Title
PREDICTIVE VALIDITY AND BIAS IN ACTUARIAL RISK ASSESSMENT INSTRUMENTS FOR CRIMINAL JUSTICE SENTENCING

Author
Champion, Christine Jacqueline

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

PREDICTIVE VALIDITY AND BIAS IN ACTUARIAL RISK ASSESSMENT INSTRUMENTS FOR CRIMINAL JUSTICE SENTENCING

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Criminology, Law and Society

by

Christine Jacqueline Champion

Dissertation Committee:
Professor Richard McCleary, Ph.D., Chair
Professor Elliott Currie, Ph.D.
Professor Charles A. Smith, J.D., Ph.D.

2022

DEDICATION

To my teachers, who enabled my love for learning
and inspired me to dream
and
To my students, who remind me the future is now
and we can always do better.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to Richard McCleary, my Ph.D. chair, advisor and mentor. I have learned so much from you about what it means to be a scholar and an educator. You have shown me such kindness and patience during this awkward phase of development.

Thank you to my committee members Elliott Currie and Tony Smith. Your input and guidance was invaluable.

Thank you to my partner, Chris. Your love and support made all of this possible.

Thank you to Brian Williams, the CLS librarian. You made it easy to come to you for help when I needed it. Thank you for putting in the time with the grad students over the years.

Thank you to the other grad students. Your companionship has helped me through tough times. Your scholarship and desire to improve the world inspires me.

VITA

Christine Jacqueline Champion

EDUCATION

Ph.D. Criminology, Law and Society **2022**
University of California, Irvine *Irvine, California*

M.A.S. Criminology, Law and Society **2010**
Law and Society University of California, Irvine *Irvine, California*

B.A. Psychology **2007**
California State University, Fullerton *Fullerton, California*

B.A. Sociology **2007**
California State University, Fullerton *Fullerton, California*

RESEARCH EXPERIENCE

Graduate Student Researcher **2013–2019**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Lecturer in Criminal Justice **2017–2022**
California State University, Fullerton *Fullerton, California*

Graduate Teaching Assistant **2013–2018**
University of California, Irvine *Irvine, California*

Graduate Reader **2016**
University of California, Irvine *Irvine, California*

SELECTED HONORS AND AWARDS

Social Ecology Graduate Mentoring Award **2012-2013**
University of California, Irvine

REFEREED BOOK CHAPTERS

Virtual currency **2018**
The SAGE Encyclopedia of Surveillance, Security, and Privacy

Apple **2018**
The SAGE Encyclopedia of Surveillance, Security, and Privacy

Consumer Fraud **2016**
The Encyclopedia of Crime and Punishment

REFEREED CONFERENCE PUBLICATIONS

False Profits: Examining the Role of Fraud in Financial Crises **Nov 2014**
American Society of Criminology

Greed is Good: The Criminogenic Bourgeois Environment of 21st Century Corporate America **Feb 2014**
Western Society of Criminology

Laissez-fairy tale: Analyzing the role of economic ideology and deregulation in white-collar crime **Nov 2013**
American Society of Criminology

A spatial analysis of increased sentencing in response to highly-publicized crime (with N. Sherman) **Nov 2013**
American Society of Criminology

PROFESSIONAL MEMBERSHIPS

American Society of Criminology
Western Society of Criminology
Association for Criminal Justice Research

ABSTRACT OF THE DISSERTATION

PREDICTIVE VALIDITY AND BIAS
IN ACTUARIAL RISK ASSESSMENT INSTRUMENTS
FOR CRIMINAL JUSTICE SENTENCING

By

Christine Jacqueline Champion

Doctor of Philosophy in Criminology, Law and Society

University of California, Irvine, 2022

Professor Richard McCleary, Ph.D., Chair

Recidivism, or the subsequent commission of a criminal offense after receiving punishment in the justice system, is a primary concern for public safety decision-makers. The quantification of an individual offender's likelihood of recidivism has been a focus of criminology for over 100 years and methods of analysis continue to develop alongside technological innovations in statistical computing. At present, machine learning approaches are being applied to criminal justice data sets in order to generate algorithms that offer predictions of recidivism by specific, individual offenders. Such algorithms underlie the risk assessment instruments that are presently used by decision-makers during the sentencing phase. Although these machine learning approaches are cutting-edge techniques of exploring patterns in data, certain strategies (especially so-called "black box algorithms") may not be appropriate for use in criminal justice applications due to their enigmatic nature conflicting with the principle of open justice. Furthermore, the underlying data used may be such a flawed reflection of recidivism that the algorithms produced by this process inevitably generate predictions which reflect and perpetuate systemic bias in the justice system.

This dissertation examines the predictive validity of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) recidivism tool, which is widely-used in the

United States at the present time. A parsimonious actuarial model (or PAM) is generated by the author, entertaining a number of state-of-the-art approaches to machine learning for its development, and employed as an alternative white-box algorithm for comparison of predictive validity. Results indicate a trivial difference in predictive validity between existing tools and open approaches. As such, a simplified and transparent model is both philosophically and practically preferable for the purpose of criminal justice applications.

Chapter 1

Introduction

The prediction of future recidivism is one of the most important functions of statistical analysis in modern correctional practice. Recidivism, or the tendency of a convicted person to re-offend, is an undesirable outcome in the criminal justice system. Ideally we don't want many "repeat customers" in our jails and prisons. Rather, one trip through the gears of the United States criminal justice system should be satisfactory for treating criminality and restoring a pro-social nature. Furthermore, a properly functioning correctional system should not create habitual offenders or otherwise induce habitual offending. An offender should exit the carceral system as a reformed person who is less likely to engage in future law-breaking.

Understanding the nature of recidivism is therefore crucial for an effective and efficient system of justice. From a financial standpoint, the cost to incarcerate a single inmate is nontrivial. For example, in 2021 the state of California paid on average a little more than \$100,000 per inmate in carceral costs for its state prisons. To this end, criminologists and statisticians study offenders in great detail and gather a wealth of information for analysis. Indeed, offenders in the United States criminal justice system are some of the best documented individuals in our society. Getting facts and figures about people in jail or prison can be almost trivial depending on the geographic location of the correctional system. For example, the state of Florida is widely known for its "Government in the Sunshine Law" which provides the public with rights to access state and local government proceedings – including arrest and jail records – with few restrictions.

Prior to the age of the Internet and statistical computing, it would be almost unthinkable for a single researcher to amass huge quantities of information on offenders, let alone engage in complex statistical modeling. Consider the impressive feats of Sheldon and Eleanor Gleuck in their data gathering on delinquent boys in the 1930s. Sociologists and criminologists often speak in reverence of this labor-intensive work by the Gluecks, recognizing the complexity of gathering longitudinal data on 1,000 individuals, painstakingly matching delinquent boys with non-delinquent boys for the sake of analysis. In the present day, it is much easier for us to compile large data sets with many individual observations and many variable of potential explanatory interest.

However, the explanatory value of a larger data set is innately tied to our technical ability to conduct analysis. Modern computing systems and statistical software packages can churn and burn through data at blazing speeds, but understanding findings and making appropriate use of them has become challenging. The primary issue at present is appropriately harnessing the power of big data in the age of machine learning. We want an understanding of how to classify similar offenders and how those similarities may indicate a specific treatment in the correctional stage. We want to separate out the offenders who have the best chances of reform from those who will be inveterate recidivists. How do we develop the best tools for predicting recidivism, given machine learning algorithms in present use? How well do current predictive instruments perform in the real world? Furthermore, can we improve upon the predictive validity of risk assessment instruments – which necessarily rely on an underlying algorithm – that are currently in use?

In this chapter, we will introduce the historical use of recidivism prediction instruments in the United States criminal justice system. Predictive approaches to quantifying relative recidivism risk have been in use during the sentencing and corrections phases since the 1920s. Initially there was interest in quantifying the relative recidivism risk posed by potential parolees if released back into the community. This conceptual approach was then applied to earlier phases within the correctional process, such as assessing the individual offender during the sentencing phase for their security and supervision needs while incarcerated. Over the last 100 years, computing and statistical technology has developed to the point where we are now able to develop sophisticated predictive algorithms to quantify recidivism risk. However, the question remains whether such "black-box" algorithms truly hold a place in a transparent legal system. We will begin our inquiry by examining the history of prediction in sentencing and corrections and discussing what precisely is meant by these scholars in terms of recidivism.

1.1 History of Prediction in Sentencing and Corrections

The history of prediction in sentencing and corrections has its roots firmly planted in the scientific study of human behavior. Conceptually, behavioral sciences have four main sequential goals of describing, explaining, predicting, and controlling behavior. In order to attempt controlling (or changing) human behavior, we must therefore first develop our ability to predict that behavior. Social scientists are particularly interested in mechanisms to control or mitigate behaviors that present public risk, such as criminal activity. At the forefront of these efforts is scientific methods of predicting recidivism, the violations of law occurring subsequent to the initial offense or "relapse" into criminal behavior. Such scientific approaches to predicting recidivism provide the criminal justice system with an evidence-based structure for addressing unlawful behavior, thereby reducing the potential for harm to society at large.

The scientific prediction of recidivism is instrumental in supporting the established goals of sentencing and corrections. The American criminal justice system is generally agreed to have four major goals in its correctional practice: retribution, deterrence, incapacitation, and rehabilitation. These goals are often in competition with one another for public support and consequently have waxed and waned in their prominence at various times. Despite this shifting landscape of conceptual priorities, the desired outcome of corrections is always to reduce unwanted criminal behavior, whatever that behavior may be. Accordingly, the aim of the scientific prediction of recidivism is to provide decision-making individuals in the criminal justice system (e.g. judges, parole boards) with relevant information that may enhance their ability to make consistent, effective judgements.

1.1.1 Prediction of Recidivism and Parole

The first meaningful attempts to gather relevant data and engage in scientific prediction of recidivism began in the 1920s and focused on offenders who were already incarcerated and potentially eligible for parole. Warner (1923) first presented statistics on recidivism and violation rates for parolees in his detailed report on the Massachusetts Reformatory, providing correlations between numerous types of information available to the Board of Parole (e.g. demographic information, criminal history, etc.), and success or failure while on parole. Warner's interpretation of this analysis was that no improvements to the parole criteria could be established on the basis of data available to the Board at the time (1923)Warner (1923).

Hart (1923) offered an opposing viewpoint, attributing Warner's interpretation to his "failure to apply accurate statistical tests to determine which of the factors involved showed significant contrasts" (see also Borden (1928)). By teasing out these factors and combining them into a "prognostic score", the Board of Parole could benefit from past experience, minimize parole violations, and promote parole for "as large a fraction as possible of the men who will succeed" (Hart (1923)). Using "the same scientific procedure employed by insurance companies" (i.e. an actuarial approach to risk), Hart suggests, makes it possible to prepare reports on individual offenders for parole boards forecasting "the probability that the man, if paroled, would violate his parole" (1923)Hart (1923). Indeed, the first wave of parole prediction instruments consisted of reporting the average violation rate in a series of "experience tables", each containing the calculated failure rates of parolees based a specific risk factor such as prior criminal record, and associated weights for predictors of recidivism (see Hart (1923); Burgess (1928); Glueck and Glueck (1930); Ohlin (1951)). The "Original Burgess Experience Table" was one of the initial attempts to build actuarial tables for recidivism risk, collecting data on parole candidates and using this information to predict which candidates would ultimately be successful when released on parole (Schuessler (1955)). This "experience table" used norm-referenced scores on 21 different variables (either present or absent) to provide an estimate of recidivism risk (Gottfredson and Gottfredson (1988)). Similar experience tables documenting parole recidivism rates were developed into the 1930s and 1940s, paving the way for formalized risk assessment instruments.

The first generation (1G) of offender risk assessment was used during the first half of the twentieth century and was administered by clinical professionals such as psychologists, psychiatrists, or social workers (Bonta and Andrews (2007)). Assessment of offender risk was also conducted by law enforcement professionals such as probation officers or prison staff who often had no mental health expertise themselves. These "experts" would use a combination of their own experience in the field and formalized professional training to make a determination about the degree of security and supervision necessary for each offender (Bonta and Andrews (2007)). These judgements of relative risk were a product of "subjective assessment, professional judgment, intuition, and gut-feelings" (Bonta et al. (1996), p. 16).

This is to say there was no structured instrument being used or formal guidelines to follow during this period of time.

The primary reason for a shift to actuarial, evidence-based approaches to evaluating risk was a need to improve the accuracy of individual assessment beyond professional heuristics or intuition. Rather, risk assessment should be rooted in scientific knowledge and these instruments should be subjected to continual assessment themselves, to improve reliability and validity of measures (Kemshall (2003)). Today we know that these first generation approaches to risk assessment provide little predictive accuracy beyond flipping a coin (i.e. random chance). The idea that using a structured risk assessment instrument significantly improves predictive accuracy has been demonstrated by many independent analyses (see Andrews et al. (2006); Grove et al. (2000)).

As such, several research teams have recommended that clinical professionals and criminal justice professionals alike should use a structured assessment instrument when gauging an offender's risk of recidivism (see Bonta and Andrews (2007); Desmarais and Singh (2013)). Latessa and Lowenkamp (2006) suggest that community-based correctional agencies use evidence-based practices such as the use of actuarial risk assessment instruments, individual tailoring of interventions, and implementing an ongoing measure of recidivism risk. Presently, there is an expectation that actuarial risk assessment tools will be employed as part of conventional criminal justice risk analysis (Skeem and Monahan (2011)). This expectation is also the case during pre-trial decision-making, such as whether a defendant will be released from custody during the pre-trial phase of their criminal case. Both the American Bar Association and the National Association of Pretrial Services have recommended the use of standardized risk assessment instruments during the pre-trial phase of criminal cases (Blomberg et al. (2010)).

Moving into the 1960s, there was an enhanced focus on the need to develop risk assessment tools that rely less on professional judgement and more upon evidence-based, statistical analysis. During this period of time, actual risk assessment instruments began to take shape, notably including the consideration of static (and primarily historical) individual traits or characteristics that have been demonstrated to increase recidivism risk. For example, an actuarial tool may consider the offender's history of substance abuse, a characteristic demonstrated to have a statistically significant relationship to increased risk of recidivism, and assign a numeric value for this particular variable. By assigning these traits or characteristics a quantitative score – such as a score of one if a risk factor is present and a score of zero if the same risk factor were to be absent – this allowed the professional to assign an overall risk score to the individual offender (Bonta and Andrews (2007)). Notably, this generation of risk assessment does not consider that various risk factors should be weighted differently based on their relative importance (Kemshall (2003)).

This second wave of recidivism risk assessment instruments heavily emphasized the use of more intricate statistical methods. Multiple regression and logistic regression were used to promote a sophisticated analysis of available data. One of the earliest second-generation (2G) risk assessment scales produced during this time was Gottfredson and Bonds' Base Expectancy Scale (1961)Gottfredson and Bonds (1961). Another example of a second-generation risk assessment scale is the Salient Factor Score, developed in the United States for use by the U.S. Parole Commission (Hoffman and Beck (1974)). The Salient Factor Score assesses evidence-based covariates of recidivism, including history of antisocial behavior and substance abuse problems, to calculate a total risk score for recidivism (Hoffman, 1996). The Salient Factor Score has remained a continuous feature of parole decision-making by the US Department of Justice since the 1970s, making it the oldest actuarial instrument still in use today (Maxfield et al. (2005)). As such, it is also one of the most well-studied actuarial risk assessment instruments. Maxfield et al. (2005)Maxfield et al. (2005) have determined that the Salient Factor Score maintains a respectable predictive accuracy (AUC =.73). Additionally, there is a similar approach to actuarial prediction of recidivism in the Offender Group Reconviction Scale (OGRS) currently used in adult correctional settings in

England and Wales (Copas and Marshall (1998)). The OGRS is now in its third iteration and though reliant entirely on the scoring of six simple demographic and criminal background factors, it demonstrates strong predictive validity characteristics (AUC =.80 when coded from centralized records) (Howard (2017)).

Although empirical studies of these statistical prediction systems revealed improved predictive efficiency when compared with clinical judgements, accuracy rates still rarely exceeded 70 percent (see Sawyer (1966); Simon (1971); Kerr (1971)). Actuarial risk assessments were nonetheless considered superior to professional judgement (even in cases where there was joint use of professional judgment and actuarial risk assessment) and so "more and more correctional jurisdictions adopted this type of assessment" (Bonta and Andrews (2007)). Common models developed at the time shared a susceptibility for significant shrinkage (Tarling and Perry (1985)). A frequent challenge in prediction studies, shrinkage is the tendency of regression predictors to have a worse fit with new data as compared with the original data (Copas (1983)). In spite of these methodological issues, there was a surge in complex statistical analyses as technological advances in computing paved the way for analysts to conduct computational statistics.

In the last 30 years, the general statistical approach to predicting recidivism among parolees has remained much the same. Researchers have instead largely turned their focus to revalidating existing offender risk instruments and improving the quality of the underlying data (Bonta (1996)). Now in the era of Big Data, criminologists and statisticians can apply adaptive machine learning procedures to data collected on potential recidivists – arguably the most well documented subpopulation in our society today – as it requires only "a conventional menu of predictors and a large enough sample to exploit them" (Berk and Bleich (2013)). The development of third and fourth generation risk instruments has more directly targeted a risk-need approach that begins application well before parole, during the sentencing and correctional intake process (Bonta and Andrews (2007)).

1.1.2 Prediction of Recidivism at Sentencing

Although there is a fairly long historical record of predicting recidivism risk for the parolee population ("back-end decision making"), the transition to predicting risk of recidivism at sentencing ("front-end decision making") is a relatively new development. This is in part due to a philosophical shift regarding the goals of corrections, now viewing sentences as essentially providing a term of intervention for offenders – incorporating rehabilitation as opposed to pure incapacitation or retribution. Third generation risk instruments like the Level of Service Inventory - Revised "were referred to as 'risk-need' instruments" and considered previously-excluded dynamic risk factors such as substance abuse or employment (Bonta and Andrews (2007)). Specifically, these risk-need instruments "were sensitive to changes in an offender's circumstances and also provided correctional staff with information as to what needs should be targeted in their interventions" (Bonta and Andrews (2007)). Theoretically, if these dynamic risks could be successfully addressed, the subsequent risk of recidivism by the offender could be reduced (Bonta (2002)).

Modern fourth generation risk instruments, such as the Level of Service/Case Management Inventory (LS/CMI) and the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), are distinct in that they are "designed to follow the offender from intake through case closure" with systematic monitoring and interventions (Fass et al. (2008)). COMPAS provides far more than a single risk score, but rather an Overall Risk Potential (including separate estimates for risk of violence, recidivism, failure to appear, and community failure) in addition to a Criminogenic and Needs Profile for the individual offender Brennan and Oliver (2000)). The Criminogenic and Needs Profile provides the offender's unique "criminal history, needs assessment, criminal attitudes, social environment, and additional factors such as socialization failure, criminal opportunity, criminal personality, and social support" (Brennan and Oliver (2000)).

Risk assessment is used in modern sentencing for three primary reasons: (1) to inform decisions regarding the imprisonment of higher-risk offenders, (2) to inform decisions regarding the supervised release of lower-risk offenders, and (3) to inform decisions designed to reduce offender risk (Monahan and Skeem (2015)). For higher-risk offenders, tools like COMPAS claim to provide an empirical estimate of recidivism risk, thereby creating a more "objective" basis for ordering incarceration. Likewise, lower-risk offenders can more easily be identified as prospects for alternative sanctions (e.g. probation, diversion to community-based drug treatment). Reducing offender risk can take the form of institutional programming, seeking to influence the aforementioned dynamic risk factors.

A key concern about these usages is that when informing sentencing decisions, risk assessment tools like COMPAS "may exacerbate unwarranted and unjust disparities" due to static factors beyond the offender's control such as their socioeconomic background (Holder (2014)). The major issue for those opposing risk assessments' use in sentencing is that risk factors such as criminal history and education may be utilized as proxies for poverty or race (Harcourt (2015); see also Starr (2014)). This sentiment is echoed in of Wisconsin v. Loomis (2016), implying that risk assessments with excessive reliance on static (versus dynamic) factors at sentencing has a disproportionate impact on black offenders, who are essentially being punished for crimes they have not yet committed. This important issue is addressed directly in this dissertation.

1.2 On Measurement and the Ethics of Forecasting Recidivism

1.2.1 Measurement of Recidivism

It should be noted that understanding what is meant by the term "recidivism" is, in and of itself, a challenging exercise. Recidivism can be described in a general sense as "the reversion of an individual to criminal behavior after he or she has been convicted of a prior offense, sentenced, and (presumably) convicted" (Maltz (1984)). Although the idea that an offender may go on to commit another offense in the future is fairly straightforward, it becomes increasingly complicated to then operationalize and (hopefully) measure its occurrence. Operationalizing recidivism may mean defining it as occurring at the time of the offense, at the time the offender has been arrested again, at the time the offender is referred back to the courts, when the offender has been reconvicted, or when the offender is placed back into the correctional environment. Each of these definitions has a different implication for measuring recidivism.

In a grand sense, it would be ideal for criminologists to know precisely in all cases when offenses have actually taken place and to whom those offenses can be truthfully attributed; however, this is not really feasible without omniscience. This of course speaks to a larger issue of measurement for the discipline of criminology known as the dark figure of crime, which broadly refers to crimes which have not (yet) been discovered or reported to authorities. Not all types of offenses suffer equally from this phenomenon, as we have learned from comparing official measures of crime (e.g. UCR, NIBRS) with measures of victimization (e.g. NCVS). For example, we have much more precise official measures of motor vehicle theft than we do of illegal drug consumption or prostitution. Before moving on to more realistic expectations of crime measurement, we must first appreciate that (1) there will always be some amount of crime that goes undiscovered/unreported and (2) the largest gap between what crimes we know happened and what crimes actually happened is in measurements of less serious and "victimless" crimes .

Another related epistemological issue here is the assumption of correctly-assigned culpability. If we are to use subsequent arrests as indicators of recidivism, there is an underlying assumption that if the offender is arrested again, it is because they are in fact guilty of violating the law for which they have been arrested. This assumption seems to be conceptually at odds with the court's presumption of innocence until one is proven guilty. To some extent, this potentially wrongful assumption of guilt, and therefore recidivism, carries on through subsequent measures (referred to court, reconviction, and incarceration) but a discussion concerning wrongful conviction is beyond the scope of this proposal. However, using reconviction as a measurement of recidivism may be more useful because at that point in the process, the offender has been determined by the court to be guilty and therefore legally responsible for the offense.

1.2.2 The Ethics of Forecasting Recidivism

There is a growing literature regarding the role of ethics in the statistical forecasting of recidivism. As touched upon previously, critics of actuarial risk assessment measures have expressed concern about the degree to which they rely upon "static" risk factors such as the offender's race or their parents' criminal histories (see Holder (2014); Harcourt (2015); Starr (2014); Ritter (2013)). Such an approach to forecasting recidivism is deeply rooted in the notion of the past as prologue and perhaps there is some truth to that notion. However, a prediction model based solely on static – and therefore, immutable – risk factors cannot entertain the possibility of the offender "changing for the better" and consequently diminishing their risk of recidivism, only that their risk of offending remains the same or increases over time (Bonta and Andrews (2007)). Forecasting that relied primarily on static factors were common in the second generation of risk assessment instruments but third and fourth generation instruments emphasize the inclusion of dynamic risk factors (Bonta and Andrews (2007)).

There are also ethical considerations in the use of actuarial risk assessment tools when they are used to justify giving offenders longer sentences based on their predicted risk scores. This is to say that a judge may sentence an offender deemed as “high risk” to a longer prison term in order to “prevent crimes that they might commit if they were not incarcerated” (Ritter (2013)). The perspective that forecasting recidivism for the purposes of sentencing allows for offenders to be punished now for crimes yet to occur has drawn comparison by critics to Philip K. Dick’s short story Minority Report. On this point it should be noted that the COMPAS risk score reports provided in pre-sentence investigation reports begin with the following:

For the purposes of Evidence Based Sentencing, actuarial assessment tools are especially relevant to: 1. Identify offenders who should be targeted for interventions. 2. Identify dynamic risk factors to target with conditions of supervision. 3. It is very important to remember that risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated.

The Wisconsin Supreme Court’s opinion in the Loomis appeal specifically underscored the importance of uniformly attaching this notice to risk assessment reports used by the court in sentencing (2016)of Wisconsin v. Loomis (2016).

Chapter 2

Legal Foundations of Risk Assessment Instrument Use

"Your scientists were so preoccupied with
whether or not they could,
they didn't stop to think if they should."
Dr. Ian Malcolm, Jurassic Park.

2.1 Introduction

In this chapter, we will discuss the use of risk assessment tools in criminal justice proceedings and touch on several legal questions that are arising in response to the newest generation of these tools. The practice of predicting future criminal events is well established within criminology and has many applications in public safety. Criminologists may be asked to predict which neighborhoods are most likely to experience burglaries during the day, so policing strategies can concentrate on those "hot-spots". We may ask criminologists or statisticians to predict how many inmates will need to be housed by the state prison system in the year 2025, helping us forecast the budgetary needs involved in providing this correctional service.

One of the most common applications of prediction is estimating the likelihood that a particular offender will recidiviate. Recidivism, or the tendency of previously-convicted persons to reoffend, is a primary concern to the correctional system and often used as a measure of effectiveness. If our justice system's correctional and rehabilitational efforts have been successful, the offender should be less likely to recidivate in the future. After all, we are not looking to create "repeat customers" in the correctional system. Since the 1990s, we have been primarily utilizing the risk-needs-responsivity (RNR) model of assessment and treatment. These tools identify specific recidivism risk factors for the individual offender that we can then attempt to address while someone is in the correctional phase. Some examples of treatable risk factors are substance addiction or an absence of educational/vocational training. We can provide evidence-based resources to the individual offender to address those needs and ideally the offender will be less likely to engage in recidivism.

A high-quality, detailed data set is extremely useful to these number-crunching criminologists. Fundamentally they are trying to solve the challenge of predicting a human being's likely future behavior. Having the best information possible, with the highest number of potentially relevant variables, can make all the difference in producing a useful predictive model. Criminologists often find themselves with access to a tremendous (sometimes unwieldy) amount of data because offenders are arguably the most well-documented members of our society. In the age of big data, statistical computing is king and algorithms are developed for every monetizable purpose. Video streaming services predict the show you'd like to binge-watch next. Marketing algorithms determine which ads are likely to loosen your purse-strings. Machine learning and artificial intelligence has found a home in the criminal justice system, just as it has virtually everywhere else in our society. However, just because we have access to a great quantity of data does not necessarily mean it is high-quality or without biases that may influence our predictive models.

2.2 What type of risk assessment instruments are used in criminal justice?

Actuarial risk assessment instruments may be employed at one or several stages of the criminal justice system, such as during the compilation of a pre-sentence investigation report, upon an inmate's arrival at a carceral institution, or in advance of a parole hearing. Classifying offenders is an ongoing exercise throughout the justice system and the history of statistical approaches to classification date back at least 100 years. However, the use of these actuarial risk assessment instruments is not without controversy among legal scholars, statisticians, and criminologists (see Starr (2014), Wisser (2019)).

Early forms of the statistical approach to estimating recidivism risk can be traced back to the efforts of scholars like Hart and Burgess in the 1920s, as well as the Gleucks in the 1930s. A more detailed discussion of these scholars' work and their methodological approach to developing actuarial models for predicting recidivism risk can be located in Section 1.1.1. The actuarial risk tools being used at this period were developed for the specific purpose of classifying potential parolees according to their relative risk of recidivism upon release back into the community. The use of actuarial prediction tools was pioneered largely by the state of Illinois from the 1930s through the 1950s, subsequently followed by the state of Ohio in the 1960s and the state of California in the early 1970s. By 2004, approximately 72% of states that maintained an active parole system utilized risk-assessment tools as a component of their decision-making process (see Harcourt (2007)). By this period of time, the actuarial approach had become the dominant and preferred standard of determination in state-level parole decisions.

In more recent years, actuarial risk assessment instruments have also gained traction as one element of the compilation of information provided to the judge during the sentencing phase in criminal cases. For example, in many states such as Wisconsin or Florida a risk assessment and classification report is a standard component of a pre-sentence investigation report. The defendant may be assessed for their relative risk of general recidivism as well as violent recidivism when compared with their peers. These risk scores are then provided to the judge in the form of bar charts with values from 1 to 10 and labeled from low to high, ostensibly to add data-driven credibility to any legal decisions being made. This risk report comes coupled with a boilerplate disclaimer, cautioning judges against relying solely upon this measurement to make their decisions. Risk assessment instruments thus have the false appearance of providing truly objective quantification of risk posed by offenders.

Recidivism risk assessment, which has traditionally been used in parole settings, has increasingly been employed in earlier stages of the criminal justice system. We may want to know how likely it is that an individual offender will recidivate *while they are incarcerated*, quantifying the risk presented to other inmates and to the correctional officers and staff. Perhaps we are entertaining whether to allow a person to serve their sentence in a community-based setting and want to know how likely they are to victimize a member of the public. Estimating post-carceral risk is one matter, but providing input to the carceral process is another matter entirely. Further, exposure to incarceration itself may inherently increase risk of recidivism, placing a high cost on overclassification for risk. In this way, recidivism risk assessment used inappropriately at early stages in the justice system may generate self-fulfilling prophecies. Inmates with a higher recidivism risk score are more likely to be placed behind bars, which in turn impacts the likelihood of recidivism. Getting sentenced to "con college" instead of community-based sanctions can have a negative impact on the individual offender's odds of recidivism.

Risk-needs-responsivity (RNR) is another form of assessment taking place in many jails and prisons. These tools are seeking to identify underlying criminogenic needs in the offender, or dynamic risk factors that may be possible to alter through therapeutic means. If an offender is a thief who steals because they are addicted to methamphetamine, then we may want to focus on dealing with the methamphetamine addiction (a treatable condition) during their incarceration. By treating the underlying criminogenic issue of methamphetamine addiction that theoretically is also driving the theft behavior, we may in turn reduce the odds of theft committed by this same individual. RNR can be thought of as a diagnostic tool for sub-types of offending behavior and may be used within a carceral setting to guide the prescription of offender treatment, based on what has historically been successful in similar cases.

2.3 So what's the problem?

There has been a great deal of criticism levied at those attempting to inject artificial intelligence measures like predictive algorithms into the US criminal justice system. These critiques come from a plethora of sources, from criminologists to lawyers to AI programming experts and other advocates of justice. Consequentially their concerns are quite diverse, running the gamut from the practical to the theoretical. In this section, we will discuss the most prominent of these criticisms: that the newest generation of these predictive algorithms lack adequate transparency, that the underlying data used to create the algorithm is innately biased, that algorithms are have an outsized influence on judicial decisions, and that the use of these algorithms reinforces existing systemic biases especially against women, people of color, and the poor.

2.3.1 Transparency

Transparency is a central issue in any democratic system of justice, promoting accountability among the many actors involved in this process. If the actions of a justice system or its representatives are shrouded in any way, this can undermine the public's perception of its legitimacy. How can the public be sure that judges are being fair rather than relying upon stereotype in their decision-making? How else can we be certain that we are being treated equally in the eyes of the law, regardless of our race, color, or creed as constitutionally guaranteed? The US justice system has a pronounced emphasis on transparency, with its open courts and published judicial opinions made publicly available. Even the Justices of the Supreme Court must explain their decisions in cases by publishing majority, concurrent, and dissenting opinions replete with citations of similar cases and relevant statutes.

Predictive algorithms, newcomers on the justice scene, may vary widely in terms of transparency. Some methods of statistical prediction are very transparent and provide a clear explanation of the relationship between the offender's characteristics and their resultant risk score. Other methods represent the polar opposite in terms of transparency – there is no way to observe "how the sausage is made" but there is more accuracy in how offenders are classified. At present time, we must make a choice between methods that provide maximal transparency and methods that provide increased accuracy of predictions. The appropriate degree of transparency for predictive algorithms has not been formally settled in law as of yet.

Furthermore, critics point out that if we are unable to understand how a predictive model functions then we are also unable to challenge the fairness of these predictions. Is this merely the modern incarnation of spectral evidence being used against the accused, under the guise of statistical computing? It is challenging to demonstrate bias in an algorithm when not even the designer themselves may understand fully what it does. It seems profoundly counter-intuitive to introduce ambiguity into the process of dispensing justice by using these black-box algorithms, even in a small part of the judgement.

Procedural Fairness

Maintaining the people's belief in the justice system is absolutely crucial for an enduring legal system. In the absence of such belief, people may become detached from social bonds that promote law-abiding behavior. They may even seek to dismantle the current justice system in order to rebuild it as something more representative of the public's desires. In Tom Tyler's classic work "Why People Obey the Law" (2006), he concludes in his analysis that people obey the law when it is viewed as a legitimate form of authority. This legitimacy, in turn, does not stem from fear of punishment or personal agreement with legal outcomes but rather from a belief in the overarching fairness of the legal process.

The perception of procedural justice arises from four cardinal principles: fairness of the process, transparency of actions, opportunity for advocacy, and impartiality of decisions. People generally want to understand what is going on throughout the criminal justice process and to this extent, it is important to make clear how and why decisions are being made. Every person wants to be treated fairly in the eyes of the law but it may be practically impossible to determine if that is actually happening when certain types of predictive models are being used. The details of how a predictive tool calculates risk scores may be considered a trade secret when it has been developed by a private corporation. How can a defendant realistically challenge the wisdom of the Colonel's Secret Recipe? How can you advocate for yourself against an ambiguously-defined threat? It is nigh impossible to challenge a process that itself is not completely understood.

2.3.2 Validity of Underlying Data

There are large-scale ongoing efforts to collect data on crime such as the FBI's annual Uniform Crime Report (UCR) – representing crimes known to the police – and the Bureau of Justice Statistics' annual National Crime Victimization Survey (NCVS) – gauging the public's experience with crime, reported or not. Although these measures are useful for estimating crime rates, they are not perfect representations of the crime that happens in reality. There are still crimes that go unreported to authorities and remain undiscovered by researchers. Criminologists refer to this phenomenon as the "dark figure of crime" - criminal offenses that occur but never make it into crime statistics. There are some experts who spend their time focused on this issue entirely. For example, criminologists may be interested in determining the true rate of sexual assault, a crime that is estimated in the NCVS to occur significantly more frequently than indicated by the UCR. Some crimes are better represented than others within this dark figure of crime, just as some neighborhoods are better represented as places where crimes are more unlikely to be reported to police officials. As a result, there is some variability to the reliability and validity of these widely-used official statistics.

There is an old maxim in programming – "garbage in, garbage out". If the information you are using to design a model is an accurate reflection of reality, the resulting model's predictions will be improved. Bias in the data, on the other hand, can diminish a model's predictive power. If we are using data on crimes known to the police, we must appreciate that not all crimes and criminals are equally known. A model based on this information may therefore not be valid for predicting *who will commit a crime* but rather *who will be caught committing a crime* or perhaps even *who will be reported to the police.*

2.3.3 Algorithmic Bias

Critics such as Starr and Harcourt argue that far from being objective, these algorithmic approaches to predicting recidivism reinforce existing inequalities in the criminal justice system. The concept of algorithmic bias is not unique to criminal justice applications of statistics and machine learning, but rather a ubiquitous challenge to all fields employing such approaches at this time. A parallel issue in emerging technology is that many current facial-recognition algorithms – including those presently in use by law enforcement in the United States – are more likely to fail their tasks when attempting to identify or match a black subject as opposed to a white subject. Consequentially, the odds are higher for a black person to be erroneously identified by such facial-recognition algorithms as a suspect with an open arrest warrant, for example.

Although these algorithms appear to be objective, they are innately driven by the analysis of underlying data and specified variables. Algorithms that are intended to accurately predict risk of recidivism will thus in the course of their design take into account information regarding the offender. Pieces of information such as the offender's gender will be used by the algorithm in order to calculate an individual risk score. In a "black-box" algorithm, it may be difficult or impossible to determine precisely which pieces of information are used by the algorithm or how that information is being used. For example, it may not be possible to determine whether or not an offender's race is taken into account during model specification if race data is provided in the dataset.

Furthermore, Harcourt (2015) and other legal scholars have argued that despite race not being explicitly included within a model, there are other variables in use that serve as effective proxies for race. That is to say that although a particular algorithm may not specifically include the variable race in its calculations, there is another variable such as prior criminal history being used that will reliably stand in the place of race. Harcourt argues that "the use of actuarial tools is likely to produce a 'ratchet effect' on all members of the higher-risk categories – whether along racial or other lines – with highly detrimental consequences on their employment, educational, familiar, and social outcomes" (Harcourt (2015), pp. 237).

2.3.4 Reinforcing Inequality

Some critics of risk assessment instruments believe that these tools promote and perpetuate disparate effects on different subgroups of offenders. These effects are particularly pronounced among specific subgroups such as women, people of color, the poor, and any combinations thereof. This is a troubling reflection of a tendency we have seen with other applications of artificial intelligence. For example, as previously mentioned there are many facial recognition systems presently in wide use for security and retail applications. These systems are more frequently deployed by retail chains in communities of color, despite the fact that people of color are more likely to be incorrectly flagged by the algorithm. This tendency may be a byproduct of the training data sets used to produce the algorithm, which typically over-represent white male faces and as a result give us a much better idea of how to classify white males than any other group. Facial recognition also more frequently fails to correctly classify individuals with unconventional or nontraditional gender expressions, reinforcing gender stereotypes.

Some of these issues may be addressed by faithfully recalibrating the algorithm using data that more closely matches the characteristics of the intended population. It is good practice with predictive algorithms to periodically maintain and revalidate its assumptions over time. Assuming that the sample used to generate the algorithm is representative of the target population, we may want to ensure that the phenomenon itself or the variables used in prediction are not changing substantially over time. However these maintenance measures will not address or eliminate the effects of an underlying systemic bias that disproportionately impacts women, people of color, and the poor.

The reinforcement of systemic inequalities arising from risk assessment may take place directly or indirectly. The direct impact of inequality on the offender, especially in the overprediction of recidivism, is that they will receive harsher treatment including jail or prison time. Exposure to incarceration may itself increase the odds of recidivism. Becoming incarcerated is often a tipping point for offenders, severing their existing pro-social bonds in the outside world. The social phenomenon of institutionalization may have a dramatic effect on the offender's behavior, leading to higher odds of recidivism upon release. People may lose gainful employment when they are incarcerated, or may break ties with friends and family. The people who are most likely to encounter overestimation of their recidivism risk are also the same people who are least capable of affording attentive legal counsel.

There are also indirect effects of overprediction that lead to increased rates of incarceration. We know for example that the children of incarcerated people may be significantly impacted by their experiences with the justice system. The functional loss of a parent is traumatic in and of itself, but we may expect these children to encounter more behavioral issues, problems in school, and a higher likelihood of future criminality. There are a number of programs designed to lessen the impact of a parent's incarceration on the child. The best strategy, of course, would be to only incarcerate the child's parent if it is necessary to do so for public safety.

2.4 Constitutional Considerations

Certainly the use of evidence-based practices that inform justice system decision-making, such as risk-needs assessment, is a noble reinvestment of human knowledge into sentencing and corrections. To the average bureaucrat, statistically validated tools lend an implicit sense of legitimacy and uniformity that is easy to accept on its face as "fairness" or "justice". Indeed, these instruments may contribute to more consistent judgements rather than those determined by the subjective application of legal heuristics or an individual's clinical assessment. However, these efforts at incorporating risk-needs tools into criminal justice-oriented decisions face increasing skepticism from both scholars and practitioners in the field (see Tonry (2014); Hamilton (2015)). As a result, there are a number of constitutional issues that have been raised about risk-needs assessments, especially in regard to their use of particular variables that qualify for heightened analysis when used in a legal setting. The most important issues, discussed herein, involve the constitutional guarantees to equal protection and prisoners' right to due process.

2.4.1 Equal Protection

Equal protection for similarly situated persons is the primary constitutional consideration in the use of risk-needs assessment tools during sentencing and corrections. The Equal Protection Clause within the Fourteenth Amendment to the United States Constitution states "nor deny to any person within its jurisdiction the equal protection of the laws" (U.S. Const. amend. XIV). Although the passage of the Thirteenth Amendment in 1865 had abolished slavery (except as punishment for a crime), the adoption of Black Codes by former Confederate states necessitated the passage of the Civil Rights Bill of 1866 to guarantee that newly freed slaves born in the United States would have "full and equal benefit of all laws and proceedings for the security of person and property, as is enjoyed by white citizens" (14 Stat. 27-30). Although distinctions between groups based on protected criteria are proscribed, the Supreme Court has also noted that "most legislation classifies for one purpose or another, with resulting disadvantage to various groups or persons" (Romer v. Evans, 1996). Therefore, it may be unclear whether disparate treatment between groups in a specific case should be considered unconstitutional. When considering the constitutionality of statutes or ordinances that explicitly call for disparate treatment, the Supreme Court indicates three levels of judicial review: rational basis review, heightened review, and strict scrutiny. Each level will be addressed in turn, as well as examples of the relevant variables for recidivism risk assessment.

Rational Basis Review

Rational basis review is applicable to the majority of equal protection claims heard before the Supreme Court and represents the least amount of scrutiny. Rational basis is the lowest level of judicial review and this test provides that the challenged policy will remain in effect if (1) it can be demonstrated to serve a legitimate public purpose and (2) making classifications is "reasonable in light of its purpose" (McLaughlin v. Florida, 1964). The government does not need to provide proof that the judgements being made are correct in such cases. If there is any other reasonable evidence in the case, its existence introduces an element of ambiguity that often prevents claimants from prevailing. Additionally, the court allows officials the discretion to make reasonable judgements for "practical considerations based on experience". To date, policies that classify on the basis of age, economic status, personality type, mental illness, mental disability, physical disability, and those differentiating treatment of drug users or alcoholics have fallen under rational basis review (see also Hamilton (2015)).

Identifying the purposes for which risk-need tools have been implemented would be a fundamental question in a rational basis review of criminal justice policy. The inclusion of risk-needs assessment has proven practical during various decision-making stages in the criminal justice system such as at sentencing, when identifying individual programming needs, or when potentially granting parole. Above and beyond these direct needs, officials may also use these classifications as a mechanism for addressing security within their facility, providing a rehabilitative opportunity to the individual prisoner, or to assure public safety. The Supreme Court has ruled that legitimate objectives such as ensuring security within the institution allow for the classification (and thus, disparate treatment) of pre-adjudicated individuals as well.

Assuming that the government can demonstrate a compelling state interest, such as "preventing crime", the legitimate purpose component of the rational basis test will be met. The second component of the rational basis test is evaluating whether risk-needs tools are being used in a way that is rationally related to said legitimate purposes. Although risk-needs tools are now being used widely in jails and prisons around the country, there is little case law to date that challenges their use on the basis of equal protection. A direct claim was made in 2013 in the case of People v. Osman regarding the use of the Static-99R sexual recidivism risk tool, wherein the claimant argued against the use of cohabitation history as a variable employed in calculating his risk score. The Static-99R instrument assigns different point values for individuals who previously cohabited with a partner versus those who had not, indicating a higher risk for the individuals with no previous cohabitation experience. This was raised as an issue at appeal because "his counsel argued that [the] defendant was a very devout follower of Islam and his religion and culture prohibit cohabitation prior to marriage" (People v. Osman, 2013, at 3).

The court determined that the use of cohabitation as a factor to distinguish between two similarly situated classes of sex offenders does not infringe on the individual offender's fundamental right to religious freedom. Rather, the court recognized that "studies have indicated that cohabitation is a relevant factor, and thus the second risk factor furthers the legitimate state purpose of predicting the potential for recidivism by convicted sex offenders" (ibid, at 10). As cohabitation history did not have a bearing on any protected classes, the court upheld its use at the time on the basis of "legitimate interests in predicting the potential for recidivism and protecting the public" (Hamilton (2015)). Based on the current state of case law, it seems likely for the time being that the use of risk-needs assessment and the bulk of factors used in determining scores will continue to survive an equal protection challenge if subjected to the rational basis test.

Additionally, some legal scholars have offered arguments supporting the elevation of socioeconomic status to heightened review or strict scrutiny. Starr (2014) maintains that although wealth is not explicitly considered to be a suspect class, “the Court has often used very strong language concerning the importance of eradicating wealth-related disparities in criminal justice”. Starr cites the Supreme Court’s broad language in Griffin v. Illinois and Bearden v. Georgia as indicating support for this position. In the case of Griffin v. Illinois, the Court ultimately struck down the state’s policy mandating payment of court costs by defendants seeking a transcript of their trial as it precluded the defendant from filing an appeal. In the Court’s opinion, they state directly that “in criminal trials, a State can no more discriminate on account of poverty than on account of religion, race, or color”.

In response to Starr, Hamilton (2015) has argued that “these two decisions do not appear adequate to sustain a broader claim that socioeconomic status can virtually never be included in a classification-oriented decision in criminal justice”. Hamilton points to rulings by judges in other cases which “clarified that the Bearden ruling merely meant that probation cannot be revoked solely because of inability to pay”. Furthermore, the Bearden opinion itself states that “such considerations [as defendant’s employment history and financial resources] are a necessary part of evaluating the entire background of the defendant in order to tailor an appropriate sentence for the defendant and crime”, but cannot be the sole reason for revoking probation. With regard to Griffin v. Illinois, Hamilton points out that the ruling “required indigency plus a complete deprivation of a right”. Lower courts abiding by these requirements consequently “have since held that wealth classifications do not qualify for heightened review and indigiency is not itself a suspect class” Hamilton (2015)). As studies have linked wealth-related factors to the risk of recidivism, Hamilton argues that equal protection law may offer an insufficient basis for eliminating wealth classification from future risk-needs assessment instruments used in the criminal justice system.

Intermediate Scrutiny

The next level of judicial review beyond the rational basis test is referred to as intermediate scrutiny or heightened review. The concept of "levels of judicial scrutiny" was first referenced in a footnote in United States v. Carolene Products (1938), one of several cases testing the constitutionality of the New Deal. Pettinga (1987) refers to intermediate scrutiny as "rational basis with bite" to describe its escalated status above the rational basis test. Equal protection challenges to risk assessment tools may be more successful though to date, only policies that classify on the basis of gender or illegitimacy have been subject to intermediate scrutiny.

Intermediate scrutiny involves a two-pronged test that requires challenged laws to (1) further an important government interest and (2) do so using means that are substantially related to that interest. The Supreme Court first created and applied the intermediate scrutiny test in Craig v. Boren (1976), a case involving an Oklahoman statute regulating the sale of "nonintoxicating" 3.2% beer such that it could only be purchased by women over the age of 18 or men over the age of 21. Notably, the argument by lower courts that the available data on drunk driving (indicating that young men are far more likely than young women to be arrested for drunk driving) was dismissed by Justice Brennan in the opinion as "far too tenuous to satisfy Reed's requirement that the gender-based difference be substantially related to achievement of the statutory objective" (at 204). This logistic tenuousness extended to the way in which the underlying data was being applied to the policy at hand. Although the gathered data attempted to quantify the correlation existing between gender and driving while under the influence, this statistical information is not put to use. The legislation at issue is not written to address driving under the influence, but rather seeking to regulate the sale of nonintoxicating beer and separating classes on the basis of gender. Simply put, there is a significant difference between buying nonintoxicating beer, consuming nonintoxicating beer, and driving under the influence of nonintoxicating beer.

Additionally, Justice Brennan remarks that the use of data in this case "this merely illustrates that proving broad sociological propositions by statistics is a dubious business, and one that inevitably is in tension with the normative philosophy that underlies the Equal Protection Clause" (at 204). Some legal scholars view this statement as indicating that the Equal Protection Clause requires individual assessments rather than assessments containing statistical generalizations (Starr (2014)). However, the Supreme Court has explicitly approved of group-based statistics being applied to decisions regarding individuals. In Schlesinger v. Ballard (1975), the Court allowed for disparate treatment between male and female officers serving in the United States Navy on the basis of "the demonstrable fact that male and female line officers in the Navy are not similarly situated with respect to opportunities for professional service". In Michael M. v. Superior Court (1981), the Court upheld a California statuatory rape law that applied only to males, stating that "young men and young women are not similarly situated with respect to the problems and the risks of sexual intercourse".

Labeling an entire group based on group-level empirical variations has also previously occurred with respect to the use of the death penalty, albeit regarding classification based on age rather than gender. In Roper v. Simmons (2005), the Court cites a number of broad findings from sociological studies on adolescents in the majority opinion and compares the cognitive and moral development of adolescents to developmentally delayed or disabled adults, and therefore less culpable for their crimes than normally-functioning adults (see also Ford v. Wainwright, 1986 and Thompson v. Oklahoma, 1988). Notably, Justice O'Connor rejects such comparisons and writes in her dissenting opinion to Roper that "the Court's analysis is premised on differences in the aggregate between juveniles and adults, which frequently do not hold true when comparing individuals" (at 601, emphasis in original). Despite this salient issue, the Supreme Court "has not banned the use of group-based statistics in equal protection analysis, nor has it required that the government treat each individual as a wholly unique case, even in criminal justice decisions" (Hamilton (2015)).

Prior to the Craig finding, there had been some recognition in case law that classification upon the basis of gender must substantially further important governmental objectives (see Reed v. Reed, 1971 and Frontiero v. Richardson, 1973). An assumption is made by some legal scholars that gender would be considered an acceptable risk factor due to ample historical and current statistical evidence of lower rates of violent crime committed by women (Monahan (2006)). Although there is no doubt that gender would survive the rational basis test as it relates to most policies concerning the assessment of violent risk, it is less clear whether gender would uniformly withstand intermediate scrutiny as a rule. Legal arguments against the use of gender often refer to the 1974 finding of United States v. Maples, where the court states that "absent any proof that rehabilitation or deterrence are more easily accomplished in the case of females rather than males, we deem the factor of sex an impermissible one to justify a disparity in sentences" (at II, 2).

Criminological research conducted since the Maples finding has demonstrated evidence of gender-based differences, including a greater deterrent effect of imprisonment versus community-based sanctions on women (see Cobbina et al. (2012); Mears et al. (2012); Durose et al. (2014); Langan and Levin (2002)). Collins (2010) also indicates a gender-based difference in sentence length as a predictor for violent recidivism, with longer sentences being a negative predictor for male offenders but a positive predictor for female offenders. As the empirical evidence of a gender-based difference continues to accumulate, legal scholars argue that female offenders' increased risk of violent victimization or medical, physical, and mental problems suggest that gender is highly relevant to risk-needs assessment aimed at promoting rehabilitation and reducing recidivism (see Wright et al. (2012)).

Intermediate scrutiny has also been applied in cases concerning classification on the basis of illegitimacy (i.e. having been born out of wedlock). In Levy v. Louisiana (1968), the Supreme Court struck down a Louisiana civil code that denied illegitimate children the right to recovery on "based on morals and general welfare because it discourages bringing children into the world out of wedlock" (at 69). Justice Douglas wrote in the majority opinion that the code in question constituted "invidious discrimination contravening the Equal Protection Clause of the Fourteenth Amendment, since legitimacy or illegitimacy of birth has no relation to the nature of the wrong allegedly inflicted on the mother" (at 70). In Trimble v. Gordon (1977), the Court revisited the standing of illegitimacy when reviewing a section of the Illinois Probate Act that allowed illegitimate children to inherit by intestate succession from their mothers only, despite legitimate children being allowed to inherit by intestate succession from both parents. Ultimately striking down the Act, Justice Powell wrote in the majority opinion that "traditional equal protection analysis asks whether this statutory differentiation on the basis of illegitimacy is justified by the promotion of recognized state objectives... [i]f the law cannot be sustained on this analysis, it is not clear how it can be saved by the absence of an insurmountable barrier to inheritance under other and hypothetical circumstances" (at 774).

Strict Scrutiny

Strict scrutiny is the highest standard of judicial review used by court in the United States. Courts have historically applied strict scrutiny to classifications made on the basis of race or ethnicity and on the basis of alienage. In equal protection analyses of such policies, there is an important legal distinction between a policy that is facially discriminatory and a policy that is facially neutral yet disparately impacts a protected group (Hamilton (2015)). The difference between the two in a practical sense involves whether the court must determine the purpose of the classification as intended by government officials. If the classification made is explicit, there is no requirement for an inquiry into the government's intent to discriminate (see Hunt v. Cromartie, 1999). However, when the classification is made as part of a facially neutral policy, strict scrutiny may be applied "only if the claimant can prove that the policy was motivated by a discriminatory purpose or object, or if it is unexplainable on any other grounds" (Hamilton (2015)). In Personnel Administrator of Massachusetts v. Feeney (1979), Justice Stewart wrote that invidious discrimination, and thus a violation of equal protection, has occurred only when a public official has "selected or reaffirmed a particular course of action at least in part 'because of,' not merely 'in spite of,' its adverse effects upon an identifiable group" (at 278). Policies subject to strict scrutiny must be narrowly tailored to achieve a compelling government purpose and use the least restrictive means possible to achieve that purpose.

Although none of the commonly-used risk-assessment tools today explicitly use race or ethnicity within their scored variables, some scholars conclude their use would ultimately be upheld as constitutional if they were among various other factors in calculating a risk score (Tonry (2014)). Tonry argues that the Supreme Court's approval use of race and ethnicity by criminal justice officials as a profiling factor has rendered constitutional law "toothless", as evidenced in decisions such as United States v. Brignoni-Ponce (1975) and McCleskey v. Kemp (1987). In McCleskey v. Kemp, the petitioners presented the Court with the Baldus study, a statistical analysis of over 2,000 murder cases from 1970's Georgia that demonstrated black defendants who kill white victims are the most likely to receive the death penalty. Although there was a statistically significant difference between the experiences of black and white defendants, Justice Powell writes "the likelihood of racial prejudice allegedly shown by the study does not constitute the constitutional measure of an unacceptable risk of racial prejudice" (at 308). While the Baldus study may have demonstrated a pattern in the aggregate, the Court's opinion was that its data was not sufficient for proving a disparate impact on individual defendants.

If there was no evidence of a statistically significant correlation between race/ethnicity and outcome measures of interest, there would be no practical or constitutional need to continue including these variables during risk assessment. This is to say, without evidence of a correlation between race or ethnicity and recidivism (or rehabilitation), there is no compelling government interest for race or ethnicity to be included in the risk assessment tool. However, previous empirical studies have indicated differences on the basis of race or ethnicity in observed recidivism rates and rehabilitation outcomes (see Beck and Shipley (1989); Lin et al. (2010); Spiropoulos et al. (2014)). A recent meta-analysis by Piquero et al. (2015) revealed a strong correlation between violent recidivism and age, sex, and race, further suggesting the implicit value of including these variables in predictive models. Additionally, there is evidence to indicate a relationship between race/ethnicity and criminogenic needs (Olusanya and Cancino (2012)). Olusanya and Cancino (2012) argue that "the deleterious social conditions (e.g., concentrated disadvantage and racial stratification) for Blacks (when compared to non-Blacks) severely undermine levels of social capital needed to overcome criminal tendencies, consequences of criminal conviction, and re-entry/reintegration into the community". A failure to consider these real social conditions will inherently hinder the predictive value of a risk-assessment tool concerned with recidivism or rehabilitation. These findings suggest that statistical models that include race or ethnicity may offer improved predictive validity as compared with models that do not include these variables.

Oleson (2011) has suggested that Supreme Court precedent would allow for the explicit use of race or ethnicity within a risk-needs assessment tool, provided that the policy in question is otherwise constructed to conform with the strict scrutiny standard. In Regents of the University of California v. Bakke (1978), the Supreme Court allowed the consideration of race/ethnicity alongside other factors considered in the college admissions process while also noting "racial and ethnic classifications of any sort are inherently suspect and call for the most exacting judicial scrutiny" (at 281). This issue is addressed again in Grutter v. Bollinger (2003), where Justice O'Connor notes that "narrow tailoring does not require exhaustion of every conceivable race-neutral alternative" (at 339). O'Connor cites Bakke in the majority decision, highlighting Justice Powell's position that "truly individualized consideration demands that race be used in a flexible, nonmechanical way" (at 334). The Court has also previously allowed for race or ethnicity to be considered explicitly in determining prison cell assignments for individual offenders in Johnson v. California (2005). Justice O'Connor writes in the majority opinion that prisons are a context in which the presence of danger "may justify racial classifications", but that strict scrutiny is the appropriate level of review for the explicit use of race as opposed to the relatively relaxed standard set in Turner v. Safley (1987) applied by lower courts.

Hessick and Hessick (2011) point out that courts have repeatedly allowed "a number of constitutionally doubtful sentencing factors" including race and ethnicity but rather than conduct a normal constitutional analysis of these dubious factors, "courts have swept constitutional concerns under the proverbial rug based on the ungrounded conclusion that the sentencing process is somehow unique and thus shielded from constitutional review" (at 57).

Due Process

-legal requirement that the state must respect all legal rights

-substantive due process and procedural due process

-5th & 14th amendments

2.5 What should we do?

2.5.1 Improving Data Fidelity

An obvious place to start is to improve the quality of the data that is being used to generate these statistical models. The best way to promote high-quality predictions is to use high-quality data in model creation and periodically reassess the model to ensure it remains a statistically valid way to classify people. The issue of data fidelity is applicable to risk assessment alone. Any application of criminal justice data, such as in tracking crime rates over time or analyzing the impact of public policy, can be impacted by inferior-quality data.

As a scientific discipline, improving our methodology for data-gathering is an ongoing focus within criminology. We want to work with data that is the truest possible reflection of reality. Human beings are often frustrating and uncooperative research subjects, even under friendlier circumstances than being caught in the gears of the United States criminal justice system. Reactivity, social desirability, and confirmation biases plague social science research. Human beings are also fallible, which can be a threat to the accuracy of officially recorded data. Research participants may want to be helpful in what they (sometimes inaccurately) perceive to be the goal of analysis and alter their behavior accordingly. Other participants may be reluctant to trust researchers fully, withholding information that would otherwise be revealing. High-fidelity data on human beings is a precious jewel obtained with agonizing persistence and care.

The ever-looming issue of the dark figure of crime is more complicated to address. Fundamentally, we must act to remove existing obstacles to the reporting of crime to police. The 2018 Police-Public Contact Survey reveals that there are significant differences in resident-initiated contact with police dependant on the resident's race or ethnicity, their gender, and their household income. White residents were 50% more likely to initiate contact with the police than any residents of any other racial/ethnic identity. Wealthy residents were more likely to initiate contact with police than any other income group. Female residents were more likely than male residents to contact the police. These differences in outreach to law enforcement may partially explain the dark figure of crime, that there are certain sub-populations that police are interacting with on a less frequent basis.

As a consequence, we have a bit of a blind spot in our data and this obscures our ability to understand how much crime is truly happening and to whom. For some circumstances, we have a pretty complete and clear view but in other situations, we encounter more cloudiness and noise in the data. When using the Uniform Crime Report's official figures of crimes known to the police, we have to appreciate this underlying issue in communities is a very real obstacle to high-quality data gathering. Criminologists and other similar experts may understand these characteristics of data but statisticians and algorithm developers from other fields may not immediately realize this. Any social scientist can tell you, human beings are trickier subjects to investigate than other natural phenomena like chemical elements. A chemical reaction will not hide itself out of shame or out of the fear of authorities. A human being certainly might.

There is room in criminology for new surveys that might help us supplement our knowledge of crime. The National Crime Victimization Survey, in use since 1973, has already given us great insight into the prevalence of unreported crimes and provides better estimates for the unspoken needs of crime victims in our communities. Other methods like self-report surveys have yielded precious additional information about the nature of the dark figure of crime from the perspective of the offender. Self-report surveys are similar to the NCVS in that we are able to discover more details about both criminal events that *were not* reported to the police and criminal events that *were* reported to police.

2.5.2 Revisiting Use In Criminal Justice Proceedings

The use of black-box algorithms in the US justice system should immediately end. The inability to produce a straightforward interpretation of these models and how the underlying variables are being used to generate algorithms is troubling to say the least. In a way, black-box algorithms are like a statistical version of the infamous Star Chamber. What happens behind these closed doors is a mystery and perhaps grossly unfair, though it is difficult to know precisely what is going on. This quality of mystery should immediately put black-box algorithms at odds with the principle of open justice, a legal principle we have enthusiastically adopted as part of the British common law tradition.

Jeremy Bentham, the quirky but revered English philosopher and jurist, once opined that open justice is "keenest spur to exertion and the surest of all guards against improbity". We hold our justice system and its actors to a higher standard of transparency in order to limit self-dealing, dishonesty, and other forms of corruption. It is important to maintain transparency in our criminal justice system, especially as it promotes the public's faith in procedural fairness. If we are unable to fully understand how or why a particular risk assessment tool works, or how an individual person's characteristics are directly translated into a specific risk score, then we are not capable of being sufficiently transparent. We should be able to specify in detail *why* a person has been assigned a particular recidivism risk score and that person should have a realistic opportunity to confront this judgement of their future risk.

Note that this suggestion of ending black-box algorithm use in the criminal justice system does not extend to other forms of data-driven prediction. There are methods of conducting machine learning and designing algorithms that are significantly easier to understand and explain. Some of these white-box model development approaches have been in use for decades, like logistic regression or decision trees. It is important to understand that there is a trade-off that exists between interpretability and accuracy in predictive models. Although white-box algorithms may have lower accuracy in their predictions, we can at the very least say how they work.

If the person designing the white-box algorithm is familiar with the phenomenon of interest (like criminal offending), they may already have a good idea of which variables will be helpful in a predictive model and what types of relationships should exist among the variables in the data set. White-box algorithms also provide a better opportunity for testing the foundational propositions of established theories in criminology, giving us a better understanding of how and why crime occurs in our society. If it is not possible to avoid the use of black-box algorithms in criminal justice, we must resolve to use the most transparent methods possible for this application.

It is also important that predictive algorithms be tuned using data that closely resembles the characteristics found in the population of interest. This may mean tailoring the functioning of the algorithm based on demographic characteristics of the jurisdiction, for example. Particular attention should be paid to the over-prediction and under-prediction of recidivism for specific subgroups. This is especially true for subgroups which may be members of legally-protected classes.

2.5.3 Judges and Algorithms: Proceed with Caution

One of the most important things we can do to combat improper use of algorithms and machine learning in the criminal justice system is to recognize the need for subject-matter expertise. These tools and their output reports are deceptively simple. When recidivism risk scores are provided to a judge, those scores have the appearance of precisely quantifying an individual offender's riskiness. There is something undeniably comforting about having a number to cling to when making such serious decisions. However, it is the rare judge who has the statistical expertise required for understanding how these risk assessment instruments fundamentally operate and the drawbacks inherent in these tools. Judges are smart, well-educated people but their experience and knowledge lies elsewhere.

There may be boilerplate disclaimers attached to these risk reports, warning judges of the limitations of prediction, but human beings are known to be susceptible to *algorithmic appreciation* – a greater reliance on algorithmic advice than other sources of non-algorithmic information. Researchers have found that this algorithmic appreciation is especially evident when the tasks being performed become more difficult. It only makes sense, therefore, that human judges may be susceptible to the same algorithmic appreciation as they engage in the difficult task of facilitating justice.

One solution to this issue is to ensure that judges have advanced training in statistics and a healthy knowledge of machine learning. This is highly impractical as a solution because both statistics and machine learning are vibrant fields of study. Expertise in artificial intelligence is a moving target. Someone who is well-versed in existing machine learning approaches may be out of their depth with newer, cutting-edge techniques. It is simply unreasonable to ask judges to cultivate such depth of knowledge on this topic, akin to asking judges to be intimately familiar with the pros and cons of different surgical interventions for a medical condition. These machine learning techniques and their practitioners are highly specialized for this purpose, and simply keeping up with developments in this rapidly-developing field is a full-time occupation.

Chapter 3

Causes of Recidivism

3.1 Introduction

Criminologists often use the term "recidivism" to broadly describe a relapse into criminal offending after a prior conviction. Although US prison populations have continued to decline from the peak of mass incarceration, there were approximately 1,471,200 people in state and federal prisons at the end of 2018 (Kang-Brown et al. (2019)). Sobol et al. (2007) indicate that approximately 700,000 of these offenders are released from federal and state prisons back into their communities each year. A significant proportion of these offenders will return to the criminal justice system within 2-3 years following their initial arrest. For example, the Pew Center on the States estimates that in the United States over 40% of former prisoners are re-incarcerated in state prisons within 3 years of their release from initial imprisonment (2011). Recidivism risk is often the outcome of interest when criminologists and other statisticians attempt to predict future criminality. In this chapter, we will explore recidivism and the causes of this phenomenon as presently understood by the social sciences.

This includes an examination of actuarial factors as well as psychological factors that are used to generate predicted recidivism rates.

It should be noted that defining precisely what is meant by the term "recidivism" is, in and of itself, a challenging exercise. Recidivism can be described in a general sense as "the reversion of an individual to criminal behavior after he or she has been convicted of a prior offense, sentenced, and (presumably) convicted" (Maltz (1984)). Although the idea that an offender may go on to commit another offense in the future is fairly straightforward, specifying when this event takes place becomes an increasingly complicated epistemological issue. Operationalizing recidivism may mean defining it as (1) occurring at the time of the offense, (2) at the time the offender has been arrested again, (3) at the time the offender is referred back to the courts, (4) when the offender has been reconvicted, or (5) when the offender is placed back into the correctional environment. Each of these definitions has a different underlying implication for measuring recidivism. Thus, how we define recidivism will shape which predictive factors become relevant to our model.

Further, it would be ideal in a grand sense for criminologists to know precisely when offenses have actually occurred and to whom those offenses can be legitimately attributed; however, this is not really feasible in a practical sense. This issue of course relates to a larger issue of measurement for the discipline of criminology known as the dark figure of crime, which broadly refers to crimes which have not (yet) been discovered or reported to authorities. Some estimates suggest as much as 60% of serious offenses are never reported to law enforcement. Not all types of offenses suffer equally from this phenomenon, as we have learned from comparing official measures of crime (e.g. UCR, NIBRS) with measures of victimization (e.g. NCVS). For example, we have much more precise official measures of motor vehicle theft than we do of illegal drug consumption or prostitution. Before moving on to more realistic expectations of crime measurement, we must first appreciate that (1) there will always be some amount of crime that goes undiscovered/unreported and (2) the largest gap between crimes known to the police and all crimes to occur is in measurements of less serious and "victimless" crimes.

Additionally, a related epistemological issue here is the assumption of correctly-assigned culpability. If we are to use subsequent arrests as indicators of recidivism, there is an underlying assumption that if the offender is arrested again, it is because they are in fact guilty of violating the law for which they have been arrested. This assumption seems to be conceptually at odds with the court's presumption of innocence until one is proven guilty. To some extent, this potentially wrongful assumption of guilt, and therefore recidivism, carries on through subsequent measures (referred to court, reconviction, and incarceration). A discussion concerning wrongful conviction is beyond the scope of this chapter, though certainly worth having. However, using reconviction as a measurement of recidivism may be more useful because at that point in the process, the offender has been determined by the court to be guilty and therefore legally responsible for the offense.

The measurement and prediction of recidivism is a chief concern in evaluating the effectiveness of correctional punishment. Recidivism itself is a somewhat common phenomenon in the United States. The Pew Center on the States (2011) estimates that as many as 40% of former state prisoners will return to the justice system within 3 years of their release. Durose et al. (2014) place their recidivism estimates as high as 67% within 3 years and 76% within 5 years of inmate release from state prisons. This relatively high rate of failure to adequately prevent future offending comes at an enormous social (and economic) cost to taxpayers. Schmitt et al. (2010) indicate that more than $75 billion is spent each year on federal, state, and local level correctional budgets, despite the fact that the rates of recidivism have remained somewhat stable over the past decade. Such high rates of recidivism also demonstrate there is ample room for improvement in the correctional system's ability to rehabilitate and reintegrate offenders back into larger society.

With these considerations in mind, let us discuss the differences between predictors and causes where they concern recidivism.

3.2 Predictors vs. Causes

A layman may believe that when a variable predicts an outcome, this is equivalent to demonstrating that the outcome is caused by that particular variable. In well-constructed models, predictors are related to the outcome variable through some causal mechanism but they do not need to be causes. This is to say that certain predictors have a correlation with recidivism, but do not have a direct impact on recidivism. These variables predict through some causal mechanism and thus only have an indirect relationship with recidivism. Though correlation is not the same as causation, there is no correlation without causation.

Consider the classic example of the correlation between homicide rates and the sales of ice cream, a familiar staple of statistics courses everywhere. Suppose we observe that when ice cream sales increase, so do homicide rates. We may hypothesize that there is a causal relationship between ice cream sales and homicide rates – that purchasing ice cream is a direct cause of being murdered or committing murder. If we had a data set that included weekly ice cream sales and the corresponding weekly homicide figures, we might attempt to construct a regression model where we predict future homicide rates given a certain number of ice cream sales.

However, the sale of ice cream itself does not cause homicide to occur. Rather, there is a spurious relationship between these variables, as both the sale of ice cream and homicide rates are significantly correlated with temperature. This is to say that a rise in temperature "causes" both a rise in ice cream sales and a rise in homicide rates. Using our regression model, we could predict the estimated weekly homicide figures given a certain amount of ice cream sales. Nonetheless it would be incorrect to claim that ice cream sales are a causal risk factor in homicide.

Additionally, causation requires a specified time ordering of events. This is to say that for A to cause B, A must occur before B in time. Returning to our ice cream-homicide example, if purchasing ice cream is a direct cause of being murdered, the purchase of the ice cream must therefore occur before the murder. An alternative explanation to the relationship between these variables is that people purchase ice cream in response to rising homicide rates. From this perspective of time-order, the murder occurs first and causes people to purchase ice cream. So if we are examining variables that predict recidivism, we may need to consider whether a risk factor is a cause or a consequence of recidivism.

This is an important distinction between predictors and causes because we may seek to translate our knowledge of recidivism into public policy. This can take the form of a structured approach to risk assessment or by designing evidence-based treatments to reduce recidivism risk. Monahan and Skeem (2014) describe a typology for these predictive risk factors, breaking them down into four categories. All types of risk factors are applicable to risk assessment, while causal risk factors are the sole focus of risk reduction. Fixed markers are risk factors that are unchangeable such as race. Variable markers such as age may change over time but are unchangeable by intervention. Variable risk factors such as employment status can be changed through intervention. Causal risk factors such as substance abuse make up a subset of variable risk factors, in that they are both changeable by intervention and have a significant impact on the risk of recidivism.

3.3 Actuarial Predictors

Consulting professionals are often employed to describe and predict human behaviors, such as the likelihood and severity of recidivism. For example, a parole board seeks to quantify the relative likelihood that a particular inmate would recidivate if he were to be allowed early release into the community, an important factor in making their determination. It is important that any consulting professionals demonstrate accuracy in their judgements in order to plan the most optimal course of treatment for the subject or to achieve public safety goals. Formation of this professional judgement may stem from either the clinical method or the actuarial method, two contrasting decision-making approaches.

The clinical method is a subjective approach, wherein judgement is based on information processed in the head of a qualified human "judge". In contrast, the actuarial method entails an objective approach based on empirically established relations between historical data and the event of interest (in this particular case, recidivism). With respect to accuracy, empirical research has repeatedly demonstrated that the actuarial method is superior to clinical judgement (Dawes et al. (1989)). Specifically, structured actuarial instruments can predict the probability of recidivism with greater accuracy and inter-rater reliability when compared to unstructured or subjective clinical approaches to risk assessment (Desmarais and Singh (2013); Mamalian (2011)). Consequently, risk assessment tools have continued to evolve over time to replace clinical judgements with more formal, evidence-based tools to estimate the likelihood of recidivism.

Initial attempts at predicting recidivism risk relied exclusively on subjective assessments such as professional judgement and discretion, or the "clinical approach". A well-documented absence of validity plaguing these measures necessitated a standardized approach to criminal prediction (see Meehl (1954); Little and Shneidman (1959); Clear and Gallagher (1985)). Since this time, actuarial predictors have been used by criminologists in statistical approaches to classify recidivism risk. The variables that are the best predictors of recidivism are those that best represent the underlying psychologically meaningful risk factors (Mann et al. (2010)). The best measures of problematic inclinations consider both static variables, such as offense history, as well as dynamic variables, such as negative attitudes toward authority. Static risk factors are helpful predictors when the underlying criminogenic tendencies are highly stable. However, dynamic risk factors are the preferred method of assessment when these underlying tendencies are prone to change. Actuarial forecasting that relied primarily on static factors was common in the second generation of risk assessment instruments but third and fourth generation instruments expanded to include dynamic risk factors (Bonta and Andrews (2007)).

3.3.1 Static Risk Factors

Static individual risk factors have historically included gender, intelligence and other neuro-psychological characteristics (Vermeiren, Schwab-Stone, Ruchkin, de Clippele, & Deboutte, 2002), young age at the start of problem behaviors, age at first conviction, as well as the intensity and length of the individual's career in delinquency (Loeber & Farrington, 1998; Carcach & Leverett, 1999; Cottle, Lee, & Heilbrun, 2001; Vermeiren, de Clippele, & Deboutte, 2000; Loeber, Farrington, Stouthamer Loeber, & Raskin White, 2008). According to Monahan and Skeem's risk factor typology, we may further subdivide these static risk factors and classify them as either fixed markers or variable markers. Fixed markers, such as race, are permanent qualities of the individual that do not change over time. Variable markers, like age, may change either spontaneously or with the passage of time but are unchangeable through direct intervention. Gendreau et al. (1996) identify ten static factors that are significant predictors of recidivism: age, criminal history (adult), antisocial behavior (preadult), family criminality, family rearing practices, family structure, gender, intellectual functioning, race, and socioeconomic status. Some of these factors, such as age, are considered well-established predictors of adult offender recidivism by empirical researchers.

In addition to static individual risk factors, there are also several static risk factors that operate on the environmental level. Examples of static environmental risk factors include physical maltreatment, conflicts with parents, and parental neglect (Benda & Tollet, 1999; Hawkins et al., 2000; Piquero, Brame, & Moffitt, 2005), as well as issuance of protective care order (Lynch, Buckman, & Krenske, 2003).

Age

The relationship between age and involvement in crime is one of the oldest and most well-established within the field of criminology (see Farrington (1986); Hirschi and Gottfredson (1983), Piquero et al. (2003); Shulman et al. (2013); Sweeten et al. (2013)). Farrington (1986) describes the age-crime curve as unimodal in shape, with the peak of crime involvement occurring approximately at age 17 and descending as the individual ages. Sampson and Lauritsen (1994) characterize age as "one of the major individual-level correlates of violent offending" in a report for the National Research Council's Panel of the Understanding and Control of Violent Behavior. Furthermore, there is an observed difference between individuals who commit property crimes versus violent crimes, with the peak offending age occurring earlier with respect to property crime (Farrington (1986)). Criminological theories, such as life course theory, social bond theory, social learning theory, and strain theory, offer differing explanations for this observed relationship.

A meta-analytic review of literature by Piquero, Jennings, Diamond, and Reingle determined that age is significantly related to violent recidivism, specifying that "lower age was associated with increased rate of violent recidivism" (Piquero et al. (2015)).

Sampson and Laub's (2003) life course theory proposes that criminal offenders will desist from crime as they continue to age, as they encounter a "turning point" that puts them on a "different trajectory". "Turning points" are major events observed to occur during the individual life span, such as getting married, having children, military service, and graduating from high school or college. Sampson and Laub suggest that these turning points result in the individual being placed on a trajectory that moves them away from criminality. A primary consideration in this process is that achievement and maintenance of these turning point markers is hampered by involvement in criminal activities. For example, maintaining a marriage may require the individual to spend additional quality time with their spouse, rather than spending that time out with friends who are involved in criminal behaviors.

This observation by Sampson and Laub of the dynamic between the individual and these turning points bears a resemblance to the underlying propositions of Hirschi's social bond theory. Hirschi (1969) suggests that the individual's relationship with society can be broken down into four categories: attachment, commitment, involvement, and belief. Attachment concerns the relationship between the individual and their families or friends. If the individual displays strong attachment, we would expect to see that they do not involve themselves in criminal activity because involvement in such acts would disappoint their family and friends. Hirschi describes commitment as a willingness to adhere to the dominant norms and laws within a society, and a resistance to criminal involvement is due to the individual accepting the theoretical value of observing these rules. Involvement requires the individual to be occupied in a prosocial manner such that they are unable to also become involved in criminal behavior. Engaging in legitimate activities limits the individual's opportunity to engage in criminal activity. Finally, belief refers to the individual's acceptance of the law and social norms as legitimate sources of authority capable of dealing out punishment in response to violations.

Farrington (1986) notes that “while the general age-crime pattern appears to hold independently of sex, the ratios of male to female offending vary substantially with age and with the type of offense”. For this reason, it would be prudent to check for interactive effects between these variables in predictive models. Some scholars have proposed that economic disadvantage may play a role in the relationship between age and crime, as the disparity between youth violence and adult violence largely disappears when poverty rates are controlled (Males and Brown, 2013). These observations stand in contrast with Hirschi and Gottfredson’s (1983)Hirschi and Gottfredson (1983) claim that the age-crime curve is consistent across various locations, times, genders, offense type, and so forth. Hirschi and Gottfredson further claim age has a direct and causal influence on crime that “cannot be accounted for by any variable or combination of variables currently available to criminology” (p. 554)Hirschi and Gottfredson (1983).

Criminal History (Adult)

-US Sentencing Commission report - Criminal History and Recidivism of Federal Offenders -criminal history is strong predictor of recidivism -rearrest rates range from a low of 30.2% of offenders with no criminal history points to a high of 85.7% for offenders with 15 or more criminal history points -prior record enhancements (PRE) and prediction in sentencing -benefits of PREs in addressing recidivism are unclear based on empirical works

Antisocial Behavior (Preadult)

3.3.2 Gender

Gender is also widely understood to be a major covariate of crime insofar that men commit crimes more frequently than women do. The FBI’s Uniform Crime Reports have consistently revealed higher rates of crime committed by men versus women. Although men and women who commit crime demonstrate similar age distributions, there is a significant "gender gap" in the types of offenses being committed. Historical data indicates that men are disproportionately more likely to be involved in violent offenses. This gender gap is notably smaller when considering minor property offenses. The gender gap also appears to be closing over time, particularly with respect to aggravated assault, robbery, and simple assault (Lauritsen et al. (2009)). Analyses that focus on violent crime have drawn their data primarily from a pool concentrated almost exclusively on male offending (Sampson and Lauritsen (1994)).

As a major covariate of crime, we will return to gender in other sections of this work. Demographic information regarding gender in the COMPAS data set is discussed in Section 4.2.1. A visual representation of the gender ratio in the COMPAS data set is portrayed in Figure 4.2. Gender is further examined as a predictive variable in Section 5.3 when developing the Parsimonious Actuarial Model.

Race

Socioeconomic Status

Another arguable flaw with the input questions is that the consideration of employment history and financial resources result in extra, unequal punishment of the poor which may violate the equal protection clause, based on the precedent case *Bearden v. Georgia* in which the Supreme Court rejected Georgia's argument that poverty was a recidivism factor that justified additional incapacitation. To prevent perpetuating a racially disparate impact, advocates are arguing for a narrow range of questions, such as strictly based on past or present criminal behavior, or an individual assessment of a defendant's conduct, mental states, and attitudes.

3.3.3 Dynamic Risk Factors

Simply put, dynamic risk factors are variables that are changeable over time. According to Monahan and Skeem (2014), dynamic risk factors may be either variable risk factors or causal risk factors. Variable risk factors, such as employment status, are not inherently permanent characteristics, the results of random processes, nor are they attributable to the passage of time. Rather, these variables may be changed through direct intervention. Causal risk factors, such as substance abuse, are a subtype of variable risk factor that when targeted have been empirically demonstrated to show a change in the risk of recidivism (Kraemer (2003)). Research that targets a specific intervention through a randomized control trial is somewhat rare but has historically demonstrated substance abuse and criminal thinking patterns as causal (Monahan and Skeem (2014)).

Criminogenic Needs

Structured Professional Judgement (SPJ)

3.3.4 Neuro-psychological Predictors

In addition to actuarial predictors, criminal justice professionals also entertain the use of psychological predictors and classification instruments. There is ample evidence to suggest that criminal cognition is related to the risk of future offending (Andrews and Bonta (2010)). We may seek to use an instrument like the Criminal Sentiments Scale (CSS) to capture and quantify the psychological status of the individual offender. Walters (1995) breaks down criminal thinking patterns into two components – criminal thought process and criminal thought content. Criminal thought process encapsulates how a criminal thinks while criminal thought content details what a criminal thinks. The criminal thought process can be broken down even further into proactive versus reactive criminal thinking. Proactive criminal thinking is deliberate, calculated, and organized while reactive criminal thinking is spontaneous, emotional, and disorganized (Walters (2016)). According to Walters and Lowenkamp (2016), "reactive criminal thinking did a better job of predicting recidivism and early offending than proactive criminal thinking" even when controlling for overlap between the two styles.

There is much more research available on the criminal thought process than there is on criminal thought content. Landau (1978) suggests that negative attitudes toward authority, acceptance of deviance, and anti-social beliefs and identity can distinguish between a prisoner and a nonoffender. Empirical research has revealed "evidence of a relationship between negative attitudes toward authority and juvenile aggression, social manipulation, and impulsivity" (Walters (2016)). Additionally, there is a significant correlation between prior criminality and antagonistic attitudes toward authority. Tangney et al. (2012) demonstrated that the Negative Attitudes Toward Authority scale on the Criminogenic Cognitions Scale (CCS) "correlated significantly with a history of violent criminality, prior incarcerations, measures of psychopathy and antisocial personality, and subsequent jail infractions".

Walters (2016) further examines the value of including measures of the criminal thought process in predictions of recidivism for adult and juvenile offenders. The Criminal Sentiments Scale (CSS), a popular 41-item instrument used as the putative measure of the criminal thought process, includes subscales for Attitudes Toward Law, Court, and Police (LCP), Tolerance for Law Violations (TLV), and Identification With Criminal Others (ICO) (Gendreau et al. (1979)). Tangney et al. (2012) determined that the LCP subscale correlated significantly with "a history of violent criminality, prior incarcerations, measures of psychopathy and antisocial personality, and subsequent jail infractions". The CSS total score as well as all three subscales "appear to have the ability to predict recidivism and inform risk assessment" (Walters (2016)). Walters notes that there is a modest to low moderate pooled effect size for the CSS total score, LCP subscale, and TLV subscale, whereas there was a small pooled effect size for the ICO subscale (2016)Walters (2016). Overall the data suggest that the CSS "may have a significant role to play in risk assessment" as a measurement of an important dynamic risk factor (Walters (2016)).

3.4 Discussion of Literature

Schmidt and Witte (1988) investigated the use of survival modeling in predicting recidivism, collecting data on inmates released from North Carolina prisons in 1978 and 1980. When using the 1978 data, Schmidt and Witte found that the most valuable predictive factors for time to recidivism were the time served by the offender (TSERVD), the number of priors the offender had (PRIORS), a dummy variable for race (WHITE), dummy for felon status (FELON), dummy for alcoholism (ALCHY), dummy for hard drug use (JUNKY), dummy for crime against property (PROPTY), dummy for gender (MALE), a natural logarithm transformation of age in months (LN(AGE)), dummy for offenders aged 20 or younger (YOUNG), an interaction between time served and age (TSERVD*LN(AGE)), and an interaction between a dummy for serving 30 months or less and a dummy for no previous incarcerations (SHORT*NOPRIOR) (Witte, 1992)Witte (1992). Witte notes that the underlying dummy variables for this interaction "SHORT and NOPRIOR affect time until recidivism only in their interaction form" (1992, pg. 13)Witte (1992).

In the final specification of the predictive model based on the 1980 Schmidt and Witte data, there are many similarities to the aforementioned 1978 model (see Fig. 1 for a side-by-side comparison of predictive variables). Here the most valuable predictors of recidivism are demonstrated to be the number of priors the offender had (PRIORS), a dummy for marital status (MARRIED), a dummy for hard drug use (JUNKY), dummy for crime against property (PROPTY), a natural logarithm transformation of age in months

(LN(AGE)), dummy for offenders aged 20 or younger (YOUNG), dummy for serving 30 months or less (SHORT), dummy for no previous incarcerations (NOPRIOR), an interaction between time served and a dummy for alcoholism (TSERVD*ALCHY), an interaction between time served and NOPRIOR (TSERVD*NOPRIOR), an interaction between LN(AGE) and SHORT (LN(AGE)*SHORT), an interaction between PRIORS and PROPTY (PRIORS*PROPTY), an interaction between a dummy for race and NOPRIOR (WHITE*NOPRIOR), and an interaction between gender and NOPRIOR (MALE*NOPRIOR). Witte remarks that this model includes a "larger number of significant interactions than we found for the 1978 data, and it is a larger number than we expected to find" (1992, p. 20)Witte (1992). Further, it is interesting to note that again there are variables which are not significant on their own but are significant as a component of an interaction term. Witte does not offer an explanation beyond noting that "we have let the data drive our specification and apparently this is what the data indicate" (1992, p.21)Witte (1992).

Predictive Variable 1978 Model 1980 Model TSERVD X

PRIORS X X WHITE X

FELON X

ALCHY X

JUNKY X X PROPTY X X LN(AGE) X X YOUNG X X MARRIED

X SHORT

X NOPRIOR

X TSERVD*LN(AGE) X

SHORT*NOPRIOR X

TSERVD*ALCHY

X TSERVD*NOPRIOR

X LN(AGE)*SHORT

X PRIORS*PROPTY

X WHITE*NOPRIOR

X MALE*NOPRIOR

X Fig. 1 Comparison of Predictive Variables Used in 1978 and 1980 Schmidt & Witte proportional hazard models. Tollenaar and Heijden (2019) compared the predictive potential of classical statistical methods with machine learning methods in censored time-to-event recidivism data. In order to promote generalizability of their findings, they applied the same models to both the Schmidt and Witte NC prison dataset as well as a Dutch recidivism dataset. They conclude that for the binary outcome criminal prediction context, there is no significant improvement for automatic models compared to traditional, manually-specified models (Tollenaar and Heijden (2019)). Although these "black box" models may not improve on the predictive value of a manually-specified model, they may be useful in identifying model misspecification or missing interactions. Tolleenaar and van der Heidjen recommend trying both statistical and machine learning approaches in order to specify the most optimal model.

Chapter 4

Methods

4.1 Introduction

This chapter outlines the techniques by which (1) the novel Parsimonious Actuarial Model (PAM) was developed and (2) we assess the predictive validity of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) and the aforementioned PAM. This model development and predictive validity assessment was conducted utilizing datasets from the Broward County Sheriff's Office and the Broward County Clerk's Office. Broward County is located in the state of Florida and is inherently a promising place for us to gather data on offenders. This promise is largely due to Florida's "Government-In-The-Sunshine Law" – in fact, a series of laws – which guarantees that any person can request public records. The scope of what constitutes "public records" is quite generously defined, including law enforcement records such as criminal history information as well as many types of prison and inmate records.

4.2 Participants in the FL COMPAS Data Set

The total sample consisted of 7,214 adult male (81%) and female (19%) criminal defendants who were assessed with the COMPAS algorithm at the pretrial stage by the Broward County Sheriff's Office. This sample includes all defendants assessed by Broward County during the years 2013 and 2014. Each pretrial defendant in Broward County was assessed for their risk of recidivism, as well as their risk of violence and risk of failure to appear. These COMPAS risk assessment scores are used by Broward County primarily for the purposes of determining whether to detain defendants prior to their trial date.

4.2.1 Demographic Information

Demographic information on criminal defendants in Broward County can be acquired by downloading criminal records from the Broward County Clerk's Office. These criminal records were then matched to COMPAS score records obtained by ProPublica via a public records request. ProPublica matched these records using the protocol used in Florida State University's 2010 validation study, matching the first and last names and date of birth. Archival records were used to collect demographic information pertaining to age at release, gender, race/ethnicity, and criminal history. This procedure yielded a total sample of 7,214 adult offenders. In a test of quality for matches (e.g. correctly matching the individual's criminal record to their COMPAS score) by ProPublica, a random sample of 400 cases yielded an error rate of approximately 3.75% (Angwin et al. (2016)). Observations were dropped if the date between the individual's arrest and the date of the individual's COMPAS assessment exceeded 30 days, as such data were likely to be erroneously matched. Table 4.1 displays the detailed demographic characteristics of the resulting COMPAS data set (N=6,172).

Table 4.1: Demographic Characteristics of FL COMPAS Data Set (N=6,172)

Variable		n (%)
Age at Release		
	Under 25	2,036 (32.99%)
	25-45	3,059 (49.56%)
	Over 45	1,077 (17.45%)
Gender		
	Male	4,997 (80.96%)
	Female	1,175 (19.04%)
Race/Ethnicity		
	White	2,103 (34.07%)
	Black	3,175 (51.44%)
	Hispanic	509 (8.25%)
	Asian	31 (0.50%)
	Native	11 (0.18%)
	Other	343 (5.56%)
Prior Offenses		
	None	2,085 (33.78%)
	One	1,129 (18.29%)
	Two	681 (11.03%)
	Three+	2,277 (36.90%)
Recidivism 2YR		
	No	3,363 (54.49%)
	Yes	2,809 (45.51%)

The COMPAS data set yields informative demographic characteristics for Broward County offenders who were assessed during the pretrial stage. Based on our discussion in Chapter 3, we are particularly interested in the following variables: the age of the offender, their gender, their race/ethnicity, the number of prior offenses they had committed, and whether they had recidivated within 2 years of their release.

Age of the Offender

As previously discussed in Section 3.3.1, there is a well-established relationship known to criminologists between age and involvement in crime. Based on the available literature, we would expect to see . Within the COMPAS data set, we actually see . In Figure 4.1, we see the age distribution observed in the COMPAS data set. It is important to note that our data is limited in terms of age representation. As this data set is drawn from adult jail records, there are no juveniles included.

Gender of the Offender

Race/Ethnicity of the Offender

Number of Prior Offenses

Recidivism within Two Years of Release Date

Training Data Set

The remaining 6,172 adult offenders were randomly assigned to either the training set or the confirmation set using Stata's pseudo-random number function. Table 4.2 displays the detailed demographic characteristics of the training data set (N=3,100).

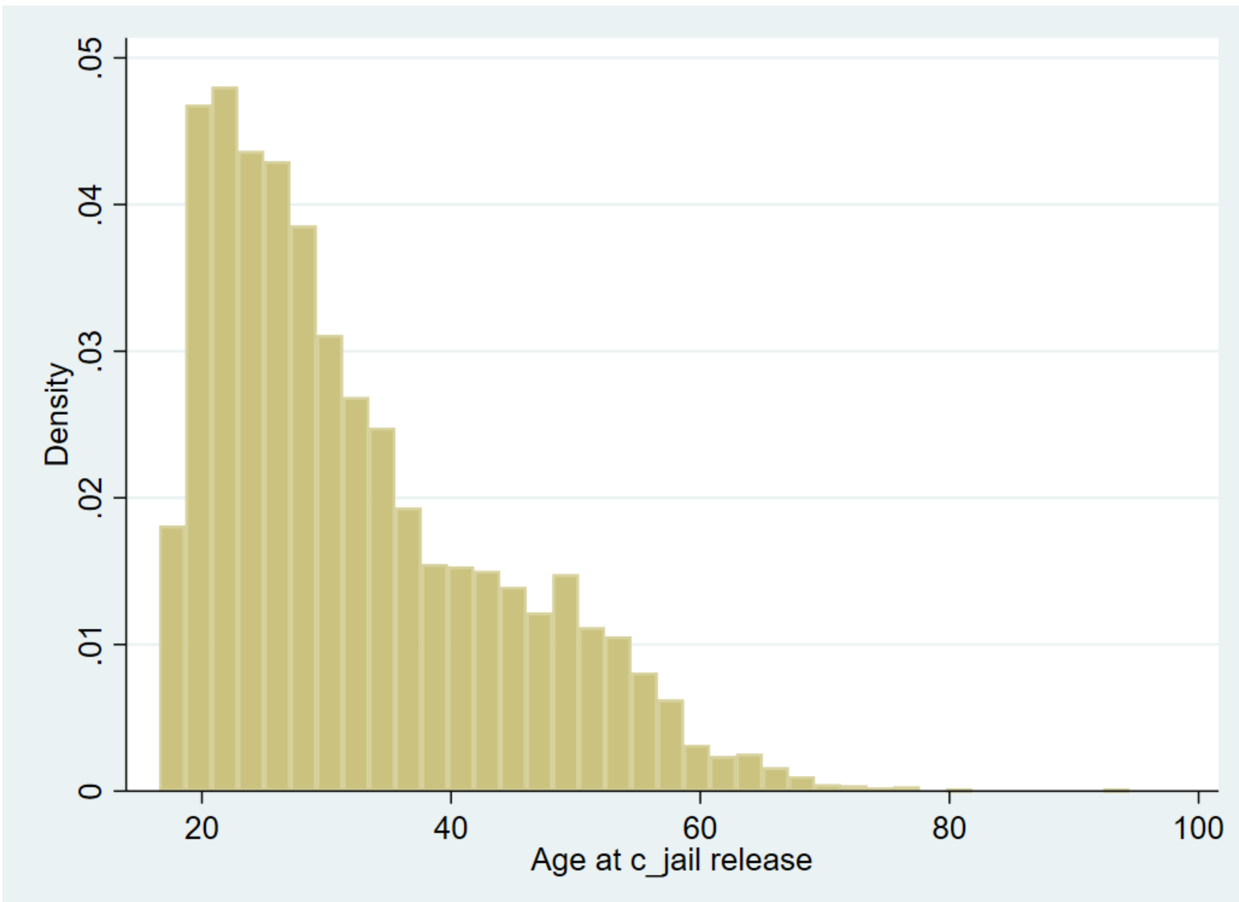


Figure 4.1: Histogram of Age at Release for COMPAS Data Set

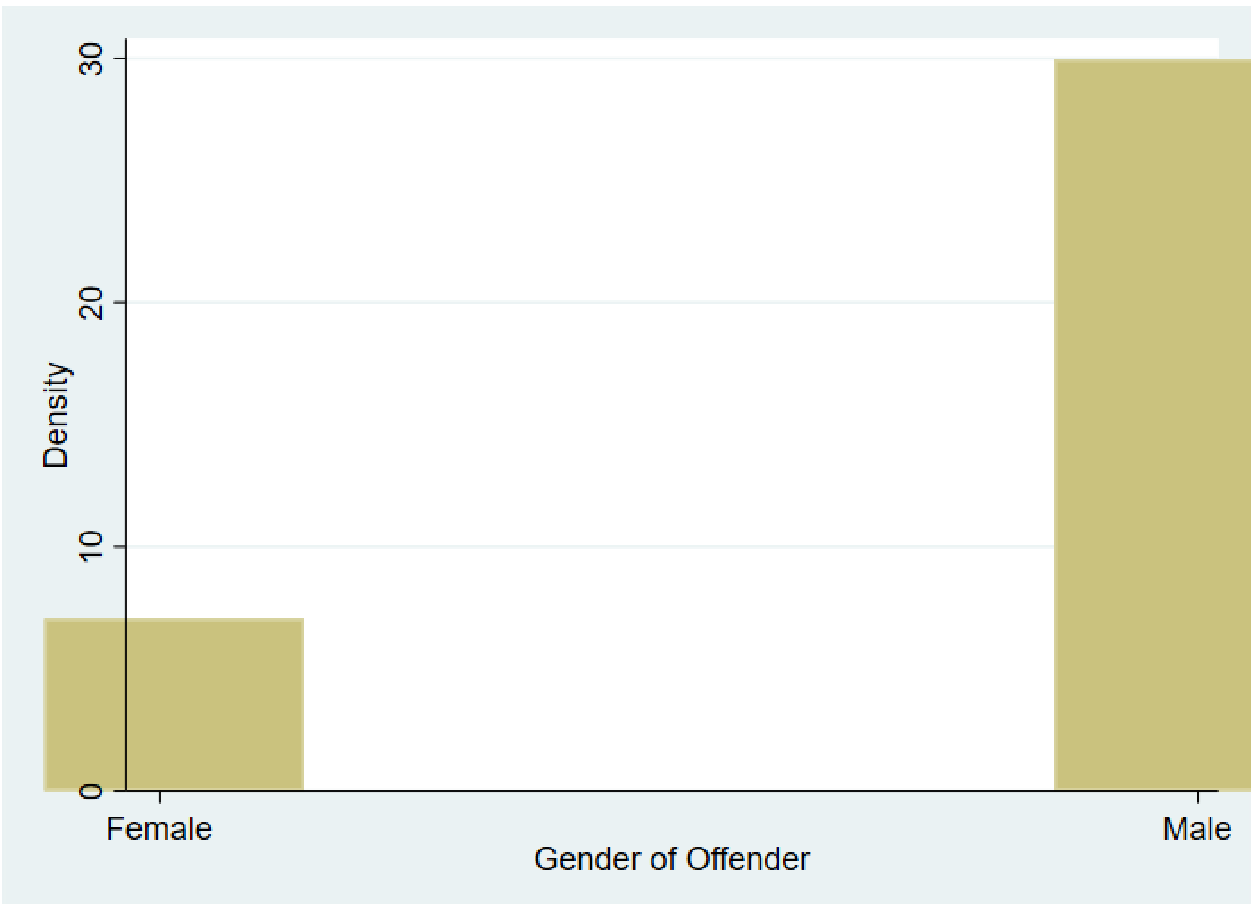


Figure 4.2: Histogram of Gender for COMPAS Data Set

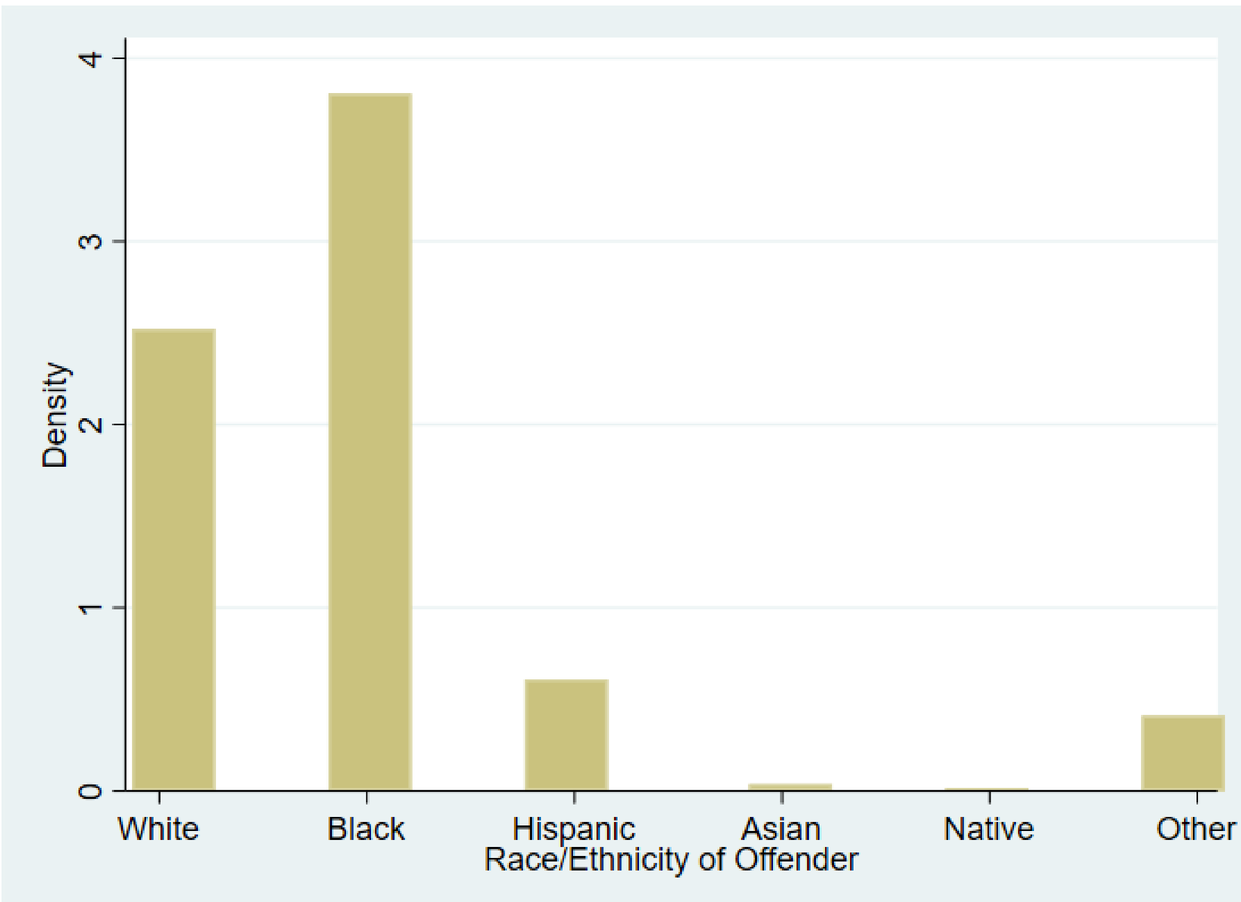


Figure 4.3: Histogram of Race/Ethnicity for COMPAS Data Set

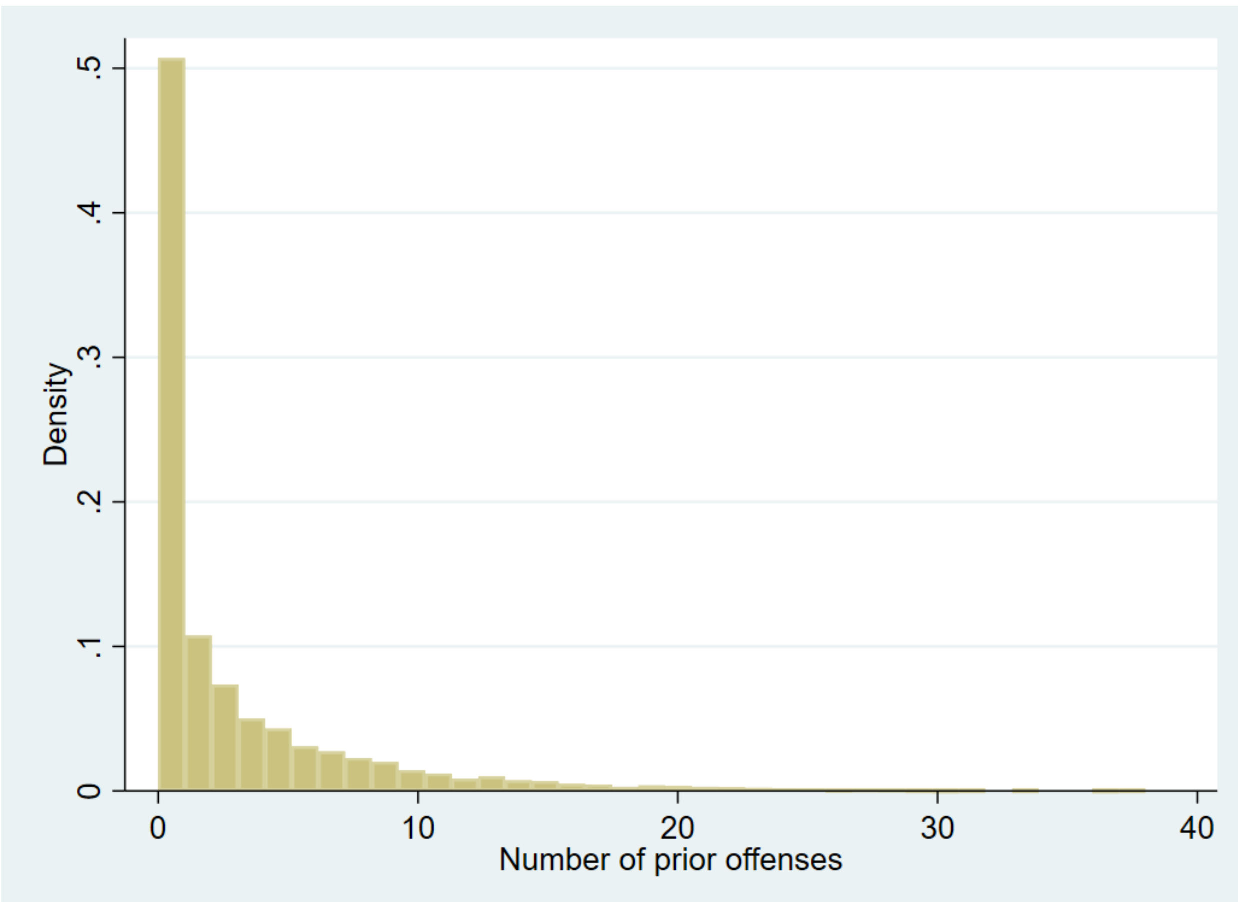


Figure 4.4: Histogram of Prior Offenses for COMPAS Data Set

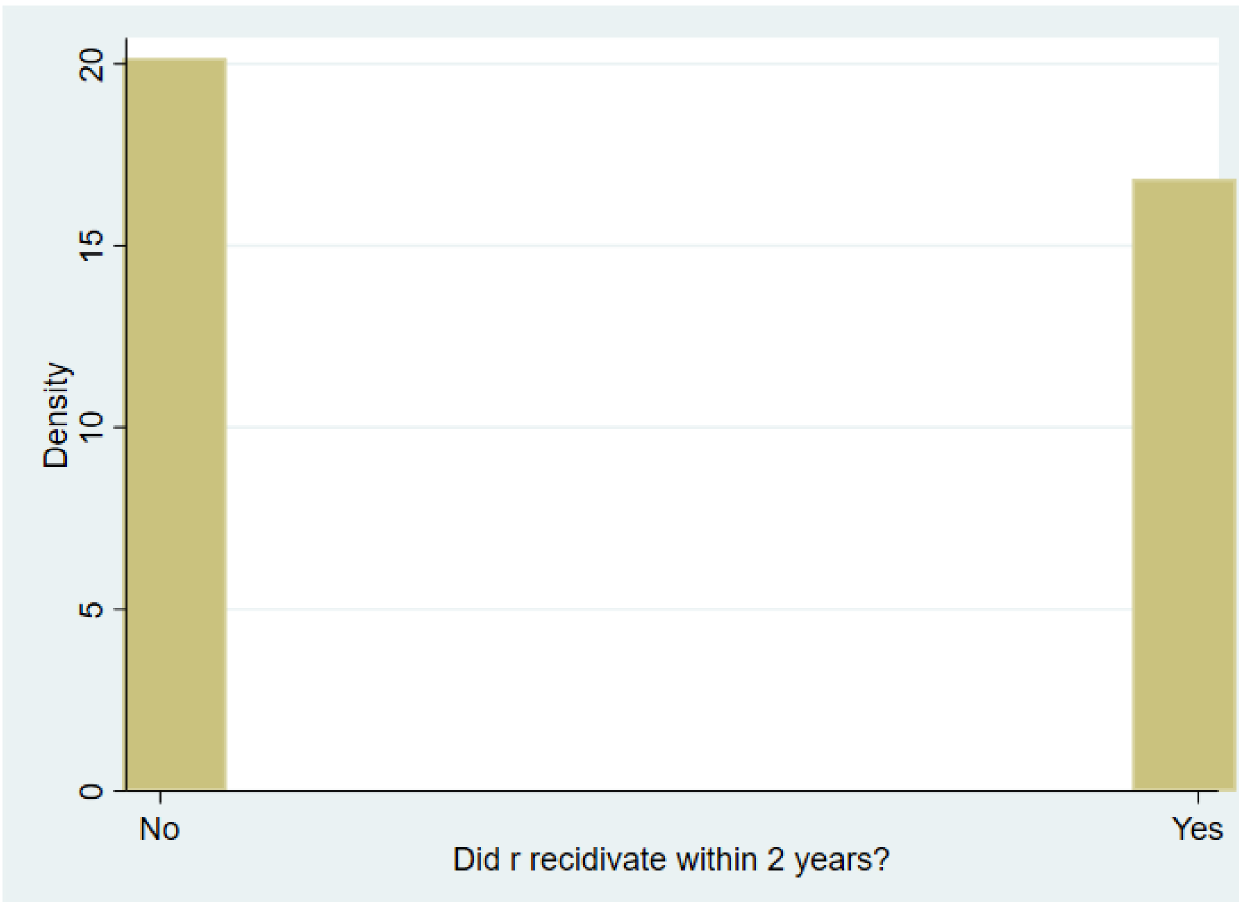


Figure 4.5: Histogram of 2-Year Recidivism for COMPAS Data Set

Table 4.2: Demographic Characteristics of Training Data Set (N=3,100)

Variable		n (%)
Age at Release		
	Under 25	1,022 (32.97%)
	25-45	1,531 (49.39%)
	Over 45	547 (17.65%)
Gender		
	Male	2,518 (81.23%)
	Female	582 (18.77%)
Race/Ethnicity		
	White	1,073 (34.61%)
	Black	1,591 (51.32%)
	Hispanic	232 (7.48%)
	Asian	17 (0.55%)
	Native	5 (0.16%)
	Other	182 (5.87%)
Prior Offenses		
	None	1,029 (33.19%)
	One	559 (18.03%)
	Two	359 (11.58%)
	Three+	1,153 (37.19%)
Recidivism 2YR		
	No	1,694 (54.65%)
	Yes	1,406 (45.35%)

4.3 Measurement and Methodological Problems

4.3.1 Criterion Events

Subsequent Arrest

Subsequent Conviction

4.3.2 Follow-up Period

Censoring

Chapter 5

COMPAS versus the Parsimonious Model

5.1 Introduction

5.2 The COMPAS Instrument

The Correctional Offender Management and Profiling Alternative Sanctions instrument, or COMPAS, is a risk assessment tool designed by Northpointe Institute for Public Management. This computerized database and analysis system was designed to assist practitioners in the criminal justice system, particularly with determining "placement, supervision, and case-management of offenders in community and secure settings" (Farabee et al. (2010)). The COMPAS is a popular actuarial risk assessment instrument used in the criminal justice system in both institution and community-based settings (Desmarais and Singh (2013)). Although the COMPAS is widely used in the United States, few studies to date have examined it and fewer still have gone on to publish the resulting analysis in peer-reviewed publications, particularly when compared with other actuarial risk assessment instruments such as the Level of Service Inventory (LSI; Andrews and Robinson, 1984).

Within the most recently released version of the COMPAS instrument there are 4 Risk scales, 19 Needs scales, and 2 Validity scales. The four Risk scales, sometimes referred to as the COMPAS Core Risk scales, include General Recidivism Risk, Violent Recidivism Risk, Recidivism Risk Screen, and Pretrial Release Risk. Northpointe (2013) states that these Core Risk scales were derived using logistic regression, drawing on other components of Core Need scales including transformations of those Needs scale scores. The Violent Recidivism Risk scale, for instance, draws on scores for History of Violence, History of Compliance, and Vocational Education Problems Needs (Northpointe (2013)). The COMPAS instrument quantifies various levels of risk by converting scale scores into decile scores, which can then be used to classify individual offenders. These various levels of risk are then given a descriptive label based upon the underlying decile scores. A "high" risk level includes decile scores between 8 and 10, a "medium" risk level is between 5 and 7, and a "low" risk level represents values between 1 and 4 (Northpointe (2013)).

Prior evaluations of the COMPAS have yielded mixed results regarding its ability to estimate an individual offender's likelihood of recidivism. The majority of published research on the COMPAS instrument has been performed by its developers at Northpointe (see Breitenbach et al. (2010); Brennan and Dieterich (2008); Brennan et al. (2004)). The only study to date that has evaluated the internal consistency and reliability of the instrument's scales was also conducted by Northpointe (Brennan et al. (2009)). In evaluations of the General Recidivism risk scale, the Receiver Operating Characteristic (ROC) curve statistics have ranged from 0.53 (Fass, Heilbrun, Dematteo & Fretz, 2008) – slightly better than chance – to a respectable 0.80 (Brennan, Dieterich, & Ehret, 2009) when the study was conducted by Northpointe. Blair, Marcus & Boccaccini (2008) have suggested that studies on risk assessment instruments, when conducted by the developers of those instruments, are subject to a particular type of bias termed the allegiance effect. The allegiance effect is indicated when validation studies conducted by the instrument's developer result in greater effect sizes than when validation studies are performed by an independent researcher or group (Blair, Marcus & Boccaccini, 2008).

5.3 The Parsimonious Actuarial Model

5.3.1 Introduction

The actuarial model developed here is designed to be both parsimonious and transparent. Whereas the COMPAS instrument requires the administration of a 137-item questionnaire, the parsimonious model attempts to achieve comparable predictive validity by employing a small number of known predictors of criminal behavior. The included predictors would be easily obtained from the offender's court and jail records. The intention here is not to develop a model with significantly better predictive validity than COMPAS. It is to determine whether the COMPAS provides comparable or additional predictive value above and beyond a basic actuarial model. Simply stated, if we want to quantify an individual's risk of recidivism, is there additional predictive value in asking 137 questions versus 4 or 5? Additionally we are interested in how the parsimonious actuarial model performs in comparison to the COMPAS instrument for sub-populations of interest.

Current practice in machine learning methods is to utilize multiple algorithmic approaches to fitting the data, then assessing the predictive validity of the resultant models. Each data set has its own characteristics, such that there is no one specific approach to model development that we should assume to be best. Here are some of the considerations we may have during our algorithm selection process:

- Size of the data set, including the number of observations as well as the number of variables
- Quality of the data set, such as missing or censored data
- Type of data, such as categorical variables or continuous variables
- Purpose of the predictive model, whether predicting a numeric value or classifying the observations
- Priority of the speed of calculating predictions versus the accuracy of predictions

5.3.2 Developing the Model

As described previously in Section 4.2.1 the 6,172 adult offenders in the data set were randomly assigned to either the training set or the confirmation set using Stata's pseudo-random number function. After reviewing the previous literature, the following variables were selected for further examination: gender of the offender, age at release, number of prior offenses, current felony charge, number of juvenile felonies, and duration of stay in jail.

Simple Linear Regression

Simple linear regression is used here to determine whether there is a predictive relationship between a single variable and an outcome of interest. Here we are examining the predictive variables associated in the literature with risk of recidivism. Each of these predictive variables is considered individually for its value in predicting time-to-recidivism, or the time the offender will be at risk in the community between release from jail and return to jail. For the outcome of interest, in this instance we will be using the time-at-risk in the community between release and re-incarceration. This is to say, we are attempting to predict the number of days someone is at risk in the community based on the value of a predictor variable. Although simple linear regression is unlikely to adequately capture a meaningful relationship between a single predictor variable and time-to-recidivism, this is a logical place to begin as the simplest modeling approach.

We being by examining the correlations between variables of interest. Variables of interest that are highly correlated to one another may cause issues in the final model. The predictor variables assessed here include age at release (*age*), gender (*male*), number of prior offenses (*priors*), current felony charge (*felony*) and length of stay in jail (*duration*). We are also examining the correlation between the predictor variables and the outcome variable of interest, time-at-risk in the community between release and re-incarceration (*time*). In Table 5.1 below are the results of that analysis, indicating no strong correlations within these predictor variables. The strongest relationship between predictor variables was between number of prior offenses (*priors*) and duration (*length of stay in jail*), where $R^2 = 0.2145$.

Table 5.1: Correlation between variables of interest

	Age	Male	Priors	Felony	Duration	Time
Age	1					
Male	-0.0077	1				
Priors	0.1173	0.125	1			
Felony	-0.0984	0.0671	0.1378	1		
Duration	0.0334	0.0698	0.2145	0.1222	1	
Time	0.1661	-0.0953	-0.3116	-0.1773	-0.1399	1

Each of the predictor variables was used in turn as the independent variable in a simple linear regression model to predict time-to-failure (*time*), or how many days will elapse between release from jail and the subsequent return to jail via recidivism. This will also allow us an opportunity to see if the relationships suggested by the literature are supported by these data. The resulting equation for each simple linear regression model follows the general form:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i$$

Table 5.2: Comparing simple linear regression models

	(1)	(2)	(3)	(4)	(5)
	time	time	time	time	time
age	5.610***				
	(0.599)				
male		-97.53***			
		(18.32)			
priors			-26.34***		
			(1.444)		
felony				-148.9***	
				(14.85)	
duration					-1.245***
					(0.158)
_cons	359.1***	621.2***	628.5***	639.1***	560.7***
	(20.76)	(16.51)	(8.316)	(12.00)	(7.506)
N	3098	3098	3098	3098	3098
R^2	0.028	0.009	0.097	0.031	0.020
adj. R^2	0.027	0.009	0.097	0.031	0.019
rmse	394.5	398.2	380.1	393.7	396.1

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Age

A simple linear regression model using age at release to predict time-at-risk in the community suggests that as the defendant's age increases, the average amount of time-at-risk in the community increases as well. This is to say that older offenders spend more time on average between initial offense and the recidivist act. This reflects the well-known relationship between age and crime within the literature. This is the resulting simple linear

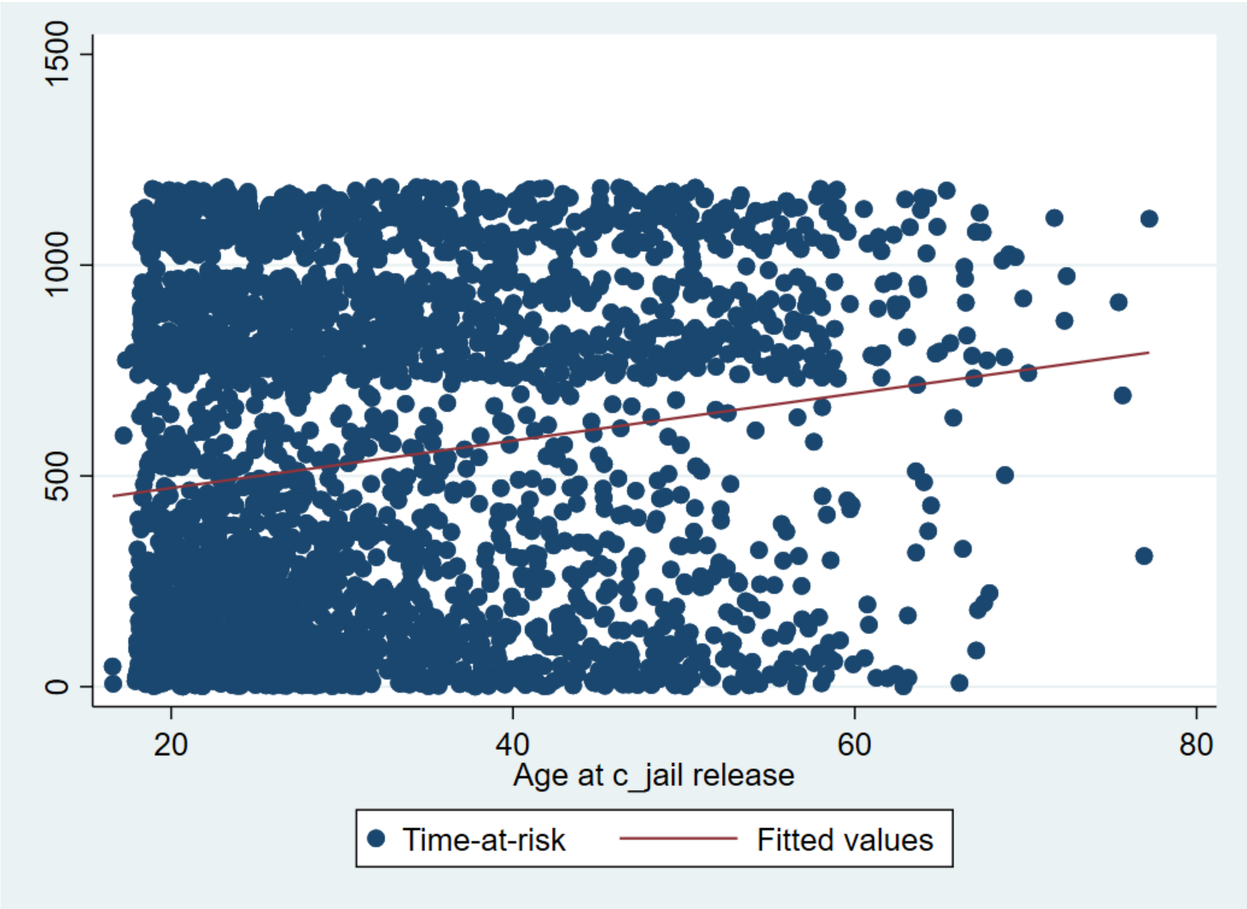


Figure 5.1: Linear regression of age at release on time-at-risk

regression model using age at release as a predictor for time-at-risk in the community, where $R^2 = 0.028$:

$$\hat{Y}_i = 359.1 + 5.61(age) + \hat{\epsilon}_i \tag{5.1}$$

A graphical representation of this model predicting time-at-risk in the community based on age at release is displayed in Figure 5.1.

Gender

A simple linear regression model using gender to predict time-at-risk in the community suggests that male offenders spend less time on average between release from jail and re-

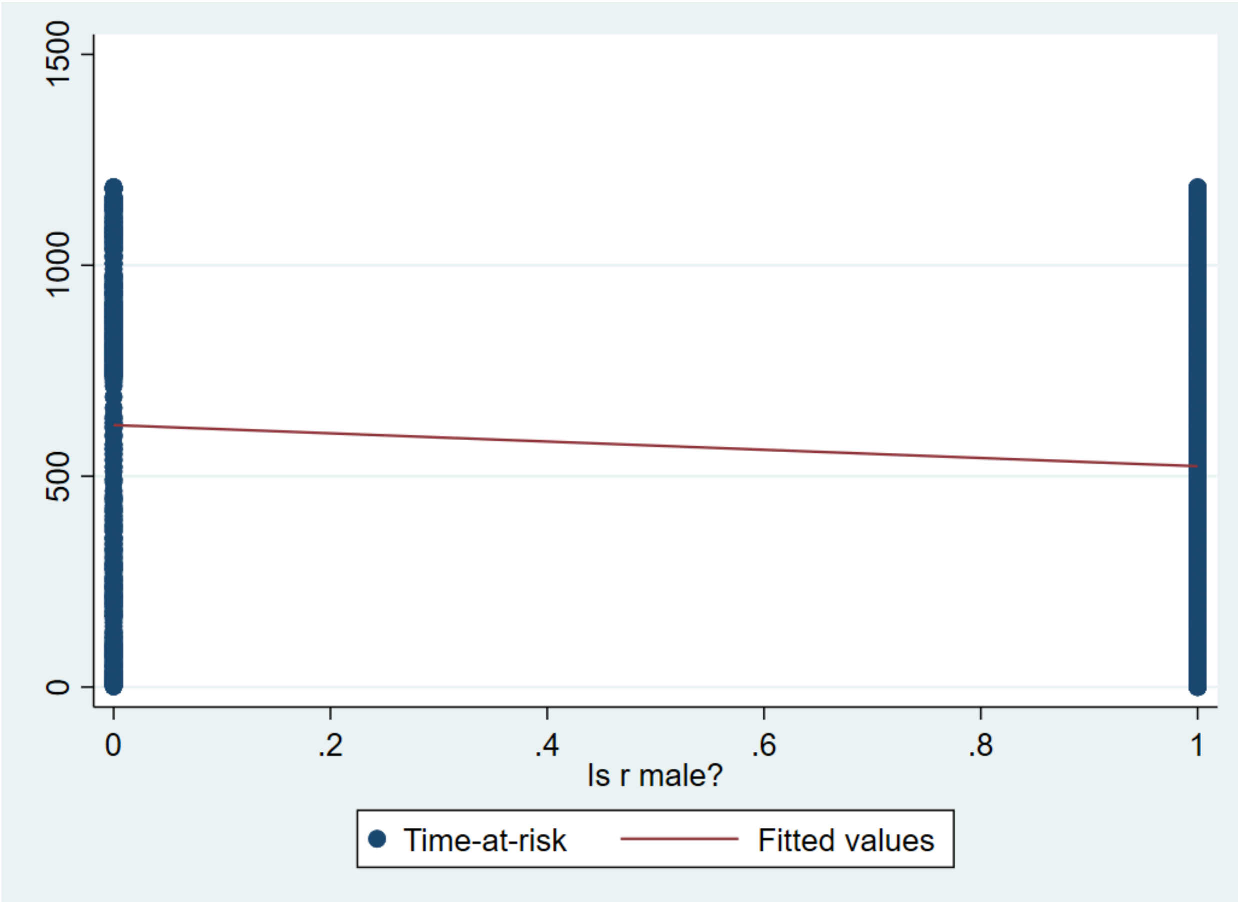


Figure 5.2: Linear regression of gender on time-at-risk

incarceration, when compared to female offenders. This is to say that older offenders spend more time on average between initial offense and the recidivist act. This reflects the well-known relationship between gender and crime within the literature. This is the resulting simple linear regression model using gender as a predictor for time-at-risk in the community, where $R^2 = 0.009$:

$$\hat{Y}_i = 621.2 - 97.53(male) + \hat{\epsilon}_i \tag{5.2}$$

A graphical representation of this model predicting time-at-risk in the community based on gender is displayed in Figure 5.2.

Prior Offenses

A simple linear regression model using the offender's number of prior offenses to predict time-at-risk in the community suggests that as the offender's number of prior offenses increases, the average amount of time-at-risk in the community decreases. This is to say that offenders with lengthier criminal histories spend less time on average between initial offense and the recidivist act. This reflects the well-known relationship between criminal history and recidivism within the literature. This is the resulting simple linear regression model using number of prior offenses as a predictor for time-at-risk in the community, where $R^2 = 0.097$:

$$\hat{Y}_i = 628.5 - 26.34(priors) + \hat{\epsilon}_i \tag{5.3}$$

A graphical representation of this model predicting time-at-risk in the community based on the offender's number of prior offenses is displayed in Figure 5.3.

Current Felony Charge

A simple linear regression model using the charge degree to predict time-at-risk in the community suggests that offenders who were charged with a felony at the time of COMPAS administration spend less time at risk in the community when compared to offenders charged with a misdemeanor. This is to say that offenders charged with a misdemeanor at the time of the COMPAS administration spend more time on average between initial offense and the recidivist act. This reflects the well-known relationship between the seriousness of the current offense and crime within the literature. This is the resulting simple linear regression model using charge degree as a predictor for time-at-risk in the community, where $R^2 = 0.031$:

$$\hat{Y}_i = 639.1 - 148.9(felony) + \hat{\epsilon}_i \tag{5.4}$$

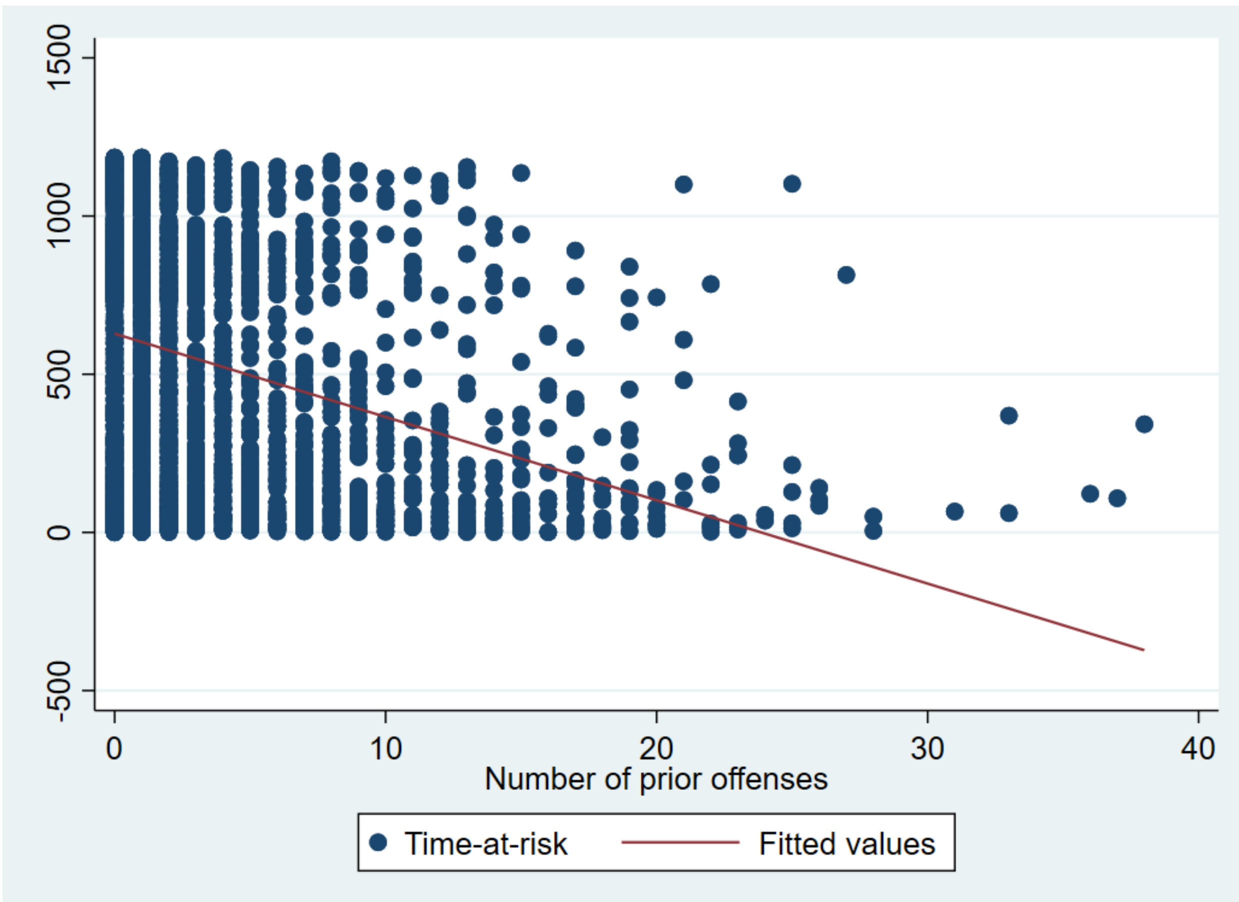


Figure 5.3: Linear regression of number of prior offenses on time-at-risk

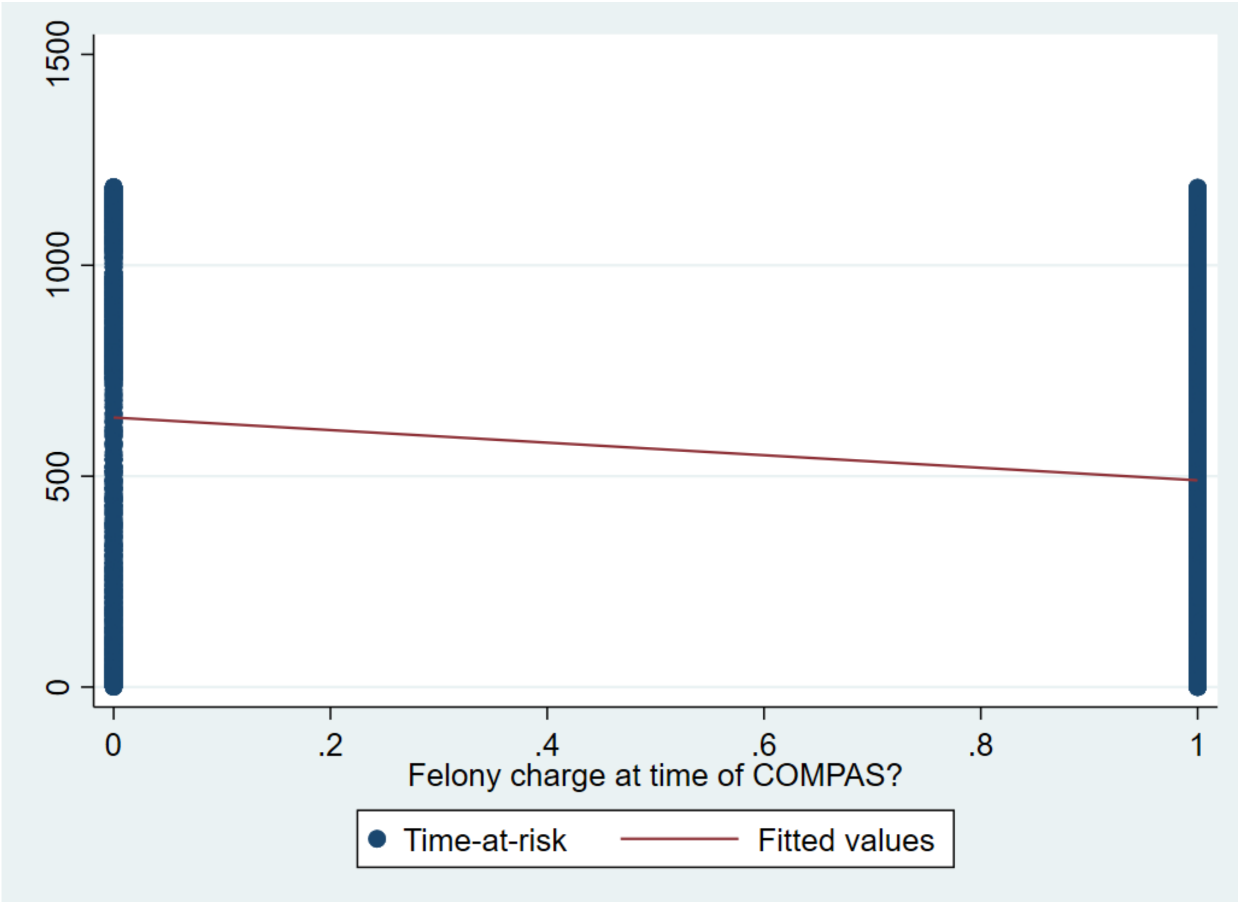


Figure 5.4: Linear regression of current felony charge on time-at-risk

A graphical representation of this model predicting time-at-risk in the community based on current felony charge is displayed in Figure 5.4.

Duration of Stay in Jail

A simple linear regression model using the duration of the offender's stay in jail to predict time-at-risk in the community suggests that as the defendant's time in jail increases, the average amount of time-at-risk in the community decreases. This is to say that on average, offenders that spend more time in jail will spend less time at risk between initial offense and the recidivist act. This reflects the well-known relationship between incarceration and recidivism within the literature. This is the resulting simple linear regression model using the duration of the offender's stay in jail as a predictor for time-at-risk in the community,

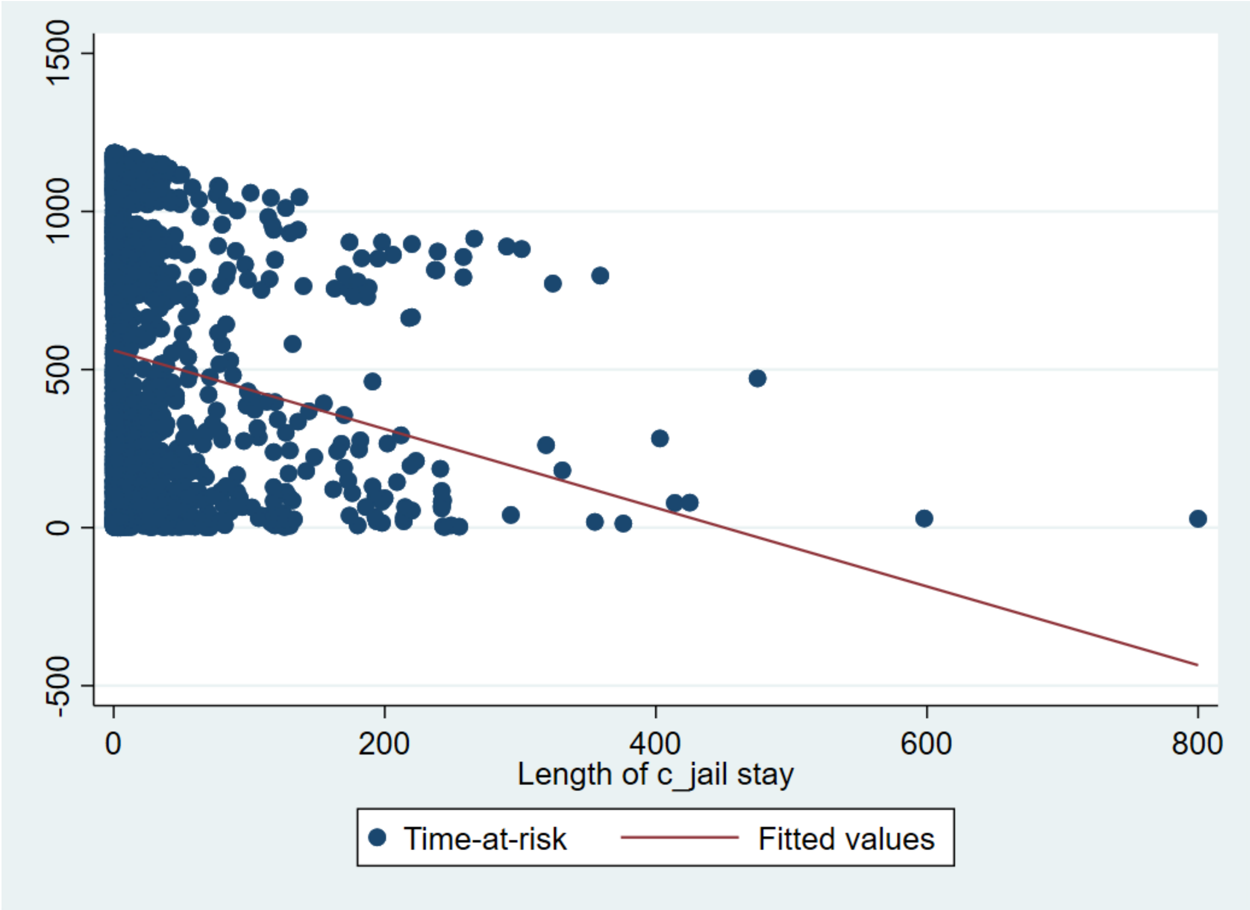


Figure 5.5: Linear regression of jail stay length on time-at-risk

where $R^2 = 0.020$:

$$\hat{Y}_i = 560.7 - 1.245(duration) + \hat{\epsilon}_i \tag{5.5}$$

A graphical representation of this model predicting time-at-risk in the community based on length of stay in jail is displayed in Figure 5.5.

Multiple Regression

Simple linear regression is unlikely to yield a sufficiently predictive model of human behavior such as recidivism. It is more likely that a combination of these predictive factors will yield a robust predictive model. Therefore we are moving on to multiple regression to determine whether there is a predictive relationship between multiple independent variable and the outcome of interest. Here we are largely re-examining the same predictive variables previously associated in the literature with risk of recidivism. We will not consider the offender's length of stay in jail (*duration*) as a predictive independent variable moving forward. This is because the parsimonious actuarial model is intended to predict recidivism at the time of sentencing. As such, the offender's length of stay in jail has not yet been determined and cannot logically be used as a component of this model. A reverse stepwise approach to fitting the final multiple regression model is employed as demonstrated in Table 5.3. Model 3 was selected as a balance between the coefficient of multiple determination and a parsimonious model. Although it would be possible to eliminate gender(*male*) from the final model, it was retained due to the well-established relationship in the literature between gender and crime.

Below we have Table 5.4 comparing the final multiple regression model to a simple regression model using only the offender's COMPAS score to predict their time-at-risk in the community until recidivism. The final multiple regression model using the offender's age, charge degree, gender, and number of prior offenses as predictors for time-at-risk in the community (in days) is as follows:

$$\hat{Y}_i = 614.9 + 114.55(age25-45) + 187.8(age45orolder) - 85.54(felony) - 29.03(male) - 123.9(logpriors) + \hat{\epsilon}_i \tag{5.6}$$

Logistic Regression

Logistic regression may be employed when the predictor variables do not fit a particular distribution and in cases where the predictor variables are both continuous and categorical

Table 5.3: Side-by-side multiple regression model comparison

	(1) time	(2) time	(3) time	(4) time	(5) time
age	5.322***	5.367***	6.464***	6.491***	6.941***
	(0.750)	(0.748)	(0.567)	(0.567)	(0.567)
priors	-26.79***	-27.00***	-26.43***	-26.91***	-28.38***
	(1.481)	(1.456)	(1.435)	(1.426)	(1.420)
felony	-93.75***	-93.77***	-94.22***	-96.16***	
	(14.13)	(14.13)	(14.14)	(14.14)	
male	-46.13**	-46.56**	-48.31**		
	(17.11)	(17.10)	(17.10)		
under 25	-43.19*	-42.79*			
	(19.08)	(19.07)			
juv felony	-12.68				
	(16.09)				
_cons	570.2***	568.9***	518.9***	481.6***	409.0***
	(34.17)	(34.13)	(25.87)	(22.28)	(19.70)
N	3098	3098	3098	3098	3098
R^2	0.155	0.155	0.154	0.151	0.139
adj. R^2	0.153	0.154	0.152	0.151	0.138
rmse	368.0	368.0	368.2	368.6	371.3

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.4: Multiple Regression PAM versus COMPAS

	(1) MLR PAM	(2) COMPAS
Age 25-45	114.55***	
	(15.07)	
Age 45 or older	187.8***	
	(19.48)	
Felony	-85.54***	
	(14.03)	
Male	-29.03	
	(17.02)	
Log priors	-123.9***	
	(5.905)	
Decile score		-53.84***
		(2.348)
Constant	614.9***	780.9***
	(19.97)	(12.36)
N	3098	3098
R^2	0.171	0.145
adj. R^2	0.170	0.145
rmse	364.5	369.9

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

in nature. Predictions in logistic regression are categorical and usually dichotomous, such as predicting whether or not a patient has a disease based on underlying risk factors. The predicted binary dependent variable in this logistic regression is whether a particular offender will *recidivate within 2 years of their release date from jail*, given their set of scores on the predictor variables.

An important issue addressed within this analysis is whether the parsimonious actuarial (PAM) model can perform as well (or better) than the COMPAS when predicting whether an individual offender will recidivate within 2 years of their release from jail. COMPAS scores are used as a component of pre-sentencing investigation reports to help the judge quantify the individual offender's relative risk to the community in the form of recidivism. Examining the relationship between COMPAS scores and subsequent recidivism will allow us to determine whether the COMPAS instrument provides predictions that are better than predictions derived by random chance. Additionally, by constructing the PAM to use as a basis of comparison, we can determine whether the relatively complex and mysterious COMPAS model provides higher quality predictions of recidivism than a simplistic approach.

We will start here by building the PAM using four predictor variables established within the literature. We have previously examined these same predictor variables when building the simple linear regression and multiple regression models that predict time-to-recidivism, or how many days will elapse between an offender's release from jail and their subsequent incarceration. These four predictor variables include the *age of the offender* at release, the *gender of the offender*, the *number of prior offenses*, and the *charge degree of the offense* during the time of assessment (i.e. felony or misdemeanor). We have transformed some of this data in light of underlying distributions. With respect to the age of the offender at their release from jail, this data has been transformed into categorical data (under 25 years old, 25-45 years old, over 45 years old) rather than utilized as continuous data. We are also transforming the data on the number of prior offenses to the *natural log of prior offenses* based on the underlying frequency distribution.

Table 5.5 shows the results of comparing nested logistic regression models that employ a combination of the aforementioned predictor variables. All 3,100 observations in the training data set were used in the analysis for each model. In Model 1, the only predictor variable is the categorical transformation of the offender's age at release. Model 2 includes the contents of Model 1 as well as the gender of the offender as predictive variables. Finally Model 3 contains the variables from Model 2 as well as the natural log of prior offenses and the charge degree for the current offense, both indicators of the offender's criminal history. All of the predictor variables of interest remain statistically significant in Model 3, which will be used as the parsimonious actuarial model (PAM) for the purposes of this study.

Table 5.5: Comparing nested logistic regression models

Variables	Model 1 Age	Model 2 Age + Gender	Model 3 All Vars
Age 25-45	-0.211***	-0.201**	-0.565***
	(0.0809)	(0.0815)	(0.0897)
Age 45 or older	-0.809***	-0.811***	-1.127***
	(0.111)	(0.112)	(0.122)
Male		0.613***	0.344***
		(0.0972)	(0.104)
Log of prior offenses			0.659***
			(0.0370)
Felony			0.247***
			(0.0839)
Constant	0.0548	-0.452***	-0.629***
	(0.0626)	(0.102)	(0.121)
Observations	3,100	3,100	3,100
Pseudo R-squared	0.0131	0.0227	0.114
LR chi2	55.80	97.08	488.4

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Let's look more closely at the parsimonious actuarial model and the relationships between predictor variables and our outcome of interest.

Table 5.6: The Parsimonious Actuarial Model

VARIABLES	(PAM) 2YR Recidivism
Age 25-45	-0.565***
	(0.0897)
Age 45 or older	-1.127***
	(0.122)
Male	0.344***
	(0.104)
Log of prior offenses	0.659***
	(0.0370)
Felony	0.247***
	(0.0839)
Constant	-0.629***
	(0.121)
Observations	3,100
Pseudo R-squared	0.114
LR chi2	488.4

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Decision Tree

Given that we have both categorical and continuous data points, a decision tree may help us efficiently and effectively classify which offenders are likely recidivists. Decision trees are a type of supervised learning algorithm used primarily for fitting classification models. They are called trees because each branch node represents a test of the data, where each branch represents a potential outcome of the test, and each leaf represents a classification label. Graphically, a decision tree appears similar to a flowchart as seen in Figure

Decision trees have both advantages and disadvantages when used for an application such as predicting recidivism. It is helpful that decision trees are relatively easy to read and understand, even for criminal justice decision-makers without rigorous statistical training. Additionally, decision trees provide a great deal of transparency regarding their function to the extent they are considered "white-box models".

Despite these advantages, employing decision trees is not a flawless approach. Small changes in the training data may result in big changes to the decision tree's optimal structure. As such, decision trees may be expected to lack in accuracy when compared to other methods. To avoid overfitting the model to our training set – a tendency of decision trees – we should follow up this particular exercise with the random forest method.

Random Forest

As mentioned in the previous section, a single decision tree is unlikely to provide us the accuracy we expect and desire from a predictive model. This is because decision trees are very sensitive to the underlying data in the training set. However, it is possible to generate many decision trees by randomly selecting new bootstrapped training sets and considering new subsets of variables. This approach generates a wide variety of decision trees, hence the name random forest. We can evaluate the predictive value of the resultant random forest and seek to optimize its characteristics by repeating this process over and over again until we have the best random forest.

Although random forests typically provide more accurate predictions than a single decision tree, the interpretation of random forest models is much more complicated. Each tree in the random forest has been trained using "bagged data" and examining any one decision tree will not give us a complete understanding the random forest's functioning.

Therefore a random forest model functions as a "black-box model" that is less responsive to our concerns regarding transparency.

Support Vector Machines

Another popular method for developing predictive classification models is to use support vector machines. Support vector machines can be used as a basic classifier to sort observations into two categories. In this specific case, we are interested in sorting the recidivists from the non-recidivists. As with all methods of machine learning, there is an inevitable tradeoff between bias and variance when choosing a threshold for classification. If we choose a threshold that is too sensitive to the training data – i.e. low bias – it is likely to perform poorly when new data is applied – i.e. high variance. Likewise, thresholds that are less sensitive to the training data will perform better when we begin to use the model with new data. By using cross-validation, we can determine the best performing threshold by using a soft margin classifier, also known as a support vector classifier. The support vector classifier fits a hyperplane to the data to maximize its predictive performance.

When we have data that includes overlapping classifications, support vector classifiers struggle to perform well. Suppose for example we are interested in the impact of the length of stay in jail on recidivism and that too short or too long of a stay in jail has a negative impact on recidivism. The parabolic shape of this relationship cannot effectively be captured on a line. However, we can use support vector machines to explore the data in higher dimensions so that we may find a support vector classifier that will effectively separate the data into two groups. In this example, we can add a y-axis where the y-axis values are equal to the length of stay in jail squared. We could then draw a support vector classifier that separates the people who recidivated from those who did not recidivate.

But how do we determine the ideal transformation for the data? Support vector machines use kernel functions to methodically investigate each support vector classifier transformation in turn. There are many different types of kernel functions that can be used to transform the data. For example, the polynomial kernel has a parameter d for the degree of the polynomial. Returning to the previous example of a parabolic shape describing the relationship between length of stay in jail and recidivism, the value of d describes the dimensional transformation of length of stay in jail. If d=2, then the ideal transformation is length of stay in jail squared. Determining the ideal value for this transformation can be done using cross-validation.

The polynomial kernel is defined as:

$$k_d(x, z) = (< x, z > +\alpha)^d \tag{5.7}$$

Using the binomial theorem, the polynomial kernel expands to:

$$k_d(x, z) = \sum_{s=0}^{d} \begin{pmatrix} d \\ s \end{pmatrix} \alpha^{d-s} \langle x, z \rangle^s \tag{5.8}$$

K-Nearest Neighbors

The k-nearest neighbors algorithm is another option for building a classification model from our data. K-nearest neighbors is a non-parametric supervised machine learning approach that assumes that similar data points will cluster together and to quantify such similarity, calculates the distance between the points on a graph. These algorithms may be quite familiar to the reader who has previously shopped online or has used a video streaming platform to watch a film. Your previous data, such as the items you have contemplated for purchase or the films you have viewed in the past, is harnessed to make suggestions for future purchases or other films you would enjoy watching. Likewise, we can apply this approach to our data gathered from offenders. If we already have a lot of data that defines distinct categories such as "recidivist" or "non-recidivist" classification, we can use that data to predict what category a new individual observation is likely to fall in.

A key issue in the use of k-nearest neighbors is determining that appropriate value of K for the dataset in question. In order to select the best value of K, we will run the algorithm repeatedly using different values of K each time. The resulting output will allow us to choose the value of K that corresponds to the lowest number of errors while still providing accurate predictions with new data. Although relatively simple to implement, the k-nearest neighbors algorithm is not without its disadvantages. Larger datasets will cause the algorithm to need significantly more time to process and slow the return of classification results.

$$\operatorname{Pr}\left(Y=j \mid X=x_0\right)=\frac{1}{k} \sum_{i \in \mathcal{N}_0} I\left(y_i=j\right) \tag{5.9}$$

$$d(x, y)=\sqrt{\sum_{i=1}^p\left(x_i-y_i\right)^2} \tag{5.10}$$

K-Means Clustering

The k-means algorithm is used to partition a given set of observations into a predefined amount of k clusters. The algorithm as described by MacQueen (1967)MacQueen et al. (1967) starts with a random set of k center-points (μ). During each update step, all observations x are assigned to their nearest center-point (see equation 5.11). In the standard algorithm, only one assignment to one center is possible. If multiple centers have the same distance to the observation, a random one would be chosen.

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - \mu_i^{(t)} \right\|^2 \leq \left\| x_p - \mu_j^{(t)} \right\|^2 \; \forall j, 1 \leq j \leq k \right\} \tag{5.11}$$

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j \tag{5.12}$$

Afterwards, the center-points are repositioned by calculating the mean of the assigned observations to the respective center-points.

The update process reoccurs until all observations remain at the assigned center-points and therefore the center-points would not be updated anymore.

Neural Networks

At the time of this dissertation, neural networks have become one of the most popular algorithms in use by the broader machine learning community. Known as a "black-box approach", neural networks can

5.4 Predicting Recidivism within 2 Years

In this section, we will examine the predictive performance of the COMPAS model and the PAM model developed in the previous section. Here we are looking to see if the complex

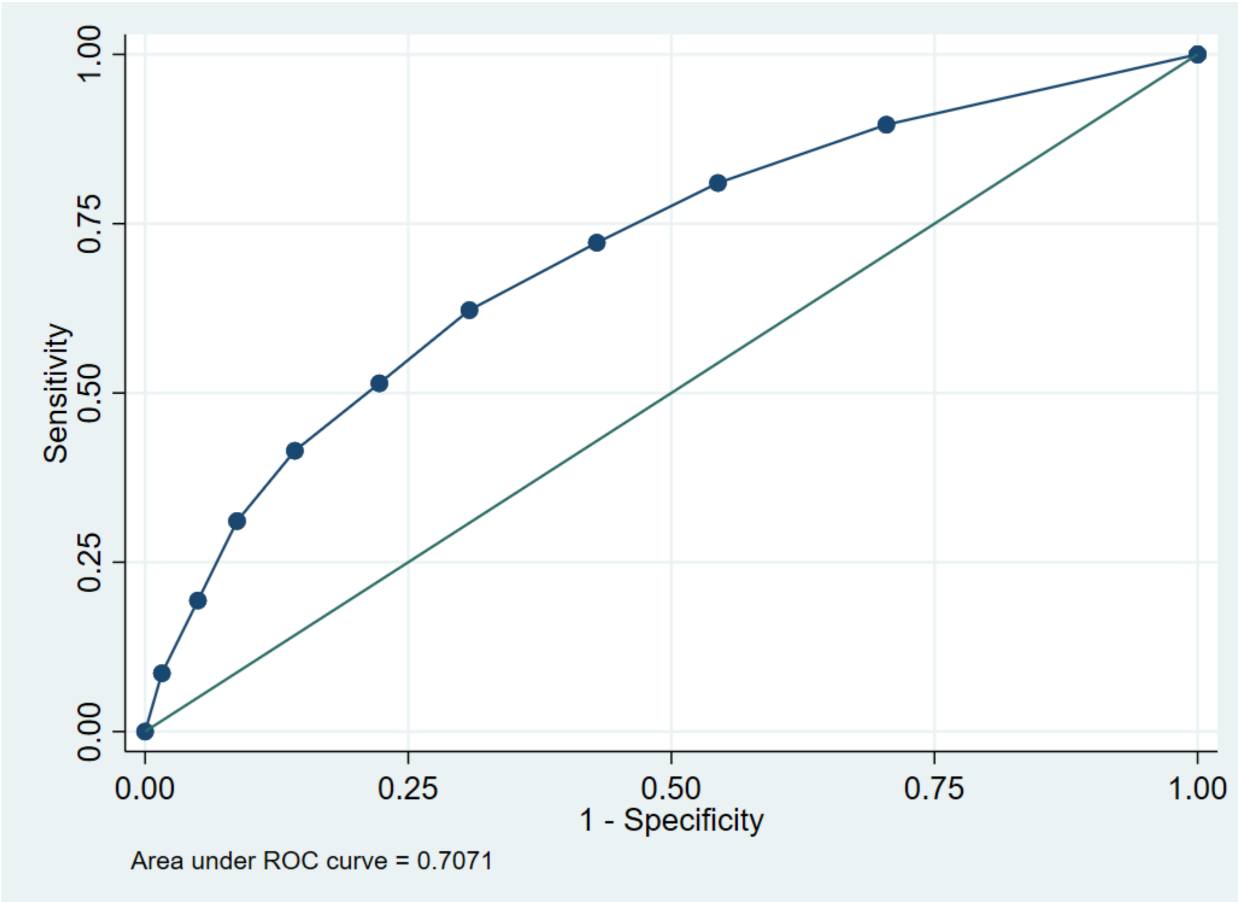


Figure 5.6: COMPAS ROC Area Under Curve

COMPAS model delivers predictions of a significantly different quality to the simple PAM model.

First, we will examine the COMPAS model by plotting the ROC area under the curve (AUC). The results of plotting this data are displayed in Figure 5.6. The ROC area under the curve value for the COMPAS instrument here is 0.7071. The generally-accepted cut-off values for predictive instruments is 0.70 so by this measure, the COMPAS does appear to have predictive value that exceeds random chance.

Next, we will examine the parsimonious actuarial model (PAM) by plotting the ROC area under the curve. The results of plotting this data are displayed in Figure 5.7. The ROC area under the curve value for the PAM instrument here is 0.7252. The PAM instrument

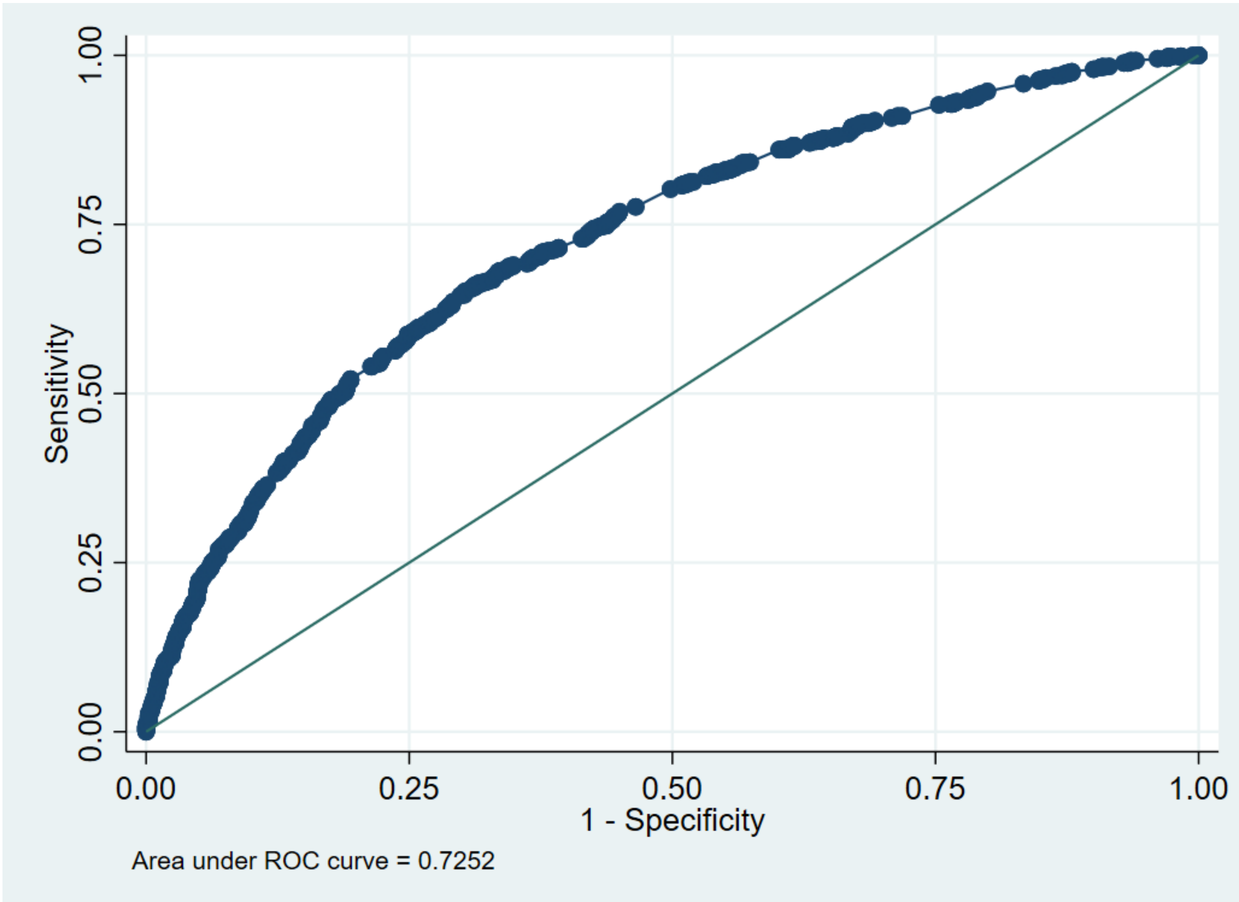


Figure 5.7: PAM ROC Area Under Curve

also appears to have predictive value that exceeds random chance as the ROC area under the curve exceeds 0.70.

Finally, we will plot both the COMPAS ROC area under the curve and the PAM ROC area under the curve. The results of plotting this data together are displayed in Figure 5.8. Here you can see that these models perform nearly identically in every decile, though the PAM appears to perform slightly better given these data.

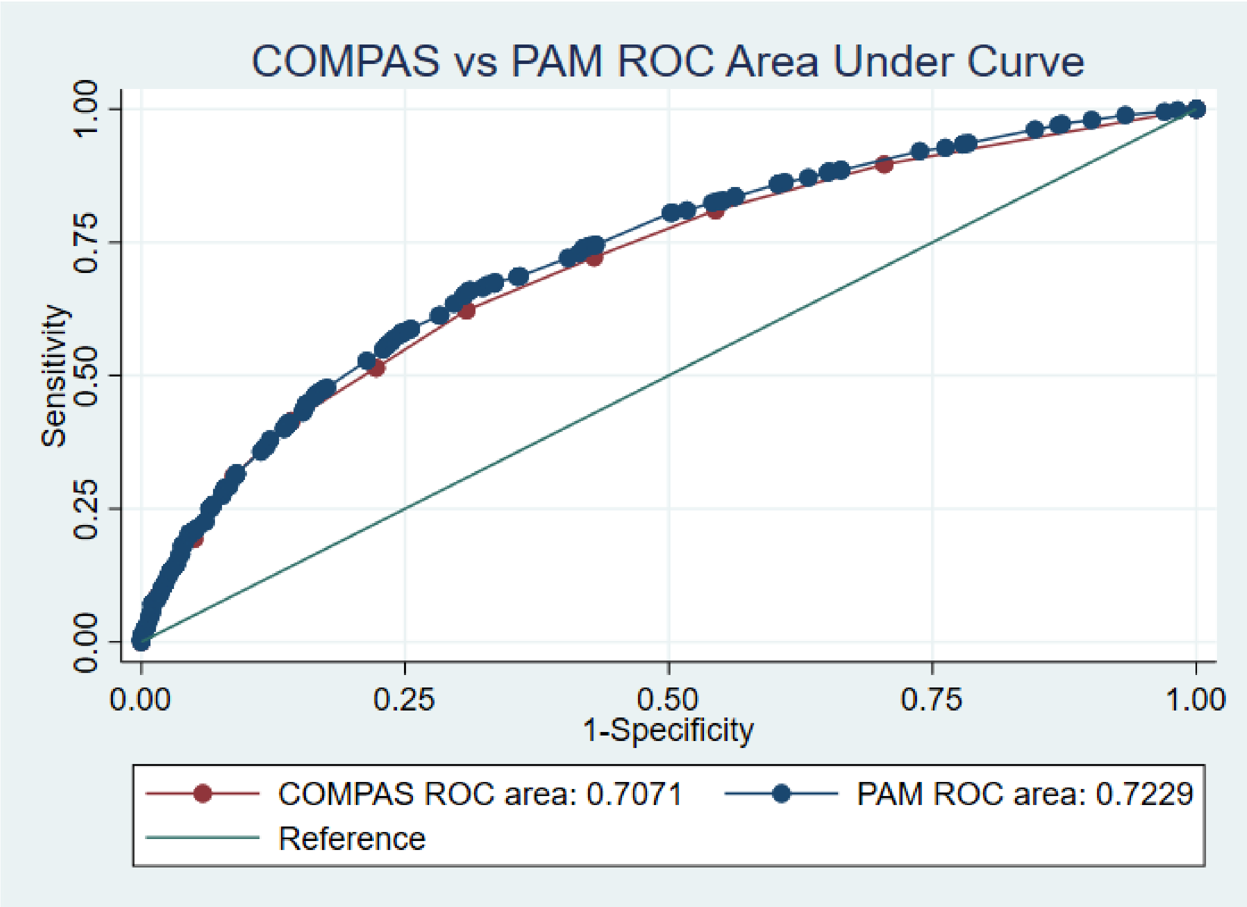


Figure 5.8: COMPAS vs. PAM ROC Area Under Curve

5.5 Predicting Violent Recidivism within 2 Years

5.6 Predicting Time-to-Recidivism

Table 5.7: Predicting time-at-risk with COMPAS
vs Parsimonious Actuarial Model

Variable	COMPAS TTR	PAM TTR
decile score	-53.84***	
	(2.348)	
under 25		-42.79*
		(19.07)
age at release		5.367***
		(0.748)
felony		-93.77***
		(14.13)
male		-46.56**
		(17.10)
priors		-27.00***
		(1.456)
_cons	780.9***	568.9***
	(12.36)	(34.13)
N	3098	3098
R-sq	0.145	0.155
adj. R-sq	0.145	0.154
rmse	369.9	368

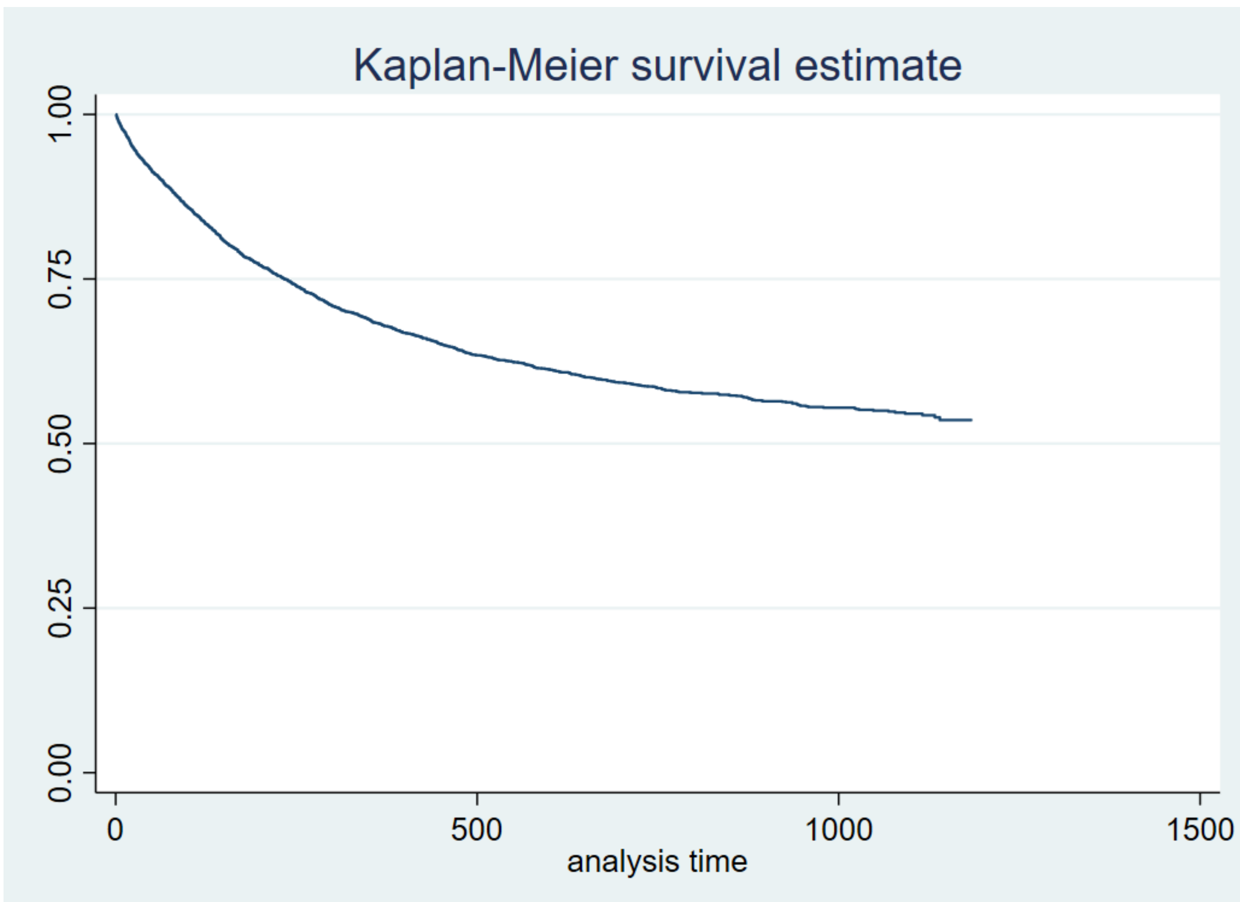
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

5.7 Non-Parametric Models

$$S(t) = \Pr(T > t) = 1 - F(t) \tag{5.13}$$

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t < T \leq t + \Delta t \mid T > t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{5.14}$$

$$H(t) = \int_0^t h(u)du \tag{5.15}$$



5.8 Cox-Proportional Hazards Model

5.9 Weibull Regression

Chapter 6

Public Safety Benefits of Prediction

6.1 Introduction

In this chapter we will discuss the public safety benefits that arise from the use of prediction. We will begin by addressing how prediction of human behavior works at present, by developing statistical models and using those models to classify or quantify future outcomes. Then we will turn to the subject of how these predictive models may be used in public safety applications. The benefits of prediction for taxpayers, offenders, and criminal justice officials will each be addressed in turn. We will then summarize these findings and propose future directions for scientific research on the subject of predictive modeling in public safety.

Many people at present, even laypersons, may be passingly familiar with the term "machine learning". However, relatively few of these people understand what constitutes machine learning or what goes into the design of a predictive algorithmic model. Yet at the same time, the average person engages with many forms of predictive algorithms on a near-constant basis. An email application may attempt to predict the words or phrases you will use next when composing a message, given the words you have already provided. A video streaming service may collect data on your viewing habits such as watch-time duration and preferred genres in order to suggest your next binge-watching indulgence. An online retailer may use your previous shopping data to categorize you as a consumer and to promote items that they think you will likely want to purchase.

However, the consequences of a poor-quality prediction may vary widely depending on the nature of what it is we are trying to predict. Your video streaming service may predict that you will enjoy a film that you turn out to loathe, yet no one will be seriously harmed by this technological misfire. On the other hand, predictive models that are used in the criminal justice system can have a real and profound impact on the individual person. In public safety, we may be using predictive models to determine who will be released back into the community pending their court date, for example. Getting it wrong here can potentially end with an innocent person or people being harmed or killed. To begin our discussion, let us first discuss how these predictive algorithms are designed and (ideally) maintained over time.

6.2 How Does Prediction Work?

The prediction of future behavior, including recidivism, shares many characteristics with more commonly familiar forms of prediction such as in meteorology. In weather forecasting, we can utilize historical data points to predict future climate events. To generate the highest quality predictions, a model must have reliable and valid underlying data as well as measurements of influential independent variables. For climate events, these may be variables such as temperature, pressure, humidity, and precipitation. Statisticians can gather large amounts of this ideally high-quality historical data and look for relationships between the independent variables and the outcome of interest (such as tomorrow's chance of rain).

Once the underlying relationships between variables can be adequately understood and mathematically captured, we can use the data to create a predictive model. This model development process is described in deep detail in Chapters 4 and then implemented in Chapter 5. To summarize, at the present time many models are developed by simultaneously using multiple machine learning techniques that employ different algorithmic approaches to generate a predictive model. Simple methods such as linear regression may be used alongside more sophisticated methods such as simulated neural networks. Multiple models result from this development process and the best model – however ”best” may be defined in its specific real-world application – is then ultimately selected from among these options for use in the field. We now have a predictive model with which we can generate predictions, given the input values of specific parameters.

Once the predictive model is generated and put into use out in the real world, it should also be maintained over time to promote the best possible accuracy and validity of its predictions. This maintenance includes two major functions: (1) evaluating the predictive value of the model given real-world data and (2) re-tuning the model as additional relevant data becomes available. With respect to evaluating the predictive value of the model, we can follow up on previously-made predictions and see if they essentially ”came true” or were accurate. If a weather model is predicting rain on certain future days, we can follow up and record when it does or does not actually rain. Then we can use this data to evaluate the predictive validity of that particular model. If a model ceases to regularly produce output of predictive value, or the predictive validity of the model falls below a certain threshold, it is no longer useful.

Re-tuning the model may be useful if there are special characteristics to the underlying data that differ substantially from the data originally used to create the model. In models predicting human behavior, there may characteristics of the original sample that do not represent the intended population with great accuracy. For example, if a model is designed with a very homogeneous sample such as 95% white people, it may not predict outcomes as accurately for out-groups within the original data such as people of color and indigenous persons. For this reason, we may want to avoid designing a single model that is used for a larger geographic area (such as the entire United States) and focus on a smaller scale such as individual states or densely-populated metropolitan areas like New York City or Los Angeles.

Additionally, there may be changes over time that occur concerning the weighted values of the underlying independent variables. Perhaps the influence of a particular variable waxes or wanes decade over decade. Returning to our earlier comparison to predicting the chance of rain, our original model may be impacted by a variable such as the rapidly-changing global average surface temperature (i.e. climate change). In a predictive model for recidivism, it may be more difficult to anticipate the precise nature of how this concern will become evident. Perhaps for example there is some shift in the definition of an independent variable itself, such as the way a specific offense's charge degree is defined as felony or misdemeanor.

6.3 Public Safety Applications of Prediction

-predictive policing/hotspots -predicting attrition from cj careers/post

6.4 Benefits of Prediction for Taxpayers

Any discussion of the benefits of prediction must begin with the humble taxpayer. Taxpayers are, after all, the ones who are fundamentally on the hook for the expenses incurred by our criminal justice system. The Bureau of Justice Statistics estimates that we spend somewhere in the neighborhood of $80 billion every year on incarceration alone. The total costs involved in the criminal justice process, from investigation to incarceration, may top $200 billion per year. In the era of mass incarceration and dwindling funding in the light of the sub-prime mortgage crisis of 2007-2008, cost-saving measures have become essential to state and federal agencies. The central question is how to "get the biggest bang for our buck" when on the taxpayer's dime.

Here in the state of California, the average cost to incarcerate an inmate in a state facility is over $100,000 per year. Such a large sum of money leads inquiring minds to wonder, what precisely are we paying for and is it working to prevent recidivism? Determining what we are paying for is a simple matter of looking at the cost breakdown helpfully provided annually by California's Legislative Analyst's Office. The primary expenditures that contribute to this massive price tag are security costs and health care services, collectively representing about 80% of the total amount spent per person. These expenses break down to an average of approximately $3,750 per month spent on security alone and an additional $2,750 on health care for a single inmate. The smallest amount of taxpayer money is allocated for rehabilitation programs such as academic education, cognitive behavioral therapy, or vocational training opportunities – a mere $300 per month is spent on these evidence-based programs that are so vital to preventing recidivism when inmates eventually return to their communities.

Indirect costs to the taxpayers also extend to the family members and friends of people who are incarcerated. Generally speaking, necessary goods and services are often substantially more expensive when behinds bars. Keeping in touch with loved ones by phone, mail, or electronic mail involves pricey middlemen. In 2021, the Washington Post reported that the average cost of a 15-minute phone call in correctional facilities is $5.74. The average hourly pay rate for prisoners varies widely by state but in California it ranges between $0.08/hr and $0.95/hr. Based on the highest possible inmate pay rate, a single call home from a CDCR facility could cost the inmate 6 hours of labor at minimum. As of 2021, inmates in CDCR facilities are now allowed one free 15-minute phone call every two weeks and the cost for paid 15-minute calls has been reduced to only \$0.38.

Electronic communication like e-mail or short video files may be available options but also come with a weighty markup in price. Special tablet computers, designed for use in jails and prisons, may be provided for free (as they presently are in California) or they may be purchased through the facility's commissary services. There are additional costs involved in actually sending or receiving electronic communications or downloading/streaming content from the internet. Inmates may need to purchase electronic "stamps" to access these services; the best bulk rate on these stamps for CDCR prisoners in 2021 was \$18.00 for 80 email stamps, or 22.5 cents per email. Video calls are available to CDCR prisoners for \$0.20 per minute after using up their free 15 minutes every two weeks. Despite these positive changes, communication with the outside world is largely available to the inmates who can afford it. Maintaining pro-social bonds through regular contact with family and friends may be crucial for resisting recidivism upon release.

Incarcerating fewer people will inevitably result on less taxpayer money being spent on carceral costs. California's jail populations have significantly declined in recent years, partly due to the passage of Proposition 47 – reclassifying certain low-level crimes from potential felonies to misdemeanors – and as a consequence of the COVID-19 pandemic.

6.5 Benefits of Prediction for Offenders

-lower risk of recidivism, not exposed to jail as a risk factor, cyclical issue, self-reinforcing

-maintaining employment and residence, family duties, schooling in the community

-better health outcomes

6.6 Benefits of Prediction for Criminal Justice Officials

-better/more accurate targeting of solutions

-additional information to help guide release decisions/custody for opt-out purposes

-correctional officers/staff

-less money spent on unnecessary incarceration means more money that can be spent on evidence-based solutions

-additional programming

-prison education

-additional training for officers/staff

Chapter 7

Conclusion

7.1 Data-driven Recidivism Prediction

In the age of big data, machine learning techniques have become popular for a wide variety of predictive applications. New approaches to methodically investigating patterns in data arrive each year. No single machine learning approach is "best" for understanding big data sets; rather, practitioners will try a multitude of options for finding a predictive model and then compare these models to one another. The selection of a particular model may be dependant not only on how successful it is at its stated task, but also on an appreciation for the specific ways in which the model tends to mispredict. Overprediction or underprediction can have a trivial or devastating impact, depending on the model's application.

-ML/algorithms have become de-facto method of prediction broadly speaking

-statistical prediction has the outward appearance of being "fair" or at least unbiased

-predictions are based on historical data, and the historical data may be innately influenced by structural bias

-historical over-representation of people of color and the poor in criminal justice data

-school-to-prison pipeline for juveniles in POC & poor communities

-consequentially these algorithms may be generating individual-level predictions that reflect a historical structural bias rather than the true nature of recidivism

-validity of predictions also may vary widely based on the individual's group membership or intersectional identity (e.g. white men vs. black women)

-in the criminal justice system, algorithmic bias and automation bias can combine to create unjust outcomes with the appearance of scientific credibility

7.2 Computer Says No

"Decision-makers increasingly face computer-generated information and analyses that could be collected and analyzed in no other way. Precisely for that reason, going behind that output is out of the question, even if one has good cause to be suspicious. In short, the computer analysis becomes a credible references point although based on poor data."

Daniel T. Brooks, Brandon Becker and Jerry R. Marlatt

In Computer Applications in Particular Industries: Securities

Figure 7.1: An issue that predates modern machine learning methods.

-garbage in, garbage out

-importance of collecting high-quality data

-reliability and validity

-comparison to racial and gender bias in facial recognition technology

-deep learning and neural networks will further complicate matters as modeling becomes more accurate but retains bias from the underlying data

7.2.1 The Dark Figure of Crime

-significant phenomenon of crimes that go unreported to the police

-value of the rates reported in the UCR undermined to some extent

-offset with the NCVS, victim self-reports capturing data that may go unknown to police or other official reporting methods

-traditional methods of gathering data on recidivism may need to be reassessed

-lessons to learn from the use of the NCVS to help estimate the dark figure of crime

7.3 Policy Considerations

-importance of transparency in criminal justice proceedings

-corporate secrecy as no place in determining or controlling a living person's liberty

-throw the doors open wide to this algorithmic star chamber

-procedural fairness and perception of legitimacy; responding to inquiries about methodology

-reviewing Loomis case and similar

-reinforcement of prejudice and bias does not result in safer communities, but can further feed into cumulative disadvantage

7.4 The Future of Recidivism Risk Assessment

Bibliography

Andrews, D. and Bonta, J. (2010). *The psychology of criminal conduct*. Matthew Bender, New Providence, NJ, 5th edition.

Andrews, D., Bonta, J., and Wormith, J. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, 52(7):7–27.

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Beck, A. and Shipley, B. (1989). Recidivism of prisoners released in 1983. us department of justice, bureau of justice statistics.

Berk, R. and Bleich, J. (2013). Statistical procedures for forecasting criminal behavior. *Criminology and Public Policy*, 12(3):25–32.

Blomberg, T., Bales, W., Mann, K., Meldrum, R., and Nedelec, J. (2010). Validation of the compas risk assessment classification instrument. Retrieved from www.criminologycenter.fsu.edu.

Bonta, J. (1996). *Risk-needs assessment and treatment*. SAGE, Thousand Oaks, CA.

Bonta, J. (2002). Offender risk assessment: Guidelines for selection and use. *Criminal Justice and Behavior*, 29:355–379.

Bonta, J. and Andrews, D. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation*. Ottawa, Public Safety Canada.

Bonta, J., Harman, W., Hann, R., and Cormier, R. (1996). The prediction of recidivism among federally sentenced offenders: A re-validation of the sir scale. *Canadian Journal of Criminology*, 38(1):61–80.

Borden, H. (1928). Factors for predicting parole success. *Journal of Criminal Law and Criminology*, 19(3):328–336.

Breitenbach, M., Dieterich, W., Brennan, T., and Fan, A. (2010). Creating risk-scores in very imbalanced datasets: Predicting extremely violent crime among criminal offenders following release from prison. In Koh, Y. and Rountree, N., editors, *Rare association rule* *mining and knowledge discovery: Technologies for infrequent and critical event detection*, page 231–254. IGI Global.

Brennan, T. and Dieterich, W. (2008). New york state division of parole compas reentry pilot study. Retrieved from.

Brennan, T., Dieterich, W., and Ehret, B. (2009). Evaluating the predictive validity of the compas risk and needs assessment system. *Criminal Justice and Behavior*, 36:21–40.

Brennan, T., Dieterich, W., and Oliver, W. (2004). *The COMPAS scales: Normative data for males and females. Community and incarcerated samples.* Northpointe Institute for Public Management, Traverse City, MI.

Brennan, T. and Oliver, W. (2000). *Evaluation of reliability and validity of COMPAS scales: National aggregate sample.* Northpointe Institute for Public Management, Traverse City, MI.

Burgess, E. (1928). Factors determining success or failure on parole. In Bruce, A., editor, *The Workings of the Indeterminate Sentence Law and Parole in Illinois*, pages 205–249. Illinois State Parole Board, Springfield, IL.

Clear, T. and Gallagher, K. (1985). Probation and parole supervision: A review of current classification. *Crime and Delinquency*, 31:423–443.

Cobbina, J., Huebner, B., and Berg, M. (2012). Men, women, and postrelease offending: An examination of the nature of the link between relational ties and recidivism. *Crime and Delinquency*, 58(3):331–361.

Collins, R. (2010). The effect of gender on violent and nonviolent recidivism: A meta-analysis. *Journal of Criminal Justice*, 38(4):675–684.

Copas, J. (1983). Plotting p against x. *Applied Statistics*, 32(1):25–32.

Copas, J. and Marshall, P. (1998). The offender group reconviction scale: The statistical reconviction score for use by probation offices. *Journal of the Royal Statistical Society, Series*, C 47:159–171.

Dawes, R., Fault, D., and Meehl, P. (1989). Clinical versus actuarial judgement. *Science*, 243:1668–1674.

Desmarais, S. and Singh, J. (2013). *Risk assessment instruments valiyeard and implemented in correctional settings in the United States.* Council of State Governments Justice Center, Washington, DC.

Durose, M., Cooper, A., and Snyder, H. (2014). *Recidivism of prisoners released in 30 states in 2005: Patterns from 2005 to 2010.* Washington, D.C, Bureau of Justice Statistics.

Farabee, D., Zhang, S., Roberts, R., and J, Y. (2010). Compas validation study: Final report. In *UCLA Integrated Substance Abuse Programs*. UCLA. Retrieved from.

Farrington, D. (1986). Age and crime. *Crime and justice: An annual review of research 7*, page 189–250.

Fass, T., Heilbrun, K., DeMatteo, D., and Fretz, R. (2008). The lsi-r and the compas: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, 35:1095–1108.

Gendreau, P., Grant, B., Leipciger, M., and Collins, S. (1979). Norms and recidivism rates for the mmpi and selected experimental scales on a canadian delinquent sample. *Canadian Journal of Behavioural Science*, 11(1):21–31.

Gendreau, P., Little, T., and Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34(4):575–607.

Glueck, S. and Glueck, E. (1930). *500 criminal careers*. A.A. Knopf, New York, NY.

Gottfredson, D. and Bonds, J. (1961). A manual for intake base expectancy scoring form (form cdc-be 61a).

Gottfredson, D. and Gottfredson, S. (1988). Stakes and risks in the prediction of violent criminal behavior. *Violence and Victims*, 3(4):247–262.

Grove, W., Zald, D., Lebow, B., Snitz, B., and Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12:19–30.

Hamilton, M. (2015). Risk-needs assessment: Constitutional and ethical challenges. *American Criminal Law Review*, 52:231–291.

Harcourt, B. (2007). *Against prediction: Profiling, policing, and punishing in an actuarial age*. University of Chicago Press, Chicago, IL.

Harcourt, B. (2015). Risk as a proxy for race: The dangers of assessment. *Federal Sentencing Reporter*, 27:237–243.

Hart, H. (1923). Predicting parole success. *Journal of the American Institute of Criminal Law and Criminology*, 14(3):405–413.

Hessick, C. and Hessick, F. (2011). Recognizing constitutional rights at sentencing. *California Law Review*, 99(1):47–94.

Hirschi, T. and Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology*, 89:552–584.

Hoffman, P. and Beck, J. (1974). Parole decision-making: A salient factor score. *Journal of Criminal Justice*, 2:195–206.

Holder, E. (2014). Speech to the national association of criminal defense lawyers 57th annual meeting and 13th state crimnal justice network conference. Retrieved from https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th.

Howard, P. (2017). Offender group reconviction scale. In Singh, J., Kroner, D., Wormith, J., Desmarais, S., and Hamilton), Z., editors, *Handbook of Recidivism Risk/Needs Assessment Tools*, page 231–240. John Wiley and Sons, Ltd, Hoboken, NJ.

Kang-Brown, J., Schattner-Elmaleh, E., and Hinds, O. (2019). People in prison in 2018. Retrieved from www.vera.org/publications/people-in-prison-in-2018.

Kemshall, H. (2003). *Understanding risk in criminal justice*. McGraw-Hill Education, Berkshire, England.

Kerr, H. (1971). Appendices to statement. In *Hearings before Subcommittee No. 3 of the Committee on the Judiciary, House of Representatives*. Government Printing Office, Washington, DC.

Kraemer, H. (2003). Current concepts of risk in psychiatric disorders. *Current Opinion in Psychiatry*, 16.

Landau, S. (1978). Thought content of delinquent and nondelinquent young adults: The effect of institutionalization. *Criminal Justice and Behavior*, 5:195–210.

Langan, P. and Levin, D. (2002). *Recidivism of prisoners released in 1994*. Washington, D.C, Bureau of Justice Statistics.

Latessa, E. and Lowenkamp, C. (2006). What works in reducing recidivism? *University of St. Thomas Law Journal*, 3(3):521–535.

Lauritsen, J., Heimer, K., and Lynch, J. (2009). Trends in the gender gap in violent offending: New evidence from the national crime victimization survey. *Criminology*, 47(2):361–399.

Lin, J., Grattet, R., and Petersilia, J. (2010). Back-end sentencing and reimprisonment: Individual, organizational, and community predictors of parole in sanctioning decisions. *Criminology*, 48(2):759–795.

Little, K. and Shneidman, E. (1959). Congruencies among interpretations of psychological test and anamnestic data. *Psychological Monographs: General and Applied*, 73:1–42.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Maltz, M. (1984). *Recidivism*. Academic Press, Inc, Orlando, FL.

Mamalian, C. (2011). *State of the science of pretrial risk assessment*. Bureau of Justice Assistance, Washington, DC.

Mann, R., Hanson, R., and Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, 22:191–217.

Maxfield, L., Harer, M., Drisko, T., Kitchens, C., Meacham, S., and Iaconetti, M. (2005). A comparison of the federal sentencing guidelines criminal history category and the us parole commission salient factor score.

Mears, D., Cochran, J., and Bales, W. (2012). Gender differences in the effects of prison on recidivism. *Journal of Criminal Justice*, 40(5):370 – 378.

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press, Minneapolis.

Monahan, J. (2006). A jurisprudence of risk assessment: Forecasting harm among prisoners, predators, and patients. *Virginia Law Review*, 92:391–435.

Monahan, J. and Skeem, J. (2014). The evolution of violence risk assessment. *CNS Spectrums*, 19:419–424.

Monahan, J. and Skeem, J. (2015). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, Virginia Public Law and Legal Theory Research Paper*, 53.

Northpointe (2013). Northpointe institute for public management. Retrieved from www.northpointe.com.

of Wisconsin v. Loomis, S. (2016). State of wisconsin v. loomis.

Ohlin, L. (1951). *Selection for parole*. Russell Sage, New York, NY.

Oleson, J. (2011). Risk in sentencing: Constitutionally suspect variables and evidence-based sentencing. *SMU Law Review*, 64(4):1330–1402.

Olusanya, O. and Cancino, J. (2012). Cross-examining the race-neutral frameworks of prisoner reentry. *Critical Criminology*, 20:345–358.

Pettinga, G. (1987). Rational basis with bite: Intermediate scrutiny by any other name. *Indiana Law Journal*, 62(3):779–803.

Piquero, A., Farrington, D., and Blumstein, A. (2003). *The criminal career paradigm*. University of Chicago Press, Chicago, IL.

Piquero, A., Jennings, W., Diamond, B., and Reingle, J. (2015). A systematic review of age, sex, ethnicity, and race as predictors of violent recidivism. *International Journal of Offender Therapy and Comparative Criminology*, 59(1):5–26.

Ritter, N. (2013). Predicting recidivism risk: New tool in philadelphia shows great promise. *National Institute of Justice Journal*, 271:4–13.

Sampson, R. and Lauritsen, J. (1994). Violent victimization and offending: Individual-, situational-, and community-level risk factors. In Reiss, A. and Roth, J., editors, *Understanding and Preventing Violence: Social Influences*, volume 1, page 17. National Academy of Sciences.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 66(3):178–200.

Schmidt, P. and Witte, A. (1988). *Predicting recidivism using survival models.* Springer-Verlag, New York.

Schmitt, J., Warner, K., and Gutpa, S. (2010). The high budgetary cost of incarceration. *Center for Economic and Policy Research.* Retrieved from www.cepr.net.

Schuessler, K. (1955). Parole prediction: Its history and status. *Journal of Criminal Law and Criminology*, 45(4):425–431.

Shulman, E., Steinberg, L., and Piquero, A. (2013). The age-crime curve in adolescence and early adulthood is not due to age differences in economic status. *Journal of Youth and Adolescence*, 42:848–860.

Simon, F. (1971). *Prediction methods in criminology.* Her Majesty's Stationery Office, London, England.

Skeem, J. and Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Sciences*, 20:38–42.

Sobol, W., Minton, T., and Harrison, P. (2007). *Prison and jail inmates at midyear 2006.* Bureau of Justice Statistics Bulletin, Washington. DC.

Spiropoulos, G., Salisbury, E., and Van Voorhis, P. (2014). Moderators of correctional treatment success: An exploratory study of racial differences. *International Journal of Offender Therapy and Comparative Criminology*, 58(7):835–860.

Starr, S. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, 66:803–809.

Sweeten, G., AR., S., and L. (2013). Age and the explanation of crime, revisited. *Journal of Youth and Adolescence*, 42:921–938.

Tangney, J., Stuewig, J., Furukawa, E., Kopelovich, S., Meyer, P., and Cosby, B. (2012). Reliability, validity, and predictive utility of the 25-item criminogenic cognitions scale (ccs. *Criminal Justice and Behavior*, 39:1340–1360.

Tarling, R. and Perry, J. (1985). Statistical methods in criminological prediction. In Ed. Farrington, D. and Tarling, R., editors, *Prediction in Criminology.* State University of New York Press.

Tollenaar, N. and Heijden, P. (2019). Optimizing predictive performance of criminal recidivism models using registration data with binary and survival outcomes. *PLoS ONE*, 14(3):1–37.

Tonry, M. (2014). Legal and ethical issues in the prediction of recidivism. *Federal Sentencing Reporter*, 26(3):167–176.

Walters, G. (2016). Predicting recidivism with the criminal sentiments scale: A meta-analysis of a putative measure of criminal thought content. *Criminal Justice and Behavior*, 43(9):1159–1172.

Walters, G. and Lowenkamp, C. (2016). Predicting recidivism with the psychological inventory of criminal thinking styles (picts) in community-supervised male and female federal offenders. *Psychological Assessment*, 28(6):652–659.

Warner, S. (1923). Factors determining parole from the massachusetts reformatory. *Journal of Criminal Law and Criminology*, 14:172–207.

Wisser, L. (2019). Pandora's algorithmic black box: The challenges of using algorithmic risk assessments in sentencing. *American Criminal Law Review*, 56:1811–1832.

Witte, A. (1992). Improving predictions of offender recidivism and patterns of offender crime. *National Bureau of Economic Research, National Institute of Justice. NIJ*, page 89– – –0010.

Wright, E., Van Voorhis, P., Salisbury, E., and Bauman, A. (2012). Gender-responsive lessons learned and policy implications for women in prison: A review. *Criminal Justice and Behavior*, 39(12):1612–1632.

Appendix A

Appendix Title

Supplementary material goes here. See for instance Figure A.1.

A.1 Lorem Ipsum

dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

"No man ever became extremely wicked
all at once."
Juvenal (c.55-c.140), Roman poet.
In Satire II, l.83.

Figure A.1: A deep quote.

“Many commit the same crime with a very different result.
One bears a cross for his crime; another a crown.”

Juvenal (c.55-c.140), Roman poet.
In Satire XIII, l.103.

Figure A.2: Another deep quote.