk Level

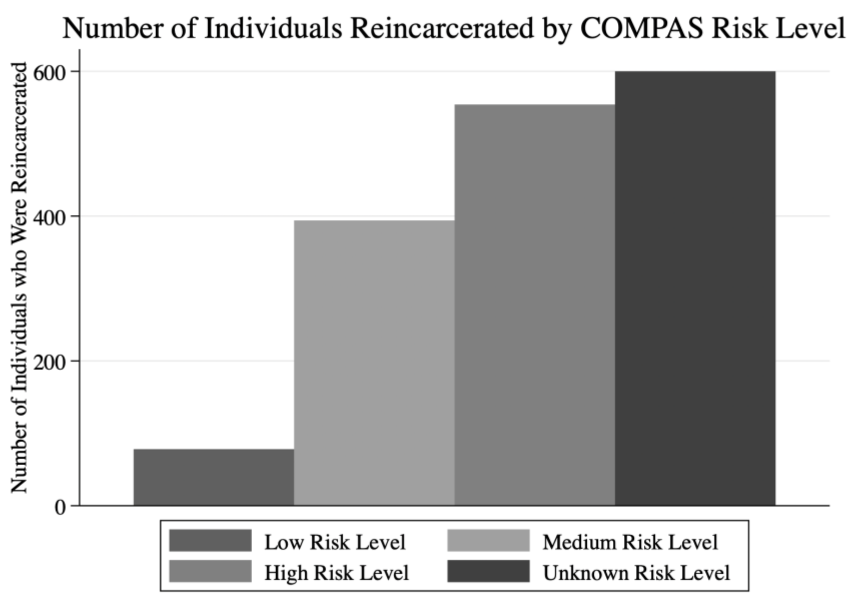


Table 8. Percent of Inmates Reincarcerated by COMPAS Risk Level at Admission

Risk Level	**Percent (of Reincarcerated)**	**Percent (of All Admitted)**
High Risk	34.07%	5.98%
Medium Risk	24.31%	4.25%
Low Risk	4.80%	0.84%

Table 9. Percent of Inmates Reincarcerated by Most Recent COMPAS Risk Level

Risk Level	**Percent (of Reincarcerated)**	**Percent (of All Admitted)**
High Risk	66.11%	11.60%
Medium Risk	28.41%	4.98%
Low Risk	5.41%	0.95%

I then found that, on average, as the COMPAS risk level increased from low risk to medium risk to high risk, the chance, of an individual being reincarcerated increased (Table 10). When controlling for demographics, age variations, and time, I found that as the COMPAS risk level that was taken on admission into the Wisconsin Department of Corrections increased, individuals were 0.0832 more likely to be reincarcerated, on average. It is important to highlight that not all individuals received a COMPAS risk assessment at admission or at all. I argue, however, that these findings still reveal the relationship between COMPAS risk level and reincarceration probability as these findings are statistically significant.

Table 10. COMPAS Risk Level At Admission and Reincarceration

	Reincarcerated
COMPAS Risk Level	0.0832*
Age at Admission	-0.0033*
Race	0.0055
Ethnicity	-0.0221*
Marital Status	-0.00431
Length of Sentence	-0.0001*
Release Date	-0.0001*
Observations	5,211

**Value is statistically significant with a p-value of 0.05 or lower*

Moreover, I then employed indicator variables for the different risk level categories to further explore the relationship between risk level and reincarceration. I found that, on average, individuals who were determined to be high risk were more likely to be reincarcerated than those who were medium or low risk (Table 11). Individuals who were low risk were less likely to be reincarcerated than those who were either medium or high risk, on average. Those who were low risk were 0.1112 less likely to be reincarcerated than those who were high risk, on average (Table 11). Those who were medium risk were 0.0363 less likely to be reincarcerated than those who were high risk, on average (Table 11). Moreover, those individuals who were low risk were more likely to not be reincarcerated than those who were medium and high risk, on average. It shall be noted that these findings were statistically significant. Overall, the presence of a high-risk individual correlates to a higher probability of being reincarcerated, on average. From this analysis, COMPAS risk levels appear to be an accurate prediction of one's risk of reincarceration, on average.

Table 11. Medium & Low COMPAS Risk Level and Reincarceration Compared to High Risk

	Reincarcerated
Medium COMPAS Level	-0.0363*
Low COMPAS Level	-0.1112*
Demographic Controls	Yes
Age Control	Yes
Sentence Length Control	Yes
Release Date Control	Yes

**Value is statistically significant with a p-value of 0.05 or lower*

Throughout my interviews, the importance of evidence-based solutions were continuously emphasized; it is important to have tools that are known to work. My interview with a law school professor who focused on recidivism risk assessment instrument research highlighted that, in their opinion, the tools did not necessarily have a good or bad impact on outcomes; there was a lack of difference in the outcomes for judges who used the tools and those that didn't. The professor, however, emphasized that some other criminal justice improvements that are made as part of programs that implement risk assessment instruments are beneficial. These include steps such as faster processing from arrest to first appearance. While in their opinion, the risk assessment instruments did not have a big impact on outcomes, other measures that were taken at the same time did.

COMPAS Risk Assessment Instruments and Bias

Throughout research process and interviews, the issue of bias in algorithms kept emerging. My interview with a law school professor argued that the most important factors in assessing one's risk of recommitting a crime are: age, gender, and presence of prior offenses. Race should not play a factor. My conversation with a lawyer working at a pro-bono legal clinic mentioned that when risk assessment instruments were first implemented, they provided promises of a fairer system that were not fulfilled. They emphasized that, even if one removes race, gender, and sexual orientation as factors being used to determine results, the algorithm within the machine can still use other aspects to put together the pieces. They drew an analogy to a picture that was just an outline of the horse and featured elements such as hay or a barn. While the horse is technically missing from the picture, anyone looking at the image knows that the horse is there. They argued that the algorithm is biased against individuals who may lack resources. The individual working for the organization involved in developing the Public Safety Assessment emphasized that the data used to generate the algorithm is often biased due to "overpolicing of communities that are black and brown." This results in an overrepresentation of black and brown people in the system, leading to biased results as the algorithm uses actuarial training data. Since risk assessment instruments are developed using real-life data that includes biases due to the over policing, there is a risk of algorithmic bias, where risk levels for demographic groups may be biased by the skewed data. It is important to understand this risk and account for it when using recidivism risk assessment instruments in the sentencing process.

In addition, my conversation with an individual working for the organization that developed the Ohio Risk Assessment System emphasized the difficulty of helping people in a system that systematically oppresses them. While the data emphasizes that the tool works, implicit bias exists. They emphasized the importance of looking at the percentage of overrides present when using the tool; as staff are not completely comfortable with the use of the risk assessment instruments, they may override and change the final determination of risk, specifically in cases of committed sexual offenders. These overrides potentially provide instances of bias as personal judgements are made. Part of my research focused on exploring forms of bias in the COMPAS risk assessment instrument that resulted in unfair outcomes.

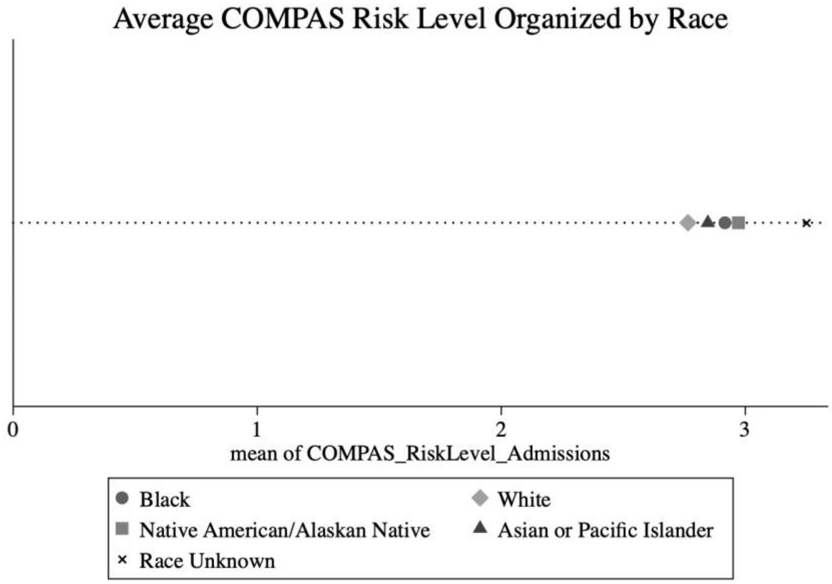
Moreover, my conversation with a law school professor whose work focused on research emphasized that validity and efficacy are important to ensuring an algorithm works well. They defined validity as “whether [the algorithm] classifies well.” Efficacy was defined as, if the algorithm is given to judges, “does it improve their decision making.” Their research had emphasized the importance of determining whether a high-risk classification indicated a higher risk of behavior. In addition, this professor emphasized the importance of asking the question of whether “racial differences in decision making get worse or better when the algorithm” is employed. The question is not about whether there is bias in algorithms, but rather whether algorithms truly mitigate this bias and provide better outcomes than would occur in the absence of the algorithm.

To better understand the racial biases in COMPAS, I began by breaking down the admitted prison population by race and then exploring if specific races had higher percentages of the population being reincarcerated than other races. I found that a higher percent of the Native American and Alaskan Native populations was reincarcerated than the other racial groups; the percent of Native American and Alaskan Natives who were reincarcerated out of the reincarcerated population is a higher percent than percent of Native American and Alaskan Natives in the general population (Table 12). Moreover, black population comprised a higher percent of the reincarcerated population than they did of the general population. The white population, however, comprised a smaller percent of the reincarcerated population than they did of the general population. Much of this effect is potentially linkable to systematic overpolicing of minority communities. These are important relationships because they will allow for a more concrete analysis of the relationships between different races and COMPAS scores, potentially revealing sources of bias.

Table 12. Reincarceration Rates and Race of Individual Admitted to the WI Prison System

Race	**Percent (of Population)**	**Percent (of Racial Population who were Reincarcerated)**	**Percent (of Reincarcerated Population)**
Black	34.52%	19.64%	40.05%
White	55.73%	14.81%	47.04%
Native American	6.56%	28.125%	10.51%
Asian or Pacific Islander	1.48%	22.63%	1.91%

Figure 4. Average COMPAS Risk Level by Race

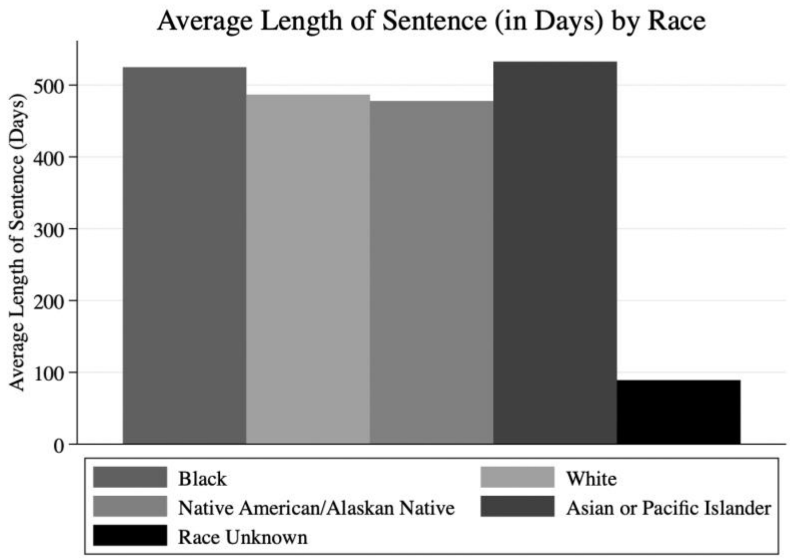


As seen in Figure 4, I then analyzed the average COMPAS risk level at admission by racial group. I found that, on average, individuals who were Native American or Alaskan Native had the highest COMPAS score. It shall be noted that the earlier analysis revealed that, Native American or Alaskan Native individuals were reincarcerated at the highest frequency and represented a larger percent of the reincarcerated population (10.51%) than of the greater population (6.56%) (Table 12). Therefore, the higher COMPAS risk level may accurately reflect reincarceration risk. From Figure 5, it was revealed that, on average, Native American or Alaskan Native individuals had the shortest average sentence length compared to the other racial groups. COMPAS risk level is not necessarily reflected in sentence length, on average, for the Native American or Alaskan Native population. This is likely an instance where judges are deviating from the COMPAS risk level recommendations when sentencing Native American or Alaskan Native populations.

Individuals who self-identified as black had the second highest average COMPAS risk level. They also had the second highest average length of sentence in days (Figure 5). Like the findings from the analysis of Native American and Alaskan Native populations, it was found that black populations comprised a larger percent of the reincarcerated population (40.05%) than of the general population (34.52%) (Table 9). This difference is likely due to systematic over policing of minority communities. Compared to Native American or Alaskan Native and Asian or Pacific Islanders, those who were black were less likely to be reincarcerated than both groups. This, however, is not reflected in the average COMPAS risk level and average length of sentence. Due to algorithmic bias on population level, the COMPAS risk assessment instrument may presume black individuals to have a higher likelihood of reincarceration than appropriate.

Asian or Pacific Islander individuals were found to have the lowest average COMPAS risk level but had the second highest percent of their population who were reincarcerated. Asian or Pacific Islanders represent 1.48% of the general population and 1.91% of the reincarcerated population, but 22.63% of their population was reincarcerated. I believe this could be potentially due to sampling bias due to a small population of individuals who identified under this category. In addition, this racial group also had the longest average sentence length compared to other racial groups.

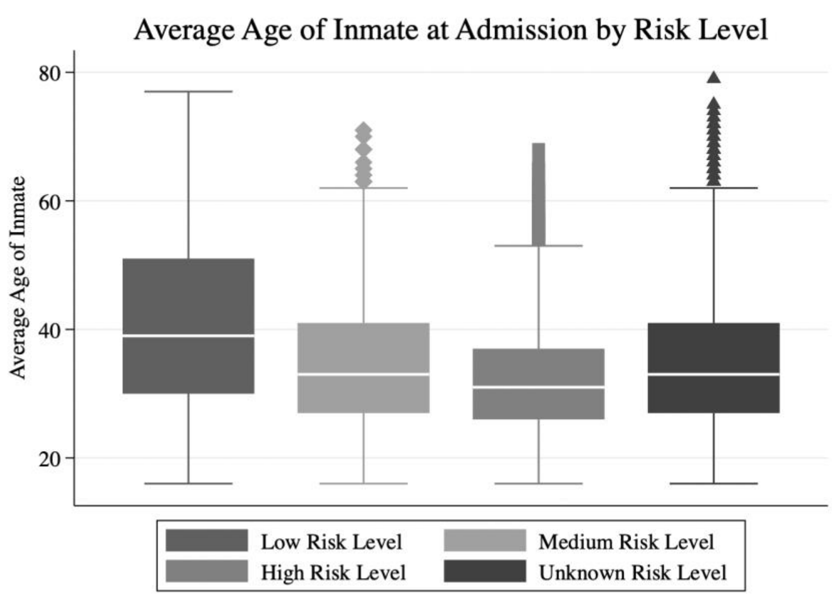
Figure 5. Average Length of Sentence in Days by Race



From my interview with a professor at a law school, it was emphasized that state or federal governments should be involved in the development of risk assessment instruments used in the criminal justice system to mitigate competing interests. Since the development of the algorithm is critical to its efficacy and the potential of bias, I would argue that it is important for governments to have an influence on the development of the algorithms. The law school professor noted that, “if a private company develops an algorithm that is truly good and revolutionary, that raises important moral questions.” In addition, my conversation with the pro-bono lawyer in Wisconsin emphasized the need to maintain a human element throughout the use of risk assessment instrument to ensure that the tools are working the way they should be. They stated, “[the risk assessment instrument] should be used as a tool, not as gospel.” My conversation with an individual at the organization involved in developing the public safety assessment emphasized that “the tools results can be considered but should never dictate a decision.” When there is a risk of overreliance on the tool, education can provide a solution to this issue. It should be noted, however, in my interview with an organization that assisted in the development of the Ohio Risk Assessment System that the state of Ohio worked directly with the organization to supervise the tools' construction. Moreover, to ensure that the tool is implemented correctly, there was a requirement that everyone involved was trained and certified before using the tools, in addition to fidelity monitoring to ensure that the assessments are completed property.

Throughout my interview process, it was emphasized that age presents a critical factor in predicting an individual's recidivism risk. From the Wisconsin DOC data, I found that, on average, risk level decreased as an individual's age increased. Discounting the ages of those who were not given a COMPAS risk level assessment at admission, I found that, those who had an older average age were more likely to be designated low risk compared to those who were younger. Those who were determined to be high risk on admission into the prison system had a younger age on average.

Figure 6. Average Age of Inmate at Admission by Risk Level



Limitations/Future Research

It is critical to emphasize that some of the findings of this paper may be limited by the Wisconsin Department of Corrections Data that was used. The paper tracked the actual and expected sentences, as well as reincarceration, of individuals admitted into the Wisconsin Department of Corrections prison system in 2019. As aforementioned, reincarceration was tracked until June 30, 2023. As a result, some of the data may be limited due to the short time frame, including which the COVID-19 pandemic occurred. Due to the pandemic, there was a changing and unknown landscape that may have impacted sentences, release, and reincarceration. In addition, while this was a robust dataset, it still was limited geographically to Wisconsin and in timeframe. I also believe it is important to highlight that this study does not account for whether the COMPAS algorithm accurately determines low, medium, and high risk in a uniform manner. For example, one wants the difference in risk level between low and medium risk to be consistent. This study does not account for the potential inconsistencies in the difference between low, medium, and high-risk levels.

In the future, it could be critical to explore the variation in COMPAS and other risk assessment tools in predicting individuals' risk over a longer period or throughout the United States. Future analysis could analyze the use of COMPAS or other risk assessment tools in different jurisdictions and over a longer time frame. This would allow for more depth related to recidivism likelihood. Robustness could be added to the study through measures such as implementing the sentence lengths as determined by the judge, rather than employing the time spent in jail. Moreover, it could be productive to compare the efficacy of these tools in providing insights for different steps of the criminal justice process, for example in sentencing compared to pre-trial release. In addition, as some tools are used differently across different states or counties, one could explore the long-term impact of this variation on the tools ability to effectively predict recidivism risk and provide the opportunity to rehabilitate people. It is also important for future research to better differentiate whether variation in COMPAS risk level and sentence length is due to judges accounting for a myriad of other factors or if this is due to mistakes. This could be achieved through a study that focused more heavily on the judges themselves.

Policy Recommendations

Bearing the findings of this paper in mind, I would propose 3 different policy recommendations that intend to provide more equitable outcomes in sentencing with recidivism risk assessment instruments. The recommendations are primarily rooted in quantitative data, with input as well from the interviews I conducted. These recommendations include:

1. **Standardized Training Program:** Implement a comprehensive training for individuals involved in administering or utilizing risk assessment result, focusing on understanding, interpreting, and applying these instruments correctly.
2. **External Validation Studies:** Conduct independent, regular validation studies to compare outcomes with and without the use of these algorithms across diverse demographic groups. This will help to ensure the instruments' effectiveness and fairness in real-world applications.
3. **Judicial Think-Aloud Exercises:** Mandate judges to verbalize their decision-making process during sentencing. This transparency will facilitate a deeper understanding of how COMPAS risk levels and other factors influence sentencing decisions.

As jurisdictions implement recidivism risk assessment instruments into the sentencing process, steps must be taken to ensure that the implementation of the algorithms are effective to minimize potential harm to particular populations. Jurisdictions must make a training program mandatory for anyone who will be either conducting the risk assessment test or using any of the results in a decision-making process. This training program must provide guidance on what risk assessment instruments are, what they seek to do, and how the results should be interpreted. In addition, this training program must establish direction and rules that oversee how the test should be administered and how the results should be used. My findings indicated that, on average, as COMPAS risk level increases, so did an individual's likelihood of reincarceration. While I found the tool to be effective in predicting likelihood to recommit a crime, I believe there is a disconnect between the recidivism risk level and the associated time spent incarcerated. Despite the data being limited to Wisconsin and within a limited time frame, a higher risk level does not correspond to longer sentence length despite a higher likelihood of reoffending. As my quantitative analysis revealed, certain racial groups are more likely to be marked as high-risk on average, but the high-risk level does not correspond to a longer sentence length. This was seen particularly in the groups of Native American/Alaskan Native and Asian/Pacific Islander inmates. In addition, I found a lack of variation in sentence term length across different COMPAS risk levels (Figure 2); a higher COMPAS risk level did not correspond to a longer sentence term length on average.

These findings point to the necessity of effectively implementing and standardizing the use of COMPAS tools in the criminal justice system to ensure that the tools are implemented successfully and consistently in the sentencing process. They must account for variation in risk level by using training programs to establish regulatory standards. Moreover, restrictions must be implemented to prevent COMPAS administrators from overriding decisions. This source of bias was revealed in my conversations with individuals working in algorithm development. If an override occurs, the instance should be flagged, with both the COMPAS algorithm's response and the administrator's response recorded. The override should not go unnoticed, as this presents concerns of bias. I also believe judges should be required to undergo training to better understand the questions asked during the COMPAS risk assessment instrument, as well as training regarding how the score should be used in the sentencing process. During my interview process, it was noted that some judges failed to understand what metrics the COMPAS algorithm used in producing a score. Without a deep understanding of what a COMPAS risk score means, judges may not use the risk score effectively in sentencing. This presents concerns of disproportionately harming minority communities.

Before risk assessment instruments are implemented on the county level in sentencing, there needs to be increased validation studies related to whether the use of recidivism risk assessment instrument provides a "better" outcome compared to when the instrument is not used. This validation study must be performed by a third-party that is independent from the developers of the algorithm and has no stake in the use of the algorithm. From my interviews, I learned that some of the risk assessment instrument developers conduct their own validation studies to see if the tool works. While I think this is a step in the right direction, I believe individual jurisdictions must conduct their own research to ensure the tool will work for their own community. As seen in Table 12, the racial breakdown in the population does not match the population that reoffended within the designated time. Certain racial groups, such as Native American populations, represent a larger percent of the recidivated population than the population. One cannot discount the systemic factors that result in particular racial groups becoming systematically more likely to end up in prison and potentially receive inequitable sentencing outcomes. Due to algorithmic bias, the recidivism risk assessment instruments may categorize individuals as being higher risk due to race. As a result, I believe that policy steps must be taken on the state-wide level to provide oversight of the implementation of recidivism risk-assessment tools in sentencing and ensure that certain factors, such as race, are not overweighted in the tools' algorithm, even implicitly. These policies must require validation studies be conducted on a county-wide or state-wide level before the algorithm is introduced. This will better ensure that the tools being implemented provide the best possible outcomes compared to in the absence of the tools. This will save jurisdictions money, energy, and potential harm.

Furthermore, through the validation studies, the conversation of bias in algorithm must shift from whether the instrument is biased to a conversation of whether the differences in outcomes for different demographics get worse or better based on the use of the tool. Native American/Alaskan Natives had the shortest sentences on average, but the highest COMPAS risk scores and highest reincarceration rates; this study found that, on average, those with the longest sentences were those who were designated to have low likelihood of reoffending. These discrepancies beg the question of if the outcomes of those whose decisions were influenced by the COMPAS risk level received better outcomes than they would without the tool being employed. As aforementioned, jurisdictions must implement validation studies that focus on whether the tools provide worse or better outcomes than in the absence of the tool for different groups and provide sentence outcomes that effectively reflect risk level. In addition, policy must be implemented into regulating how the tool is used in sentencing and restricting the judge's jurisdiction in determining how the COMPAS risk score should play in; regulating the use of these tools could be established through mandatory training that was mentioned before. Across the board, judges must use the COMPAS risk score as a single data point among many and ensure that the score plays a consistent role in their decision-making process. This could be achieved through increased education, laws, and oversight.

Another critical policy recommendation would be requiring judges to dictate their thought process for their sentencing decision through think-aloud exercises. This explanation would indicate with words what factors, such as COMPAS risk level, personal circumstance, etc, played into their decision-making process. This would hold judges more accountable in how they were using a defendant's COMPAS risk level in the decision-making process. Moreover, this would allow for a greater understanding of what additional factors contribute to judge's sentencing decisions. From the quantitative analysis, it was revealed that sentencing decisions did not reflect people's COMPAS risk level and reincarceration risk. As mentioned before, this is either because judges are making mistakes if their single goal is to minimize reincarceration or judges are accounting for more factors than just reincarceration risk when making decisions. Think-aloud exercises would require judges to share their thought processes openly, providing insight into the additional factors that are impacting their sentencing decisions. During my interview with the lawyer working at a pro bono legal clinic in Wisconsin, they highlighted that, with the pandemic, new judges who lacked extensive experience may overly rely on the algorithm in the decision-making process. In addition, older judges who are unfamiliar with the technology may unduly dismiss the algorithms findings. Requiring that judges explain their decision-making process would provide transparency, accountability, and potentially consistency. Moreover, it will provide greater insight into the sentencing decision process. Requiring judges to justify their use of the COMPAS risk score in the sentencing process ensures intentionality in their decision making process. A judge would be less able to overly rely on the algorithm or dismiss its findings in the face of personal bias since they will have to describe orally their thought process. Through oversight, the misuse of the algorithm could be better prevented.

Conclusion

While recidivism risk assessment tools may accurately determine an individual's likelihood of reoffending, this study determines that the implementation of the risk assessment tools falls short. While the algorithm accurately reflected one's risk reincarceration, this was not necessarily reflected in the average length of sentence. The use of the tool may not be providing a better outcome for specific groups than would occur in the absence of the tool due to inconsistent implementation and systematic issues that result in increased imprisonment for particular groups. Moreover, I find that judges are deviating from risk assessment instruments scores either due to mistakes or because they are accounting for factors beyond one's reincarceration risk in their sentencing decisions. The proper implementation of risk assessment tools, continued oversight, and requiring judges to conduct think-aloud exercises while making sentencing decisions may allow for better outcomes than would occur in the absence of the use of the tool, but outcomes will vary depending on individual demographic factors.

While this study works to provide a unique perspective on recidivism risk assessment instruments and sentencing in Wisconsin, I believe there is room for future research. First, researchers need to explore the efficacy of risk assessment instruments in sentencing in more jurisdictions and over longer periods of time as this would allow for a broader perspective to be drawn. Second, more analytic research needs to focus on whether recidivism risk assessment instruments truly result in more fair outcomes than occur in the absence of the instruments across different demographic groups. I believe more research should be conducted to better understand the real-life impact of racial differences that appear in the algorithm. Lastly, a major finding of this research focused on judge's deviating from risk scores during sentencing decisions, either due to mistakes or because they are accounting for more factors beyond recidivism risk. I believe that more research must be conducted to analyze what factors are playing into a judge's sentencing decisions beyond the COMPAS risk score.

References

Blomberg, Thomas, William Bales, Karen Mann, Ryan Meldrum, and Joe Nedelec. 2010. "Validation of the COMPAS Risk Assessment Classification Instrument." Validation Study. Florida State University.

Chatziathanasiou, Konstantin. 2022. "Beware the Lure of Narratives: 'Hungry Judges' Should Not Motivate the Use of 'Artificial Intelligence' in Law." *Germal Law Journal* 23 (4): 452–64.

Chohlas-Wood, Alex. 2020. "Understanding Risk Assessment Instruments in Criminal Justice." Brookings Institute. https://www.brookings.edu/articles/understanding-risk-assessment-instruments-in-criminal-justice/.

Crawford, Kim, and Jason Schultz. 2019. "AI Systems as State Actors." *Columbia Law Review* 119 (7): 1941–72.

Deeks, Ashley. 2019. "The Judicial Demand for Explainable Artificial Intellgience." *Columbia Law Review* 119 (7): 1829–50.

Donohue, Michael E. 2019. "A Replacement for Justita's Scales?: Machine Learning's Role in Sentencing." *Harvard Journal of Law & Technology* 32 (2).

"Economy at a Glace." 2023. United States Bureau of Labor Statistics. September 2023.

Fagan, Frank, and Saul Levmore. 2019. "The Impact of Artificial Intelligence on Rules, Standards, and Judicial Discretion." *Southern California Law Review* 93 (1): 1–36.

Gacutan, Joshua, and Niloufer Selvadurai. 2020. "A Statutory Right to Explation for Decisions Generated Using Artificial Intellgience." *International Journal of Law and Information Technology* 28 (3): 193–216.

Glaze, Kurt, Daniel Ho, Gerald Ray, and Christine Tang. 2022. "Artificial Intelligence for Adjudication: The Social Security Administration and AI Governance." In *The Oxford Handbook of AI Governance*, Online Edition. Oxford Academic.

"Guidelines and Statutes: Risk Assessment." 2020. Pennsylvania Commission on Sentencing. July 2020.

Henman, Paul W. Fay. 2021. "Administrative Justice in a Digital World: Challenges and Solutions." In *The Oxford Handbook of Administrative Justice*, Online, 459–80. Oxford Academic.

"History of Risk Assessment: Public Safety Risk Assessment Clearinghouse." n.d. Government Website. US Department of Justice Bureau of Justice Assitnace. https://bja.ojp.gov/program/psrac/basics/history-risk-assessment#gcov4e.

Hyatt, Jordan, and Jordan Chanenson. 2016. "The Use of Risk Assessment at Sentencing: Implications for Research and Policy." *Villanova University Charles Widger School of Law*, December.

James Austin, Dana Coleman, Johnette Peyton, and Kelly Johnson. 2003. "Reliability and Validity Study of the LSI-R Risk Assessment Instrument." Final. Institute on Crime, Justice, and Corrections: George Washington University.

Jon Kleinberg, Lakkaraju Humabindu, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics* 133 (1): 237–93.

Mazur, Joanna. 2021. "Can Public Access to Documents Support the Transparency of Automated Decision-Making? The European Union Law Perspective." *International Journal of Law and Information Technology* 29 (1): 1–23.

Meserole, Chris. 2022. "The SAIS Review Interviews Dr. Chris Meserole: Artificial Intelligence and the Future of Governance." *The SAIS Review of International Affairs* 42 (2): 113–20.

Nishi, Andrea. 2019. "Privatizing Sentencing: A Deleagation Framework for Recidivism Risk Assessment." *Columbia Law Review* 119 (6): 1671–1710.

Ostrom, Brian, Matthew Kleiman, Fred Chessman II, and Randall Hansen. 2002. "Offender Risk Assessment in Virginia: A Three-Stage Evaluation."

Price, Nicholson. 2018. "Big Data and Black-Box Medical Algorithms." *Science Translational Medicine* 10 (471).

"Quick Facts: United States of America." n.d. United States Census Bureau. United States Census Bureau.

"Quick Facts: Wisconsin." n.d. United States Census Bureau. United States Census Bureau.

Reiling, A. D. 2020. "Courts and Artificial Intelligence." *International Journal for Court Administration* 11 (2): 1–10.

Spaulding, Norman W. 2020. "Is Human Judgement Necessary? Artificial Intelligence, Algorithmic Governance, and the Law." In *The Oxford Handbook Fo Ethics of AI*, Online, 375–402. Oxford Academic.

Starr, Sonja B. 2014. "Evidence-Based Sentencing and the Scientific Rationalization of Discriminiation." *Stanford Law Review* 66 (803): 803–72.

"State of Iowa v. Guise." 2018. Justia. December 14, 2018. https://bja.ojp.gov/program/psrac/basics/history-risk-assessment#gcov4e.

State of Wisconsin, Plaintiff-Respondent v. Eric L. Loomis, Defendant- Appellant. 2016. Supreme Court of Wisconsin.

Sullivan, Yulia, and Samuel Fosso. 2022. "Moral Judgements in the Age of Artifiial Intelligence." *Journal of Business Ethics* 178 (4): 917–43.

Valle-Cruz, David, and Rodrigo Sandoval-Almazán. 2022. "Role and Governance of Artificial Intelligence in the Public Policy Cycle." In *The Oxford Handbook of AI Governance*. Oxford Academic.

"Vulnerable Populations - Prisoners." 2022. Iowa State University. https://compliance.iastate.edu/wp-content/uploads/sites/4/pdf/vulpop-prisoners.pdf.

"What Is Public Safety Assessment?" 2020. Government Website. New Mexico Courts. December 2020.

"What Is Risk Assessment." n.d. US Department of Justice Bureau of Justice Assistance.

Wisconsin Department of Corrections. n.d. "Parole Information: About the Parole Process." Wisconsin Department of Corrections.

Appendix

I coded the variables for my analysis under the following numerical system:

RACE	
Black	0
White	1
Native American/Alaskan Native	2
Asian or Pacific Islander	3
Unknown	4

ETHNICITY	
Not Hispanic or Latino	0
Hispanic or Latino	1
0 or African American	2
American Indian	3
Unknown	4
African	5
Alaska Native	6
Asian Indian	7
Cambodian	8
Central American	9
Chinese	10
European	11
Filipino	12

Great Britain	13
Haitian	14
Hmong	15
Indonesian	16
Japanese	17
Korean	18
Malaysian	19
Mexican	20
Middle Eastern	21
Pacific Islander	22
Native Hawaiian	23
Puerto Rican	24
Spanish Origin	25
Thai	26

MartialStatusAtAdmission

Single, Never Married	0
Married	1
Divorced	2
Separated	3
Unknown	4
Other	5
Common Law Marriage	6
Un1 Partner	7
Widowed	8

COMPAS_RiskLevel_Admissions

Low	1
Medium	2
High	3
Unknown	4

REINCARCERATION

Yes	0
No	1
Still Incarcerated	2
Alternative to Revocation Admission - Excluded from Recidivism Data	3
Unknown	4
Deceased - Unable to recidivate	5

I generated these following variables:
**COMPAS Risk Level at Admissions

```
gen HighRisk= COMPAS_RiskLevel_Admissions==3
gen MediumRisk= COMPAS_RiskLevel_Admissions==2
gen LowRisk= COMPAS_RiskLevel_Admissions==1
**COMPAS risk Level at Most Recent time
gen HighRisk_recent= COMPAS_RiskLevel_Recent==3
gen MediumRisk_recent= COMPAS_RiskLevel_Recent==2
gen LowRisk_recent= COMPAS_RiskLevel_Recent==1
**Reincarcerated Indicator
generate Reincarcerated_Indicator=Reincarceration==0
generate Releasedate = date(ActualReleaseDate, "MDY")
```