THE COLOR OF RISK:
UNPACKING THE IMPLICATIONS OF ACTUARIAL RISK
PREDICTION AT SENTENCING

A Dissertation in
Criminology
by
Julia A. Laskorunsky

©2017 Julia A. Laskorunsky

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

August 2017
The dissertation of Julia A. Laskorunsky was reviewed and approved* by the following:

Jeffery Ulmer
Professor of Criminology and Sociology
Dissertation Advisor
Chair of Committee

Eric Baumer
Professor of Criminology and Sociology

Christopher Zorn
Professor of Political Science

Gary Zajac
Director, Justice Center for Research
Department of Sociology and Criminology

Jeremy Staff
Professor of Criminology and Sociology
Director, Graduate Program in Criminology

*Signatures are on file in the Graduate School
Abstract

This dissertation examines the practice of using actuarial risk prediction to inform sentencing decisions and assesses its implications for racial disparity in sentencing. I use a large sample of serious offenders (N=7,935), convicted and released in Pennsylvania, to construct and validate a risk assessment instrument and analyze the relationship between race, risk score, and recidivism. In the first part of the analyses, I introduce an approach to develop a static risk assessment instrument that judges can use at sentencing. Using the development sub-sample, a series of bivariate and multivariate logistic regression analyses are used to identify significant covariate patterns among convicted individual with a new arrest or parole revocation within three-years of release. I identify eight parsimonious risk factors associated with recidivism and use a modified Burgess scoring method to construct an 11-point additive risk scale. Scores are collapsed into Low, Medium, High, and Very High risk levels to maximize interpretability. Finally, the instrument is validated on the validation sub-sample. The area under the curve values (AUC) for the risk score are .72 for the full validation sample and .70 for both the White and Black offender sub-samples — showing a medium to strong ability to discriminate between recidivist and non-recidivism.

In the second part of the analyses, I conduct disparate impact and instrument bias analyses to determine the instrument’s predictive utility across race. I use a series of bivariate and multivariate logistic regression models to examine the extent to which race is associated with recidivism, net of other factors, and to test for instrument fairness and interaction effects between risk factors and race. Results show that Black offenders receive higher average risk scores on the instrument than White offenders, mostly due to differences in their criminal histories. However,
the instrument slightly under-predicts the recidivism rate for Black offenders, and over-predicts
the recidivism rate for White offenders, because race is significantly correlated with recidivism,
but not included in the scoring model. I situate the consideration of risk in the focal concerns
perspective of sentencing and discuss the policy implication of using actuarial risk assessment
tools to structure sentencing decisions.
# Table of Contents

List of Tables ......................................................................................................................... viii
List of Figures ......................................................................................................................... ix
Acknowledgments ................................................................................................................... x

Chapter 1: Introduction — Actuarial Risk Prediction in Sentencing ........................................ 1
  Background and Importance of the Problem ........................................................................... 2
  Gaps in Research ..................................................................................................................... 5
  Research Questions ................................................................................................................ 9
  Project Significance ............................................................................................................... 10
  Conclusion .............................................................................................................................. 14

Chapter Two: Literature Review ............................................................................................. 15
  Overview ................................................................................................................................. 15
  History of Risk Assessment in Sentencing ............................................................................ 17
  Modern Sentencing Risk Assessment Tools and Policies ...................................................... 19
  The Risk-Needs-Responsivity Model in the Courtroom ........................................................ 22
  Integrating Risk Assessment into Sentencing ....................................................................... 24
  Judicial Philosophy and Consideration in Sentencing .......................................................... 28
    A Case Study ......................................................................................................................... 31
  Race, Risk, and Criminal History .......................................................................................... 33
  Risk Assessments: Disparate Impact and Bias ....................................................................... 36
  Risk Assessment Under Sentencing Guidelines .................................................................... 40
  Conclusion .............................................................................................................................. 42

Chapter Three: Data ................................................................................................................. 43
  Sample ................................................................................................................................. 43
  Missing/Incomplete Data ....................................................................................................... 43
  Justification for Sample ........................................................................................................ 44
    Level 5 Offenders ................................................................................................................. 45
    Follow-up Time .................................................................................................................... 46
  Pennsylvania's Sentencing System ....................................................................................... 47
    Pennsylvania Sentencing Guidelines ............................................................................... 47

Chapter 4 - Risk Assessment Development and Validation ................................................... 49
  Overview ............................................................................................................................... 49
Chapter 5 – The Relationship Between Race, Risk and Recidivism

Literature Review ........................................................................................................... 104

Disparate Impact and Criminal History ......................................................................... 104

Predictive Bias ................................................................................................................ 107
List of Tables

Table 4.1: Predictor Variables........................................................................................................65
Table 4.2: Descriptive Statistics (Full Sample) ..................................................................................70
Table 4.3: Bivariate Analysis for Predictor Identification....................................................................73
Table 4.4: Multivariate Logistic Model of Recidivism (Development Sample) ...............................76
Table 4.5: Risk Instrument Scale (0-10) .............................................................................................81
Table 4.6: Risk Score Distribution (Development Sample) ................................................................82
Table 4.7: Average Risk Score by Failure (Development Sample) ....................................................82
Table 4.8: Risk Levels (Development Sample) ....................................................................................84
Table 4.9: Multivariate Logistic Model of Recidivism (Validation Sample) ......................................86
Table 4.10: Risk Score Distribution (Validation Sample) ...................................................................87
Table 4.11: Average Risk Score by Failure (Validation Sample) .......................................................88
Table 4.12: Risk Levels (Validation Sample) .......................................................................................89
Table 4.13: AUC-ROC Between Risk Score and Re-arrest ...............................................................92

Table 5.1: Select Descriptive Statistics (Black/White Sample)............................................................111
Table 5.2: Recidivism by Risk Factor (Black/White Sample) .............................................................112
Table 5.3: Multivariate Logistic Model of Recidivism by Race ..........................................................114
Table 5.4: Multivariate Logistic Model of Recidivism with Interaction Effect .................................116
Table 5.5: Average Risk Score by Race (Black/White Sample) .........................................................117
Table 5.6: Mean Score Differences by Race for Each Risk Factor (Black/White Sample) ................119
Table 5.7: Failure Rate, AUC, for Recidivism (White/Black Sample) ..............................................121
Table 5.8: Failure Rate, AUC, for Recidivism (White/Black Sample) ..............................................122
Table 5.9: Average Risk Score by Failure (Black/White Sample) ....................................................124
Table 5.10: Logistic Model Predicting Recidivism and Testing Racial Fairness by Risk Score ........125
List of Figures

Figure 4.1: Recidivism Rate at 1, 2, and 3 Years ................................................................. 71
Figure 4.2: Recidivism by Risk Score (Development Sample) ............................................. 83
Figure 4.3: Recidivism by Risk Level (Development Sample) .............................................. 85
Figure 4.4: Recidivism by Risk Score (Validation Sample) .................................................... 88
Figure 4.5: Recidivism by Risk Level (Validation Sample) .................................................... 90
Figure 4.6: Survival Curve by Risk Level (Validation Sample) ............................................. 91
Figure 4.7: AUC-ROC Curve (Development Sample) ........................................................... 92
Figure 4.8: AUC-ROC Curve (Validation Sample) ............................................................... 93

Figure 5.1: Risk Level Membership by Race ........................................................................ 118
Figure 5.2: Failure Rate by Risk Level ................................................................................ 122
Figure 5.3: Adjusted Probabilities of Recidivism by Race .................................................... 126
Acknowledgments

I received assistance from a great number of people along my path into academia. This dissertation is a testament to their time, effort, and patience. First, I want to thank my advisor and committee chair, Jeff Ulmer, for taking me on as a graduate student and supporting my efforts during this long process. Not only has he provided invaluable feedback on my work, but he has also shaped my research interests as a scholar. Thank you for your unwavering support along the way. I look forward to years of future collaborations as your colleague.

I am also thankful to the rest of my committee members, Eric Baumer, Chris Zorn, and Gary Zajac, who were always available to provide feedback and suggestions on the dissertation, the job market, and academic life in general. It has been a privilege to bring this work to fruition having the attention of such an esteemed group of scholars. I’ve learned much from you along the way.

I am also grateful to Doris MacKenzie who first supported my interest in risk assessment, and applied research in general. Thank you, Doris, for your guidance along the way and for having impeccable taste in postdocs. I am also indebted to the Justice Center for Research, first under Doris MacKenzie, then, Gary Zajac, for serving as my home base for the past 6 years. The opportunities I’ve received and the connections I’ve made as a member of the Center have been invaluable.

I have been lucky enough to make a number of life-long friends during my time at Penn State. They have served as my cheerleaders, confidants, and much needed distractions from the intensity of grad school. Thank you for being there during both the stressful times and the brighter days. Thank you also to my Mom and my sister who have had to bear the brunt of my absence. They kept me grounded and provided unwavering support.

Finally, much of this would not have been possible without my partner in crime, Matthew DeMichele. If I listed all the ways he has helped and influenced me during my graduate career, the acknowledgments section would be as long as the dissertation. Thank you for your constant support, kinds words, and patience along the way.
Chapter 1: Introduction — Actuarial Risk Prediction in Sentencing

"Here in Pennsylvania and elsewhere, legislators have introduced the concept of "risk assessments" that seek to assign a probability to an individual's likelihood of committing future crimes and, based on those risk assessments, make sentencing determinations. Although these measures were crafted with the best of intentions, I am concerned that they may inadvertently undermine our efforts to ensure individualized and equal justice. By basing sentencing decisions on static factors and immutable characteristics – like the defendant's education level, socioeconomic background, or neighborhood – they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society."

Attorney General Eric Holder
Philadelphia, PA 2014

Scholars have documented the disproportionate representation of Black and Hispanic offenders in the criminal justice system (Alexander, 2012; Pettit & Western, 2004) and have identified the sentencing stage as one point of disparity (Mitchell, 2005; Spohn, Beichner, & Davis-frenzel, 2012; Ulmer, 2012; Zatz, 2000). Swayed by the possibility of lowering incarceration through efficient and individualized offender management, states have increasingly relied on actuarial risk assessments in the sentencing process. These instruments allow courts to statistically assess an offender’s likelihood of recidivism based on a number of risk factors, such as criminal history and offender demographics. In response, recent research has focused on the role actuarial risk assessments play in sentencing decisions, with some scholars presenting a generally favorable view of the practice (Monahan & Skeem, 2014; Skeem & Lowenkamp, 2016) and others urging caution (Hannah-Moffat, 2013; Starr, 2014). The recent expansion in consideration of risk in punishment decisions raises questions about the predictive utility of sentencing risk assessment tools and the potential for disparate impact on minorities. Examining the role of both actuarial and informal risk assessment in legal decision-making is important for understanding the over-representation of minority offenders in the criminal justice system. As a
new framework for structured sentencing, risk assessments adoption also creates tension with existing sentencing guidelines and provides a new, and possibly superior, way to inform judges’ focal concerns.

This dissertation explores the implications of moving toward a sentencing structure with an actuarial consideration of risk by putting risk assessment into a broader context of judicial decision-making and racial sentencing disparity. Using a sample of serious offenders who were convicted and released in the state of Pennsylvania from 2001-2005, I create and validate a judicial risk assessment instrument and analyze the relationship between risk score, race, and recidivism. This manuscript is laid out as follows: this chapter provides an overview of the problem, the aims of this study, and significance of the project. Chapter Two broadly summarizes the current research and debate surrounding the consideration of recidivism risk in the courtroom. Chapter Three includes the data source used for analysis and provides an overview of sentencing practices in the state of Pennsylvania. In Chapter Four, I identify which case and offender characteristics best predict recidivism at the sentencing stage and constructs a risk assessment instrument that judges would be able to use at sentencing. In Chapter Five, I outline the relationship between race, instrument risk score, and recidivism and provide insight on whether the use of actuarial risk assessment has the potential to affects levels of racial disparity in sentencing. Finally, Chapter Six, provides theoretical and policy implications for the use of actuarial risk prediction at sentencing and makes suggestions for future research.

Background and Importance of the Problem

Since the 1970s the U.S. has experienced an unprecedented growth in incarceration. At its peak in 2008, one of every one hundred people in the US was incarcerated (Gelb, 2017). This
incarceration rate dwarfs the rates found in Western European countries, despite comparable rates of criminal victimization (Lappi-Seppälä, 2008). Mass incarceration has not brought the crime reduction benefits once promised (National Research Council, 2014), but rather has exalted a significant financial and human toll on the American public. Over the last 15 years, states have slowly begun to implement policies to reduce their reliance on incarceration. From 2010-2015, for the first time in decades, the corrections population was reduced (Gelb, 2017), leading some scholars and activist to wonder if the start of "mass de-incarceration" had begun. However, a 2016 uptick in the homicide rate, isolated to several major cities, complicated the push for large scale de-incarceration (M. Friedman, Grawert, & Cullen, 2016). Policy makers questioned the wisdom of letting more potential future offenders go free. In particular, Attorney General Jeff Sessions issued a memo instructing federal prosecutors to charge offenders with “…the most serious, readily provable offense.”, including those that carry mandatory minimum sentences (Sessions, 2017).

One way in which states have attempted to reduce their correctional population is by relying on "smart sentencing" (Etienne, 2009; Peters & Warren, 2006; Virginia Criminal Sentencing Commission, 2004). Smart sentencing generally focuses on the effectiveness of sanctions in reducing both recidivism and cost of incarceration. It is part of a larger evidence-based movement in criminal justice, which has until recently been largely missing from the courtroom. A myriad of studies assessed the effects of prison on recidivism, the effectiveness of alternative sanctions, and how well juveniles fare when sentenced as adults (e.g., Bishop, Frazier, Lanza-Kaduce, & Winner, 1996; Nagin & Snodgrass, 2013; Padgett & Bales, 2006). Thus, some major policy changes at the state levels, such as the slow repeal of mandatory minimum laws and the increase of minimum juvenile age for waivers have been research-
supported. And while supporting research is limited, states have also turned to the use of actuarial risk assessment instruments in the courtroom to better match individual offenders with sanctions.

Currently, about a dozen states use actuarial recidivism risk assessment instruments to inform sentencing for the general offender population (e.g., Utah, Virginia, Wisconsin, Missouri) and many other states use instruments to sentence sex offenders or domestic violence offenders (Hannah-Moffat, 2013; Starr, 2014). Both the construction of these instruments and the policies that dictate their use vary from state to state. However, the overarching purpose is the same — to inform judges of offenders’ risk of recidivism prior to sentencing. How judges use that information has been a source of much debate (Ruback, Kempinen, Tinik, & Knoth, 2016; Skeem & Lowenkamp, 2016; Starr, 2014). Some scholars warn that the consideration of risk scores in the courtroom leads to harsher sentences for minority offenders, thus increasing racial disparity in sentencing and in the criminal justice system (Harcourt, 2015; Starr, 2014). Others contend that minorities already receive harsher sentences as a result of legal factors and informal consideration of risk, thus the effect of using a formal risk assessment instrument are indeterminate (Skeem & Lowenkamp, 2016). The general accuracy of these instruments has also been questioned (Angwin, Larson, Mattu, & Kirchner, 2016; Harcourt, 2015). For example, questions remain about the best methods to construct a risk assessment instrument for the sentencing stage and how well risk of recidivism can be predicted at the sentencing stage.

Ultimately, state governments want to decrease rates of incarceration without reducing

---

1 It is difficult to know exactly how often actuarial risk assessments play a role in judicial decisions. Many states include risk assessment information from prior system involvement in pre-sentencing reports, but do not explicitly tie it to sentencing.
public safety. Risk assessment instruments offer the hope of differentiating offenders based on their predicted level of re-offending at the sentencing stage. In theory, if the riskiest offenders received the most severe punishment, while those on the other end of the spectrum received the least severe punishment (within the limit prescribed by law), states could focus their de-incarceration efforts on those least likely to re-offend. However, in reality, changes to sentencing policy and practice rarely work as intended and can have a number of unintended consequences.

Gaps in Research

Increasingly, courts are using actuarial risk assessment tools to differentiate between offenders based on their risk of committing a later crime (Starr, 2014), particularly for serious offenders who would normally end up incarcerated (Virginia Criminal Sentencing Commission, 2004). Despite the proliferation of actuarial sentencing risk assessments, questions remain about which factors best predict risk of recidivism at the sentencing stage and the relationship between risk score and race. Legal scholars, in particular, worry that the proliferation of this trend will increase disparities in the criminal justice system (Hannah-Moffat, 2013; Monahan & Skeem, 2014; Starr, 2014). To this end, I will outline the two research areas that remain under-examined: the instrument creation process and the relationship between instrument score and race.

Actuarial risk assessments have been used in criminal justice since Burgess (1928) developed a 21-factor instrument for the Illinois State Parole Boards to determine parole release. Yet, the reliance on of actuarial instruments during the sentencing process has only expanded in the last twenty years (Harcourt, 2015; Monahan & Skeem, 2014). While a handful of states (e.g., Pennsylvania, Virginia, Missouri), developed their own risk assessment instruments based on their offender population, most states use previously validated "off the shelf" risk assessments,
such as the Level of Service Inventory-Revised [LSI-R] and Correctional Offender Management Profiling for Alternative Sanctions [COMPAS]. Given the resources needed to develop a unique validated instrument, this decision to use a generic instrument is understandable. However, but there are several issues with this approach.

First, generic risk assessment instruments were developed using a non-native offender population which may not match the offender population in the state or jurisdiction for which the instrument is adopted. While the instruments may (and should) be validated on the population of that particular state, they are not reflective of the demographic makeup of each state's offender population and do not reflect the criminal justice processes unique to each state. Second, the structure of the instrument may not align with the purpose of each state's risk assessment policy. For example, some states use risk assessments to reduce recidivism, a task which requires a more complex instrument than one intended to simply predict recidivism. Other states may focus their risk assessment efforts only on serious offenders, making a risk assessment instrument developed using a general offending population inappropriate. Finally, many states rely on risk assessment tools which use proprietary algorithms to determine risk scores. This is the case with both the COMPAS and the LSI-R, both of which are popular instruments to use at sentencing. These instruments are owned by private companies which sell licenses to use the tool and accompanying software. Although jurisdictions buy the right to use the tools, they don't actually have a right to examine how each risk factor is weighted and how the final risk score is determined. This presents legal and ethical challenges in a court of law, because these tools fall outside of the usual evidentiary rules of discovery.

Methodological literature on the development of risk assessments for violence and re-offending is well established, particularly in the field of psychology (e.g., Douglas & Skeem,
2005; Monahan & Steadman, 1996). However, the literature on risk instrument creation for use at the sentencing stage is limited to two major projects: 1) Virginia's development of a risk assessment instrument to divert low-risk, serious offenders from sanctions (Ostrom, Kleiman, Cheesman, Hansen, & Kauder, 2002), and 2) efforts by the Pennsylvania Commission on Sentencing in the creation of Pennsylvania's risk assessment instrument (Pennsylvania Commission on Sentencing, 2016). To that end, the first research issue surrounding the use of these instruments is the lack of guidance in creating actuarial risk assessment instruments that are specific to native jurisdictions. Jurisdictions would benefit from creating localized instruments, yet the lack of research on the methodological process used for instrument construction hinders future development efforts. For example, the creation of a sentencing risk assessment instrument would allow for the identification of offender and case characteristics that best predict risk at the sentencing stage — as opposed to relying on more generic correlates of recidivism. Criminology, and related fields, has a broad understanding of which factors are associated with criminality (e.g., parental abuse), but these factors may not be available at the sentencing stage. Furthermore, factors that are unique to the sentencing stage (e.g., conviction offense) may be important for predicting offending behavior, but lack a strong underlying research base.

The issue of predictive accuracy also remains a much-debated topic in this area, to which the creation of a new risk assessment instrument can contribute. Because no risk assessment can predict offender behavior with 100% accuracy, a question to address is, "How accurate is accurate enough?" Actuarial risk assessments are expected to predict at least better than chance, and scholarship on this topic shows that they predict better than informal risk prediction (S. D. Gottfredson & Moriarty, 2006; Meehl, 1954). However, because there are so few examples of sentencing risk assessment instrument construction and validation efforts, we lack a good
understanding of what acceptable or expected predictive validity is.

Any changes to sentencing policy, such as the adoption of a risk assessment tool at sentencing, has the potential to affect racial disproportionality in sentencing by affecting the sentencing structure and judicial discretion. Critics of these assessment tools contend that basing sentences on recidivism risk factors will lead to greater racial disparity because marginalized populations – particularly racial and ethnic minorities — will score higher on risk assessment instruments (M. Hamilton, 2015; Hannah-Moffat, 2013; Harcourt, 2015; Starr, 2014). Despite this criticism, no states in which sentencing risk assessment technology has been adopted have conducted a post-adoption disparity analysis. Thus, the second major problem in the area of actuarial risk assessment at sentencing is a dearth of research on the relationship between actuarial risk score and race (however, see Flores, Bechtel, & Lowenkamp, 2016 and Skeem & Lowenkamp, 2016 for two recent examples). Two separate, but related, questions need to be answered: First, to what extent are minority offenders receiving higher risk scores compared to White offenders on risk assessments used in the courtroom? The answer to this question would inform what effect risk assessments may have on judicial perceptions of risk and sentencing decisions. Second, how well do scores reflect risk of recidivism across racial groups? The experiences of Black and White offenders in the criminal justice system are qualitatively different. Decades of research has shown that Black offenders are not only over-represented in offending, but are punished more harshly for similar offenses and targeted at a greater rate than White offenders (Frase, 2014; National Research Council, 1986b; Sampson & Laub, 1997; Ulmer, Painter-Davis, & Tinik, 2016). Thus, the relationship between offender and case characteristics and risk may be different for Black
and White offenders.

In light of intense criticism, the criminal justice system has faced regarding racial disproportionality, it is surprising that the use of actuarial tools in the courtroom has not been better evaluated. The sentencing stage has been identified as a point of disparity for minority offenders (Hagan & Bumiller, 1983; Mitchell, 2005; Spohn, 2000; Steffensmeier, Ulmer, & Kramer, 1998), thus changes to sentencing policy and practice should be accompanied by research on potential disparate effects. Instead, states have embraced this new technology without first considering how it may perpetuate longstanding inequalities evidenced in the criminal justice system.

**Research Questions**

To address gaps in knowledge outlined in the previous section, the purpose of this research project is to create and validate a recidivism risk assessment instrument for serious offenders and to analyze the relationship between race, risk score, and recidivism. This two-part analysis is organized around five main questions:

**Part 1: Risk Assessment Development and Validation**

1. Which offender and case characteristics best predict re-offending at the sentencing stage?

2. How accurately can risk of re-offense be predicted at the sentencing stage?

Part one of the study involves creating a static risk assessment instrument using variables that judges in Pennsylvania have access to at sentencing. Using a large sample of serious offenders convicted and released in the state of Pennsylvania, I identify case and offender
characteristics that have a significant relationship with recidivism and build a risk prediction algorithm in order to classify offenders based on their likelihood to reoffend. Using a separate sample of offenders, I validate the predictive accuracy of the instrument to determine how well it discriminates between recidivists and non-recidivists.

Part 2: The Relationship Between Race, Risk, and Recidivism

1. What is the relationship between risk score and race?
2. Which risk factors contribute most to mean score differences between Black and White offenders?
3. How well does the instrument predict recidivism for Black and White offenders?

Part two of the study will outline the relationship between risk score, race, and recidivism by analyzing mean group differences in risk score and examining the extent to which race is correlated with high risk scores. Subsequently, the study will identify which individual risk factors that explain the differences in average risk scores between Black and White offenders. Finally, I determine how well the instrument predicts recidivism among White and Black offenders and compare racial parity in the relationship between risk score and recidivism.

Project Significance

This dissertation makes a number of contributions to sentencing and corrections research, and provides empirical evidence for criminal justice policy decisions. First, it responds to a number of calls by legal and criminology scholars and practitioners to examine risk assessment
at sentencing, particularly as a potential contributor to racial disparity (Angwin et al., 2016; Hannah-Moffat, 2013; Harcourt, 2015; Holder, 2014; Skeem & Lowenkamp, 2016; Starr, 2014). Risk assessment tools have been lauded as way to reduce prison populations, but these reductions may not come without consequence. Research shows that marginalized groups tend to score higher on risk assessment instruments because they present more risk factors (Hannah-Moffat, 2013; Skeem & Lowenkamp, 2016), and legal scholars caution that the introduction of a risk score into the sentencing process will encourage judges to sentence those with high risk scores (who are disproportionately minorities) to harsher sentences than normal (Starr, 2014). As such, it is possible that overall reductions in the incarceration population may accompany increases in sentencing disparity. However, sentencing research has also shown that the use of criminal history in sentencing decisions disadvantaged minorities because it is unequally distributed by race. As a risk factor, criminal history is highly predictive of reoffence, and as such has become a “bedrock” of actuarial assessments (Hamilton, 2015). Thus, there may be significant overlap between the riskiest offenders and those who receive the harshest sentences. This study will inform this critique by establishing the relationship between risk score, race, and recidivism and by identifying which risk factors most account for mean group differences in instrument scores.

Second, by outlining the process of instrument development and validation, the study provides an example of instrument construction which may be adapted for use in other locations and for different populations. There are few examples and almost no "best-practices" established for sentencing risk assessment instruments. Most states rely on "off-the-shelf" instruments — the proprietary nature of which does not easily lend itself to evaluation (e.g., the construction process for these instruments is kept private to safeguard proprietary information). Each actuarial
risk assessment instrument is a culmination of a myriad of decisions which are ultimately consequential in determining instrument fairness and predictive accuracy. This study relies on a well-established method of instrument construction which provides a computationally easy way to create a risk scoring algorithm. The Burgess (1928) method of using categorical factors to arrive at an additive risk score does not require the use of complicated statistical techniques or special computer programs, making it ideal for replication. This process also provides a comparison to existing risk assessment instruments by establishing a parsimonious set of static predictors that judges would typically have access to at sentencing. Additionally, validating the instrument for predictive validity provides a measure of offender discrimination to use as a comparison in future studies.

Third, this study adds to the burgeoning criminological and statistical literature on instrument bias as a separate phenomenon from mean race differences in scores (Chouldechova, 2016; Flores et al., 2016; Kleinberg, Mullainathan, & Raghavan, 2016; Larson, Mattu, Kircher, & Angwin, 2016; Skeem & Lowenkamp, 2016). When risk assessments produce mean score differences for Black and White offenders, it is usually due to the unequal distribution of significant risk factors. In other words, properly validated instruments can accurately reflect Black offenders’ higher risk of recidivism, resulting in the potential for disparate impact. However, it is also possible that risk instruments are better at differentiating White recidivists from White non-recidivists than Black recidivists from Black non-recidivists (i.e., differences in AUC statistic by race), or that the same risk score is associated with significantly different rates of recidivism for Black and White offenders (i.e., omitted variable bias related to race). In particular, if risk scores are associated with similar rates of re-offending across race, it indicates that individual risk factors affect Black and White likelihood of re-offending in the same way.
New research suggests that two prominent risk assessment instruments (i.e., PCRA and COMPAS) predict re-offense with similar accuracy for both Black and White offenders, but, on average, assign higher scores to Black offenders (Flores, et al. 2016; Skeem and Lowenkamp, 2016). This results in unintentional bias, where a racially-neutral process results in disparate impact — similar to other sentencing policies, such as the use of criminal history in sentencing (Bushway & Piehl, 2011; Crow, 2008; Frase, 2009). However, a risk assessment instrument that over-estimates Black offenders’ risk of recidivism, is a case of direct bias on the part of the instrument. This research will explore both unintentional and direct bias as a result of the instrument by assessing the predictive accuracy of risk assessment prediction for both Black and White offenders, and by assesses the classifications of low, medium, and high risk offenders by race.

Fourth, the study will contribute to the limited literature which focuses on the recidivism patterns of serious (i.e. level 5) offenders. Previously, most sentencing and recidivism studies focused on the general population of offenders because it provides for a larger sample and allowed for generalization to a greater offending population. However, by studying serious offenders in particular, criminologists can provide information about a population that is of particular interest to the criminal justice community. These serious offenders are almost always slated to receive sentences of incarceration, and efforts to lower state incarceration rates must focus on them, as opposed to lower-level offenders who are more likely to receive community sanctions or short jail terms. As such, the present study addresses the National Institute of Justice's 2015 Dear Colleague Letter (Sabol, 2014) which identified alternatives to incarceration (ATIs) for serious offenders as a major topic of interest at the state and federal level. Identifying low-risk offenders who have been convicted of serious crimes can inform criminal justice policy-
makers and practitioners interested in reducing mass incarceration.

Conclusion

Numerous scholars, judges, and sentencing professionals have called for the adoption of evidence-based sentencing (e.g., Kern & Bergstrom, 2013; Ostrom et al., 2002; Peters & Warren, 2006). For example, The National Center for State Courts called for states to "get smarter about sentencing" by using risk assessments and predictive instruments to assist judges in selecting the most effective sentencing option (Marcus, 2009). The potential to use social science research to assist judges in making the best use of scarce correctional resources is an attractive proposition for most criminologists. However, the practice of using actuarial risk prediction at sentencing has surprisingly little empirical evidence regarding its effectiveness - and even less attention has been paid to the potential effects on racial disparity. This study aims to contribute to the evidence-based sentencing movement by exploring the implications of using risk assessment in the courtroom.
Chapter Two: Literature Review

Overview

A rich body of literature has established that the sentencing stage is a point of disparity for Black and Hispanic offenders, and that these disparities are highly variable based on context and location (Mitchell, 2005; National Research Council, 1983; Spohn, 2000; Ulmer, 2012; Zatz, 2000). A meta-analysis of 71 published sentencing studies concluded that Black offenders were significantly more likely to receive prison sentences compared to similarly situated White offenders (Mitchell, 2005). Mitchell (2005) also reported a small effect of race on sentence length, with Black offenders receiving slightly longer sentences (see also Ulmer, 2012). Research has also shown that gender and age have an effect on sentencing decisions (Doerner & Demuth, 2010; Steffensmeier & Demuth, 2000; Steffensmeier et al., 1998), and that race/ethnicity, gender, and age interact and result in more severe sentences for young, minority, offenders (Doerner & Demuth, 2010; Steffensmeier et al., 1998). Disparities in sentencing contribute to the stark racial disproportionality found in the prison population (Mauer, 2011; Tonry & Melewski, 2008).

Changes in sentencing policy, such as the adoption of sentencing guidelines, mandatory minimum laws, and use of risk assessment instrument, have the potential to affect racial disproportionality in sentencing by structuring judicial decisions and through inadvertent disparate impact on racial groups (Ulmer, Painter-Davis, & Tinik, 2016). The effect of the adoption of sentencing guidelines, which were enacted to reduce disparities in sentencing, is mixed. Some scholars report that the adoption of guidelines improved racial and ethnic disparity in sentencing (in Pennsylvania: Kramer & Ulmer, 2009; Ulmer et al., 2016), but that the effect
was conditioned by level of judicial discretion and changed over time. Alternatively, mandatory minimum laws have had a strong and consistent effect on sentencing outcomes by increasing the length and severity of sentences across the board which disproportionately affected minority offenders (Fischman & Schanzenbach, 2012; Kramer & Ulmer, 2009; National Research Council, 2014). More recently, some states have adopted policies that either require or allow for the actuarial consideration of risk for criminal sentencing, which may also disproportionately affect minority offenders. The expansion of actuarial instruments into the courtroom has spurred debate about the role formal risk assessments should play in punishment decisions (Hannah-Moffat, 2013; Skeem & Lowenkamp, 2016; Starr, 2014), but many important questions remain unanswered because of a lack of empirical studies on the subject.

The use of formal risk assessment to sentence offenders signifies a departure from a sentencing structure that emphasizes uniformity to one which places greater importance on efficient offender management (Silver & Miller, 2002). In the face of soaring incarceration costs and high recidivism rates, states have become more focused on reducing incapacitative sanctions among lower risk offenders and incapacitating higher-risk offenders for the safety of the community. Formal consideration of risk can be at odds with the purpose of sentencing guidelines, which courts have adopted as a way to increase sentencing uniformity and reduce disparity. As such, for many states, sentencing risk assessments juxtapose two potentially conflicting policy and legal mandates: controlling disparity and managing recidivism.

In the last chapter, I identified questions which surround the use of actuarial risk assessments in the courtroom and showed how the present study will address the gaps in knowledge. This chapter begins with an overview of how risk assessment fits into structured sentencing, explains how risk assessment tools group offenders based on risk of recidivism, and
explores how legal and extra-legal factors affect sentencing outcomes through both formal and informal evaluations of risk. Finally, issues of bias and disparate impact and the incongruence between sentencing guidelines and risk assessment tools is discussed.

**History of Risk Assessment in Sentencing**

Assessments of risk of recidivism have been used in the criminal justice system, particularly to determine sentence length on the back end (e.g., parole release), for over a century (e.g., see Burgess, 1928). In the 1960s and 1970s risk assessment at sentencing focused on identifying mentally ill offenders for incapacitation, with the help of clinicians and occasionally with the use of diagnostic tools (Simon, 2005). In the 1970s, this practice sustained political, judicial, and academic attacks which questioned the accuracy of assessing individuals’ risk of reoffending, as well as the legality of holding persons for preventative reasons. The demise of this earlier-kind of risk assessment in sentencing also coincided with rise of "mass incarceration", as the United States shifted to a more retributive approach to punishment (Monahan & Skeem, 2014; Tonry, 2013). This approach ushered in a new era in sentencing, and in the mid-1970s liberal reform movements focused on increasing sentencing uniformity and reducing the effects of extra-legal variables through the use of sentencing guidelines. The first presumptive guidelines were established in the 1970s and from 1978 to 1996, 20 states and the federal government established their own versions.

Tonry (2013) outlined the transition away from indeterminate sentencing and rehabilitation that began in the 1970s. Martinson’s (1974) highly regarded "nothing works" report turned policymakers and many academics away from supporting rehabilitation, and judges and scholars criticized indeterminate sentencing, claiming that they were unpredictable and gave
too much power to parole officials. Without the promise of rehabilitation, and lacking well-developed risk prediction tools (Ulmer & Spencer, 1999), it became less important to assess risk at sentencing. If the best that the criminal justice system could do is incapacitate offenders, it seemed easier to increase penalties across the board rather than attempt to identify offenders for selective incapacitation. By the mid-1980s, the US was fully engaged in “collective incapacitation” and sentences became longer due to mandatory minimums, three strikes laws, and an increase in life sentences without the possibility of parole (Tonry, 2013; Ulmer and Spencer, 1999).

Since the late 1990s, interest in actuarial risk assessment steadily increased as part of the evidence-based movement in sentencing and corrections (Casey, Warren, & Elek, 2011; Etienne, 2009; Marcus, 2009). Even more recently, public support for tougher policies has fallen and risk assessment tools for criminal offending have appeared as a potential tool to aid in the efforts to de-incarcerate. Numerous states (e.g., Pennsylvania, Virginia, Missouri, Wisconsin, Kentucky, Ohio, Oklahoma, Arizona, Utah, Maryland), are currently using or exploring the use of statewide sentencing risk assessment tools, while the use of risk assessments instrument for sentencing sex offenders is almost ubiquitous (Starr, 2014).\(^2\) The goals of risk assessment policies vary from state to state, as do the individual tools, but all rely on offender and case characteristics to estimate the likelihood of recidivism.\(^3\) Theoretically, risk and needs assessments can assist judges in identifying which offenders require more incapacitative

\(^2\) It is difficult to estimate how many jurisdictions consider actuarial risk assessments at the sentencing stage, because so many use them informally. Assessments used to assist with pretrial release or to assess probation risk often show up in a pre-sentence reports, even though states may not have explicit risk assessment policies.

\(^3\) This is commonly defined as arrest, but violent arrest or a new conviction post-release are also common outcomes.
sentences, which should receive community services, and which may be diverted to less punitive sentences.

**Modern Sentencing Risk Assessment Tools and Policies**

The increased use of actuarial risk assessments at sentencing has recently received attention from policy-makers and academics (e.g., Holder, 2014; Monahan & Skeem, 2014; Starr, 2014). Critics of this trend warn of the potential effects on racial and ethnic disparity and of constitutional challenges of using extra-legal variables such as race, gender, and age to determine sentencing decisions (Hannah-Moffat, 2013; Holder, 2014; Starr, 2014). For example, Hannah-Moffat (2013, p. 281) warns that "marginalized groups unavoidably score higher on risk instruments because of their increased exposure to risk, racial discrimination, and social inequality"— and this may lead to longer or more severe sentences for minorities and the poor. However, it is still unclear how the increased reliance on actuarial risk assessments has affected courtroom decision making and what effects, if any, such reliance will have on sentencing disparity. While a few states have closely integrated risk assessment procedures into their sentencing process, other states use assessments in an advisory fashion or for certain offenders (e.g., sex offenders) only. These variations in tools and policies, and the contextual and place-based differences in sentencing disparity have contributed to gaps in research on this topic (Monahan & Skeem, 2014).

It is important to note that there is a difference between risk assessment technology — that is, the actuarial tool itself and risk assessment policy, which outlines how and for what purposes the tool should be used. Tools and policy vary from state to state, and these variations have implications for potential effects on racial, ethnic, and gender disparity, judicial discretion,
and possible constitutional challenges. At a basic level, all risk assessment instruments determine
the likelihood of an offender recidivating based on factors that were found to correlate with re-offending. Pennsylvania, Missouri, and Virginia stand out in this arena because each state has
created an original risk assessment instrument for use at sentencing. Pennsylvania Commission
on Sentencing created a tool that uses age, gender, and variations of criminal history markers to
predict likelihood of recidivism (Ruback et al., 2016). Pennsylvania’s tool does not use many of
the questionable variables associated with structural inequality such as race, employment status,
neighborhood, or level of education to determine risk. Missouri’s tool includes both employment
and education, but not gender when assessing risk (Wolff & Oldfield, 2010). And, Virginia's
tool assesses gender, age, employment, and marital status (Ostrom et al., 2002). The use of
criminal history and case characteristics to predict recidivism is less controversial, and less likely
to be legally challenged, than using markers that denote membership in a protected class (Starr,
2014). However, some researchers warn that over-reliance on criminal history to predict risk
may exacerbate racial inequalities found in the system (Skeem & Lowenkamp, 2016).

There are also policy differences among the three states. For example, when judges begin
using the tool in 2018, Pennsylvania's risk assessment scores will be included with the
sentencing guideline recommendations. The scores are intended to encourage judges to request a
pre-sentence report if offenders fall on either tail end of the risk score range. Most importantly,
scores are meant to provide judges with more information about offender risk and are not
directly tied to sentence recommendations. In Missouri, risk assessment results are integrated
into the voluntary sentencing guidelines. That is, offenders who score in the "Good" range are

---

4 Most states do not create their own risk assessment tools and instead rely on "off the shelf" tools that have
previously been validated (e.g., COMPAS, LSI-R).
recommended for mitigated sentences and offenders who score in the "Poor" range are
recommended for aggravated sentences. Virginia's policy was implemented with the specific
goal of reducing incarceration and is only used for felony, non-violent offenders. Offenders
scoring in the low range of the scale are recommended for diversion (i.e., a non-specific, non-
custodial sentence). Judges in Virginia follow these recommendations in two out of three cases
(Ostrom et al., 2002). Piloted in 1997, Virginia's risk assessment policy is the best evaluated of
the three (see Kleiman, Ostrom, & Cheesman, 2007; Ostrom et al., 2002), however evaluations
have not focused on disparate impact.

States interested in using risk assessment during sentencing should take note that the
details of the accompanying policy are particularly important for warding off potential
constitutional challenges. For example, the ACLU challenged the state of Virginia over the use
of their sex offender risk assessment at sentencing, noting that it recommended longer sentencing
ranges based on offender's age, employment history, and level of education – which have no
bearing on the offense committed (Brooks v. Commonwealth of Virginia, 2012). The ACLU
argued that this procedure created a system in which the most marginalized offenders also
received the harshest sentences. The State of Virginia won the case largely on the grounds that
judges were allowed to ignore risk assessment recommendations, although research showed they
usually did not (Ostrom et al., 2002). Similarly, the Wisconsin Supreme Court (State of
Wisconsin v. Loomis, 2016) ruled that the use of Northpointe Inc.'s COMPAS risk assessment
tool at sentencing does not violate the due process rights of a defendant as long as judges merely
consider, but do not rely on, the risk assessment score to determine sentencing. Rather, the
assessment can be used as an "observation" to "corroborate" the sentence imposed by the court
(p.43-44). The Supreme Court of Wisconsin agreed that handing out sentences based on
improper factors, such as gender or socioeconomic status was unconstitutional, but that considering these factors for the purpose of statistical calculation was appropriate.

The Risk-Needs-Responsivity Model in the Courtroom

The psychology behind the use of risk assessment tools in the criminal justice system is best explained by the Risk-Needs-Responsivity (RNR) model. This model, developed by Andrews, Bonta, and colleagues (Andrews & Bonta, 2010a; Andrews, Bonta, & Hoge, 1990) suggests that offenders' criminogenic characteristics (i.e., their risks and needs) should be met with appropriate supervision and services to reduce reoffending. In the courts, this means low-risk offenders should be given less incarcerative sentences, and vice versa. It also suggests that offenders who score high on individual needs, such as drug addiction, are given resources to address those needs. The underlying rationale behind these principles is that high and low risk offenders are less likely to recidivate when they are treated differently. Supporting research shows that giving low-risk offenders a high level of supervision and giving high-risk offenders a low level of supervision results in a higher level of recidivism than if offenders were supervised in line with their risk level (Andrews & Bonta, 2010a; Lowenkamp & Latessa, 2004).

The correctional system has relied on the RNR model to differentiate among offenders once they enter the system. The goals of correctional offender management pair well with the use of risk assessments because the tools are used to categorize individuals into specific categories. New inmates are assessed by their risk level and are matched with the appropriate rehabilitative programming. However, using this method to inform sentences is more complex. On one hand, classifying offenders earlier in the correctional process can better reflect the RNR principle. For example, offenders assessed as low risk after being sentenced to prison may receive a low level
of inmate supervision, but they will never receive a non-custodial sentence, which may have resulted in a better outcome. On the other hand, sentencing has multiple goals beyond appropriate offender management. The consideration of risk may be in contrast to goals of deterrence and retribution. Issues of fairness and legal ethics also play a more prominent role in the courtroom. By basing sentences on the RNR model, punishment decisions are affected not only by unchangeable offender characteristics (e.g., gender), but also by life choices that are not illegal (e.g., being unemployed).

It should be also be noted that some risk assessment instruments used at the pre-trial and sentencing stage do not fully reflect the Risk-Needs-Responsivity principles - nor, arguably, do they need to. Assessing risk may be done in a relatively simple manner using just a few static predictor variables from agency data (i.e., using a 2nd generation risk assessment instruments (Andrews, Bonta, & Wormith, 2006). The few jurisdiction-specific instruments created for the sentencing stage overwhelmingly rely on static factors for prediction. (However, employment, education, and marital status appear in Missouri’s and Virginia’s risk assessment instrument.) Assessing dynamic needs and risk factors (i.e., 3rd generation instruments, such as the LSI-R) require trained practitioners to administer the test and time for individualized interviews. Such investments may not be pragmatic for the hundreds of thousands sentencing decisions made every year in any given state. Also, responding to risk and needs (i.e., 4th generation instruments, such as COMPAS) is a highly-structured, resource-intensive process, which may not be appropriate for the sentencing stage. States in which judges have limited options to sentence offenders in a way that directly responds to offenders’ risks and needs, may not be interested in using such an extensive assessment.
Integrating Risk Assessment into Sentencing

Some proponents of using risk assessment in the courtroom hope that the process will simply inform judicial decisions without constraining discretion (Casey et al., 2011; Hyatt, Mark, & Steven, 2011; Kern & Bergstrom, 2013). For policy changes, this is a low bar, but perhaps one fraught with the least amount of controversy. In fact, many states already include an informal assessment of risk in the pre-sentence reports, which are generally completed for more serious offenders. Results from actuarial risk assessment instruments are also sometimes included in the reports, but — almost certainly — there are ad-hoc, narrative assessments of the offender’s risk of reoffending. Probation officers, who are tasked with monitoring offenders on community supervision, are appropriately concerned about offenders who are at high risk for failing. Thus, pre-sentence reports have a utilitarian focus — to assess an offender's likelihood of doing well with less restrictive punishment. These reports provide valuable information for judges, who consider community protection, as well as more retributive considerations, into account to arrive at a sentencing decision. Including actuarial risk assessment results in the pre-sentence report, or as standalone information, is a way to integrate social science knowledge into judicial decision making and is supplemental to the informal assessment of risk already being done by probation officers and the judges. One benefit of this loose approach is that by avoiding the coupling of risk assessment results with policies intended to affect judicial discretion, jurisdictions may be likely to avoid many of the constitutional challenges associated with using extra-legal variables in sentencing.

For example, Pennsylvania's risk assessment tool was developed under a mandate from its state legislature and will be used as a way to flag high- and low- risk offenders for closer judicial scrutiny. The risk assessment score, which is calculated using static administrative
variables, is supposed to be included with the sentence guideline range recommendations. Offenders who fall on the tail ends of the range (i.e., having few or many risk factors), will be flagged to receive a comprehensive pre-sentence report for judges to review. While the process is not yet “live”, preliminary research using simulated sentencing scenarios shows that the instrument’s effects on judicial decisions varies based on type of conviction offense and that judges tend to over-estimate offender risk before being presented with the risk score (Ruback et al., 2016). In this way, a risk assessment score (when used in a purely advisory capacity) may be no different than any other piece of case information provided to a criminal justice professional.

Indeed, there are benefits to using risk assessment at the sentencing stage, which offers substantially more constitutional protections for offenders compared to parole hearings or correctional settings (McGarraugh, 2012). The sentencing stage provides a "controlled environment" in which the accuracy and the legality of using an instrument can be challenged. Despite constitutional challenges, one proposed way to use risk assessment at sentencing is to focus it on identifying individuals for diversion, rather than increasing the sentences for higher risk offenders. Research shows that incarcerating low-risk offenders, rather than providing a community-based sanction, may increase recidivism (Lowenkamp & Latessa, 2004). However, focusing the risk assessment on diverting people from incarceration does not address the problem of potential disparate effects on minorities. In absolute terms, this option reduces the number of people behind bars, but in relative terms it increases sentencing inequality by picking winners and losers (Oleson, 2011). These "losers", or offenders who score on the higher ends of the scale, are disproportionately male, younger, and of a racial/ethnic minority (DeMichele & Laskorunsky, 2014; Skeem & Lowenkamp, 2016; Skeem, Monahan, & Lowenkamp, 2016).

---

5 Ultimately, the judge decided whether to order the pre-sentence report.
Thus, while the use of risk assessment instruments at sentencing may be used to achieve some policy-oriented goals (e.g., reduce levels of incarceration), it may undermine other goals (e.g., reducing racial disparity at sentencing).

Another challenge of using risk assessment tools is that they must be adopted to meet the sentencing goals (e.g., risk reduction or risk prediction) of each jurisdiction (Heilbrun, 2009). For example, a tool designed to reduce risk of recidivism will be structured differently than one that is used to simply predict risk of recidivism. A tool used for prediction can be fairly easy to administer by incorporating a handful of static predictors accessed from official records (e.g., select demographics and criminal history). However, if the goal of using a risk assessment tool is to reduce offenders’ risk of future offence, dynamic and causal risk factors (e.g., substance abuse, unemployment) must be identified to match offenders with services. A tool meant to reduce risk would need to be more extensive and incorporate information from semi-structured interviews with offenders. The two types of approaches present tradeoffs in the areas of effectiveness and efficiency. Administering an extensive risk assessment questionnaire would provide jurisdictions with the ability to match offenders with community services, but it also requires substantially greater agency resources. Extensive assessment would be a waste of resources if results (i.e., identified risk and need factors) were not closely linked to sentencing outcomes. Because most jurisdictions have limited resources to address risks and needs through sentencing, courts have overwhelmingly focused on using risk assessment tools to predict, not to decrease, recidivism.

Perhaps the biggest unknown in this field is how judicial behavior changes as a result of risk assessment consideration. Will judges respond to risk assessment information by giving more serious sentences to the riskiest offenders, and vice versa? Evaluations of Virginia's
sentencing risk assessment instrument showed that judges sentenced in accord with the risk assessment recommendation about two-thirds of the time (Ostrom et al., 2002), and other research showed that higher overall risk scores were associated with greater likelihood of receiving a custodial sentence (Carlson, Daniels, & Stein, 2015). In order to provide a simulation of how the inclusion of a risk score would change informal estimates of risk, Starr (2014) used a sentencing vignette to compare differences in sentencing outcomes. She reported that study participants (law school students) who were given a sample sentencing case where an offender had a high risk score not only saw those offenders as more likely to recidivate, but also sentenced them to longer sentences, compared to the same offenders with no accompanying risk score. However, a similar study by Ruback and colleagues (2016) with Pennsylvania judges found that the effect of considering the risk score substantially varied by conviction offense type.

In sum, the increasing popularity of using actuarial risk assessments in the courtroom has the potential to alter judicial decision-making by structuring the ways in which judges arrive at their final assessment of the offender's risk of re-offense. The overarching purpose of considering risk at sentencing is that judges will give similarly situated lower risk offenders less restrictive sentences, and vice versa (within the limits allowed by law). However, as discussed in the following section, the consideration of actuarial risk assessment tools in sentencing must be compared to the status quo (i.e., baseline sentencing practices). Instruments which aim to predict risk of re-offense include the same legal and extra-legal factors already consider by judges in determining offender risk and in making sentencing decisions (Clair & Winter, 2016; Gottfredson, 1999; Steffensmeier et al., 1998). The voluntary nature of most sentencing guidelines provides for wide discretion on the parts of judge, and risk assessment instrument integration is unlikely to change this.
Judicial Philosophy and Consideration in Sentencing

To assess the implications of using an actuarial risk assessment instrument at sentencing, it is important to first understand the role of offender risk in sentencing decisions. While juries sometimes play a role in determining the outcome of a case, the decision to incarcerate, and for how long, rests largely with the judge (and increasingly with prosecutors through plea bargaining and guilty pleas). Steffensmeier, Ulmer, and Kramer (1998) outlined how judges use "focal concerns" — offender blameworthiness, protection of the community, and practical constraints and consequences — to make sentencing decisions. Both legal and extra legal considerations filter through individual beliefs and ideologies to influence the interpretation and prioritization of these concerns (Savelsberg, 1992; Ulmer & Kramer, 1996). This perspective incorporates Albonetti’s (1991) causal attribution theory, which explains how judges develop patterned responses to deal with sentencing decisions. Because of time constraints of the judicial process and the constraints of gathering comprehensive information, judges rarely have complete facts about the cases they sentence. Even when time and information is more readily available, judges and other court officials face uncertainty related to the likelihood of future criminal behavior (Albonetti, 1991; Kramer & Ulmer, 2009). To compensate for a lack of information and other sources of uncertainty, judges develop a personal and perceptual shorthand for assessing whether an offender is blameworthy, dangerous, or a good candidate for incarceration (Albonetti, 1991; Kramer and Ulmer, 2009). The "patterned responses" are based on legal variables, and offender characteristics (e.g., age) and social statuses (employment status), which help explain disparities

---

6 Sentencing guidelines and mandatory minimum sentencing laws have taken away some of the power of judges to make sentencing decisions. For a more thorough explanation, see Ulmer, Light, and Kramer (2011).
in sentencing.

The determination of offender blameworthiness is in line with a retributivist model of sentencing in which the main factor concerning judges is the offender's culpability. Offense severity is a basic indicator of blameworthiness, with serious crimes causing more harm. As blameworthiness increases, an offender is perceived to be more deserving of punishment — usually in the form of a longer or a more restrictive sentence. Perceptions of offender blameworthiness can also be influenced by the length and seriousness of an offender's prior record and in addition to mitigating or aggravating circumstances, such mental capacity or relationship to victim. While risk assessment tools do not assist with determining offender blameworthiness (which is a value judgment), prior record and offense gravity and type are both important predictors of recidivism and are included in the major sentencing risk assessment instruments (Ostrom et al., 2002; Pennsylvania Commission on Sentencing, 2016).

Actuarial risk assessment instruments are better suited to assist with more utilitarian concerns of sentencing, such as community protection and practical considerations (e.g., efficient use of correctional resources). Community protection is directly tied to an offender's risk of committing a new crime. Offenders perceived to be at high-risk for re-offense may require more restrictive or lengthier punishment, and offenders who are perceived to have an elevated risk of committing violence may need to be incapacitated. Judges mainly rely on criminal history to make informal predictions about an offender's risk of recidivism (Hamilton, 2015), and to a lesser extent, they also rely on the seriousness of their conviction offense, and their potential to succeed using less restrictive sanctions, such as probation (Clair & Winter, 2016; Gottfredson, 1999; Ulmer et al., 2016). Offender demographics also play a role in molding perceptions of risk (Clair & Winter, 2016; Gottfredson, 1999; Steffensmeier & Demuth, 2006). For example, judges
viewed younger offenders as more likely to re-offend (Gottfredson, 1999) and perceived Black offenders as more likely to fail on community sanctions (Clair & Winter, 2016). An actuarial risk assessment take into account offender and case characteristics to inform judges of an offender's likelihood to re-offend upon release, but with more accuracy than informal prediction (Meehl, 1954; Skeem & Monahan, 2011).

Finally, as noted above, courtroom workgroups consider practical constraints and consequences in sentencing decisions. Since they are at least loosely coupled to other criminal justice agencies (Hagan, 1983), judges might take into account a variety of factors, such county resources (e.g., space available to house the offender), local politics, and case flow. Judges may also be wary of incarcerating an older, sick offender or one who is a primary caretaker of children. Risk assessments assist judges with making efficient use of correctional resources by identifying the riskiest offenders for selective incapacitation. In general, offender characteristics used to inform practical concerns overlap with instrument risk predictors. For example, Ulmer and Kramer (1996) reported that judges were reluctant to send White offenders to a predominantly Black prison, or to incarcerate female offenders if they had dependents.

As shown, instrument risk factors overlap with the same offender and case characteristics used by judges to assess focal concerns. Given this overlap, how does actuarial assessment of risk differ from the status quo – that is, the informal prediction of risk? While judges may consider a greater variation of offender and case characteristics than captured by a risk assessment instrument (e.g., offender demeanor), they are also less likely to be accurate about an offender's likelihood of re-offending (Gottfredson, 1999; Meehl, 1954; Skeem & Monahan, 2011). Because judges already have a broad sense of which offender characteristics are associated with future offending, risk assessments generally offer evaluations that are
complimentary to the clinical assessment of risk. For example, judges perceive offenders who are younger and those with a lengthy criminal record to have a higher likelihood of recidivism, both of which are identified as risk factors for recidivism in actuarial instruments (Gendreau, Little, & Goggin, 1996; Gottfredson, 1999). This complementary overlap can serve to confirm existing judicial perceptions about which offenders should receive the most severe punishment.

However, one legal factor that may be problematic is offense seriousness, because as a predictor variable, it often has an inverse relationship with recidivism (DeMichele & Laskorunsky, 2014; Durose, Cooper, & Snyder, 2014). Judges view offenders convicted of serious crimes as more blameworthy (Kurlychek & Johnson, 2004; Steffensmeier & Demuth, 2000) and more likely to recidivate (Gottfredson, 1999), yet offenders convicted of more serious crimes are less likely to reoffend than similarly situated offenders convicted of less serious crimes (DeMichele & Laskorunsky, 2014; Gottfredson, 1999). Serious offenders, particularly violent ones, are also singled out by sentencing guidelines to receive the most severe sentences (Kramer & Ulmer, 2002). In this way, retributivist sentencing goals that aim to punish the most blameworthy offender are contrasted against utilitarian goals of sentencing (i.e., incarcerating the highest risk offender). If risk assessment is tied to sentencing recommendations, judges who view sentencing primarily as a way to impose "just deserts" style punishment may discount the results. As explained in the following section, public backlash toward seemingly unjustly light sentences is also an issue.

**A Case Study**

The recent trend in sentencing has been to consider more utilitarian concerns, which may explain the renewed interest in the development and use of risk assessment tools (Lawrence,
While criticism of risk assessment has largely focused on the possibility of longer sentences for poor or minority offenders (Hannah-Moffat, 2013; Starr, 2014), sentences which appear too lenient for "privileged" offenders also elicit strong public responses. For example, a 2016 California rape case involving a Stanford competitive swimmer, Brock Turner, made headlines when the judge handed down a 6-month jail sentence based on the offender's likelihood of re-offending and (in)ability to do time in prison (Stack, 2016). Partially in response to the probation officer's recommendation, Santa Clara County Superior Court Judge Aaron Persky declined to sentence the offender to prison because, "...a prison sentence would have a severe impact on him...I think he will not be a danger to others". The first-time offender — a 20-year old Stanford student from a wealthy background — scored in the low to moderate range on California's sex offender risk assessment scale. While a sexual assault conviction typically garners prison time, both the probation officer and the judge felt that a period of prolonged incapacitation was unnecessary for Turner. Public backlash was severe. Nationwide opinion pieces derided the "light sentence", a recall petition for Judge Persky was formally launched, and he was subsequently removed from all criminal cases and transferred to preside over civil court. Following the public outcry, the California legislature passed a mandatory sentencing law that made the minimum sentence for sexual assault three years in prison (Chokshi, 2016).

The Turner case illustrates what can happen when utilitarian goals of sentencing conflict with retributive goals of sentencing. On one hand, there has been a concerted effort to employ "smart-sentencing" as a way to divert low-risk offenders from incarcerative sentences. Proponents of utilitarian tools such as risk assessments claim that in order to reduce unsustainable prison populations, diversion efforts should be focused on the lowest risk individuals. However, this sentiment clashes with the retributive goals for sentencing. Law, and
by extension the punishment for breaking it, is a reflection of the fundamental values of society (Black, 2010; Friedman, 1975). A society reifies these beliefs by punishing offenders for harming not just the victim, but also the community. Judges have a difficult job of forming these concerns into, "…a single value on a continuous dimension of sentencing severity" (Monahan & Skeem, 2014, p. 33).

**Race, Risk, and Criminal History**

Simply stated, race and risk of reoffending are statistically correlated (Harcourt, 2015; Kopf, 2015; Skeem & Lowenkamp, 2016). On average, Black offenders score higher on risk assessment instruments because they have more risk factors of recidivism (Durose et al., 2014; Skeem & Lowenkamp, 2016). Reports indicate that race has a weak, but significant, relationship with re-arrest even when other risk factors are controlled for (Skeem & Lowenkamp, 2016). Minority race status serves as a stand-in for levels of exposure to structural inequality and, consequently, a variety of criminogenic factors that are difficult to control for in a risk algorithm. Some scholars report that criminal history mediates the relationship between race and risk (Skeem & Lowenkamp, 2016), while others further contended that criminal history is a proxy for race (Harcourt, 2015).

Research also shows that the seriousness of the offense and the offender's criminal history already explained most of the variation in sentencing outcomes, particularly between White and Black offenders (Mitchell, 2005; Spohn, 2000; Ulmer et al., 2016). Thus, Black offenders are more likely to be sentenced to prison because, on average, they have more substantial criminal histories, and to a lesser extent, are charged with more serious crimes (Chiricos & Crawford, 1995; Rehavi & Starr, 2014). However, it should be noted that Black
offenders are more likely to be sentenced to prison, and for slightly longer terms, even when legally relevant factors are controlled (Chiricos & Crawford, 1995; Spohn, 2000; Ulmer, 2012). Courts, in turn, often determine whether an offender is likely to re-offend largely based on their criminal history (Monahan & Skeem, 2014), which means that criminal history is a non-actuarial stand-in for risk (although it also affects perceptions of offender blameworthiness). Sentencing guidelines also reflect the importance of criminal history by structuring presumptive sentences mainly based on the severity of the offense and the offender’s criminal history. Thus, the informal consideration of risk in sentencing already has a racially disparate impact on Black offenders, because they tend to have more substantial criminal histories.

Measures of criminal history are ubiquitous in criminal risk assessments (Hamilton, 2015), because prior record is a consistent and significant predictor of recidivism (Gendreau et al., 1996; Monahan & Skeem, 2016). Criminal history helps to explain much of the differences in risk assessment scores between Black and White offenders. For example, Skeem and Lowenkamp (2016) found that 66% of the difference in the Post-Conviction Risk Assessment scores between Black and White offenders was explained by differences in criminal history. Thus, criminal history is an important factor in explaining both differences in risk score and differences in sentencing outcomes. In states that rely on actuarial risk assessments to sentence offenders, offenders with substantial criminal histories may experience the negative effects of being singled out by sentencing guidelines for the harshest sentence and experience the negative effects of having a high-risk score. However, it is also possible that a high-risk score would confirm judicial stereotypes regarding the offender's recidivism risk of recidivism.

While exact proportions of contributing factors that explain racial differences in criminal history are difficult to determine, scholars agree that differential involvement in crime (Beck &
Blumstein, 2017; Sampson & Lauritsen, 1997), differential selection into the criminal justice system (Alexander, 2012; Tonry, 1995), and differential treatment and impact at various stages in criminal justice system (Cole, 1999; Spohn, 2013) all contribute to Black/White differences in criminal history. Black offenders are overrepresented in both violent offending and victimization because of longstanding patterns of inequality and disadvantages among communities of color (Sampson & Lauritsen, 1997). Differential exposure to structural risk factors such as concentrated poverty, racial segregation and bias, and community disorganization help explain racial differences in criminal violence (Peterson & Krivo, 2005; Sampson & Wilson, 1995). Studies also show that once minority individuals enter the court system, they are treated slightly more punitively than similarly situated White offenders (Cole, 1999; Spohn, 2013) and are disparately impacted by policies that claim to be race neutral, such as mandatory minimum sentencing (Tonry & Melewski, 2008).

Differential selection into the criminal justice system is also a significant issue. Research has shown how differences in selective enforcement result in disparate impact for minority offenders. For example, Black drivers are significantly more likely to be pulled over by the police and both Black and Hispanic drivers are more likely to be searched when pulled over (Langton & Durose, 2013). Blacks are more likely than Whites to be arrested for marijuana possession, despite having comparable rates of marijuana use (ACLU, 2013). Black students are more likely to receive official institutional punishment for the same infractions as White students (Nicholson-Crotty, Birchmeier, & Valentine, 2009; Skiba, Michael, Nardo, & Peterson, 2002) and "Stop and Frisk" policies have overwhelmingly targeted minorities at rates above their involvement in crime (Gelman, Fagan, & Kiss, 2007). Selective enforcement and charging are a bigger issue among property and public order crimes in which police and prosecutors have more
enforcement discretion (Piquero & Brame, 2008). Differences in sentencing outcomes resulting from criminal history is considered "warranted disparity", yet the processes of how offenders come to have these variations in risk factors underscore the effect of cumulative disadvantage associated with race and ethnicity (Sampson & Laub, 1997).

**Risk Assessments: Disparate Impact and Bias**

While risk assessment instruments have been used in a variety of criminal justice decisions for nearly a century, criticism regarding their use in the courtroom has been particularly strong (e.g., Harcourt, 2015; Holder, 2014; Larson et al., 2016; Starr, 2014). These criticisms focus on the possibility of instrument bias and also on disparate impact on minority offenders, among other concerns relating to the constitutionality of using extra-legal factors to make punishment decisions. For example, a recent study from the investigative outlet *ProPublica* (Angwin et al., 2016) examined Northpointe's COMPAS risk/needs assessment instrument, which is used in a variety of criminal justice settings, including at pre-trial and sentencing. The authors reported that the instrument was equally as successful at predicting which offenders recidivated regardless of race. However, it incorrectly predicted failure rates for Black offenders at about the same rate as incorrectly predicting survival rates for White offenders. By returning a higher rate of false positives for Black versus White offenders (at a ratio of 2 to 1), use of the instrument results in a higher proportion of Black offenders experiencing the consequences of having a higher risk score (e.g., being denied pre-trial release), despite not offending. Angwin and colleagues (2016) suggested that the instrument is “biased against Blacks” (pg.1). Their conclusion was heavily contested by Northpointe's researchers (Dieterich, Mendoza, & Brennan, 2016) and the topic of rebuttal paper (Flores, Bechtel, &
Lowenkamp, 2016). Both papers showed that the instrument predicted recidivism equally well for Black and White offenders, and thus, did not meet the definition of "bias". Importantly, however, and as explained below, even an instrument without bias can have disparate impact (Chouldechova, 2016).

The discussion about predictive bias and disparate impact revolves around two issues. First, although race is not included in COMPAS's algorithm (nor, to the author’s knowledge, is it included in any of the algorithms used to predict risk of recidivism in the criminal justice system) risk factors are unevenly distributed by race. In other words, Black offenders, on average, have more risk factors and thus, end up scoring higher on the risk assessment instrument (and this correlates to differences in the reoffending rate). Thus, Hannah-Moffat’s (2013) criticism of sentencing risk assessments is generally correctly in that minorities tend to score higher on risk assessment instruments, even when race is excluded from the algorithm.

Second, a missing piece from Larson and colleagues' (2016) analysis is that the model error differences (i.e., the difference in the false positive and false negatives between Black and White offenders) are inevitable because of the difference in base rates between Black and White offenders (for a more technical explanation see Imrey & Dawid, 2015). In other words, it is not possible to have equal model error tradeoffs for two groups whose risk score distribution and base rates differ (Dieterich et al., 2016; Kleinberg et al., 2016). As the recidivism base rate increases, the false positive rate (i.e., offender who were predicted to fail, but didn’t) increases as well, so a higher portion of Black offenders are marked as high-risk regardless of whether they recidivate. Corbett-Davies and colleagues (2017) utilized the COMPAS instrument and the same data as Larson et al. (2016) to illustrate both of these points. The authors reported that maximizing public safety (i.e., detaining all individuals deemed high-risk) yielded stark racial
disparities, yet statistical solutions to mitigate disparities resulted in detaining low-risk defendants while reducing public safety.

It should be noted, that Larson et al.’s (2016) decision to dichotomize the instrument’s three risk levels (i.e., collapse medium risk offenders into the high-risk level) in order to compare the false positive and false negative rate was a methodologically questionable way of assessing the instrument’s performance (Flores et al., 2016). This essentially created one cut-off score, with everyone above the score predicted to fail and everyone below the score predicted to survive (i.e., not re-offend). In the medical field, this method is more common because all probabilities of having an illness are related to one decision: treat the patient for the illness, or don’t. However, imposing a dichotomous outcome on a risk instrument that is designed to show the probability of reoffending across multiple risk levels is unnecessary and can create misleading results.

Recently, Skeem, and Lowenkamp (2016) used a sample of federal probationers matched on gender, age, and conviction offense to test the Post-Conviction Risk Assessment (PCRA) instrument for potential bias and disparate impact. Their study showed that while Black offenders, on average, received higher scores than White offenders, the PCRA instrument had similar predictive accuracy across race. The authors concluded that while the PCRA did not show test bias, it may contribute to disparate impact due to Black offenders’ higher risk scores. The authors also reported that the majority (66%) of the difference in the risk scores was due to differences in criminal history between Black and White offenders, concluding that legal factors already imbedded within sentencing guidelines contribute to most of the difference in scores. Skeem and Lowenkamp (2016) subsequently suggested avoiding an over-reliance on risk assessment instruments in which criminal history is weighted heavily, while maintaining the
predictive utility (i.e., validity) of the instrument.

Given that criminal history has shown to influence risk score differences between Black and White offenders, it is important to consider whether the use actuarial risk prediction instruments in sentencing actually represents a change from the status quo. Decades of research have shown that the sentencing stage is already a point of disparity for minority offenders (Spohn, 2000; Ulmer, 2012; Ulmer et al, 2016). For example, Ulmer and colleagues (2016) show that legally relevant factors like prior record score and eligibility for mandatories yield big differential race impacts (Ulmer et al, 2016). Steffensmeier, Ulmer, and Kramer (1998) show that offenders who are young, minority, and male receive the harshest sentences – and that the interaction between these statuses is higher than the additive effects of each status (see also, Nowacki, 2016). Presumably, this is because young, minority, men are perceived to be more blameworthy and more dangerous, and judges have fewer reservations about sentencing them to incarceration. Conversely, gender, age, and neighborhood (i.e., a stand -in for race) are also used in many risk assessment instruments because they are highly predictive of recidivism (Gendreau et al., 1996; Oleson, 2011), placing young, minority, males in higher risk categories. Even when risk assessment instruments exclude demographic factors, other variables highly correlated with minority racial status, such as criminal history and conviction offense, can serve as stand-ins for race. (Likely at the cost of predictive accuracy [Oleson, 2011]). As such, basing sentences on an actuarial prediction of risk may exacerbate racial disparity in sentencing outcomes, but it may also merely perpetuate the disparate patterns already in place. The consideration of risk of reoffending, using actuarial or informal methods of prediction, tends to creates disparate impact on minorities, specifically Black offenders. In other words, if judges and actuarial risk assessments are equally as accurate in their assessment of risk for both Black and White
offenders, the inclusion of risk as a factor in sentencing would likely disproportionately impact Black offenders.

**Risk Assessment Under Sentencing Guidelines**

Another issue for consideration is the incongruence between sentencing guidelines and actuarial risk assessment. Sentencing guidelines were enacted to establish a uniform set of sentencing standards, to reduce unwarranted sentencing disparity, and to make sentencing more consistent (Kramer and Ulmer, 2009). Voluntary sentencing guidelines, first appearing in multiple states in the mid-1970s, were set based on past sentencing norms in that jurisdiction and were non-binding. (National Research Council, 1983). Presumptive guidelines, which appeared later, typically involved legislature-appointed sentencing commissions, and expressed more explicit goals – mainly to inflict "just punishment" on convicted offenders (Parent, Dunworth, McDonald, & Rhodes, 1996). In theory, guidelines were supposed to reduce, or at least structure judicial discretion, thereby reducing sentencing disparity resulting from subjectivity and extra-legal factors. Under guidelines, offenders with a similar criminal record, convicted of a similar crime should receive a similar sentence — regardless of their gender, age, or race/ethnicity. This logic stands in stark contrast to the purpose of using actuarial risk assessment at sentencing, which is to individualize sentencing recommendations based on differences in correlates of reoffence (i.e., legal and extra-legal factors).

Few sentencing guidelines were adopted with the explicit purpose of reducing risk of reoffense (Kauder & Ostrom, 2008), however, some guidelines states (e.g., Virginia, Utah) now use both sentencing guidelines and actuarial risk assessment at sentencing. This poses a potential problem, because risk assessments are in some ways antithetical to sentencing uniformity. For
example, in 1991, Virginia's sentencing guidelines were adopted with the explicit goals of reducing unwarranted sentencing disparity (i.e., sentencing disparity based on extra-legal factors) (Farrar-Owens, 2013). Similar to other states, the guidelines were adopted to increase consistency (predictability and proportionality) and fairness (non-discrimination), by reducing the effect of extra-legal variables, such as offender’s gender, age, and education (Casey et al., 2011). In 2004, Virginia formally adopted an actuarial risk assessment into sentencing, with the purpose of diverting low-risk, non-violent offenders from incarcerative sentences. This instrument includes measures of offender gender, age, and employment status, and is directly tied to sentencing outcomes (i.e., diversion eligibility). Thus, the Virginia risk assessment instrument puts forth a sentencing recommendation based on the same offender characteristics that Virginia’s sentencing guidelines sought to reduce the effects of. And while an evaluation of the risk assessment policy shows that use of the instrument reduced rates of incarceration overall, the effect on sentencing racial disparity was not assessed (Kleiman et al., 2007).

In reality, the widespread adoption of sentencing guidelines has not eliminated racial, ethnic, or gender disparities in sentencing (Doerner & Demuth, 2010; Everett & Wojtkiewicz, 2002; Kramer & Ulmer, 1996), although some scholars report that the guidelines had a mitigating effect (Kramer & Ulmer, 2009; Pfaff, 2006; Stolzenberg & Alessio, 2008). Pfaff’s (2006) overview showed that even voluntary guidelines were successful in reducing variation in sentencing among judges, and, to a lesser extent, in reducing the effects of offender gender and race. As more states adopt risk assessments for use at the sentencing stage, some of these gains may be undermined. Sentencing uniformity is achieved when similarly situated offenders are sentenced similarly despite their personal characteristics such as age, gender, and employment. Thus, sentencing uniformity, at least in the traditional sense, cannot be achieved if sentences are
also based on these same characteristics.

**Conclusion**

Risk assessment instruments are part of a broader movement to embrace evidence-based sentencing, and more states are considering including them in sentencing (e.g., Maryland). A National Center for State Courts survey showed that policy makers and judges were concerned with the ineffectiveness of traditional probation supervision and lack of appropriate sentencing alternatives, and that they were interested in using risk and needs assessment tools to reduce recidivism (Peters & Warren, 2006). Risk assessments provide an evidence-based way to structure judicial decision making by providing empirical evidence of an offender's likelihood of re-offense. Despite practical benefits of assessing risk at sentencing, there are still many unanswered questions about the impact the use of these tools will have in the long run. Studies evaluating Northpointe's COMPAS (Flores et al., 2016) and the PCRA instrument (Skeem and Lowenkamp, 2016) present an example of a potential consequence of using recidivism risk assessments at the sentencing stage. On average, risk assessments will classify Black offenders as higher risk than White offenders, because Black offenders tend to have more risk factors. If judges use assessment scores to sentence high risk offenders to more serious sentences than they normally would have, racial disparity in sentencing may be exacerbated. These issues are similar to those that policymakers have grappled with regarding other "racially-neutral" criminal justice policies, such as mandatory minimum and repeat offender laws. Unbiased policies can have disparate effects due to the divergent circumstances of minority populations. The possibility of increasing racial disproportionality in incarceration due to the reliance on these instruments is a consequence worth weighing against the benefits of using actuarial risk assessment.
Chapter Three: Data

Sample

This chapter includes descriptions of the data sources used for the analyses in the study and provides an overview of sentencing practices in the state of Pennsylvania. Three datasets will be used to conduct this study: The first is a sentencing dataset provided by the Pennsylvania Commission on Sentencing (PCS), the second is an incarceration dataset from the Pennsylvania Department of Corrections (DOC), and the third is an arrest dataset provided by Pennsylvania State Police. The PCS data are Pennsylvania sentencing records ($N = 7,935$) from 2001 to 2005 for all offenders sentenced within the highest sentencing level in the state (level 5) (See Appendix G for the sentencing matrix that would have been in use during the study period). The sentencing data will establish the research sample and include demographic markers (i.e., offender age, race, gender) and case characteristics (e.g., conviction offense type, criminal history; refer to the Methods section in Ch.4 below, for a detailed description of individual variables). The DOC dataset will be linked to establish the offender release date and indicates whether the offender's parole was revoked. Criminal history reports from Pennsylvania state police will indicate if and when an offender was arrested post-release and also include individuals’ lifetime arrest history. The datasets used for this project were collected, appended, and recoded by the author during multiple visits to the Pennsylvania Commission on Sentencing in 2016, and for a prior project in 2012 (DeMichele and Laskorunsky, 2014).

Missing/Incomplete Data

All three datasets were linked by a State Identification number (SID), which is similar to
a social security number, and is used to track offenders across agencies in Pennsylvania. During the initial gathering of data on level 5 offenders, several hundred cases included missing SIDs. PCS was able to locate and verify most of these numbers using the offenders’ name and date of birth. Similarly, the DOC was not able to match some of the SIDs provided, but was successful in identifying some offenders by way of other personal markers. After linking data from the three datasets, 501 offenders originally included in the sentencing data were not able to be matched with the DOC and/or the state police data. These offenders were removed from the sample for analyses. No other variables presented substantial issues regarding missing information. Because it was possible to manually cross-check missing variables across the three datasets, variables that may have been missing in one dataset could be filled in using another dataset.

There were 14,026 level 5 offenders convicted in PA between 2001-2005, of which 7,935 met the requirements for sample selection. 2,880 offenders were still incarcerated, and an additional 2,067 had been released for less than three years by the cut-off time established for the study. Another, 1,144 offenders either died before the end of the study period, were transferred to mental institutions, or were excluded due to missing data⁷.

**Justification for Sample**

Sentencing, corrections, and recidivism data must be used in order to construct a recidivism risk assessment instrument. The combination of three agency datasets provides a longitudinal sample of serious offenders sentenced in criminal court and followed after serving their sentence. The combined dataset includes information that judges would have access to at sentencing, the offenders’ release date, and measures of recidivism, thus making it possible to

---

⁷ This includes the 501 offenders with missing SIDs.
construct a risk assessment instrument and to analyze the relationship between risk score, race, and reoffence.

The sample years were selected and are appropriate for three reasons. First, the PCS made changes to their data collection system in 2000 when they adopted a computer based system to improve recording keeping accuracy. Second, level 5 offenders receive the longest sentences in the state, and, therefore, these years provided enough time to allow for many of the offenders to serve out their sentence and to be released from prison. Third, the selected years provide five years of sentencing data, which lend some assurances that the findings are not unduly biased by unobservable temporal trends.

Level 5 Offenders

The study will focus on recidivism patterns and predictors of level 5 offenders, who are the most serious offenders to be sentenced under the Pennsylvania sentencing guidelines.\(^8\) This group of offenders was selected for two main reasons. First, the presumptive sentence for level 5 offenders is state incarceration (i.e., prison), although about a quarter are sentenced to county jail or to probation. A major goal of using actuarial risk assessment at sentencing is to use it to divert some offenders from incarceration in order to minimize the harm of prison and to maximize cost savings for the state. In Pennsylvania, and in the rest of the US, serious offenders take up a disproportionate number of government resources because the usually receive longer and more incapacititative sentences. Thus, assessing serious offenders whose presumptive sentence is prison is more important than assessing the general offending population, who is more likely to receive jail or community sanctions. Second, level 5 offenders have been convicted of serious crimes

\(^8\) Pennsylvania guideline levels rank from 1-5, from lower to higher presumptive sentence.
such as robbery, sexual assault, and large scale drug distribution, yet the majority are released within 5 years — and almost always on parole (DeMichele & Laskorunsky, 2014). Instances of recidivism within this population reflect on sentencing authorities and the parole board, making legal authorities averse to risk. As evidenced, the population of level 5 offenders is of particular interest to policy makers and court and correctional practitioners.

Level 5 offenders are considered the most serious offenders, mainly given the nature of their crimes. 9 Level 5 offenders are made up of offenders who were convicted of crimes in offense gravity levels 9-14, as well as some offenders, with long history records, who were convicted of crimes in offense gravity levels 7 and 8 (see Appendix G for the sentencing matrix). Crimes at offense gravity levels 9-14 are include drug distribution, sexual assault, and robbery (see Appendix C for a full list of convictions coded for the analysis). The presumptive sentence for level 5 offenders is state prison with the standard minimum sentence ranging from 12 months to 240 months. 10 Younger offenders are also eligible for state boot camp.

Follow-up Time

Offenders began their sentence between January 1, 2001 and December 31, 2005 and were followed up to 11 years after release — however recidivism analysis only includes the immediate three years after release. Release dates was calculated depending on the sentence. For offenders who received a prison sentence (63% of the sample) recidivism tracking began on the date of their release, which was obtained from the DOC. For offenders sentenced to probation or

9 Level 5 does not include Murder 1 and Murder 2 offenses, which fall outside of the sentencing guidelines. These offenders receive either lifetime imprisonment or the death penalty, and therefore are not included in the analysis.

10 Offenders in Pennsylvania are not eligible for parole until their minimum sentence is up.
other community sanctions (9% of the sample), recidivism tracking began on their sentence date. I was not able to obtain exact release dates for offenders sentenced to jail (28% of the sample), because jail records are not centralized in Pennsylvania and obtaining them would have required obtaining individual datasets from 67 counties. Thus, offenders who received a jail sentence were tracked beginning on the expiration of their minimum sentence date. In Pennsylvania, the majority of county jail inmates are released at the expiration of their minimum sentence.

**Pennsylvania's Sentencing System**

Pennsylvania is a large, diverse state. It is particularly well suited for this study because the state has a long history of guidelines sentencing and well-documented levels of racial disproportionality in incarceration (R. L. Austin & Allen, 2000). With the passage of Act 95 of 2010 (42 PA.C.S. §2154.7), the Pennsylvanian legislature also mandated that the Pennsylvania Commission on Sentencing develop an instrument for judges to use at sentencing, and this policy spurred the interest in this study. Finally, Pennsylvania maintains excellent data on its sentencing and corrections systems and processes and both the Pennsylvania Commission on Sentencing and the Pennsylvania Department of Corrections make this data available to researchers.

**Pennsylvania Sentencing Guidelines**

Pennsylvania has had voluntary sentencing guidelines since 1978. Guideline ranges are determined by the Offense Gravity Score (1-14) and the Prior Record Score (0-5, and Repeat Felony Offender and Repeat Violent Offender designations) (see Appendix F for the current guideline matrix). The recommended sentencing ranges present a sentence minimum in months and the presumed mode of punishment (state prison, county jail, probation, etc.). Aggravating or
mitigating circumstances (defined by the law) are integrated into the guidelines and can increase or decrease a minimum sentence up to 12 months. The maximum sentence must be at least double the minimum sentence. Convicted offenders must serve their minimum sentence before being eligible for either county or state parole. Judges are free to ignore the sentencing guideline recommendations as long as they provide a reason for departing from the guidelines. These sentences are recorded as "departures above" and "departures below" to indicate whether the departure goes above or below the recommended guideline range. The guideline range is calculated by court officials after conviction and is given to the judge before the sentencing stage. According to the PCS, judges sentence within the guideline range in 90% of their rulings (Pennsylvania Commission on Sentencing, 2012).

As of this writing, Pennsylvania is in the process of integrating actuarial risk assessment into sentencing. It has constructed multiple recidivism risk assessment instruments for judges to use during sentencing, conducted some preliminary testing on the tools integration. Current efforts on refining the instrument are continuing and the instrument is slated to be released in 2018. It should be noted that the risk assessment instrument produced for this dissertation was constructed with PCS input, but independent of agency efforts to create an instrument for official use. PCS’s proposed instrument for level 5 offenders was developed and validated on a different sample of offenders, operationalizes variables in a different way (e.g., conviction offense), and uses a smaller set of variables for prediction risk of re-arrest.

11 PCS has produced multiple reports about this process which are available at: http://pcs.la.psu.edu/publications-and-research/research-and-evaluation-reports/risk-assessment.
Chapter 4 - Risk Assessment Development and Validation

Overview

Evidence-based sentencing is the process of using social science methods to identify offender characteristics that are associated with desired sentencing outcomes (Casey et al., 2011). Scholars, judges, and policy makers have called for the adoption of evidence-based sentencing at the state and federal level (Chanenson, 2003; Gottfredson, 1999; Kern & Bergstrom, 2013; Silver & Chow-Martin, 2002; Vigorita, 2003; Wolff, 2006). A central element to evidence-based sentencing is the assessment of offender risk at the sentencing stage. The National Center for State Courts has called for states to "get smarter about sentencing" by adopting risk assessment instruments to assist judges in selecting sentencing options that protect the public, hold offenders accountable, and reduce recidivism (Casey, Warren, Elek, 2011). Research shows that jurisdiction-specific instruments are best at accomplishing these goals (Z. Hamilton et al., 2016; Latessa, Lovins, & Makarios, 2013). However, there is a lack of guidance on the creation of actuarial risk assessment instruments which are specific to a native population. There have been few efforts to create instruments specifically designed for use during the sentencing stage, and only the Virginia Criminal Sentencing Commission and the Pennsylvania Commission on Sentencing have openly documented the processes of those efforts (Kleiman et al., 2007 - for non-violent offenders; Pennsylvania Commission on Sentencing, 2016). About a dozen states currently use actuarial risk assessments at the sentencing stage, and several others are considering adopting the practice (Starr, 2014). Most of these states use instruments that were previously created for use at different points in the criminal justice system (e.g., Level of Service Inventory-Revised [LSI-R], Correctional Offender Management Profiling for Alternative
Sanctions [COMPAS], and based on populations which may not be representative of the offending population in that state. Using instruments not validated for the sentencing stage may cause inaccuracies in assessment measures, resulting in misclassification of offenders as it relates to recidivism (M. Hamilton, 2015). Furthermore, because private companies sell access to proprietary risk assessment instruments, interested parties are not able to examine exactly how given risk scores are calculated. The creation and testing of a risk assessment instrument for use at the sentencing stage would provide guidance on methodological issues and serve as an example for future development efforts.

In the previous three chapters I situated the importance of sentencing risk assessment within the broader movement to address growth in correctional populations, outlined the goals of this research, summarized previous research on sentencing, risk, and racial disparity, and described the data used for these studies. This chapter will identify which factors predict re-offending in a group of serious offenders convicted in Pennsylvania and outline the development and validation of a risk assessment instrument which judges can use at the sentencing stage. In the next section, I present a brief review of prior research that motivates the statistical methods used for instrument construction and cross-validation. The results section presents models that were used to identify factors significantly related to recidivism, the risk group classification procedure, and validation efforts. The concluding chapter reviews the methodological, theoretical, and practical contributions of this analysis.

**Literature Review**

**The Actuarial Risk Assessment Instrument**

At a basic level, all risk assessments predict the likelihood (i.e., risk) of an event
occurring. Andrews & Bonta (2010b) described 1st generation assessments as consisting of clinical intuition: a clinician uses his or her professional opinion to predict whether someone may be prone to violence after release. Alternatively, actuarial risk assessment instruments use statistical methods on large datasets of criminal offending rates to determine the different levels of offending associated with group traits - and on the basis of those correlations, predicts the criminal behavior of a particular individual (Harcourt, 2008, p. 16). Offenders with certain offense characteristics or demographics recidivate at unequal rates and these risk factors are identified for the purpose instrument creation. Most risk instruments used in criminal justice classify offenders according to their risk score (i.e., the number of risk factors and the importance of those risk factors in predicting recidivism), with higher risk scores correlating to a greater chance of failure (i.e., a higher percentage of that group recidivating).

These scores are often regrouped to produce a manageable number of categories, which correlate with distinct differences in risk of re-offense. Subsequently, when this assessment is used in practice, it provides a likelihood estimate of failure for an offender based on how similarly situated offenders have behaved in the past. The output is less of a "prediction" (i.e., will fail/won't fail) and more of a "post-diction" in which offenders are classified into groups based on measures of past behavior (Auerhahn, 2006). In general, offenders with more risk factors have a higher chance of recidivating because they share characteristics with offenders for whom the re-offending base rate is known. While risk assessments are given individually, predictions are based on aggregate levels of recidivism. For example, risk assessment instruments can predict that a certain number of offenders in a risk group will recidivate, but it cannot determine which individual offender will. Risk assessment scores are often related to terms such as low-, med- or high-risk, however, there are no standards for classifying according
to these terms. For example, offenders who have a 50% probability of recidivating may be considered high risk in one jurisdiction and medium risk in another.

States Using Actuarial Risk Assessment at Sentencing

The use of risk assessment tools to guide criminal justice actors' decision making is not new. Burgess (1928) worked with the Illinois State Parole Board to develop a parole release instrument that relied on an additive binary assessment instrument of 21 factors to predict which offenders were most likely to succeed and fail on parole. As a comparison, three trained psychiatrists also made predictions on a subsample of the 3,000 offenders involved in Burgess' study. The clinicians were found to be slightly more precise at identifying parole successes, but significantly less accurate in predicting parole failures. Reviewing Burgess' research, Grove and Meehl (1996, p. 293) stated that the "...conclusion was clear that even a crude actuarial method such as this was superior to clinical judgment in accuracy of prediction."

More recently, there have been efforts to create jurisdiction-specific risk assessment instruments for use at the sentencing stage. The Virginia Criminal Sentencing Commission was the first to develop and adopt a state-wide risk assessment instrument for use with the general criminal population (Virginia Criminal Sentencing Commission, 2004). The Commission’s goal was to divert low-risk offenders away from incarceration as a way to reduce prison populations. Using survival analysis, they found 11 statistically significant factors predicting recidivism (i.e., re-conviction or arrest for a new felony or misdemeanor offense) at the sentencing stage and assigned scores based on the relative importance of each of the factors. The National Center for State Courts (Ostrom et al. 2002) also published a report on the instrument creation process and found that the instrument was able to classify recidivists and non-recidivist with moderate
accuracy. Their pre- and post- evaluation also showed a cost benefit to the state of Virginia due to diverting offenders.

The Missouri Sentencing Advisory Commission (MSAC) also developed a risk assessment tool for judges to use at sentencing. This Commission also identified 11 factors that were associated with a new conviction or a return to prison, and assigned scores based on their relative importance. Their assessment is integrated into the voluntary sentencing guidelines and low, medium, and high risk scores correspond to mitigated, presumptive, aggravated sentencing ranges (Wolff & Oldfield, 2010). The Missouri Sentencing Advisory Commission has not published the analysis and methods used to create their risk assessment instrument.

Finally, the Pennsylvania Commission on Sentencing (PCS) published research on its efforts to develop a sentencing risk assessment, first using lower-level offenders (Pennsylvania commission on Sentencing, 2012), then developing one assessment for each offense gravity score (Pennsylvania Commission on Sentencing, 2015). The Commission used eight factors to predict risk of re-arrest within three years of release. Their plan is to include risk assessment scores with each sentencing guideline recommendation in order to encourage judges to order a pre-sentence report for offenders on the high and low end of the scale. As of this writing, PCS is preparing to revise the most current iteration of the scale to exclude any arrests which were dismissed at the preliminary hearing.

While there are many benefits to developing and validating a jurisdiction-specific tool, the effort takes time, financial resources, and access to research expertise on this topic. Thus, the majority of states using risk assessments at sentencing use instruments that were developed for recidivism prediction at a different stage in the criminal justice system and/or on non-local populations. A survey of state practices by the Electronic Privacy Information Center (2016)
showed that while many state statues recommend or authorize the use of risk assessment at sentencing, only Missouri and Virginia currently use tools that have been created or specifically adopted for use at the sentencing stage. Other states’ assessment tools rely on proprietary instruments such as the COMPAS and the LSI-R. For example, Colorado requires the inclusion of LSI-R results in pre-sentence reports, which are generally ordered for more serious offenders and compiled by the probation department (Casey et al., 2013) and Wisconsin uses the COMPAS to inform judges prior to sentencing (State of Wisconsin v. Loomis, 2016). Several states rely on state-specific instruments that were developed for use at a different point in the criminal justice system. For example, Indiana uses the Community Supervision Screener in pre-sentencing reports, which assesses the offender's likelihood of success on community supervision. The tool was adopted from the Ohio Risk Assessment System (ORAS) and validated using offenders on community supervision in Indiana (Latessa et al., 2013).

**Utilizing Risk Factors for Prediction**

Risk assessments typically include two types of data that are important to consider when making criminal justice decisions: static and dynamic predictors of recidivism (Andrews et al., 2006). Static risk factors are offender characteristics that either do not change or change only in one direction (e.g., gender, age), and dynamic predictors are offender characteristics that may change with time (e.g., employment, education). Both static and dynamic risk factors are already, informally, considered in courtroom decisions, on top of legally relevant factors such as seriousness of offense and criminal record (Steffensmeier and Demuth, 2006; Ulmer et al., 2016). Actuarial risk assessment tools provide a standardized way to assess these risk factors as they relate to an offender's risk of recidivism.
There is substantial overlap in the risk factors used to predict recidivism from instrument to instrument (Desmarais & Singh, 2013). While there are many known risk factors predictive of criminal involvement (Gendreau et al., 1996), information that is unavailable or not easily obtainable at the sentencing stage cannot be included in an instrument. The handful of risk assessment instruments created specifically for use at the sentencing phase have focused more on static risk factors that are easily obtained from administrative records and less on dynamic measures such as employment and marital status. The inclusion of the latter generally requires some level of investigation or offender input. Unsurprisingly, the most common group of factors included in risk assessments across the criminal justice system are measures of criminal history: past arrests, prior incarceration, juvenile record, etc. Offender demographics, such as gender, age, and place of residence also appear with inconsistency across instruments. Alternatively, risk assessments are sometimes created for men and women (e.g., the Minnesota Screening Tool Assessing Recidivism Risk; Duwe, 2014) and for juvenile and adult offenders, separately. For pre-trial and sentencing assessments, measures of offense type and severity are also common (e.g., serious violent offense).

Interview-based risk assessment instruments, such as COMPAS and LSI-R, go into depth by measuring multiple dimensions of risk, such as association with criminal peers and criminal thinking patterns, and they identify needs, such as substance abuse issues. However, the use of instruments that identify both risk and need factors may not be necessary at the sentencing stage during which judges are mainly concerned with predicting risk of re-offense. These instruments require substantially more resources to administer than instruments that rely mainly on administrative data. Furthermore, studies have shown that adding dynamic risk factors to an instrument does not significantly improve the predictive validity of an instrument and may be
unnecessary if the purpose of the instrument is only to predict recidivism (Austin, 2006; Gendreau et al., 1996).

Many risk factors found in risk assessment instruments are grounded in criminological theory and prior research. For example, research shows that individuals with previous criminal offenses are more likely to commit future crimes (Andrews et al., 2006; Gendreau et al., 1996). That is, the number of arrests and convictions have a positive relationship with risk of arrest (Kleiman et al., 2007; Silver & Chow-Martin, 2002). Both arrest and convictions are important to prediction of recidivism because convictions may be affected by process issues (e.g., charges may be dropped due to lack of evidence or police misconduct), whereas arrest provides a more general indication of the level of contact with the legal system. Additionally, research shows that offenders who have been arrested for domestic violence have an elevated risk of being arrested again (D. Wilson & Klein, 2006). Having contact with the juvenile justice system is also a strong predictor of recidivism (National Research Council, 1986b), and the younger a person is involved with the criminal justice system the more likely they are to continue their criminality into adulthood (Piquero, Brame, & Lynam, 2004).

Prior research also routinely finds associations between crime and demographic characteristics. Two of the strongest relationships between individual characteristics and crime are gender and age. First, criminological research and crime incidence data demonstrate that males are significantly more likely to engage in crime, to be under some form of criminal justice system control, and to be a victim of crime (Harrison & Beck, 2002; Steffensmeier & Allan, 1996). Further, research has shown that men display higher rates of violence, aggression, and criminality than have women across temporal or spatial locations (Archer, 2000; Simons & Burt, 2011). Second, the age-crime curve is one of the more generally agreed upon relationships...
within criminology, as younger individuals commit more crime than their older counterparts (i.e., the age-crime curve follows a normal distribution from 12 to around 25 years of age). Other research proposes that as people progress through life, informal control mechanisms (e.g., marriage, work obligations) reduce rates of offending (Laub & Sampson, 1993; Toby, 1957). Sampson and Laub (2003) also argued that everyone eventually ages out of crime.

Research shows differences in recidivism rates of sex, violent, property, and drug offenders. For example, federal data show that property offenders consistently recidivate at a higher rate than other types of offenders (Beck & Shipley, 1989; Langan & Levin, 2002), and that property crimes are committed more often than personal crime (Blumstein & Cohen, 1979). Researchers also consistently shows that the recidivism rate is lower for sex offenders than for other types of offenders (Hanson, Scott, & Steffy, 1995; Langan, Schmitt, & Durose, 2003; Sample & Bray, 2003). Research on drug offenders and violent offenders is mixed. Lo, Kim, and Cheng’s (2008) retrospective study suggested that both drug and violent offenders were more likely to experience long periods of no arrest. Data from the Bureaus of Justice Statistics (Langan and Levin, 2002) showed that drug offenders were less likely to recidivate than property offenders, but were also more likely to recidivate than sex and violent offenders. However, studies on serious drug offenders suggest that they are more likely to recidivate than other types of offenders (e.g., Holleran & Spohn, 2002). Likewise, some studies on violent offenders show that they are less likely to return to prison than non-violent offenders (Schwanert, 1998), while other studies report that violent offenders have a lengthy criminal history compared to non-violent offenders and commit a disproportionate amount of crime (Farrington, 2003; Moffitt, 1993).

Another important criminal offense characteristic is the seriousness of the conviction
offense. While offenders with more serious offenses (i.e., offenses that cause more harm) are more likely to receive harsher sentences, the relationship between offense gravity and recidivism is more complicated. For example, the PCS's Report 2 (Pennsylvania Commission on Sentencing, 2011) reported that the relationship between offense gravity and recidivism was negative; that is, those sentenced for more serious crime were less likely to commit another offense after release (see also, DeMichele & Laskorunsky, 2014). Other risk assessments research also found that lower-level offenses were linked with a higher rate of recidivism (Texas Parole Board, Ohio-ORAS). Lower level crimes, such as theft or drug possession, require less pre-mediation and are often done by younger, risk-seeking individuals who are likely to recidivate. Other markers of offense seriousness, such as use of a gun during the offense, also have mixed research findings. For example, Huebner, Varanos, and Bynum (2007), reported that gun use was not associated with an increased risk of post-release recidivism, while Daniel (2010) reported that firearm involvement increased the risk of recidivism for gang offenders only.

Race as a Predictive Factor

The over-representation of racial and ethnic minorities in the correctional populations has been linked to the effects of social characteristics related to disadvantage and inequality that minority groups experience, as well as to structural factors that expose minorities to greater criminal justice control (Sampson & Groves, 1989; Shaw & McKay, 1942; Wilson, 1987). Prior risk assessment research has shown that minorities, especially Blacks, displayed a higher likelihood of re-offending (Kleiman et al., 2007; Monahan, Steadman, Silver, & Appelbaum, 2001). For example, Kleiman et al. (2007) reported that being Black explained a significant portion of the variance in both re-conviction and re-arrest. However, modern actuarial risk
assessment instruments typically do not include race or ethnicity as predictor variables, because of ethical considerations when using race to predict criminal behavior (Tonry, 1987) and potential legal issues of using a constitutionally protected status to make punishment decisions (Starr, 2014). Some critics argue that neighborhood and criminal history serve as stand-ins for race (Harcourt, 2008; Starr, 2014) and, thus, should not be included in risk assessment instruments. One way researchers address the significance of race in predicting recidivism is by including race in the model building procedure for the risk assessment instrument, but excluding it from the scoring phase. While this process may lower the overall predictive utility of the instrument, it prevents other explanatory variables from soaking up the explanatory power of race/ethnicity.

Options for Operationalizing Recidivism

For the purposes of risk assessment development, recidivism may be operationalized in several forms, such as arrest, re-conviction, and re-incarceration. Those three categories can be further broken down by type of offense, such as re-conviction for a violent or sex crime. Because many jurisdictions use generic risk assessment tools, recidivism is already operationalized in a particular way. Ideally, however, the outcome measure should be based on the needs of a particular jurisdiction and the purpose of the tool. Arrest for a felony or a misdemeanor crime is the most common outcome metric and is the least conservative measure of recidivism. Brame (2016, p. 69) writes that arrest is the “combined product of a criminal act, a victim's reporting of that crime to the police (unless the police observe it directly), and a clearing of the probable cause evidentiary standards for police to take action.” The criminal act is the only measure indicative of public safety. The likelihood of detection and enforcement may be reflective of
“patterns or biases that have little to do with public safety.”

For this reason, some jurisdictions prefer to use recidivism measures which provide more of a check against disproportionate minority contact. For example, re-conviction and re-incarceration require a greater standard of evidence that a crime has occurred. Focusing on felony re-convictions, or re-convictions for a violent or sex crime, further focuses the assessment on events that directly affect public safety. For example, Pennsylvania stopped short of using re-conviction as an outcome variable for their sentencing risk assessment, but is in the process of changing the outcome measure from simple arrest to arrest upheld by a minor court (e.g., at a preliminary arraignment). A benefit of using stricter measures of recidivism is that it minimizes the effect of false arrests, which have been criticized for increasing racial disparity at the arrest stage. However, using more conservative measures also ensures that many offenders who commit a crime are not identified due to procedural rules or other system process issues. For example, research using charging data from the Los Angeles Police Department and the Los Angeles Sheriff’s Department showed that there was substantial attrition in the processing of serious sexual offense cases, mainly due to the difficulty of prosecuting sex crimes — and that only 30% to 60% of offenders arrested for rape or attempted rape were incarcerated for their crimes (Spohn, White, & Tellis, 2014).

Prediction Methods for Classification

Most risk assessment instruments in use in the criminal justice system are a form of additive scales. Factors related to recidivism are identified and weighted (e.g., scored) relative to their importance. These scores are summed up to create a composite score for each individual offender, which relates to their likelihood of recidivism. The Burgess (1928) method of using
unweighted binary risk factors (i.e., scoring 1 when the risk factors is present, 0 when it is not) is a conceptually simple way to construct a risk assessment scale and remains the most popular method for constructing criminal justice risk assessment scales. While it has been shown to be just as predictive as more complicated weighted and non-binary methods (Andrews et al., 2006; Gottfredson, 1986), critics argue that it is imprecise because each risk factor is given equal treatment regardless of its effect strength (Duwe & Kim, 2016). Thus, the Burgess method has been further refined to assign weights to each risk factor (e.g., Glueck & Glueck, 1950) and, more commonly, to change the predictor measure unit to a non-binary scale. For example, for a risk factor such as arrest, an ordinal scale could be constructed (0 arrest = 0 points, 1-3 arrests = 1 point, 4 or more arrests = 2 points, etc.). Most modern risk assessment instruments were developed by pairing a Burgess or a Burgess-like weighting technique with regression methods to identify risk factors (Duwe and Kim, 2016). Logistic or Cox regressions can be used to develop multivariate prognostic models to identify parsimonious predictor variables, and regression coefficients can be used to assign weights. However, weights and ordinal scales can also be constructed using simple bivariate analysis (J. L. Johnson, Lowenkamp, & VanBenschoten, 2011).

Other, less popular, methods of risk prediction have appeared in the last several decades. For example, data mining and machine learning approaches show promise in increasing the predictive accuracy of risk assessment instruments (Berk, Sherman, Barnes, Kurtz, & Ahlman, 2009; Duwe & Kim, 2016). These techniques test different combinations of risk factors many times, each time refining the model to better fit the data. Two major differences set these methods apart from more traditional approaches. First, the risk prediction models are atheoretical, meaning that risk factors are identified solely based on their predictive utility. In
fact, Richard Berk, a proponent of these methods for recidivism prediction, writes that if shoe size predicts criminal behavior, it should be included in the model (Berk & Bleich, 2013). The second major difference is that an assessment using these methods can only be administered using specialized software and, thus, is not as transparent or convenient as simpler methods. Classification Tree Analysis (which sometimes counts as a form of machine learning) is another method that has appeared in risk assessment research. In particular, The MacArthur Study of Mental Disorder and Violence (Monahan et al., 2001; Silver & Chow-Martin, 2002) utilized a multiple model approach to improve prediction of violence over traditional regression methods. This approach partitioned the data into smaller and smaller groups (nodes) based on differences in risk level and capitalized on the notion that each person can be better partitioned using the unique combination of risk factors they have. While this approach also involves the use of specialized software to administer, it does not suffer from the same interpretability problems as more complicated machine learning approaches. However, it is still not preferred over computationally simpler approaches that require less resources to develop and administer.

**Evaluating Accuracy Using the AUC Statistic**

At a minimum, risk assessment instruments must be able to predict recidivism better than chance, and ideally better than informal prediction. Researchers report that actuarial methods are superior to informal predictions of risk (i.e., clinical judgement) (Grove & Meehl, 1996) and meta-analyses show that there are relatively small differences in predictive accuracy between instruments (Yang, Wong, & Coid, 2010). However, some evidence indicates that more computer intensive techniques, such as machine learning, are superior to traditional regression and Burgess style methods (Duwe & Kim, 2015). The area under the receiver operator
characteristics curve (AUC) has been widely used assess the predictive discrimination of risk assessment instruments (Taxman & Dezember, 2017). For a standard instrument which groups offenders into multiple “bins” that vary along a spectrum of risk, the AUC statistic indicates the chance that an offender who recidivates will obtain a score higher than one who does not. A value of 0.5 indicates that the instrument score predicts no better than chance, while a score of 1 indicates that the instrument score discriminates between recidivists and non-recidivists perfectly. The AUC statistic is unaffected by base rate differences in offending. A meta-analysis by Yang and colleagues (2010) found that the AUC values across nine validated violence prediction instruments in circulation ranged from 0.56 and 0.71. This a range that is generally associated with "moderate" to "strong" predictive utility (Rice & Harris, 2005). It is rare to observe AUC values greater than 0.8 for recidivism prediction (Yang et al., 2010). Perhaps most importantly, the AUC measure is more meaningful when it is calculated during a validation process — which involves developing and validating the instrument on separate groups of offenders. In particular, jurisdictions that use generic instruments, should not rely on the AUC values obtained from administering the instrument using a non-representative sample, as the predictive validity may vary from population to population.

**Research Questions**

The purpose of this study is to create and validate a risk assessment instrument using static variables to which judges in Pennsylvania have access during sentencing. Two overarching questions motivate this research:

1. Which offender and case characteristics best predict re-offending for serious adult offenders at the sentencing stage?
2. How accurately can risk of re-offense be predicted at the sentencing stage?

Analytical Methods

Using a large sample of serious offenders (N=7,935), I use bivariate analysis to test the relationship between case and offender characteristics and recidivism. The sample is then split into a development group (N=5,260) and a validation group (N=2,675). Using the development sample, I use multivariate logistic regression and block testing analysis to identify which factors best predict an arrest or parole revocation within three years of release. I use a modified Burgess method to compile risk factors into a risk assessment algorithm and classify offenders according to their risk of reoffence. Once the risk assessment model is constructed, I use the validation sample to estimate the AUC statistic of the risk scale in order to measure predictive validity.

Predictor Variables

A hallmark of a thorough recidivism study is one that tests a large variety of risk-related factors in order to minimize omitted variable bias (Schwalbe, 2007). The risk related measures tested for their explanatory power include three main groups: demographics (e.g., age, race, gender), criminal history (e.g., number of prior arrests) and offense related variables (e.g., offense gravity). Items used to develop this instrument were based on prior research in predicting offending behavior (e.g., Andrews et al., 2006; Gendreau et al., 1996; Kleiman et al., 2007) and were available at the sentencing stage. As stated earlier, the focus is on creating a risk instrument using static factors which could easily be obtained from administrative data, to maximize predictive utility, interpretability, and ease of use. Table 4.1 describes the independent variables used in the bivariate and multivariable logistic analysis.
Table 4. 1: Predictor Variables

**Independent Variables**

**Offender Demographics**
- Gender: Male/Female
- Race/Ethnicity: White, Black, Hispanic and other race\(^{12}\)
- Age at sentencing
- Sentencing County: Rural, Other Urban, Philadelphia, Allegheny\(^{13}\)

**Criminal History**
- Total prior arrests (at date of sentence)
- Juvenile Arrest: offender was arrested or convicted of a crime prior to 18
- Prior Record Score (PRS): 0-5, Repeat Felony Offender (6), Repeat Violent Offender (7)\(^{14}\)
- Domestic Violence (DV) Arrest: Offender had a prior arrest for domestic violence
- Prior Conviction Type: sex, drug, violent, property, and/or miscellaneous
- Specialist Offender\(^{15}\)

**Offense Related Variables**\(^{16}\)
- Type of conviction offense: sex, violent, drug, property, or miscellaneous\(^{17}\)
- Offense Gravity Score (OGS): 7-14\(^{18}\)
- Multiple Counts: offender was convicted of multiple counts
- Use of gun during offense
- Inchoate Offense: crime was not completed (e.g., attempted murder)

The regression models used to develop the instrument include race/ethnicity and county variables as controls, but race/ethnicity and county are not included in the instrument scoring algorithm. Race/ethnicity and county are included in the model building procedure so that other

---
\(^{12}\) Hispanic ethnicity and Asian, Pacific Islander and Native American races were combined as a category because they had a similar rate of recidivism (36% and 41%, respectively). Race/ethnicity is not included in the final instrument and only serves as a control.
\(^{13}\) Pennsylvania’s 67 counties were separated into urban and rural counties according to population density. Philadelphia and Allegheny counties were kept separate because they qualify as the only two “large courts” in the state - which, may affect their sentencing practices (Ulmer and Johnson, 2004).
\(^{14}\) The PRS measures the number and seriousness of an offender’s past convictions. Convictions for any serious misdemeanors (e.g., firearm offenses) or any type of felony are counted in the total. To get a detailed look at the calculation see [http://www.pacode.com/secure/data/204/chapter303/s303.7.html](http://www.pacode.com/secure/data/204/chapter303/s303.7.html)
\(^{15}\) Offender specializes in sex, violent, property, or drug crimes. I used a 75% specialization threshold of prior and current convictions.
\(^{16}\) Most serious offense within a judicial proceeding.
\(^{17}\) Appendix D provides a summary of offense categorization. The overwhelming majority of miscellaneous offenses in the sample were convictions for the illegal possession of a firearm.
\(^{18}\) The OGS is a measure of offense seriousness. To see a full categorization of offenses by score see [http://www.pacode.com/secure/data/204/chapter303/s303.15.html](http://www.pacode.com/secure/data/204/chapter303/s303.15.html)
variables do not soak up their explanatory power. County of sentencing is not included in the risk assessment algorithm for two reasons. First, because of Pennsylvania's demographics, urban counties (Philadelphia and Allegheny,\(^{19}\) in particular) are highly correlated with race – they have a high rate of minority residents. Second, including county of sentencing in the scoring algorithm would mean offender assessed in higher risk counties would have a point added to their risk score (and vice versa for lower-risk counties), thereby reducing the between person comparative utility of the instrument.

Several measures of criminal history were tested, but ultimately not included in the initial model building procedure because of collinearity with other variables. For example, the total number of prior convictions is highly collinear with the prior record score (PRS), but is a less powerful predictor of recidivism. Similarly, separating the specialization variable into types of specialization (i.e. sex, violent, drug, and property) did not significantly predict recidivism over the general specialization variable.

**Outcome Measures**

The outcome measure is a binary variable indicating whether or not the offender recidivated. Recidivism is measured as a new arrest for a felony or misdemeanor within 3 years of release from prison or jail, or after the start of a probation sentence.\(^{20}\) Arrests that occurred as a result of a municipal (non-criminal) violation or citation are not included. Serious technical violations that result in a return to prison (i.e., revocation) are regarded as failure. Recidivism data were obtained over an 11-year period; thus, some offenders were monitored well past the

---

\(^{19}\) Pittsburgh is located in Allegheny county.

\(^{20}\) Out of state or federal arrests are not included.
three-year mark for follow-up. However, it is important to equalize offender’s time at risk, thus each offender had a minimum follow-up of up to three years. The majority of offenders who recidivated within the total 11-year study period (60.4%), did so within the first three years of release (46.5%). Appendix A shows a Cox regression model for all offenders who recidivated within the study period, with the risk of failure decreasing quickly after the 1-year mark.

**Analysis**

The analysis starts with descriptive statistics for the entire dataset (N=7,935). Then, I conduct bivariate analysis to identify which individual risk factors are statistically related to recidivism using a Chi-Squared test ($\chi^2$). Through bivariate analysis and a visual check of the data, I also choose the best cut-points for individual continuous variables. For example, age is separated into three categories which have below average (Age 41+), average (Age 22-40), and above average (Age 14-21) associations with recidivism. Next, the total sample is split into two random sub-samples: the development sample (N=5,260) (66%) and a validation sample (2,675) (34%). I develop the scoring model in the development sample and test its predictive utility in the validation sample to minimize the likelihood that the findings are a matter of over-fitting the data (Hastie, Trevor, Tibshirani, Robert, Friedman, 2009).

The development sample is used to investigate numerous relationships between the independent variables and recidivism using logistic regression and block testing analysis. Logistic regression calculates the probability of an event occurring and is usually the most appropriate approach to use when an outcome is binary and time at risk is equalized (J. L. Johnson et al., 2011). The benefits of using logistic regression are that it requires few statistical assumptions and that it can generate odds ratios for ease of interpretability. From the full
sample bivariate analysis, I rank the risk predictors in terms of the strength of their relationship with recidivism. Then, beginning with the control variables (county and race), I add each individual variable from strongest to weakest, until model fit is no longer improved.\(^{21}\) Once the final model is obtained, I systematically rotate the reference categories for variables that have more than two categories (i.e., current offense type, arrest, and age) to make sure that the significance of the relationship does not change due to the effect of the reference variable. To construct the risk assessment scale, I use a modified Burgess (1928) method to classify offenders — assigning either 1 or 2 points for each risk factor present.\(^{22}\) Once individual scores are assigned, I construct four risk levels - Low, Medium, High, and Very High - which correspond to risk of recidivism. Finally, to determine the predictive validity of the instrument, I use the validation sample to assess the AUC statistic of the risk score model.

**Results**

**Descriptive Statistics**

Table 4.2 displays the descriptive statistics for the entire sample \((N=7,935)\). Almost half (46.5\%) of offenders fail (i.e., re-offend) within 3 years of their release. This rate is lower than the average three-year recidivism rate for all offenders sentenced in Pennsylvania, which generally hovers in the 60\% range depending on the year (Pennsylvania Commission on Sentencing, 2012). Most likely this difference may be attributed to the absence of lower-level offenders from the sample who typically have higher re-offending rates. The overwhelming

\(^{21}\) Log likelihood test is used to compare model fit.

\(^{22}\) Bobko, Roth, and Buster (2007, p. 693) noted that unit weights and regression weights performed similarly in cross-validation studies
majority of the sample was male (90.3%), which is fairly typical for serious offenders. The sample had an almost equal number of White (45.6%) and Black (42.9%) offenders, with a small portion of Hispanics (8.7%) and other races/ethnicities (2.8%). The average age is 30 years old, with the youngest offenders being 14 years old and the oldest 78 years at sentencing. Very few offenders (14.9%) were sentenced in rural counties. The majority are from Philadelphia county (28.4%), Allegheny county (14.5%), or other urban counties (42.2%). Just over a quarter of all offenders have been arrested as a juvenile or have a juvenile adjudication on their record.

Offenders have an average of 5 arrests at the sentencing stage, with the distribution being fairly right skewed (s=4.8.) Offenders had average prior record score (PRS) of 1.5 - which also displayed a right skewed distribution (s=2.1). Almost 20% of offenders have a prior drug or miscellaneous offense and only 1% have a prior sex offense. A small portion also have a past property (7.4%) and violent (8.2%) offense. About 6.4% of offenders specialize in sex, violent, drug, or property crime and 10% have been arrested for domestic violence at some point in their lives. The overwhelming majority of offenders (68.6%) were convicted of a violent crime, an additional 16% were convicted of a drug offense, and approximately 12% were sex offenders. A very small amount of property offenders (2.1%) and miscellaneous offenders (1.4%) were also in the sample. Because level 5 offenses tend to be the most serious, property and drug offenses comprise only a small portion of the sample. The average offense gravity score was 10, which rests in the middle of the level 5 guideline grid. About one-third of the sample had multiple convictions during the most current judicial proceeding and about 4% used a firearm during the commission of their offense. Seven percent of offenders were convicted of an inchoate crime, which means that their crime was attempted.
Table 4.2: Descriptive Statistics (Full Sample)
N=7,935

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Percent</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recidivated</td>
<td>46.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

**Independent Variables**

*Offender Demographics*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (reference)</td>
<td>90.3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Female</td>
<td>9.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>White (reference)</td>
<td>45.6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Black</td>
<td>42.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Other Race</td>
<td>2.8</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Age^23</td>
<td>—</td>
<td>30.0</td>
<td>14.7</td>
<td>78.6</td>
<td>10.2</td>
</tr>
<tr>
<td>Allegheny County</td>
<td>14.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Philadelphia County</td>
<td>28.4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Other Urban County</td>
<td>42.2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rural County</td>
<td>14.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Offender Criminal History*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Record Score</td>
<td>1.5</td>
<td>0</td>
<td>7</td>
<td>2.1</td>
<td>—</td>
</tr>
<tr>
<td>Prior Arrest Total</td>
<td>4.9</td>
<td>1</td>
<td>63</td>
<td>4.8</td>
<td>—</td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>25.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prior Sex Offense</td>
<td>1.1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prior Violent Offense</td>
<td>8.2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prior Drug Offense</td>
<td>19</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prior Property Offense</td>
<td>7.4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Prior Misc. Offense</td>
<td>19.2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Specialist</td>
<td>6.4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>10</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Case Characteristics*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Percent</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offense Gravity Score (OGS)</td>
<td>—</td>
<td>10</td>
<td>7</td>
<td>14</td>
<td>1.2</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>11.9</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Violent Offense</td>
<td>68.6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Property Offense</td>
<td>2.1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Drug Offense</td>
<td>16</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>1.4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Multiple Convictions</td>
<td>30.2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Use of Gun</td>
<td>4.3</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Inchoate</td>
<td>7.1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

^23 In Pennsylvania, juveniles as young as 14 can be waived up to adult court for certain offenses. There are 150 juveniles in the sample.
Recidivism at 1/2/3 Years

Figure 4.1 presents the percent of offenders who recidivated at 1, 2, and 3 years after their release. At year 1, 23% percent of offenders had reoffended. By year 2, this portion jumps up to 38% and by the end of the their third year 46% percent had been arrested or failed on parole at least once.

Figure 4. 1: Recidivism Rate at 1, 2, and 3 Years
(N=7,935)

Bivariate Analysis

Using the development sample (N=5,260), I conduct bivariate analysis between individual risk factors and reoffence. Using a visual check of the data I break up continuous
variables into ordinal categories to simplify the scoring procedure, while maximizing risk
discrimination by the instrument. Age, number of prior arrests, PRS, and OGS were broken
down into ordinal variables which maximized differences in the strength of the relationship with
recidivism. For example, offenders who have 1-2 arrests have a lower than average arrest rate,
those who have 3-4 arrests have average arrest rates, and those who have 5 arrests or more have
a higher than average arrest rate. Table 4.4 shows the percent of offenders who recidivated at
each level of a predictor variable and which variables produced a significant relationship to
recidivism at \( p = 0.05 \) or lower (\( \chi^2 \)). Most variables were significantly predictive of recidivism
when no other variables were controlled for. All the demographic factors (race, age, and gender)
have a significant relationship to recidivism. All measures of criminal history are significantly
and positively related to recidivism, with the exception of offense specialization and having a
prior sex conviction. Finally, all offense characteristics are predictive of recidivism, except for
the use of a firearm during the conviction offense and having multiple counts at conviction.
Variables that are not significant have recidivism base rates that are close to the average (46%).
Not significant variables are unlikely to make it into the final scoring model, although they were
included in the subsequent model building procedure to maximize model fit (Johnson et al.,
2011).
Table 4.3: Bivariate Analysis for Predictor Identification  
(Development Sample) N=5,260

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Failure %</th>
<th>$\chi^2$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Offender Demographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>48%</td>
<td>49.15</td>
<td>***</td>
</tr>
<tr>
<td>Female</td>
<td>32%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>41%</td>
<td>49.51</td>
<td>***</td>
</tr>
<tr>
<td>Black</td>
<td>55%</td>
<td>103.98</td>
<td>***</td>
</tr>
<tr>
<td>Hispanic and Other Race</td>
<td>39%</td>
<td>23.2</td>
<td>***</td>
</tr>
<tr>
<td>Age 14-21</td>
<td>57%</td>
<td>87.7</td>
<td>***</td>
</tr>
<tr>
<td>Age 22-40</td>
<td>46%</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Age 41+</td>
<td>29%</td>
<td>113.56</td>
<td>***</td>
</tr>
<tr>
<td>Allegheny County</td>
<td>55%</td>
<td>22.89</td>
<td>***</td>
</tr>
<tr>
<td>Philadelphia County</td>
<td>50%</td>
<td>11.55</td>
<td>**</td>
</tr>
<tr>
<td>Other Urban County</td>
<td>43%</td>
<td>18.83</td>
<td>***</td>
</tr>
<tr>
<td>Rural County</td>
<td>42%</td>
<td>9.36</td>
<td>**</td>
</tr>
<tr>
<td><strong>Offender Criminal History</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Record Score=0</td>
<td>38%</td>
<td>205.46</td>
<td>***</td>
</tr>
<tr>
<td>Prior Recode Score=1+</td>
<td>57%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arreets 1-2</td>
<td>30%</td>
<td>356.28</td>
<td>***</td>
</tr>
<tr>
<td>Arrest 3-4</td>
<td>47%</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Arrest 5+</td>
<td>62%</td>
<td>343.17</td>
<td>***</td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>65%</td>
<td>259.81</td>
<td>***</td>
</tr>
<tr>
<td>No Juvenile Arrest</td>
<td>40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Sex Offense</td>
<td>39%</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>No Prior Sex Offense</td>
<td>46%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Violent Offense</td>
<td>57%</td>
<td>18.12</td>
<td>***</td>
</tr>
<tr>
<td>No Prior Violent Offense</td>
<td>48%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Drug Offense</td>
<td>58%</td>
<td>56.72</td>
<td>***</td>
</tr>
<tr>
<td>No Prior Drug Offense</td>
<td>44%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Property Offense</td>
<td>65%</td>
<td>51.88</td>
<td>***</td>
</tr>
<tr>
<td>No Prior Property Offense</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Misc. Offense</td>
<td>62%</td>
<td>114.76</td>
<td>***</td>
</tr>
<tr>
<td>No Prior Misc. Offense</td>
<td>43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialist</td>
<td>47%</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Not a Specialist</td>
<td>47%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>59%</td>
<td>32.68</td>
<td>***</td>
</tr>
<tr>
<td>No Domestic Violence Arrest</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>51%</td>
<td>90.45</td>
<td>***</td>
</tr>
<tr>
<td>OGS 11-14</td>
<td>36%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent Offense</td>
<td>51%</td>
<td>87.94</td>
<td>***</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>28%</td>
<td>91.95</td>
<td>***</td>
</tr>
<tr>
<td>Property Offense</td>
<td>67%</td>
<td>18.57</td>
<td>***</td>
</tr>
<tr>
<td>Drug Offense</td>
<td>37%</td>
<td>36.42</td>
<td>***</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>64%</td>
<td>8.28</td>
<td>**</td>
</tr>
<tr>
<td>Entity Type</td>
<td>Percentage</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Multiple Counts</td>
<td>48%</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>One Count</td>
<td>45.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Gun</td>
<td>49%</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>No Use of Gun</td>
<td>47%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inchoate Offense</td>
<td>37%</td>
<td>14.21</td>
<td>***</td>
</tr>
<tr>
<td>Completed Offense</td>
<td>47%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pearson's Chi Squared Test
* p < .05  ** p < .01  *** p < .001

**Multivariate Analysis**

Using the development sample (N=5,260), I use multivariate logistic regression and block testing analysis to arrive at a scoring algorithm for the risk assessment instrument. All predictor variables are first ranked by the strength of their relationship to failure. The model building procedure begins with race/ethnicity and county controls, and each predictor variable, from strongest to weakest, is individually added into the analysis until the model fit no longer improved. A likelihood ratio test compares the fit of each iteration of the model to the previous model and predictors are retained only if they significantly improve the model fit. Appendix B shows the block testing procedure, and Table 4.4 Model 1 shows the full model with all predictor variables. The multivariate model produced some important differences from the bivariate analysis (Table 4.3). For example, offenders who were of Hispanic ethnicity or other race were significantly less likely to recidivate according to results of the bivariate analysis, but they were not significantly less likely to recidivate than White offenders according to the multivariate analysis. Similarly, all counties were significantly correlated to arrest in the bivariate analysis, but compared to offenders from urban counties, only offenders from Allegheny County showed slightly higher odds of recidivating in the full model. This means that there is something unique about offenders sentenced in Allegheny County that increases their odds of recidivism compared to offenders from other counties, even when other risk factors, including race, are controlled for.
It’s possible that the policing and re-entry practices in Pittsburgh differ from those in other urban counties and this is reflected in higher recidivism odds for offenders across the board.

Finally, having an inchoate offense was no longer predictive of recidivism when controlling for other variables. Having a prior sex offense is associated with a decreased risk of recidivism ($or = 0.42; p<.01$) compared to having no prior sex offenses, however, this variable was removed from the final model (Model 2). While prior sex offense could be negatively scored on the risk assessment instrument (thus putting offenders who had a prior sex offense into a lower risk category), the small number of prior sex offenses in the sample (1% of the total sample), as well as the non-significant difference in the recidivism base rates from the bivariate analysis (39% sex prior vs 46% no sex prior, $\chi^2:0.25$), may be cause for questioning the robustness of the relationship. Using a firearm during commission of an offense, having multiple counts, and being a specialist offender remained insignificant and were removed from subsequent models.
Table 4. 4: Multivariate Logistic Model of Recidivism (Development Sample) N=5,260

<table>
<thead>
<tr>
<th>Offender Demographics</th>
<th>(1) Full Model</th>
<th></th>
<th>(2) Final Model</th>
<th></th>
<th>(3) County &amp; Race Removed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1.36 0.15 **</td>
<td>1.36 0.14 **</td>
<td>1.38 0.15 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>1.37 0.1 ***</td>
<td>1.34 0.09 ***</td>
<td>1.38 0.15 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic and Other Race</td>
<td>0.93 0.1</td>
<td>0.92 0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 14-21</td>
<td>3.33 0.4 ***</td>
<td>3.29 0.38 ***</td>
<td>3.3 0.38 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 22-40</td>
<td>2.1 0.2 ***</td>
<td>2.09 0.19 ***</td>
<td>2.06 0.19 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allegheny County</td>
<td>1.21 0.11 *</td>
<td>1.23 0.11 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philadelphia County</td>
<td>1.03 0.08</td>
<td>1.05 0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural County</td>
<td>1.07 0.1</td>
<td>1.08 0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offender Criminal History</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Recode Score=1+</td>
<td>1.32 0.12 **</td>
<td>1.36 0.01 ***</td>
<td>1.39 0.1 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrest 3-4</td>
<td>1.61 0.14 ***</td>
<td>1.6 0.14 ***</td>
<td>1.63 0.14 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrest 5 +</td>
<td>2.75 0.24 ***</td>
<td>2.78 0.24 ***</td>
<td>2.96 0.26 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>1.44 0.12 ***</td>
<td>1.44 0.12 ***</td>
<td>1.46 0.12 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Sex Offense</td>
<td>0.42 0.12 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Violent Offense</td>
<td>0.91 0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Drug Offense</td>
<td>1.0 0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Property Offense</td>
<td>1.22 0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Misc. Offense</td>
<td>1.18 0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specialist</td>
<td>0.75 0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>1.66 0.17 ***</td>
<td>1.66 0.17 ***</td>
<td>1.63 0.17 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>1.42 0.11 ***</td>
<td>1.42 0.1 ***</td>
<td>1.38 0.1 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Offense</td>
<td>0.99 0.13</td>
<td>1 0.13</td>
<td>0.98 0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent Offense</td>
<td>1.23 0.12 **</td>
<td>1.31 0.12 **</td>
<td>1.33 0.12 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Property Offense</td>
<td>1.72 0.42 *</td>
<td>1.85 0.42 **</td>
<td>1.78 0.4 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>1.76 0.5 *</td>
<td>1.8 0.5 *</td>
<td>1.91 0.52 *</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Counts</td>
<td>0.96 0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Gun</td>
<td>0.83 0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inchoate Offense</td>
<td>0.93 0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

-2 Log Likelihood                      | -3192.6        | -3204.84 | -3221.69        |          |                          |          |
Pseudo R2                               | 0.12           | 0.12     | 0.11            |          |                          |          |
AUC-ROC                                 | 0.73           | 0.73     | 0.72            |          |                          |          |

* p < 0.05 ; ** p < 0.01 ; *** p < 0.001.
Reference groups: female, White, Age 41+, Other Urban County, Arrest 1-2, Drug Offense, OGS 11-14
Table 4.4 Model 2 shows a model of recidivism with non-significant variables removed. Compared to Model 1, no major changes to effect direction or effect size occurred as a result of removing non-statistically significant variables. Most of the demographic predictor variables have significant associations with recidivism. Males have 36% higher odds of recidivating than females ($or=1.36; p<.01$). Black offenders have 34% higher odds of recidivating than White offenders ($or=1.34; p<.01$). Further, offenders who were under age 22 are over three times as likely to recidivate, ($or=3.29; p<.001$) and offenders between ages 22 and 40 are twice as likely to recidivate ($or=2.09; p<.001$), as offenders over age 40.

Compared to offenders in other urban counties, those from Allegheny county have 23% higher odds of recidivating ($or=1.23; p<.05$). Most of the criminal history variables also significantly predict recidivism and all were in the expected directions. Offenders with a prior record score of 1 or more have 36% higher odds of recidivating than offenders who had a prior record score of 0 ($or=1.36; p<.001$). Offenders who were arrested 3 to 4 times have 60% higher odds of recidivating than offenders who only had 1 or 2 arrests ($or=1.6; p<.001$), and those arrested 5 times or more have almost three times the odds of recidivating ($or=2.78; p<.001$).

Offenders who had a juvenile arrest or conviction are also more likely to recidivate than those who did not ($or=1.44; p<.001$). Additionally, offenders who had a prior domestic violence arrest have 66% higher odds of recidivating than those who do not, when controlling for the number of arrests ($or=1.66; p<.001$). As a category, a smaller portion of the case characteristics is predictive of recidivism compared to the demographic and criminal history variables. Offenders who committed less serious crimes are more likely ($or=1.42; p<.001$) to recidivate than those who are convicted of more serious crimes.

Offenders convicted for a sex crime are no more likely to recidivate than those sentenced
for drug crimes \((or=1; \text{NS})\). However, offenders convicted of a violent \((or=1.31; p<.001)\), property \((or=1.85; p<.01)\), or miscellaneous \((or=1.8; p<.05)\) offense, are more likely to recidivate than offenders convicted of a drug offense. A likelihood ratio test between Model 1 and Model 2 shows a slight, but significant, loss in model fit \((\chi^2: p<.001)\). Similarly, a test of the difference in the AUC-ROC statistic shows a slight loss in the discrimination between offenders \((p<.001)\).\(^{24}\) However, it is more useful to include several strong risk predictors in an instrument than to include a large variety for an incremental increase in model fit.

To show how the control variables changed the strengths of relationships between the predictor variables and the outcome variable, I run a model that excludes race and county as predictors. Table 4.4 Model 3, shows that removing the control variables race/ethnicity and county does not change the direction or substantially change the size of any of the predictors from Model 2. A likelihood ratio test showed that the removal of the controls slightly, but significantly, decreased the fit of the model \((p<.001)\). Similarly, a test of the difference in the AUC values between Model 2 and Model 3 showed a slight loss in the discrimination utility \((p<.001)\). (If county and race/ethnicity were not excluded from the scoring procedure for ethical and practical reasons, Black offenders and offenders from Allegheny county would each receive 1 point in the scoring algorithm.) To complete the final scoring algorithm, I rotated the categorical predictor variables (i.e., age, arrest, and type of offense variables) between models to ensure that the differences between the variables was not caused by a specific reference category. No significant changes from Model 2 were required.

In sum, the bivariate analysis and the multivariate analysis answer Research Question

\(^{24}\) To compare the area under the ROC curve between models, I use logistic linear predictors and the roccomp command in Stata 14.
1— Which offender and case characteristics best predict recidivism at the sentencing stage? The best predictors of recidivism are measures of criminal history: prior arrest total, having contact with the juvenile justice system, having a prior felony or serious misdemeanor convictions (i.e., PRS >1+) and having a domestic violence arrest. Perhaps unsurprisingly, having increased contact with the criminal justice system, in a variety of ways, is related to a higher recidivism rate. Although not as predictive as criminal history measures, demographics (i.e., age, gender, and race) also show a significant relationship with recidivism. Younger offenders, Black offenders, and male offenders are more likely to recidivate. Finally, two specific case characteristics are significant predictors of recidivism: type of offense and the seriousness of the offense. Offenders who are convicted of a violent, property, or miscellaneous type offense, as well as those convicted of a lower-gravity (OGS 7-10) level 5 offense are more likely to recidivate.

Item Scoring

Table 4.5 presents the scoring procedure for the risk assessment instrument developed for this study. The classic Burgess method (1928) of adding 1 point for each risk factor present remains the most popular methods for constructing risk assessment scales (Kim & Duwe, 2017). This method is conceptually simple and has been shown to be just as predictive as more complicated methods (Andrews & Bonta, 2010b; Gottfredson, 1986), however it can be imprecise because it ignores effect strength. For example, the presence of a risk factor may increase the likelihood of recidivism twice as much as a different risk factors, yet a classic Burgess method would treat both the same. To address this issue, I use a modified Burgess procedure to create an additive scale which encompasses the significant risk predictors from
earlier analysis (Table 4.4, Model 2). For each factor present, the offender receives one or two points, depending on the strength of the effect size for that predictor. All risk factors with a significant odds ratio of 1.2-2.2 would receive 1 point, and all risk factors with a significant odds ratio of 2.2-3.2 would receive 2 points. As discussed in the literature review, there is no well established system of scoring risk assessment items, however similar Burgess-like techniques have been used to construct risk assessments instruments in Pennsylvania (Ruback et al., 2016), Virginia (Kleiman, et al., 2007), and Missouri (Wolff and Oldfield, 2010).

The instrument is an 11-point scale in which the total risk scores may range from 0 to 10. For example, a 41-year-old female offender with 2 prior arrests, a PRS of zero, no juvenile contact with the system, no domestic violence arrests, and who was convicted of distributing 1,000 grams of cocaine would score a zero to reflect her low risk of recidivism. Alternatively, a 20-year-old male offender with 5 prior arrests, contact with the juvenile system, a domestic violence arrest, a prior record score of 2 and who was convicted of a home invasion would score the maximum 10 points.

---

25 The only variable that falls above 3.2 odds is Age 14-21 (OD: 3.29), however I decided to score it a 2 in order to avoid over-complicating the scoring procedure.

26 Note that offenders can have one type of conviction offense, thus they would receive one point if they had a violent, property, or miscellaneous offense.
Table 4.5: Risk Instrument Scale (0-10)

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1</td>
</tr>
<tr>
<td>Age 14-21</td>
<td>2</td>
</tr>
<tr>
<td>Age 22-40</td>
<td>1</td>
</tr>
<tr>
<td>PRS 1+</td>
<td>1</td>
</tr>
<tr>
<td>Arrest 3-4</td>
<td>1</td>
</tr>
<tr>
<td>Arrest 5+</td>
<td>2</td>
</tr>
<tr>
<td>Juvenile Arrest or Conviction</td>
<td>1</td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>1</td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>1</td>
</tr>
<tr>
<td>Violent Offense</td>
<td>1</td>
</tr>
<tr>
<td>Property Offense</td>
<td>1</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>1</td>
</tr>
</tbody>
</table>

Risk Score Analysis

Table 4.6 displays the percent of offenders receiving each risk score and the failure rate distribution by risk scores for the development sample. Figure 4.2 shows a visual presentation of the percent of offenders recidivating at each risk score. First, the sample distribution is slightly left skewed, meaning that more offenders have risk scores in the higher end of the risk scale (6-10, n=1,776) than the lower end of the scale (0-4; n=2,563). Second, risk scores have a clear positive relationship with failure, with a greater proportion of offenders recidivating at each subsequent risk score. The only two exceptions are the tail ends of the scale. There are only 10 offenders with a risk score of 0 and 17 offenders with a risk score of 10, thus these groups are too small to serve as stand-alone risk categories. Table 4.7 shows that there is a significant difference (t: p <0.001) in the average risk score between offenders who recidivated (6.1) and...

---

27 One way to deal with this in practice would be to present the findings after collapsing risk scores 0 and 1 and risk scores 9 and 10, which would retain the upwards trajectory of the scale. Another way would be to reorganize the entire risk score scale into levels - which I provide an example of in the next section.
offenders who did not recidivate (4.6).

Table 4. 6: Risk Score Distribution (Development Sample) 
N=5,260

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>N</th>
<th>%Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>10%</td>
</tr>
<tr>
<td>1</td>
<td>107</td>
<td>6%</td>
</tr>
<tr>
<td>2</td>
<td>375</td>
<td>15%</td>
</tr>
<tr>
<td>3</td>
<td>555</td>
<td>25%</td>
</tr>
<tr>
<td>4</td>
<td>729</td>
<td>29%</td>
</tr>
<tr>
<td>5</td>
<td>921</td>
<td>45%</td>
</tr>
<tr>
<td>6</td>
<td>962</td>
<td>55%</td>
</tr>
<tr>
<td>7</td>
<td>901</td>
<td>66%</td>
</tr>
<tr>
<td>8</td>
<td>482</td>
<td>70%</td>
</tr>
<tr>
<td>9</td>
<td>201</td>
<td>75%</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>76%</td>
</tr>
</tbody>
</table>

Score Lower 95% CI Upper 95% CI
Succeed 4.6 4.5 4.7
Recidivate 6.1 6 6.2

T-test significant at the p <0.001 level
Risk Levels

As the risk scale becomes more dispersed there are fewer and fewer substantive differences between the risk scores. In this instrument’s scale, several score groups have similar rates of recidivism. For example, it may not be meaningful for a judge to know that an offender who scored a 4 is associated with a group that has a 29% rate of recidivism, while an offender who scored a 3 is associated with a group with a 25% rate of recidivism. This is particularly true at the tail ends of the scale in which the small number of offenders makes it difficult to make statistical generalizations to other offenders with similar scores. One way to address this issue would be to restructure the scores into broader risk levels. There are few accepted rules on how to group risk scores, but ideally, jurisdictions would choose groupings that are substantively
meaningful (e.g., labelling everyone who has a 50% or greater chance of recidivism as “high-risk”). Another way would be to determine risk levels based on the distribution of offenders in across the levels (e.g., labeling the bottom 30% of all offenders as low risk and eligible for diversion). Table 4.9 shows an example of the latter grouping (see Duwe, 2014) for another example) and Figure 4.3 shows the positive relationship between recidivism and risk level. The risk levels correspond to the distribution of offenders in the scale. The 20% of the lowest risk offenders (scores 0-3) have an overall recidivism rate of 19% and the top 15% of the highest risk offenders (scores 8-10) have an overall recidivism rate of 72%.

Table 4. 8: Risk Levels (Development Sample)
N=5,260

<table>
<thead>
<tr>
<th>Risk Level</th>
<th>N</th>
<th>%Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0-3)</td>
<td>1,047</td>
<td>19%</td>
</tr>
<tr>
<td>Med (4-5)</td>
<td>1,650</td>
<td>38%</td>
</tr>
<tr>
<td>High (6-7)</td>
<td>1,863</td>
<td>60%</td>
</tr>
<tr>
<td>Very High (8-10)</td>
<td>700</td>
<td>72%</td>
</tr>
</tbody>
</table>

Failure base rate = 46%
Chi2(3) = 683.64  p < 0.001
Validation of Multivariate Analysis

To validate the findings, and to determine the discriminatory power of the instrument, I re-run the final multivariate logistic regression model on the validation sample (N=2,675). Table 4.9, Model 1 shows similar significance and strengths of relationships between most of the predictor variables and the outcome variable when compared to Table 4.4, Model 2 (Final Model in the development sample). Two noticeable differences are found. First, property offenders in the validation sample have almost 3 times higher odds of recidivating compared to drug offenders (or= 2.83; p<0.01), versus 75% higher odds in the development sample (or=1.75; p<0.01). Second, offenders convicted of a miscellaneous offense were no more likely to recidivate than were drug offenders in the validation sample (or= 1.16; NS). Table 4.9, Model 2 shows the predictive model with no control variables. Similar to the results in the development sample, removing the county and race/ethnicity variables resulted in a slight, but significant, loss of model fit (lr: p<.001). However, removing the control variables did not significantly change...
the strength or direction of other predictor variables. In sum, the risk prediction model developed for the development sample is also a good fit for predicting recidivism in a different sample of offenders.

**Table 4.9: Multivariate Logistic Model of Recidivism (Validation Sample)**

<table>
<thead>
<tr>
<th>N=2,675</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>(1) Final Model</th>
<th>(2) Final w/No Race and County</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Offender Demographics</strong></td>
<td><strong>Odds</strong></td>
</tr>
<tr>
<td>Male</td>
<td>1.72</td>
</tr>
<tr>
<td>Black</td>
<td>1.31</td>
</tr>
<tr>
<td>Hispanic and Other Race</td>
<td>1.02</td>
</tr>
<tr>
<td>Age 14-21</td>
<td>2.78</td>
</tr>
<tr>
<td>Age 22-40</td>
<td>1.48</td>
</tr>
<tr>
<td>Allegheny County</td>
<td>1.36</td>
</tr>
<tr>
<td>Philadelphia County</td>
<td>1.89</td>
</tr>
<tr>
<td>Rural County</td>
<td>1.07</td>
</tr>
<tr>
<td><strong>Offender Criminal History</strong></td>
<td></td>
</tr>
<tr>
<td>Prior Recode Score=1+</td>
<td>1.32</td>
</tr>
<tr>
<td>Arrest 3-4</td>
<td>1.53</td>
</tr>
<tr>
<td>Arrest 5+</td>
<td>2.67</td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>1.5</td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>1.47</td>
</tr>
<tr>
<td><strong>Case Characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>1.63</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>1.16</td>
</tr>
<tr>
<td>Violent Offense</td>
<td>1.49</td>
</tr>
<tr>
<td>Property Offense</td>
<td>2.83</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>1.16</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.06</td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>-1618.81</td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.12</td>
</tr>
<tr>
<td>AUC-ROC</td>
<td>0.73</td>
</tr>
</tbody>
</table>

*p < 0.05 ; ** p < 0.01 ; *** p < 0.001.

Reference groups: female, White, Age 41+, Other Urban County, Arrest 1-2, Drug Offense, OGS 11-14.
Validation of Risk Score Model

Table 4.10 and Figure 4.4 show how well the risk score algorithm created using the development sample discriminates between offenders in the validation sample (N=2,675). There is a clear upward trend in the failure rate by risk score. However, as with the development sample, few offenders had risk scores at the tail ends of the scale (i.e., risk scores 0 and 10). These categories would either need to be collapsed into the rest of the scale in order to provide meaningful recidivism estimates, or collapsed into a larger risk level. The distribution of offenders is similar as well — with more offenders situated at the high end of the scale (6-10; n=1,297) than the low end (0-4; n=907) (mean risk score: 5.3). Table 4.11 shows a significant difference (t: p < 0.001) in mean risk score for offenders who recidivate and those who succeed.

Table 4. 10: Risk Score Distribution (Validation Sample)
N=2,675

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>N</th>
<th>%Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>0%</td>
</tr>
<tr>
<td>1</td>
<td>52</td>
<td>8%</td>
</tr>
<tr>
<td>2</td>
<td>201</td>
<td>17%</td>
</tr>
<tr>
<td>3</td>
<td>280</td>
<td>21%</td>
</tr>
<tr>
<td>4</td>
<td>368</td>
<td>34%</td>
</tr>
<tr>
<td>5</td>
<td>471</td>
<td>41%</td>
</tr>
<tr>
<td>6</td>
<td>493</td>
<td>53%</td>
</tr>
<tr>
<td>7</td>
<td>468</td>
<td>66%</td>
</tr>
<tr>
<td>8</td>
<td>231</td>
<td>73%</td>
</tr>
<tr>
<td>9</td>
<td>94</td>
<td>83%</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>63%</td>
</tr>
</tbody>
</table>
Table 4.11: Average Risk Score by Failure (Validation Sample)  
N=2,675

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Succeed</td>
<td>4.6</td>
<td>4.5</td>
<td>4.7</td>
</tr>
<tr>
<td>Recidivate</td>
<td>6.1</td>
<td>6.0</td>
<td>6.2</td>
</tr>
</tbody>
</table>

T-test significant at the p <0.001 level

Figure 4.4: Recidivism by Risk Score (Validation Sample)  
N=2,675

The red line shows the mean risk score (5.3) and the square outline shows where the majority of offenders are situated on the scale.

Validation of Risk Level Model

Table 4.12 shows that the risk level scheme developed using the development sample
may be applied to the validation sample with no resulting differences in risk score distribution. For example, offenders with risk scores of 0-3 still represent the bottom 20% of all offenders, and offenders who scored 8-10 still represent the top 15% of the sample. Furthermore, compared to the development sample, the failure rate within each risk level is within 3% percentage points. Figure 4.5 shows a visual distribution of the risk scores and associated levels of recidivism for the validation sample. The validation analyses suggest that the risk level scheme developed when using the development sample can provide meaningful discrimination between recidivists and non-recidivists outside of the sample.

### Table 4. 12: Risk Levels (Validation Sample)

<table>
<thead>
<tr>
<th>Risk Level</th>
<th>N</th>
<th>% Fail</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0-3)</td>
<td>539</td>
<td>18%</td>
<td>Bottom 20%</td>
</tr>
<tr>
<td>Med (4-5)</td>
<td>839</td>
<td>38%</td>
<td>20%-50%</td>
</tr>
<tr>
<td>High (6-7)</td>
<td>961</td>
<td>59%</td>
<td>50% - 85%</td>
</tr>
<tr>
<td>Very High (8-10)</td>
<td>336</td>
<td>75%</td>
<td>Top 15%</td>
</tr>
</tbody>
</table>

Failure base rate = 46%

Chi2(3) = 683.64  p < 0.001
Survival Analysis by Risk Level

Figure 4.6 shows the survival curves associated with each risk level within the study’s timeframe. The survival curve provides a statistical picture of the above analysis by showing the fraction of offenders surviving per risk score level over time. Offenders in the Very High risk group (scores 8-10) display a steep drop in survival with 50% recidivating by year 1 and only 25% remaining arrest-free at year 3. In contrast, the Low risk group (scores 0-3) display a much more moderate decrease in survival with approximately 6% recidivating at year 1 and more than three-quarters still remaining arrest-free at year 3. The Medium risk group (scores 4-5) and the High risk group (scores 6-7) has survival curves that fit in the middle of the Low and Very High risk groups, without overlap in confidence intervals.
Comparison of AUCs

The final part of instrument validation process includes identifying the Area Under the Curve (AUC) for both the development and the validation sample. Because the risk scoring algorithm is not an exact replica of the final multivariate risk prediction model, the AUC value is calculated after regressing failure on the risk score. The AUC value measures the instruments score’s ability to discriminate between recidivists and non-recidivists. For context, Yang and colleagues’ (2010) meta-analysis on the efficacy of popular violence prediction tools (e.g., LSI-R, Static-99) showed that validation AUC values ranged from 0.56 to 0.71, with the majority of tools falling between 0.65-0.69 (p.753). Table 4.13 shows an AUC value of 0.72 for both the
development and the validation sample. A visual inspection of the curves in Figures 4.6 and 4.7 show an almost identical curve which lays between the upper left hand corner (perfect discrimination) and the diagonal line (no discrimination). Essentially, in both samples, there is a 72% probability that a randomly selected recidivist will have a higher risk score than a randomly selected non-recidivist. In the behavioral and social sciences, this value can be interpreted as having large predictive effect (Rice & Harris, 2005). *Thus, in answering Research Question 2, I find that an actuarial risk assessment instrument using static predictors and agency-level data can predict risk of re-offense moderately well.*

### Table 4.13: AUC-ROC Between Risk Score and Re-arrest

<table>
<thead>
<tr>
<th>Sample</th>
<th>AUC</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>0.716</td>
<td>0.703</td>
<td>0.73</td>
</tr>
<tr>
<td>Validation</td>
<td>0.721</td>
<td>0.702</td>
<td>0.74</td>
</tr>
</tbody>
</table>

### Figure 4.7: AUC-ROC Curve (Development Sample)

*Area under ROC curve = 0.7168*
Discussion

The purpose of this study was to create and validate an actuarial risk assessment instrument which judges can use at the time of sentencing and to provide an overview of the methods used in the development of the instrument. Using a sample of serious adult offenders who were sentenced in Pennsylvania between 2001-2005, analyses identified 10 distinct risk factors that significantly predicted recidivism at the sentencing stage – 8 of which were included in the risk instrument: gender, age, prior record score, number of arrests, juvenile arrest or conviction, domestic violence arrest, offense gravity score, and being convicted of a violent, property, or miscellaneous conviction. Two other risk factors — race and county — were also

---

28 Almost all of these convictions were for the unauthorized possession of a firearm.
predictive of recidivism, although neither were included in the final scoring model. Using a sample of 7,935 offenders, and using a series of binary and logistic multivariate analyses on static, agency-level data, it was possible to create a risk assessment instrument with strong predictive validity ($AUC: 0.72$). Collapsing the individual risk scores into larger risk levels provided a way to identify four distinct categories of risk (Low, Medium, High, and Very High) which loosely corresponded to the lowest 20%, 20-50%, 50-85%, and top 15% highest risk offenders in both the development and validation samples. Several implications stem from these findings.

First, while results suggest that the new instrument is valid and robust for level 5 offenders in Pennsylvania, it serves as a methodological example, rather than a suggested model, for assessing risk at sentencing in other states, for lower-level offenders, or for offenders at other stages in the system. Many of the risk factors in the instrument are specific to Pennsylvania, including offense gravity score and prior record score. While it is likely that the same risk factors would be significantly associated with recidivism within other offender samples, variable coding and scoring would need to be adjusted for non-Pennsylvania or non-sentencing stage samples of offenders. Furthermore, states differ — sometimes dramatically — in sentencing patterns, average recidivism rates, and parole practices, which makes the external validity of this instrument outside of Pennsylvania unknown. Instrument validation should be performed on a representative, jurisdiction specific sample of offenders, prior to making any claims of generalizability. Thus, an implication of these findings is that — when possible — risk assessment instruments should be created and validated using the population and specific sentencing system of each state. Barring this option, well-researched “off-the-shelf” instruments, such as the LSI-R, should be validated on a jurisdiction specific sample prior to adoption.
Second, as suggested by previous researchers (e.g., Austin, 2006; Gendreau et al., 1996), the use of a relatively small number of static risk factors available from official offender records provides sufficient information for the instrument to discriminate between recidivist and non-recidivist. Therefore, if states are looking for an instrument to assess risk of recidivism — rather than to identify offender needs — a short, non-interview based instrument may be sufficient. This stands in contrast from modern risk assessment processes which evolved to include longer and more dynamic risk assessment standards (e.g., 3rd and 4th generation instruments, Andrews and Bonta, 2010b). Ease of implementation and the use of system resources are major consideration within any proposed policy change. Many proprietary instruments require one-on-one interviews, interviewer certification and continued training, and use fees for assessment. In contrast, the use of this instrument would require integrating 8 points of data from official agency records, which can be done easily and with little investment of agency resources.

Third, the findings add to the limited research on recidivism patterns of serious offenders. While serious offenders make up only a small portion of all convicted offenders, they take up a disproportionate share of correctional resources as a result of receiving longer terms of incarceration and requiring higher security levels. These findings show that even within a sample of level 5 offenders, rates of recidivism differ substantially: a full two-thirds of the 15% highest risk offenders (i.e., risk score 8-10) were arrested or returned to prison for a parole violation within three years. However less than 20% of the bottom fifth (i.e., risk scores 0-3) riskiest offenders recidivated during the same period. In line with some research (Langan and Levin, 2002), but in contrast to other research (Spohn and Holleran, 2002), this study found that serious drug offenders had significantly lower rates of recidivism than to property, violent, and miscellaneous offenders. This finding has clear policy implications in that drug offenders, who
are generally considered less deserving of incarceration than other types of offenders, are good candidates for less restrictive punishment and shorter sentences despite the high gravity of their crimes. Sex offenders also had significantly lower rates of recidivism than property, violent, or miscellaneous offenders. This finding is well supported by previous research (Hanson et al., 1995; Holleran & Spohn, 2002; Langan & Levin, 2002; Sample & Bray, 2003). Sex offenders generally have low rates of recidivism, yet they have high rates of incarceration and high levels of supervision. While it is unlikely that de-incarcerating sex offenders will become a policy focus, sanctions that prioritize incapacitation and practices that use valuable resources on heavy monitoring of sex offenders could be reconsidered.

Even within a group of serious offenders, the gravity, or seriousness, of the offense is still important. Offenders who were convicted for offenses at OGSs 7-10, were more likely to recidivate than those convicted for offenses under OGSs 11-14. Offenders convicted of offenses under OGSs 7-10 committed serious, but more common, offenses, such as burglary, drug distribution, and indecent assault. However, offenders sentenced under OGSs 11-14 committed the most serious, and most rare, offenses such as rape, robbery and distribution of large quantities of drugs. This describes only 26% of the total sample and these offenders do not fit a typical repeat offender profile. For example, 65% of offenders sentenced under OGSs 11-14 had a prior record score of 0, versus almost half of those sentenced under OGSs 7-10. From a policy perspective, it may be beneficial for sentencing judges to leverage system resources on offenders that have a high risk of re-offending — whether that be in the form of longer or more incarcerative sentences. Similarly, some lower-risk offenders may be good candidates for less

---

29 This is sexual assault that does not include rape. This is often charged for inappropriate sexual touching.
severe sentences, despite the seriousness of their crime.

Fourth, the use of demographic factors (e.g., race, gender, and age) to predict recidivism in the courtroom, has received its share of criticism — yet block testing analysis shows that the addition of these factors substantially improves the model fit. The use of demographic factors (minus race) to predict recidivism is common in other parts of the criminal justice system. However, some legal scholars claim that the practice is unethical — and possibly unconstitutional — at the sentencing stage (Starr, 2014). Because the results of a risk assessment instrument are assumed to inform sentencing decision, critics claim that the use gender, race, age, or place of residence to predict risk of recidivism is a way of punishing people for social statuses over which they have no control (Hannah-Moffat, 2013; Holder, 2014).

While race is not included in the final scoring model (despite its significance in the bivariate and multivariate analysis), both age and gender are. Of course, it is possible to also exclude gender and age from the scoring model, but prior research shows that the removal of parsimonious significant risk factors from the prediction algorithm results in a reduction of instrument accuracy. For example, Skeem and colleagues (2016) reported that removing gender from an actuarial risk assessment instrument resulted in the over-estimation of women’s recidivism rates. Thus, practitioners and policy makers should weigh the costs of using ethically and constitutionally suspect variables against the benefit of increased instrument accuracy.

Fifth, while sentencing county was excluded from the scoring algorithm, it was a significant predictor variable in the model building procedure. In particular, offenders from Allegheny (Pittsburgh) county had about 20% higher odds of recidivating than offenders from
other urban counties. This was true even when race and other significant factors were controlled for. Thus, apart from individual level risk factors, structural risk factors such as location matter for individual risk. If we presume that most offenders sentenced in Allegheny county resume living there upon release, this finding shows that there is something unique about Allegheny county that is correlated with a higher likelihood of recidivism. In other words, released offenders have better luck avoiding re-arrest and revocation outside of Allegheny county. This could be due to a number of factors, such as difference in policies regarding the revocation of parole/probation, differences in the levels of police presence and levels of discretion, and differences in what constitutes an arrestable offense. It’s also possible that other urban counties provide better re-entry services for offenders, such as job training and substance abuse programs, and thus reduce the likelihood of arrest for offenders in those counties. The importance of variation in state and county-level crime control policies and practices is well established in criminological research (e.g., Garland, 2001; Johnson, 2006; Weidner & Frase, 2003). Similarly, scholars have outlined the ways in which penal response affects arrest rates (Black, 1970) and levels of incarceration (Stemen & Rengifo, 2011). Despite this, there is almost no research on how localized differences in policy and practice may affect the validity of criminal justice risk assessment instruments. Not only may “county” serve as a risk factor, there may also be difference in which individual-level risk factors matter between counties (see Vigorita, 2003). As state-wide instruments become more common, this is an area that will warrant more attention.

30 There were no significant differences between Allegheny county and rural counties or Philadelphia county, not between any of the other combinations of counties.
Limitations

The present study has a number of limitations and identified areas for future research. First, while the study period was longer than average recidivism studies (i.e., over 11 years total), the offender sample used in the analyses only included offenders who were released from prison in time to experience a minimum three-year follow-up period. Of the 14,026 level 5 offenders convicted between 2001 and 2005, 7,935 were included in the study. As such, offenders who received longer sentences, offenders who had time added to their sentence because of institutional misbehavior, offenders who were transferred to mental health institutions or who died before release were not included in the sample. On one hand, assessing risk of recidivism of offenders who are sentenced to long prison terms is less important because their risk level may change substantially while in prison due to age or institutional effect. However, if risk assessment instruments will be used to inform sentencing decisions, the development sample should be as representative of the population being sentenced. Thus, future research on serious offenders should focus on extending the study period to include offenders who are released following lengthy prison terms.

Second, the outcome measure for this study was arrest for a felony or misdemeanor crime, which is the most generic measure of re-offending. In future research, additional risk assessment instruments can be created to capture risk predictors for re-convictions, violent arrest, or re-incarceration in order to measure outcomes that are of greater consequence to the public. This approach would also minimize some disparity in the differential selection of Black and offenders into the system. If Black offenders appear to recidivate at higher rates because they are more likely to be arrested on charges that do not progress past the first hearing, requiring a
higher standard of proof that an offense occurred (e.g., conviction) would minimize differences in recidivism.

Third, the potential impact the use of this type of instrument will have on sentencing decisions is not a product of the instrument or its construction. In the context of structured decision making, the actual instrument is just one part of the process. How sentencing practices change when assessments are adopted has more to do with the accompanying risk assessment policy and practitioner response. For example, without knowing how the informal consideration of risk affects sentencing outcomes, it may be premature to assume that, due to the instrument, high risk offenders will receive longer sentences than lower risk offenders. Or, that high-risk offenders will receive longer sentences than they normally would have if the instrument were not in use.

Protection of the community is one of several considerations during sentencing, thus we do not know how other factors (e.g., blameworthiness or practical constraints) interact with risk to affect sentencing. Thus, future research should assess how sentencing practices change after the instrument is adopted by a jurisdiction. This research would also inform how risk of recidivism changes as a result of the using instrument. If, due to the use of the instrument, certain offenders received probation instead of prison, and vice versa, their risk of recidivism would potentially change. Thus, new information would continually need to be incorporated into the instrument in order to retain its validity.

Conclusion

Creating an actuarial risk assessment instrument is a task that involves a multitude of consequential decisions. Some decisions, such as which risk scoring method or predictive
validity statistic to use, can be made by reviewing prior risk assessment literature and looking at
the benefits and downsides of each option. Other decisions, such as whether to include
demographic factors or to pick certain cut-off scores for sentencing decisions, are best made in
reflection of the jurisdiction-specific purpose of the instrument. States interested in reducing
their prison populations will continue to have interest in using actuarial risk assessment tools at
the sentencing stage. These tools can assist judges in making difficult decisions about offender
management, and with the advent of electronic criminal justice records, may be compiled using
few additional agency resources. The current study provides a clear example of the development
and validation process for creating a jurisdiction-specific instrument for the sentencing stage,
which can inform future efforts.
Chapter 5 – The Relationship Between Race, Risk and Recidivism

The potential benefits of using actuarial risk assessment during the sanctioning process are clear. Risk assessment instruments offer important information to judges who seek to make more informed sentencing decisions, and it offer policy-makers a way to structure sentencing decisions using a data-driven approach. Intuitively, it is difficult to argue against providing judges with more information about the offenders they are sentencing. And while some retributive purists contend that sentencing should be based on the offender’s blameworthiness alone, many policymakers, judges, and advocates believe that sentences should reflect likelihood of recidivism (Bergstrom & Mistick, 2010; Casey et al., 2011; Chanenson, 2003, 2004). Decades of research have identified race as a significant factor in explaining criminal justice outcomes (Sampson & Lauritsen, 1997; Walker, Spohn, & DeLone, 2012). Yet, because of ethical and legal considerations, no current criminal justice risk assessments use race as a risk factor to predict recidivism. How does this universal omission of a significant risk factor affect the utility and fairness of recidivism risk assessment instruments? This study uses a previously constructed sentencing risk assessment instrument to explore the relationship between risk score, race, and recidivism.

Despite scholarly criticism regarding the use of risk assessment instruments in the courtroom (Hannah-Moffat, 2013; Harcourt, 2012; Starr, 2014), more states are exploring the option (e.g., Maryland). Perhaps the criticisms that have received the most attention are issues of disparate impact and instrument bias (Flores et al., 2016; Hannah-Moffat, 2013; Harcourt, 2012; Holder, 2014; Skeem and Lowenkamp, 2016; Starr, 2014). While these criticisms are closely associated, they are actually two distinct issues. The first criticism, that of potential for disparate impact, has received more attention in research. Disparate impact can occur when mean
instrument score differences between Black and White offenders result in different sentencing outcomes. For example, if White offenders receive diversion, and similarly situated Black offenders receive incarcerative sentences due to a higher risk score. While research shows that Black offenders do indeed score higher on risk assessment instruments (Flores et al, 2016; Skeem and Lowenkamp, 2016), only a couple of studies have attempted to pinpoint which individual risk factors accounted for the difference. For example, Skeem and Lowenkamp (2016) wrote that differences in levels of criminal history account for most of the differences in risk scores between Black and White offenders in the federal PCRA. However, the PCRA was not designed to be considered (nor is considered) during sentencing decisions, thus the generalizability of findings remain limited.

The second criticism of instrument bias has less support — largely due to a lack of research on the topic. Yet, this criticism has recently garnered a considerable amount of attention, partly in response to a report by the investigative news outlet ProPublica (Larson et al., 2016), which claimed that the COMPAS instrument over-classifies Black non-recidivists as recidivists. Some of Larson and colleagues’ methods were criticized (see rebuttals: Flores et al. 2016; Dieterich, et al., 2016), yet the high-profile piece spurred debate on the topic of instrument bias in risk assessments. Broadly, instrument bias occurs when an instrument miss-classifies Black offenders as higher risk than they actually are (or vice-versa) – in other words, when the instrument is not racially-neutral. Two studies tested the COMPAS risk assessment instrument and the federal PCRA tools for racial bias, and the respective authors concluded that the tools were "bias-free" (Flores et al, 2016; Skeem and Lowenkamp, 2016).\(^{31}\) Despite the fact that

\(^{31}\) This author is concerned with the strong conclusions reached by both the papers (i.e., that the instruments are "bias-free"). The analysis with the unmatched sample of offenders in the Skeem and Lowenkamp (2016) paper showed that the PCRA instrument over-classified whites for both any arrest and violent arrest. Similarly, the Flores et al. (2016) paper showed that
"inherent bias" criticisms are generally lobbied against all actuarial risk assessment instruments in criminal justice (e.g., Angwin et al., 2016; Starr, 2014), results from the assessment of two instruments offers no guarantees about other instruments used in the system. As documented in the previous chapter, the instrument represents a culmination of decisions on the part of the researcher tasked with constructing the instrument. Decisions such as identifying which risk factors to include in the analysis or which outcome variable is the best measure of recidivism can have major implications for how well the tool performs across racial groups.

In Chapter 4, I constructed and validated a risk assessment instrument which judges can use during sentencing. In this chapter, I use the newly-developed risk instrument to explore the relationship between race, risk score, and recidivism. First, I determine mean group differences in risk score between Black and White offenders and assess the likelihood of being in a high-risk group. Second, I identify which risk factors most influence the mean score differences between Black and White offenders. Third, I determine whether the instrument is free from predictive bias by assessing whether the instrument predicts equally well for Black and White offenders, and by determining whether the inclusion of race in the risk assessment model improves the predictive utility of the instrument.

**Literature Review**

**Disparate Impact and Criminal History**

The over-representation of Black offenders in the criminal justice system is well documented (Alexander, 2012; Carson, 2015; Pettit & Western, 2004). Research also shows that, while the probability slopes of Black and White offenders both increase with the decile score, the predicted probabilities showed consistently larger recidivism predictions for Black offenders. Also, the base rate differences in risk level range from 2% and up to 9% for Black and White offenders.
on average, Black offenders consistently score higher on actuarial risk assessment instruments because they have more risk factors related to recidivism (Dieterich et al., 2016; Eisenberg, Bryl, & Fabelo, 2009; Gendreau et al., 1996; Hannah-Moffat & Maurutto, 2003; Harcourt, 2008; Kleiman et al., 2007; Skeem & Lowenkamp, 2016). Life circumstances of minority populations, particularly individuals who are involved with the justice system, differ dramatically from those of White offenders. There are many examples illustrating these differences: Black offenders are more likely to have been previously involved with the juvenile justice system (Ulmer & Laskorunsky, 2016; Redding, 2002), more likely to have extensive criminal histories (Frase, 2009; Harcourt, 2015; Tonry, 2014), and they are typically more likely to return to neighborhoods of high minority composition after release (Huebner & Bynum, 2008). Black offenders applying for parole are also more likely to be unmarried, to have lower educational achievement, to have weaker employment records and family bonds, and to have less residential stability than White offenders (Hoffman & Beck, 1985). Additionally, Black offenders not only commit more violent crime, but are also more likely than Whites to be arrested for the same type of crimes (National Research Council, 2014; Tonry, 2014). Thus, mean differences in risk assessment scores between Black and White offenders reflect longstanding patterns of inequality that adversely affect minority populations.

The use of criminal history measures is almost ubiquitous in recidivism risk assessments (Hamilton, 2015). Research shows that prior convictions, prior incarceration, juvenile arrest, and the number of prior arrests are all significant predictors of recidivism (Gendreau et al., 1996; Monahan & Skeem, 2016). Harcourt (2015) argues that because it is no longer appropriate to use race as a predictor variable, actuarial tools have reduced the number of risk factors included in the scoring procedure and have increasingly focused on criminal history as a predictor
(Harcourt, 2015). In fact, his argument is that race and criminal history are so highly correlated, risk has become a proxy for race. Skeem and Lowenkamp (2016) reports that 66% of the difference in Post-Conviction Risk Assessment scores between Black and White offenders is explained by differences in criminal history. Particularly, Black offender have a substantially higher number of arrests (which is weighted heavily in the scoring model).

How, or if, mean score differences in risk assessment instruments translate into sentencing disparities has been debated, and the answer may not be universal. Criminal history is embedded in all state sentencing guidelines and is considered to be an important sentencing factor in all non-guideline states (Frase, Roberts, Hester, & Mitchell, 2015). For example, research shows that the seriousness of the offense and the offender's criminal history already explain most of the variation in sentencing outcomes, particularly between White and Black offenders (Frase et al., 2015; Mitchell, 2005; Spohn, 2000; Ulmer et al., 2016). The correlation between criminal history and minority race status means that Black offenders are already disproportionately affected by the consideration of criminal history in sentencing decisions (Frase et al., 2015). While recent research has begun to untangle the ways in which states use criminal history enhancements to increase sentences for offenders (e.g., Frase et al., 2015; Roberts & Ylincak, 2014) there is no comparative research on the diversity of criminal history measures used in actuarial instruments. In fact, “criminal history” has no standard definition and could include a variety of measures — the prevalence of which differs across populations. For example, the use of prior arrest to predict recidivism in an actuarial instrument does not completely overlap with the use of prior convictions for criminal history enhancement. In fact, it is possible that people with zero convictions have a number of arrests.
Predictive Bias

Two recent articles (Skeem and Lowenkamp, 2016; Flores et al., 2016) address the issue of racial bias in recidivism risk assessment which refer to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) for testing criteria and definitions. According to the APA’s testing standards, for an instrument to be racially-neutral it should predict with similar accuracy across racial groups and should show that a given score of X will indicate a similar recidivism rate of Y across racial groups. Similar standards have been applied to testing the predictive validity of recidivism risk instruments across gender (Desmarais, Johnson, & Singh, 2016; Skeem et al., 2016). For example, the LSI-R was found to have similar levels of predictive validity for males and females across multiple samples (Desmarais et al., 2016), although the form of the association was not tested. Skeem et al. (2016) reported that the PCRA strongly predicted recidivism for both genders, but also overestimated women's likelihood of recidivism.

Skeem and Lowenkamp (2016) tested the predictive utility and the predictive fairness of the PCRA across races. Using a large sample of federal probationers, the authors found that the test strongly predicted both general and violent arrests for Black and White offenders (AUC statistics 0.71-0.75). The authors also tested the form of the relationship between PCRA scores and arrest as a function of race, with slightly less robust results. While the authors conclude that the risk score has "essentially the same meaning - that is, the same probability of recidivism" for each given score (pg. 21), a closer look at the results reveals that this is only true for general arrest and only true for the matched sample of offenders. For example, results show that the score provided "modest overestimation of violent recidivism for White offenders" in the matched sample (pg. 13) and that "the intercept of the relationship between PCRA scores and both violent
and any arrest was significantly lower for unmatched White than for Black offenders...which suggests overestimation of arrest for White offenders" (pg. 16). The process of matching Black and White offenders on age, gender, and conviction offense removes the natural variation in the proportion of risk factors found between Black and White offender groups in the criminal justice system.

Flores and colleagues (2016) conducted similar analyses on the COMPAS, which is used in some states at the pre-sentencing and sentencing stages. However, their analysis was completed using a sample of pre-trial defendants. The authors reported similar AUC values across race (0.65 -0.70) for both violent and general arrest, indicating that the assessment had moderate to strong predictive utility. However, while the probability slopes of Black and White offenders both increased with the decile score, risk levels showed consistently larger associated recidivism rates for Black offenders - ranging from up to 6% higher for general arrest and up to 9% higher for violent arrest. Nonetheless, Flores and colleagues conclude that race does not add to the predictive utility of the mode and that "a given COMPAS score translates into roughly the same likelihood of recidivism, regardless of race" (pg. 44). In reality, it appears that their instrument marginally, but consistently over-estimated arrest for White offenders, similar to the results reported by Skeem and Lowenkamp (2016).

Research Questions

The purpose of the present study is to analyze the relationship between risk score, race, and

32 Given the difference in the predicted probabilities by race, it’s surprising that the inclusion of race into the prediction model yielded no significant improvement in model fit. The Flores et al. (2016) paper makes no mention of the level of significance used, nor does it provide confidences intervals for its measures, making evaluation difficult.
recidivism. Three overarching research questions guide the study:

1. **What is the relationship between instrument risk score and race?**

2. **Which risk factors contribute most to mean score differences between Black and White offenders?**

3. **How well does the instrument predict recidivism for Black and White offenders?**

**Methods**

**Sample**

Disparate impact and instrument bias analysis is conducted using the risk assessment instrument constructed in Chapter 4. Comparison of AUC scores shows that the instrument has strong predictive validity for both the validation sample (0.72) and the development sample (0.72). For the following analyses, I use a large sample of serious offenders convicted in Pennsylvania between 2001 and 2005 (see Chapter 3 (Data) for more information about the construction of the sample). The starkest differences in the criminal justice system are found between White and Black offenders, thus these analyses exclude Hispanic offenders and those of Native American, Pacific Islander, or Asian descent. Exclusion of these groups from the full sample yields a total sample of 7,022 level 5 offenders with 3,045 Black offenders and 3,607 White offenders.

**Predictor Variables and Outcomes Measure**

The outcome measure - arrests for a felony or misdemeanor crime, or a revocation of parole within 3 years of release - remains unchanged from the analysis described in Chapter 4. The instrument risk score and race are used as main independent variables to determine
associations with recidivism. Chapter 4, pg. 91 detailed the scoring procedure for the instrument and Chapter 4, pg. 77 provided the coding for individual risk items.

**Analytic Approach**

Analysis begins with a selection of descriptive statistics for Black and White offenders in the sample. For the disparate impact analysis, I first present differences in mean risk score by race, in addition to differences in risk level categorization by race. Second, I determine which risk factors contribute most to mean score differences between Black and White offenders. Finally, I conduct mediation analysis to determine the relationship between recidivism, race, and measures of criminal history.

For the predictive bias analysis, I first compare AUC statistic for Black and White offenders to determine if the instrument is equally as predictive across race. Second, I determine if the direction and strength of the instrument score changes for Black and White offenders. Third, I test whether the form of the slope and intercept differs for Black and White offenders, or as a function of the risk score.

**Results**

**Descriptive Statistics**

Table 5.1 presents the descriptive statistics for the White and Black offending sample (N=7,022). Black offenders were more likely to recidivate than were White offenders (54% vs 41% respectively; t (-11.13) = -13%, p <0.001). Black offenders were also overrepresented across most of the risk factors used in the instrument (all t-test mean differences were significant at p <0.001). In particular, Black offenders, on average, have a slightly higher prior record score
than White offenders ($m=1.9$ versus $m=1.3$, respectively), are much more likely to have a juvenile arrest (32% versus 20%) and have about one and a half more arrests than the average White offenders ($m=5.9$ versus $m=4.3$). They are also much more likely to have a violent conviction offense (73% versus 67%) and are less likely to be convicted for a sex offense (8% versus 16%). Black offenders are also slightly more likely to be male (92% versus 88%) and are slightly younger ($m=29.4$ versus $m=30.1$).

Table 5. 1: Select Descriptive Statistics (Black/White Sample)

N=7,022

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Black (N=3,405)</th>
<th>White (N=3,607)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recidivated</td>
<td>54%</td>
<td>41%</td>
</tr>
</tbody>
</table>

Independent Variables

Offender Demographics

<table>
<thead>
<tr>
<th></th>
<th>Black</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>92%</td>
<td>88%</td>
</tr>
<tr>
<td>Age (SD)</td>
<td>29.4 (9.8)</td>
<td>30.1</td>
</tr>
<tr>
<td>Allegheny County</td>
<td>18%</td>
<td>14%</td>
</tr>
<tr>
<td>Philadelphia County</td>
<td>42%</td>
<td>12%</td>
</tr>
<tr>
<td>Other Urban County</td>
<td>34%</td>
<td>49%</td>
</tr>
<tr>
<td>Rural County</td>
<td>6%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Offender Criminal History

<table>
<thead>
<tr>
<th></th>
<th>Black</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Record Score</td>
<td>1.9 (2.2)</td>
<td>1.3 (2)</td>
</tr>
<tr>
<td>Prior Arrest Total</td>
<td>5.9 (5.4)</td>
<td>4.3 (4.3)</td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>32%</td>
<td>20%</td>
</tr>
<tr>
<td>Domestic Violence</td>
<td>11%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Case Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Black</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offense Gravity</td>
<td>10 (1.3)</td>
<td>9.9 (1.2)</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>8%</td>
<td>16%</td>
</tr>
<tr>
<td>Violent Offense</td>
<td>73%</td>
<td>67%</td>
</tr>
<tr>
<td>Property Offense</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>Drug Offense</td>
<td>16%</td>
<td>13%</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>2%</td>
<td>1%</td>
</tr>
</tbody>
</table>

All differences between groups significant at the $p<0.001$ level, except domestic violence arrest.
Recidivism by Risk Factor

Not only are Black offenders overrepresented in each risk factor, Table 5.2 shows that Black offenders consistently fail at higher rates than White offenders at each risk factor. Furthermore, some differences in the recidivism proportions are quite substantial. For example, out of offenders who had a juvenile arrest or conviction, 72% of Black offenders failed, while only 55% of White offenders failed. Similarly, out of offenders with a miscellaneous conviction offense, 71% of Black offenders failed, compared to only 42% of White offenders. Each risk factor contributes to the overall risk score; thus, these findings suggest that a higher proportion of Black offenders fail at every point on the instrument scale.

Table 5.2 Recidivism by Risk Factor (Black/White Sample)
N=7,022

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Points</th>
<th>%Black Fail</th>
<th>%White Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1</td>
<td>56</td>
<td>42</td>
</tr>
<tr>
<td>Age 14-21</td>
<td>2</td>
<td>65</td>
<td>49</td>
</tr>
<tr>
<td>Age 22-40</td>
<td>1</td>
<td>52</td>
<td>43</td>
</tr>
<tr>
<td>PRS 1+</td>
<td>1</td>
<td>61</td>
<td>53</td>
</tr>
<tr>
<td>Arrest 3-4</td>
<td>1</td>
<td>52</td>
<td>42</td>
</tr>
<tr>
<td>Arrest 5+</td>
<td>2</td>
<td>65</td>
<td>58</td>
</tr>
<tr>
<td>Juvenile Arrest/Conviction</td>
<td>1</td>
<td>72</td>
<td>55</td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>1</td>
<td>65</td>
<td>52</td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>1</td>
<td>58</td>
<td>45</td>
</tr>
<tr>
<td>Violent Offense</td>
<td>1</td>
<td>57</td>
<td>46</td>
</tr>
<tr>
<td>Property Offense</td>
<td>1</td>
<td>71</td>
<td>68</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>1</td>
<td>71</td>
<td>42</td>
</tr>
</tbody>
</table>
Multivariate Logistic Regression

Table 5.3 displays results of the final multivariate logistic regression model used to validate the instrument (Table 4.9) for both Black and White offenders. Table 5.3, Model 1 includes Black offenders only \((N=3,405)\) and Model 2 represents the prediction model with White offenders only \((N=3,607)\). Several differences between the models are evident. First, prior record score is only significant in Model 2, suggesting that having a prior record score of 1 or more is associated with a greater likelihood or arrest for White offenders only \((or=1.5, p<0.001)\). Second, juvenile arrest is associated with an 88% increase in the odds of recidivating for Black offenders \((or=1.88, p<0.001)\), but is not significant for predicting recidivism for White offenders. Third, Black offenders who have a conviction for a miscellaneous offense have twice the odds of recidivating than offenders who have a conviction for a drug offense \((or=2.1; p <0.05)\). Yet, in Model 2, there is not a significant difference between the recidivism rate of sex offenders and miscellaneous offenders for White offenders \((or=0.87; NS)\). Model fit is comparable between the two sub-samples \((AUC: 0.71 \text{ for both samples})\). This finding suggests that while the prediction model has similar discriminatory utility for both Black and White offenders, individual level items may differ in their relationship across race. Because split sampling is not an ideal approach to showing heterogeneous variable effects, the following section compares model fit using the addition of race as an explanatory variable, and explores the possibility of interaction effects between risk factors and race.

\[33 \text{ Bivariate correlation analysis indicated no significant collinearity between variables for both Black and White offenders.}\]

\[34 \text{ Split sampling is analogous to running a fully interactive model, so the significance of any one variable depends on other covariates.}\]
Table 5.3: Multivariate Logistic Model of Recidivism by Race

<table>
<thead>
<tr>
<th></th>
<th>(1) Final Model for Black Offenders (N=3,405)</th>
<th>(2) Final Model for White Offenders (N=3,607)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Offender Demographics</strong></td>
<td><strong>Odds</strong></td>
<td><strong>SE</strong></td>
</tr>
<tr>
<td>Male</td>
<td>1.87</td>
<td>0.27***</td>
</tr>
<tr>
<td>Age 14-21</td>
<td>2.49</td>
<td>0.36***</td>
</tr>
<tr>
<td>Age 22-40</td>
<td>1.59</td>
<td>0.18***</td>
</tr>
<tr>
<td><strong>Offender Criminal History</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Recode Score=1+</td>
<td>1.18</td>
<td>0.1</td>
</tr>
<tr>
<td>Arrest 3-4</td>
<td>1.54</td>
<td>0.16***</td>
</tr>
<tr>
<td>Arrest 5+</td>
<td>2.89</td>
<td>0.3***</td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>1.88</td>
<td>0.19***</td>
</tr>
<tr>
<td>Domestic Violence Arrest</td>
<td>1.63</td>
<td>0.2***</td>
</tr>
<tr>
<td><strong>Case Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>1.41</td>
<td>0.12***</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>1.16</td>
<td>0.19</td>
</tr>
<tr>
<td>Violent Offense</td>
<td>1.38</td>
<td>0.15**</td>
</tr>
<tr>
<td>Property Offense</td>
<td>2.2</td>
<td>0.73*</td>
</tr>
<tr>
<td>Miscellaneous Offense</td>
<td>2.1</td>
<td>0.61*</td>
</tr>
<tr>
<td>Constant</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>-2 Log Likelihood</td>
<td>-2100.17</td>
<td></td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>AUC-ROC</td>
<td>0.71</td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05; ** p < 0.01; *** p < 0.001.
Reference groups: female, Age 41+, Arrest 1-2, Drug Offense, OGS 11-14

Table 5.4, Model 1 shows a multivariate logistic model of recidivism for the Black/White sample (N=7,022) which retains all significant predictors from the newly-developed instrument model. All relationships between risk factors and recidivism were in the expected direction, and were generally similar to the results using the full sample of offenders (see Table 4.9, Model 1; with Hispanics and other races included). Table 5.4 Model 2, shows that even when controlling for all risk factors included in the instrument, being Black is still associated with a 36% higher odds of recidivating (or = 1.36, p <0.001) and adds a small amount of predictive utility to Model.
1 ($\chi^2(1) = 34.34, p<0.001$). To determine whether the effect of any one risk factor varied in its relationship with recidivism based on race, I ran a series of logistic regression models to determine interaction among each risk factors with race (models with non-significant interaction effects not shown). Table 5.4 Model 3 shows that the interaction of race and juvenile arrest (black x juvenile arrest) was significant ($OR=1.74, p<0.001$) and added, minimally but significantly, to the predictive utility of the Model 2 ($\chi^2 (1) = 21.31, p<0.001$).\(^{35}\) Confirming the findings from Table 5.3, this finding indicates that the effect of juvenile arrests on recidivism only increases the odds of recidivism for Black offenders ($OR=1.74, p<0.001$). In sum, the relationship between juvenile arrest and recidivism differs for Black and White offenders. However, this interaction cannot be accounted for in the risk assessment instrument because, for ethical reasons, race cannot be included as a predictive factor.

\(^{35}\) No other interaction variables were significant at the P<0.001 level.
To determine whether the use of the previously constructed risk instrument has the potential for disparate impact, I assessed the difference in average risk scores between White and Black offenders. Consistent with prior research, Black offenders displayed higher average scores on the risk assessment instrument than did White offenders. Table 5.5 shows a 0.7 difference in
the average score between Black and White offenders on an 11-point scale ($t$ (-15.51) -0.7, $p <0.001$). Cohen's $d$ analysis, which is used to indicate the standardized difference between two means, demonstrates that the overall difference in scores is small to medium ($d=0.37; 27\%$ non-overlap).\footnote{According to Cohen (1992) a $d$ of 0.2, 0.5, and 0.8 is defined as a small, medium, and large effect, respectively.}

Table 5.5: Average Risk Score by Race (Black/White Sample)
N= 7,022

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>5.73</td>
<td>5.67</td>
<td>5.8</td>
</tr>
<tr>
<td>White</td>
<td>5.03</td>
<td>4.97</td>
<td>5.09</td>
</tr>
</tbody>
</table>

T-test significant at the $p <0.001$ level

Figure 5.1 shows the distribution of White and Black offenders across risk levels (with the corresponding risk scores). Black offenders compose 43\% of the total sample, yet they are under-represented in the Low-risk level (32\%) and over-represented at both the High (55\%) and Very High (60\%) risk levels. Conversely, the opposite pattern is observed with White offenders. From another perspective: 10\% of all White offenders are in the Very High risk category, compared to 17\% of all the Black offenders (Appendix H presents the sample distribution by risk level). Similarly, one-quarter of all White offenders are in the Low risk category, compared to 13\% of Black offenders. In sum, these analyses address the first Research Question to find that Black offenders score higher on the risk assessment instrument and are more likely to appear in the High and Very High risk categories than White offenders.
Mean Score Differences by Race Across Risk Factors

Group mean differences in risk scores between Black and White offenders are the result of an unequal distribution of risk factors, in addition to the weight of said risk factor in the scoring algorithm (which is dependent upon effect size in the development sample). In the following analysis, I determine which risk factors contribute most to mean score differences across race. Table 5.7 shows the mean score differences and standard deviations across each risk factor. As shown, Black offenders score higher on every risk category, with the exception of offense gravity score. Criminal history appears to be the strongest contributor to the 0.7 mean score difference between White and Black offenders. In particular, differences in the number of prior arrests contributes to almost half (46%) of the total mean score difference between Black and White offenders, which translates to a small to medium standardized difference ($d=33$; 21% non-overlap). Differences in prior record score and juvenile arrest contribute to another 19% and 17% of the difference, respectfully. Race differences in scores across the other risk factors are
small. Thus, this analysis answers the second Research Question of Chapter 5 to find that criminal history measures, and particularly prior arrest, contribute most to the mean score differences between Black and White offenders.

Table 5.6: Mean Score Differences by Race for Each Risk Factor (Black/White Sample)
N=7,022

<table>
<thead>
<tr>
<th></th>
<th>Black (n=3,405)</th>
<th>White (n=3,617)</th>
<th>Difference</th>
<th>% Attributable to Score Diff.</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>5.73</td>
<td>1.82</td>
<td>5.03</td>
<td>1.98</td>
<td>.7</td>
</tr>
<tr>
<td>Male</td>
<td>.92</td>
<td>.27</td>
<td>.88</td>
<td>.32</td>
<td>.04</td>
</tr>
<tr>
<td>Age</td>
<td>1.14</td>
<td>0.64</td>
<td>1.07</td>
<td>0.64</td>
<td>.07</td>
</tr>
<tr>
<td>PRS 1+</td>
<td>0.54</td>
<td>0.5</td>
<td>0.41</td>
<td>0.49</td>
<td>0.13</td>
</tr>
<tr>
<td>Arrest</td>
<td>1.21</td>
<td>0.85</td>
<td>0.89</td>
<td>0.88</td>
<td>0.32</td>
</tr>
<tr>
<td>Juvenile Arrest</td>
<td>0.32</td>
<td>0.47</td>
<td>0.2</td>
<td>0.4</td>
<td>0.12</td>
</tr>
<tr>
<td>Dom. Viol. Arrest</td>
<td>0.12</td>
<td>0.31</td>
<td>0.1</td>
<td>0.29</td>
<td>0.02</td>
</tr>
<tr>
<td>OGS 7-10</td>
<td>0.73</td>
<td>0.44</td>
<td>0.77</td>
<td>0.42</td>
<td>-0.04</td>
</tr>
<tr>
<td>Conviction Offense</td>
<td>0.76</td>
<td>0.42</td>
<td>0.71</td>
<td>0.45</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Dichotomous and trichotomous variables were collapsed into single risk domains.

Differences in the number of arrests, prior record, and juvenile arrest between White and Black offenders drive the difference in the average risk scores. Harcourt (2015) argues that criminal history has become a "proxy" for race — essentially serving as a stand-in for race in actuarial risk assessment instruments that are no longer able to include race as a risk factor due to ethical reasons. To examine the relationship between different measures of criminal history, race, and recidivism apart from their relation to the risk score, I conduct a series of logistic regression analyses and conduct mediation analysis to determine how much of the effect of race on recidivism is mediated by criminal history. First, criminal history (arrest, prior record score, and

---

37 Total does not equal 100% because of rounding.
juvenile arrest) and race are correlated ($r = 0.16, 0.15, and 0.13, respectively). Second, a comparison of AIC and BIC values\textsuperscript{38} between a model predicting recidivism with race and three separate models predicting recidivism with arrest, prior record score, and juvenile arrest shows that each measure of criminal history predicts arrest better than does race. Third, binary mediation analysis showed that arrest mediates 28%, prior record score mediates 19%, and juvenile arrest mediates 26% of the total effect of race on recidivism.\textsuperscript{39} Fourth, race still has a small, but significant, effect on recidivism even after arrest, prior record score, and juvenile arrest are controlled for in a logistic regression model ($\chi^2(1)=34.01$, p<0.001). Thus, while criminal history is correlated with race, it does not serve as a "proxy" for race in the relationship between race and recidivism.

**AUC by Race**

To test the predictive fairness of the risk instrument by race, I compare the AUC statistics and the failure rate by risk level across race. Table 5.8 shows that the overall AUC statistic for the risk score is 0.71, indicating that the instrument has "strong" discriminatory power (Rice and Harris, 2005).\textsuperscript{40} While there is a significant difference in the recidivism base rate between Black and White offenders (54% versus 40%, respectively), the instrument also has medium to strong predictive validity for each race group (0.70 AUC for both groups). Thus, the instrument shows no significant difference in predictive utility by race. Substantively, this finding indicates that if one were to select a random recidivist and a random non-recidivist individual from the Black or

\textsuperscript{38} Estimates the quality of each model, relative to each of the other models. Generally used for model selection.

\textsuperscript{39} I used the binary mediation package in Stata 14 (Ender, 2011).

\textsuperscript{40} Minimum AUC of 0.56, 0.64, and 0.71 correspond to small, medium, and large effects, respectively (Rice and Harris, 2005) and are not affected by base group differences in recidivism.
White group, there would be a 70% chance that the recidivist would have a higher risk score.

Table 5.7: Failure Rate, AUC, for Recidivism (White/Black Sample)
N=7,022

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>White</th>
<th>Black</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0-3)</td>
<td>20%</td>
<td>19%</td>
<td>23%</td>
</tr>
<tr>
<td>Medium (4-5)</td>
<td>38%</td>
<td>36%</td>
<td>41%</td>
</tr>
<tr>
<td>High (6-7)</td>
<td>60%</td>
<td>56%</td>
<td>64%</td>
</tr>
<tr>
<td>Very High (8-10)</td>
<td>72%</td>
<td>65%</td>
<td>77%</td>
</tr>
<tr>
<td>Base Rate*</td>
<td>48%</td>
<td>40%</td>
<td>54%</td>
</tr>
<tr>
<td>AUC</td>
<td>0.71</td>
<td>0.70</td>
<td>0.70</td>
</tr>
</tbody>
</table>

*Pearson chi2(1) = 121.86,  p < 0.001 (between black white recidivism base rate).

Table 5.8 also shows that the recidivism rates for offenders classified at each risk level (Low, Medium, High, and Very High) increase monotonically with risk level and that this pattern occurs for both Black and White offenders. Figure 5.2 shows a positive relationship between recidivism and risk level for both Black and White offenders. Similarly, Table 5.9 shows that failure rates increase with higher scores for both groups (with the exception of score 0 and score 10 which contain too few cases to be of value). Thus, the degree of the relationship between risk score and recidivism does not vary by race.
Figure 5. 2: Failure Rate by Risk Level

Table 5. 8: Failure Rate, AUC, for Recidivism (White/Black Sample) N=7,022

<table>
<thead>
<tr>
<th>Risk Score</th>
<th>White N</th>
<th>% Fail</th>
<th>Black N</th>
<th>% Fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>1</td>
<td>99</td>
<td>5</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>310</td>
<td>17</td>
<td>143</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>470</td>
<td>23</td>
<td>249</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>555</td>
<td>30</td>
<td>397</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>632</td>
<td>41</td>
<td>609</td>
<td>46</td>
</tr>
<tr>
<td>6</td>
<td>595</td>
<td>50</td>
<td>733</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>574</td>
<td>62</td>
<td>678</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>263</td>
<td>64</td>
<td>385</td>
<td>75</td>
</tr>
<tr>
<td>9</td>
<td>99</td>
<td>69</td>
<td>163</td>
<td>82</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>60</td>
<td>14</td>
<td>86</td>
</tr>
</tbody>
</table>
**Predictive Fairness**

Next, I tested whether the relationship between risk score and recidivism varies by race. To achieve predictive fairness, an instrument score of Y should correspond to the same rate of recidivism for both Black and White offenders. Table 5.8 shows that while recidivism rates increased with risk levels for both Black and White offenders, there were small to substantial differences in the corresponding recidivism base rate between the two groups. For the lower end of the scale (Low and Medium risk levels) the difference appears to be small, with Black offenders having 3%-5% higher rates of recidivism than White offenders. However, for offenders in the High and Very High levels Black and White recidivism rates differ by 9% and 12%, respectively. This suggests that the recidivism rate for White offender may be overestimated, and the recidivism rate for Black offenders may be underestimated by the model.

Table 5.9 provides a similar comparison by score, showing that Black offenders recidivate at higher rates at most points in the scale.

Table 5.10 shows that, on average, both White and Black recidivists score 1.3 points higher on the risk assessment scale than non-recidivists. However, both Black recidivists and non-recidivists also score half a point higher than White recidivists and non-recidivists, respectively.
Table 5. 9: Average Risk Score by Failure (Black/White Sample)  
N= 7,022

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Black</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Succeed</td>
<td>5</td>
<td>4.9</td>
<td>5.1</td>
</tr>
<tr>
<td>Recidivate</td>
<td>6.3</td>
<td>6.3</td>
<td>6.4</td>
</tr>
<tr>
<td><strong>White</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Succeed</td>
<td>4.5</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Recidivate</td>
<td>5.8</td>
<td>5.8</td>
<td>5.9</td>
</tr>
</tbody>
</table>

T-test significant at the p <0.001 for all scores

Predictive Fairness by Risk Score

Prior analysis shows that the *strength* of the relationship between risk score/risk level and recidivism does not vary by race. That is, rates of recidivism both increase monotonically with risk score and risk level for both Black and White offenders. To test whether the *form* of the relationship between the risk score and recidivism varied by race, I estimated a series of logistic regression models predicting recidivism with risk score, race, and an interaction of race and risk score. As shown in Table 5.11, Model 1 and Model 2 both the instrument risk score and race are significantly associated with failure, although risk score provides better model fit. A comparison of Model 2 to Model 3 shows that race added small, but significant, incremental utility to the risk score in predicting recidivism ($\chi^2(1) = 914.92, p <0.001$). Specifically, Model 3 shows that even after controlling for risk score, Black offenders have 36% higher odds of recidivating than White offenders. Finally, a comparison of Model 3 to Model 4 shows that there is no significant interaction between risk score and race, meaning that the slope of the relationship between risk score and reoffence is similar for Black and White offenders ($\chi^2(1) = 3.05, NS$). Essentially, this finding indicates that the effect of risk score on recidivism does not vary based on race.
Table 5.10: Logistic Model Predicting Recidivism and Testing Racial Fairness by Risk Score (Black/White Sample) N=7,022

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th>Model 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds</td>
<td>SE</td>
<td>Odds</td>
<td>SE</td>
<td>Odds</td>
<td>SE</td>
<td>Odds</td>
<td>SE</td>
</tr>
<tr>
<td>Risk Score</td>
<td>1.5</td>
<td>0.02</td>
<td>***</td>
<td>—</td>
<td>1.52</td>
<td>0.02</td>
<td>***</td>
<td>1.48</td>
</tr>
<tr>
<td>Black</td>
<td>—</td>
<td>1.7</td>
<td>0.08</td>
<td>***</td>
<td>1.36</td>
<td>0.07</td>
<td>***</td>
<td>1.02</td>
</tr>
<tr>
<td>Risk Score X</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.05</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.08</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>Chi-Square</td>
<td>1001.73</td>
<td>122.19</td>
<td>1037.11</td>
<td>1040.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-</td>
<td>-</td>
<td>-340.1</td>
<td>-</td>
<td>4338.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4357.79</td>
<td>4797.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.1</td>
<td>0.01</td>
<td>0.11</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.3 shows the predicted probabilities of recidivism at each risk score by race (Table 5.4, Model 3). First, the slope of the relationship between risk score and recidivism does not differ for Black and White offenders, in that the predicted probability of arrest for both groups increased monotonically with risk score. Second, the intercept of the relationship for Black offenders is higher than for White offenders, meaning that the instrument overestimates recidivism for White offenders and underestimates recidivism for Black offenders. Third, the difference between the two slopes increases with risk score, indicating that over-estimates for White probabilities of recidivism are more prominent at the higher end of the scale. In these analyses, I answer the third Research Question of Chapter 5 – what is the relationship between risk score, race, and recidivism? I find that the while the direction and strength of the
relationship between risk score and recidivism does not vary by race (recidivism increases monotonically with the risk score), the instrument over-estimates the recidivism rate for White offenders, and under-estimates the recidivism rate for Black offenders due to the exclusion of race from the scoring algorithm.

**Figure 5. 3 Adjusted Probabilities of Recidivism by Race**  
N=7,022

![Adjusted Prediction of Failure by Risk Score for Offenders by Race](image)

**Discussion**

The universal omission of race from recidivism risk assessment instruments is coupled with decades of research showing the importance of minority racial status for criminal justice outcomes. The purpose of the present study was to explore the relationship between risk score, race, and recidivism by determining mean score differences by race, identifying which risk factors contribute most to differences in risk score, and testing the instrument for predictive fairness. Using a large sample of Black and White offenders (N=7,022) sentenced and released in
the state of Pennsylvania, findings showed that Black offenders score higher on the risk assessment instrument — due in large part to differences in criminal history — and that the instrument under-classifies Black offenders on the risk scale. These findings are discussed in more detail below.

First, I find that although race was not included in the risk scoring model, Black offender still receive slightly higher average risk scores on the risk instrument (i.e., a 0.7 difference on an 11-point scale) and are more likely to appear in High and Very High risk levels than White offenders. This finding is in line with prior research showing that Black offenders typically have more risk factors related to recidivism (Dieterich et al., 2016; Eisenberg et al., 2009; Flores et al., 2016; Gendreau et al., 1996; Hannah-Moffat & Maurutto, 2003; Ostrom et al., 2002; Skeem & Lowenkamp, 2016) and score higher on risk assessment instruments (Flores, et al., 2016; Hannah-Moffat & Maurutto, 2003; Ostrom et al., 2002; Skeem & Lowenkamp, 2016). These findings inform the debate regarding actuarial risk assessment and the potential effects on sentencing disparity, but they do not provide evidence of disparate impact. The effect of this risk instrument, or any similar static risk assessment instruments, on racial disparities in sentencing will depend on baseline sentencing practices, the accompanying risk assessment policy, and practitioner response. Certainly, one could think of a scenario in which an instrument of this type could have disparate impact. If, for example, sentences were not already reflective of risk (i.e., judges did not consider protection of the community in their sentencing decisions), the use of this tool in sentencing decisions would exacerbate the differences in sentencing outcomes for Black and White offenders. Presumably, Black offenders would be identified for more severe sentences because of their higher risk status, and vice versa for White offenders.

However, this scenario is unlikely to be a baseline sentencing structure in the US because
consideration of criminal history, which serves as an informal indicator of risk, is ubiquitous in state and federal sentencing structures. This fact provides context for the second main finding of the paper — that the majority of the difference between Black and White risk scores was due to differences in criminal history. Black offenders had significantly higher scores on three measures of criminal history: prior arrests, prior record score, and juvenile arrest. The combination of these variables accounted for 82% of the total difference in scores between Black and White offenders. Arrest, alone, accounted for almost half the total difference. These scoring differences exist because of two reasons: First, Black offenders have higher average prior records scores, more prior arrests, and are significantly more likely to have a juvenile arrest. Second, arrest is weighted more heavily in the instrument algorithm to reflect the large effect size in the development model (Table 4.4, pg. 88). The unequal distribution of specific risk factors interacts with decisions about instrument construction to produce higher mean score for Black offenders, underscoring the importance of transparency in the development process.

Skeem and Lowenkamp (2016) also found that differences in criminal history accounted for the majority of the racial difference in scores on the PCRA. The authors argued that criminal history is "already embedded in the sentencing guidelines" (pg. 1), challenging the notion that the consideration of risk scores would exacerbate racial disparities in sentencing. Given the ubiquity of criminal history enhancements (Frase et al., 2015), this is an important contention, although one that requires thoughtful consideration. For example, in the present study and in Skeem and Lowenkamp’s assessment of the PCRA, it is the difference in the number of prior arrests that is the most powerful explanatory variable for mean race difference in score. Yet, in the context of sentencing it is often the number and seriousness of prior convictions that serves as a relevant marker of criminal history (Frase et al., 2015). While there is plenty of research
indicating how various measures of criminal history affect sentencing outcomes (e.g., Frase et al., 2015; Rehavi & Starr, 2014; Ulmer & Laskorunsky, 2016), there is little research examining the effect of arrest history on the formal and informal consideration of risk, or sentencing outcomes. Indeed, we know little about the consistency with which judges have access to rap sheet information at sentencing. Thus, the introduction of actuarial recidivism prediction instruments into the courtroom may magnify previously obscured differences in arrest history, with indeterminate consequences.

Results also show that Black offenders are not only more likely to have a contact with the juvenile system (32% vs 20% for White offenders), but that the effect of juvenile arrest increased the odds of recidivism for Black offenders only, after controlling for other relevant risk factors. In fact, a full 72% of Black offenders who have a juvenile arrest recidivated within 3 years (compared to 55% of White offenders). Juvenile arrests and convictions (i.e., adjudications) are an important measure of criminal history, but they are also qualitatively different than adult involvement with the system. Research shows that having contact with the juvenile system is a strong predictor of adult recidivism (National Research Council, 1986a). While 40%-60% of juvenile offenders desist from crime by early adulthood, those who continue offending increase the severity of their offending (Farrington, 2003). Young adults who desists from crime usually do so as they gain "stakes in conformity", such as getting married and becoming employed (Horney, Tolan, & Weisburd, 2012; Sampson & Laub, 1990). Thus, adult offenders who have previous involvement with the juvenile justice system represent offenders with few protective factors.

In the U.S., racial segregation and concentrated disadvantage has hindered Black men’s abilities to participate in conventional norms, such as getting married and having stable
employment (Peterson, Krivo, & Browning, 2008; Sampson & Wilson, 1995). Given the unequal distribution of opportunities for "stakes in conformity", Black offenders with juvenile records represent a particularly high-risk group of offenders. Interestingly, research also shows that juvenile convictions elicit different responses for Black and White adult offenders at sentencing (Ulmer and Laskorunsky, 2016). In particular, having a juvenile conviction increases the odds of imprisonment for Black offenders but not for White offenders — even while controlling for relevant risk factors. In Pennsylvania, Black youth are arrested at almost four times the rate of White youth, and are significantly less likely to be diverted away from the justice system (Griffith, Jirard, & Ricketts, 2012). Thus, Black offenders, particularly Black men, receive the cumulative effects of disadvantage and disparity in system response. Ideally, risk assessment instruments would be reflective of the way risk factors vary by race. However, given the exclusion of race as a scoring criteria in modern risk assessment instruments, it is not possible to vary the effect of individual risk factors for White and Black offenders.

In testing the predictive fairness of the instrument across race, results for the present study showed that the instrument scores had "medium" (borderline “strong”) predictive utility for both Black and White offenders (AUC: 0.70) (Rice and Harris, 2005). This indicates that the instrument does well at discriminating between recidivists and non-recidivist for both Black and White sub-samples. Additionally, the degree of the relationship between risk score and recidivism is similar for Black and White offenders: rates of failure increased monotonically with both risk score and risk level across race. While race still showed a significant relationship with arrests even after controlling for risk score, it was not included in this instrument's scoring procedure. Thus, this finding shows that it is possible for an actuarial risk assessment instrument to both exclude race as a risk factor and to differentiate between offenders based on their risk of
re-offense with moderate strength.

The final set of analyses shows that the instruments is not racially-neutral in its form of prediction. Ideally, a risk score (or risk level) of X should indicate a similar predicted probability of recidivism for both Black and White offenders. However, this is only possible if the recidivism rate for both groups is somewhat similar at each point of the scale. As it stands, Black offenders recidivate at a small, but consistently higher, rate at each risk point in the scale. The risk instrument averages the recidivism rates at each point in the scale to produce predicted probabilities of reoffering, thus the result is an over-prediction of risk for White offenders and an under-prediction of risk for Black offenders. While these differences are small at the lower end of the scale, they become problematic at the High and Very High levels in which the difference between Black and White offender recidivism rates is more pronounced. To be clear, the risk score does not interact with race: the slope of the relationship between the score and arrest is similar for Black and White offenders (Figure 5.3). However, the intercept of the relationship between the score and recidivism is higher for Black offenders than for White offenders. Given a particular risk score, a Black offender is still 1.52 times more likely than a White offender to recidivate after release.

The remaining effect of race on recidivism, accounting for major risk factors, may indicate two things. First, it could be that minority race status serves as a proxy for a variety of associated risk factors that are difficult to identify at the sentencing stage. For example, by the time they become involved with the criminal justice system, Black offenders have experienced a level of disadvantage and structural inequality not typically experienced by White offenders (Sampson and Lauritsen, 1997; Sampson and Wilson, 1995). These differences manifest themselves through Black racial status and affect risk of recidivism. Second, it could be that skin
color serves as a risk factor for police attention. Research shows that Black offenders are more likely to be monitored by the police and to receive differential system response for behavior similar to White offenders. For example, Black students are more likely to receive official institutional punishment for similar infractions as White students (Nicholson-Crotty et al., 2009; Skiba et al., 2002). Black motorists are more likely to be pulled over and to have their car searched than are White motorists (Langton & Durose, 2013), and “Stop and Frisk” style policies have overwhelmingly been directed toward minority individuals (Gelman et al., 2007). In all likelihood, both these reasons contribute to the significant effect of race on recidivism, and future research efforts should focus on delineating these effects.

Can actuarial instruments be designed to accurately estimate the risk of recidivism of two groups whose base rates substantially differ? The findings from the present study, and previously published papers (Monahan, Skeem, & Lowenkamp, 2017; Skeem & Lowenkamp, 2016; Skeem et al., 2016), indicate this may be a challenge. For example, two studies tested the predictive fairness of the PCRA for men and women (Skeem, Monahan, and Lowenkamp, 2016) and for younger and older offenders (Monahah, Skeem, and Lowenkamp, 2017) and found similar results: In both studies, the instrument over-estimated the recidivism rate of the lower-risk group (i.e., women and older offenders), and vice versa. Interestingly, the instrument over-predicted the recidivism rate for older offenders despite the inclusion of age in the scoring algorithm. A third study testing the predictive fairness of the PCRA across race did not find prediction bias for the matched sample, but did find that the instrument over-predicted White recidivism rates for the unmatched sample (Skeem and Lowenkamp, 2016).41 Differences in risk

---

41 The matching procedure included matching offenders on the age, gender, and offense type, reducing the natural variation in base rates found between Black and White offending groups.
score distribution and base rates differences make it difficult to create an instrument that is equally fair to different groups (Kleinberg et al., 2016). Thus, it appears that statistical realities may undermine criminal justice ideals of predictive fairness. The foundation of actuarial risk prediction relies on probabilistic estimates which can obscure complex phenomena such as the divergent experiences of Black and White offenders in the criminal justice system.

Limitations

Several limitations of the present study should be noted. First, the generalizability of findings to other offending samples or to the use of a different instrument are unknown. The focus on testing risk assessment instruments for predictive bias and potential disparate impact is relatively new in criminal justice. With few established trends and many different instruments in use, more research is needed to make generalizations outside of the present context. Second, the actual impact a risk assessment instrument, such as the one used in the present study, would have on racial disparities in sentencing can only be determined by evaluating its use within a real sentencing environment. Explorative research can assess the potential for disparity, but it cannot account for structural and practitioner responses to policy changes, which would have substantial implication for how the instrument is used. A third limitation to note is that it is not possible to know how well the instrument reflects real differences in criminal offending. An arrest does not confirm that an offender committed a crime. Similarly, and perhaps more importantly, the absence of an arrests does not mean that an individual did not engage in criminal activity. This understanding is particularly important when evaluating the over-representation of Black offenders in the criminal justice system. Differential targeting and enforcement of criminal laws may over-estimate Black offenders’ actual criminal activity and under-estimate Whites
offenders’ involvement with crime.

Conclusion

In line with some of the criticism regarding the use of actuarial risk assessments in the courtroom (Hannah-Moffat, 2013; Starr, 2014), this study finds that Black offenders do indeed receive higher average risk scores than White offenders. The reliance on criminal history to predict recidivism in modern risk assessment instruments magnifies the differences in prior arrests between Black and White offenders, and allows for the consideration of a less conservative measure of criminal involvement in the courtroom. In some scenarios, this may increase racial disparities in sentencing, but more research is needed. Paradoxically, results from this study also show that that Black offenders could benefit from the use of actuarial instruments in the courtroom. The exclusion of race from the risk assessments algorithms creates recidivism estimates that obscure Black offenders’ higher reoffending base rates. Depending on how the risk assessment information is integrated into the sentencing process, as a group, Black offenders would be treated as slightly lower-risk than they actually are. Consequentially, a similarly constructed instrument is also likely to over-predict the recidivism rate of White offenders. These findings provide nuance to the debate about actuarial risk prediction instruments in the courtroom and prod policy makers and practitioners to consider statistical realities along with ethical and legal parameters.
Chapter 6- Conclusion

Summary of Findings

This dissertation uses an original agency dataset to examine an important trend in criminal justice: the use of actuarial risk assessment instruments during sentencing. Using a sample of serious offenders convicted and released in Pennsylvania, ten distinct risk factors were identified as significant predictors of recidivism at the sentencing stage. Using eight of the identified factors, I created a risk prediction algorithm that would allow judges to evaluate offenders’ likelihood of re-offense. Subsequently, the instrument was validated for predictive utility on a separate sample of offenders and evaluated for possible racial bias and potential for disparate impact. The instrument displayed “strong” (i.e., AUC:0.72) predictive utility for the validation sample and “medium” (i.e., AUC: 0.70) predictive utility for the Black and White sub-samples (Rice & Harris, 2005). Results showed that, on average, Black offenders received higher risk scores on the instrument than did White offenders, mostly as a result of differences in their criminal histories. However, the assessment instrument slightly under-predicted the actual recidivism rate for Black offenders, and over-predicted the recidivism rate for White offenders, particularly at the higher end of the scale. These finding suggests that some uses of the instrument could disparately affect Black offenders, by targeting them for more severe punishment due to higher average risk scores. However, the instrument’s use could also benefit higher risk Black offenders whose recidivism rate would be under-predicted by the instrument.

42 This study found that being Black, from Allegheny county, young, male, having a larger number of arrests, a higher prior record score, a domestic violence arrest, a juvenile arrest, and being convicted of less serious property, violent, or miscellaneous offense was associate with a higher recidivism rate. Race and county were not included in the final scoring algorithm.
Theoretical Implications

Previous research on race, gender, and age disparities in sentencing has mainly relied on the focal concerns perspective (Steffensmeier & Demuth 2000; Steffensmeier et al. 1998) and Albonetti’s (1991) causal attribution theory to explain processes behind judicial reasoning during sentencing decisions. The focal concerns perspective suggests that judges are guided by three focal concerns in making sentencing decisions: the offender’s blameworthiness, protection of the community, and practical implications of the sentencing decision. And, Albonetti’s (1991) causal attribution theory states that to reduce their level of uncertainty, judges rely on social status characteristics (e.g., gender, race, and social class) to make determinations about offenders’ likelihood of recidivism (Steffensmeier & Demuth, 2000). Judges determine offender blameworthiness mainly using the seriousness of the conviction offense and assess offender’s risk for future criminality mainly using prior history. The use of actuarial risk assessment instruments in the courtroom mainly speaks to protection of the community.

Theoretically, the inclusion of an actuarial instrument into the sentencing process should reduce judicial uncertainty about an offender’s likelihood of recidivism. If, to reduce their uncertainty, judges are likely to rely on offender stereotypes to make informal assessments about an offender’s risk, the use of a formal risk assessment instruments during the sentencing process should also reduce judges’ reliance on offender stereotypes for making sentencing decisions. With the use of actuarial instruments, judges would be relying less on informal stereotypes related to risk and relying more on the statistical associations between risk factors and recidivism. Interestingly, because of overlap between predictors of recidivism and factors judges use to make sentencing decisions, judges would generally be considering the same factors – albeit in a different form. Several of the factors found to have a positive effect on recidivism in
the present study, such as age and gender, overlap with factors previously identified in research on criminal stereotypes (e.g., Clair & Winter, 2016; Gottfredson, 1999; Steffensmeier & Demuth, 2006). Thus, not only would actuarial risk assessments instrument utilize similar variables, it’s likely that risk assessments instruments would also categorize offenders in a way that overlapped with the patterned responses judges have developed to assess cases. For example, instrument would assess younger offenders as higher risk, which is in line with informal judicial assessment.

Similarly, if judges mainly use measures of criminal history (i.e., the prior record score under sentencing guidelines) to informally assess offender risk, consideration of an actuarial risk assessment tool should reduce judges’ reliance criminal history as an indicator or risk. In fact, in the context of focal concerns, it would be worth considering reworking the guideline matrix to replace the prior record score with an instrument risk score. Because of the extent that risk assessment algorithms rely on measures of criminal history to predict reoffence, there would be substantial overlap between the two scores. However, this study showed that the composite instrument score is a better predictor of recidivism than multiple measures of criminal history. As such, the score would provide judges concerned with protecting the community more accurate assessments of offenders’ risk.

Actuarial risk assessment instruments are often perceived as objective and superior to the way in which judges currently assess risk - that is, by using stereotypes and cognitive filters. An oft repeated benefit of using risk assessment is that they minimize the influence of extra-legal variables and other biased factors. In reality, actuarial risk assessment instruments are reflective of the criminal justice system, including the inherent biases and stereotypes that produce the outcomes within it. The instrument risk score is an amalgam of legal and extra-legal factors, as
well as criminal stereotypes that have statistical support.

**Policy Implications**

Risk assessments are of interest to government agencies because they offer a data-driven, evidence-based approach to structuring sentencing decisions. Rather than adopting across-the-board reductions in the length and severity of sentences, actuarial risk assessments provide rationality for focusing reduction efforts on specific offenders. This is an attractive approach for cash-strapped legislatures still weary of appearing "soft on crime". With the correct policy in place, risk assessment tools may assist in reducing prison and jail populations, although this is not an inherent function of the tool. For example, the Virginia Sentencing Commission adopted a state-wide risk assessment tool for use at the sentencing stage and a post-adoption evaluation showed a significant decrease in the state’s prison population (Ostrom et al., 2002). This outcome was achieved because the state's goal for the use of the risk assessment instrument was to identify lower risk offenders for diversion. It is possible to craft a policy in which use of the exact same tool would have no effect on the correctional population, or even increase it. For example, mass incarceration may increase if actuarial tools are used as justification to increase the sentences of higher risk offenders, without comparable reduction in the sentences of lower-risk offenders.

The present study suggests that if the newly-created tool, or a similarly constructed tool, were used to inform sentencing decisions, Black offenders would fare comparably worse than White offenders. Due to their higher average risk scores Black offenders could be singled out for harsher sentences, or, at the very least, be denied access to diversion. Even if the overall prison population decreased as a result of the tool, the Black prison population would decrease
proportionally less. What findings from the present study cannot conclude is whether differences in mean scores reflected in this, and several other, risk assessment instruments would exacerbate the existent inequalities in sentencing outcomes. In both guideline and non-guideline states, Black offenders are already singled out for the harshest sentences, due in part to differences in the conviction offense, and more so, their higher criminal history scores (Frase et al., 2015). In their assessment of Minnesota's criminal history enhancements Frase and colleagues (2015, pg. 115) wrote that "offenders who are black pay again and again for their prior crimes. Their higher criminal history scores cause them to have much higher recommended executed-prison rates, somewhat higher recommended executed prison durations, much higher average recommended prison months (combining prison-commitment and prison-duration presumptions), and lower rates of downward dispositional departure from recommended executed-prison terms". In that context, an increase in racial disparity would only occur if, as a result of the tool, Black offenders received longer and more severe sentences than they would have otherwise received. One could imagine that the consideration of an actuarial risk assessment instrument at sentencing, in addition to the consideration of criminal history and offense gravity, would do little to change to change the existent longstanding patterns of disparity at sentencing.

Results from the present study show that the exclusion of race from the scoring algorithm results in under-estimating the recidivism rate for Black offenders, and over-estimating the recidivism rate of White offenders. Assessments of the PCRA suggest that this not a feature unique to this risk assessment instrument (Monahah, Skeem, and Lowenkamp, 2017; Skeem and Lowenkamp, 2016; Skeem, Monahan, and Lowenkamp, 2016). There are several potential solutions to this issue. One approach for making the instrument more accurate is to include race in the prediction algorithm. The recent Wisconsin court case (Wisconsin v. Loomis, 2016)
challenging the use of actuarial risk assessment in sentencing set a precedent for the use of "improper" factors, such as gender and socioeconomic status for the purpose of statistical prediction. If states were to avoid explicitly linking the risk score to sentence recommendations, there may be room to include race into prediction algorithms for the purpose of yielding more precise estimates of recidivism.

A second way to address a biased instrument, is to create a separate risk assessment for Black and White offenders. Given the large differences in recidivism base rates, this approach is ideal from a methodological perspective. Typically, this option has been employed to address the unbalanced effects of gender on recidivism. For example, the Minnesota Department of Corrections created separate risk assessment instruments for male and female prisoners (Duwe, 2014) as a way to account for differences in significant risk factors. However, the task of creating a race-specific instrument for use at the sentencing stage would be me with substantial ethical and legal challenges.

A third option for addressing instrument bias, is to continue excluding race as risk factor in the instrument model, but include offender characteristics that account for the effect of race on recidivism. This approach would require identifying mechanisms through which minority racial status affects recidivism. In fact, multi-dimensional, interview based instruments, like the COMPAS, already do a better job at identifying these underlying mechanisms — which may explain why race added no utility to the risk score in explaining recidivism in the Flores et al. (2016) study. For example, the COMPAS questionnaire includes 137 items assessing, among others, the offender’s residential stability, employment status, criminal peers, and even family criminal history (e.g., Has one of your siblings ever been arrested? [ProPublica, _ND]). The intention of such questions is not to create stand-in variables for race/ethnicity, but given the
unequal distribution of risk factors across populations, the result is similar. There are two downsides to this option. First, jurisdictions would no longer receive the benefit of using a short, static instrument that can be compiled with the use of agency-level data. It is unlikely that many jurisdictions would invest sufficient resources required to administer an interview-based assessment to every convicted offender, so the scalability of the tool would be limited. Second, education, employment, and family characteristics are the same “suspect” risk factors that have received the bulk of criticism regarding actuarial risk assessment (Hannah-Moffat, 2013; Starr, 2014). On the basis on fairness, it may be easier to argue for determining risk at the time of sentencing by using variables that are already considered for sentencing outcomes (i.e. offense and criminal history). The ethics of considering immutable social circumstances, such as family members' criminal behavior, or constrained life choices, such as level of education, to make punishment decisions are murky at best. Of course, similar criticisms can be, and have been, made regarding the use of criminal history variables in recidivism prediction (Harcourt, 2015).

The present study also highlights the importance of transparency in the instrument building process, and in the scoring algorithm. Decisions made during the model building process and the scoring procedure have substantial consequences for the predictive utility of the instrument and its potential effect on racial disparity. By relying on proprietary and undisclosed formulas to determine risk scores (as is the case for jurisdictions that utilize patented instruments), court officials, policy makers, and the defendant are not provided with an opportunity to examine how algorithms are weighing different data points. For example, the importance of criminal history - and in particular, prior arrest - in driving the difference between mean instrument scores for Black and White offenders is due in part to the weights assigned to arrest in the scoring algorithm. Presumably, judges may be interested in knowing that prior
arrests, not convictions, play a major role in determining risk scores, and defense attorneys may be interested in flagging any potential discrepancies between the score and official data.

Conversely, the effect of criminal history on mean risk score differences is important information for policy makers who are concerned with disparate impact. The popularity of private risk assessment instruments in the criminal justice system (and lack of alternatives) has forced government agencies to rely on the accuracy and perceived relevance of trademarked risk assessment algorithms. While there may be no inherent issues in using private companies to supply research and tools for the criminal justice system, the process of creating an instrument consequential to determining the loss of one’s freedom works best when the assessment process includes oversight and transparency.

Future Research

Given the interest in reducing the overrepresentation of minorities in the criminal justice system, the potential effects of using actuarial instruments to sentence offenders will require additional careful and thorough research. Most important, researchers must establish baseline sentencing practices to make inferences about disparate impact or changes to sentencing outcomes. While decades of research have focused on the effect of race/ethnicity and gender on sentencing outcomes, we know surprisingly little about the role of risk in those same sentencing decisions. In other words, are judges sentencing the riskiest offenders to the longest sentences? Researchers should examine how closely sentences conform to offender risk of recidivism and identify how judges evaluate the concept of risk. This process should take into account localized differences in sentencing practices.

Future research should also focus on several tangential questions. First, although the
presence of risk factors may be assessed at the sentencing stage, court officials generally have little understanding of the processes behind those risk factors. Thus, they cannot assess the likely complex selection processes by which offenders come to have or not have risk related factors. As an example, while differential offending contributes to differences in criminal history for Black and White offenders, we don’t know how much of that difference is also attributed to differential selection into the system. And, we know even less about the feedback loop between social circumstance, differential offending, and victimization that occurs over an offender’s life. The collateral consequences of being involved with the system, such as the inability to attain employment, perennially ripple through the lives of offenders. Risk assessment instruments measure the presence of risk factors, but researchers can help outline the complexity of the processes behind them.

Second, future research should focus on assessing how risk assessment instruments affect judicial behavior in a real sentencing context. Only two studies have attempted to assess how judicial behavior changes as a result of being presented with risk assessment information — neither of which used real sentencing events (Ruback et al., 2016; Starr, 2014). Researchers should explore whether, and how much, judges conform to risk assessment recommendations and if the inclusion of risk assessment information affects sentencing outcomes. Future research should also evaluate how the introduction of a risk score into the sentencing process changes the behavior of other courtroom actors. For example, could risk-related variables become a factor in plea bargaining?

Third, researchers have little understanding of how risk assessment scores interact with other case and offender characteristics in shaping judicial perceptions of risk or in what types of scenarios risk is important. In other words, when does offender risk matter? For example, risk of
recidivism may not be a strong consideration for serious crimes in which the harm of the offense outweighs more utilitarian consideration of efficient offender management. Qualitative research could also help to explore how judges evaluate risk factors which are generally within an offender's control and those that arise from ascribed status, and whether this has an effect on sentencing decisions.

Fourth, pre-sentence reports play an important role in shaping judicial perceptions of offender risk, yet we know little about how often these reports are filled out, how most states structure their reports and whether, or how closely judges review them. Do these differences have an effect on the final sentencing decisions? For example, in Pennsylvania, pre-sentence reports are not standardized, and include a variety of information which may shape the informal consideration of risk, including an offender’s employment status, upbringing, and substance abuse. Other states may include scores from risk assessment instruments.

Finally, more research is needed regarding local variation in policies and practices as it affects the relationship between risk factors and recidivism. A factor that predicts reoffending in one location may not increase the likelihood of failure in another location. Similarly, policies and practices unique to certain counties may serve as structural risk factors and affect individual likelihood of recidivism (e.g., Allegheny county for the present risk assessment instrument). For example, differences in policy typically dictate what constitutes as failure on probation and parole, with some agencies placing more importance on sanctioning offenders for recreational drug use than others agencies. Thus, the presence of certain risk factors, such as substance abuse issues, may affect the probability of offender failure differently across contexts. One possible line of research would be to use existing risk assessment instruments to compare predicted probabilities of arrest across counties. Population and policy differences between counties may
affect the strength with which specific variables predict recidivism and create the need for more localized instruments.

**Conclusion**

A major goal of criminal justice reform is to reduce mass incarceration while avoiding an increase in crime and victimization. As of this writing, bipartisan support for criminal justice reform remains strong — particularly at the state levels (Crutchfield, 2017; Lampard, 2016). Sentencing risk assessment tools can help to identify where to focus de-incarceration efforts, and, barring any major constitutional challenges, they are here to stay. The responsibility of criminologists will be to assess any collateral consequences of using these instruments, particularly as they might affect racial disparities in sentencing. It is now time to progress beyond estimating the potential effects of using such instruments, and to assess their effects in a real sentencing context. Multiple states consistently use risk assessments at the sentencing stage, and several more states are moving in that direction (e.g., Maryland). There are now opportunities to conduct pre- and post-evaluations on sentencing instruments’ effects on sentencing outcomes, changes in incarceration levels, offender recidivism, and levels of racial disparity in sentencing. Further, any blanket statements regarding the effects of using these risk instruments should be moderated by acknowledging localized differences in risk assessment policy, risk assessment construction, and baseline sentencing practices.

This scholarship adds to the debate regarding the practical and ethical implications of using actuarial risk assessment at sentencing. The current study’s findings show the importance of transparency in instrument construction and provide a replicable example for future development efforts. Specifically, I found that risk of recidivism can be predicted moderately
well using only a handful of static, agency-level factors. However, the findings regarding racial differences in risk score and prediction bias should serve as a caution to the criminal justice field. The use of actuarial instruments in the courtroom brings with it the potential to increase racial disparities in sentencing and will require continued focus from the field of criminology.
Bibliography


DeMichele, M., & Laskorunsky, J. (2014). Sentencing Risk Assessment: A Follow-up Study of
the Occurrence and Timing of Re-Arrest among Serious Offenders in Pennsylvania.

University Park, PA: Pennsylvania State University Justice Center for Research.


I(3), 155–176.


Chapman Journal of Criminal Justice, 1, 43.


Savelsberg, J. J. (1992). Law that does not fit society: Sentencing guidelines as a neoclassical


Directions in Psychological Science, 20(1), 38–42.


Appendices
Appendix A: Hazard of Recidivism

Full sample with 11 year follow-up time (N=7,935)
Appendix B: Model Building Procedure

Multivariate Logistic Models of Recidivism (Block Testing) Development Sample N=5,260

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
</tr>
<tr>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
</tr>
<tr>
<td>Alleghany***</td>
<td>Alleghany</td>
<td>Alleghany</td>
<td>Alleghany</td>
<td>Alleghany</td>
<td>Alleghany</td>
</tr>
<tr>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
</tr>
<tr>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
</tr>
<tr>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
</tr>
<tr>
<td>PRS 1+*</td>
<td>PRS 1+*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex Prior *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viol. Prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug Prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. Prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Misc. Prior</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Chi-Squared:** 126.69*** 483.34*** 617.25*** 628.92*** 650.61***

**Pseudo R2:** 0.02 0.07 0.08 0.09 0.09

**Log-Rank Test** N/A *** *** ** **
<table>
<thead>
<tr>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
</tr>
<tr>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>Philadelphia</td>
<td>Philadelphia</td>
<td>Philadelphia</td>
<td>Philadelphia</td>
</tr>
<tr>
<td>Alleghany</td>
<td>Alleghany *</td>
<td>Alleghany</td>
<td>Alleghany</td>
<td>Alleghany</td>
</tr>
<tr>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
</tr>
<tr>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
</tr>
<tr>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
</tr>
<tr>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
</tr>
<tr>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
</tr>
<tr>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
</tr>
<tr>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
</tr>
<tr>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
</tr>
<tr>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>Sex Offense</td>
<td>Sex Offense</td>
<td>Sex Offense</td>
<td>Sex Offense</td>
</tr>
<tr>
<td>Male**</td>
<td>Male**</td>
<td>Male**</td>
<td>Male**</td>
<td>Male**</td>
</tr>
</tbody>
</table>

**Chi-Squared:** 788.43*** 824.47*** 841.32*** 850.36*** 876.1***

**Pseudo R2:** 0.11 0.11 0.012 0.12 0.12

**Log-Rank Test*** *** * * ***
<table>
<thead>
<tr>
<th></th>
<th>Model 11</th>
<th>Model 12</th>
<th>Model 13</th>
<th>Model 14</th>
<th>Model 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
<td>Black***</td>
</tr>
<tr>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
<td>Hisp./Other</td>
</tr>
<tr>
<td>Alleghany*</td>
<td>Alleghany *</td>
<td>Alleghany *</td>
<td>Alleghany *</td>
<td>Alleghany *</td>
<td>Alleghany*</td>
</tr>
<tr>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
<td>Rural</td>
</tr>
<tr>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
<td>Arrest 3-4***</td>
</tr>
<tr>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
<td>Arrest 5+***</td>
</tr>
<tr>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+**</td>
<td>PRS 1+***</td>
</tr>
<tr>
<td>Sex Prior**</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
<td>Sex Prior **</td>
</tr>
<tr>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
<td>Drug Prior</td>
</tr>
<tr>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
<td>Age 14-21***</td>
</tr>
<tr>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
<td>Age 22-40***</td>
</tr>
<tr>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
<td>OGS 7-10***</td>
</tr>
<tr>
<td>Sex Offense</td>
<td>Sex Offense</td>
<td>Sex Offense</td>
<td>Sex Offense*</td>
<td>Sex Offense*</td>
<td>Sex Offense</td>
</tr>
<tr>
<td>Male***</td>
<td>Male**</td>
<td>Male**</td>
<td>Male**</td>
<td>Male**</td>
<td>Male**</td>
</tr>
<tr>
<td>DV Prior***</td>
<td>DV Prior***</td>
<td>DV Prior***</td>
<td>DV Prior***</td>
<td>DV Prior***</td>
<td>DV Prior</td>
</tr>
<tr>
<td>Inchoate</td>
<td>Multi</td>
<td>Gun</td>
<td>Specialist</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chi-Squared:</td>
<td>876.41***</td>
<td>876.59</td>
<td>877.76***</td>
<td>879.2***</td>
<td>857.84***</td>
</tr>
<tr>
<td>Pseudo R2:</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>Log-Rank Test</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>*</td>
</tr>
</tbody>
</table>
Appendix C: Conviction Offense Coding

**Violent Offenses**

Aggravated Assault - Cause S.B.I.
Aggravated Assault - Attempt S.B.I.
Aggravated Assault - Cause S.B.I. Police, etc.
Aggravated Assault - Attempt S.B.I. Police, etc.
Aggravated Assault - Cause or Attempt B.I. Police, etc.
Aggravated Assault - Cause or Att B.I. w/Deadly Weapon
Aggravated Assault - Fear S.B.I.
Neglect Care-dependent Person (Cause S.B.I.)
Arson - Endangering Persons; Person in Bldg. or B.I. results
Arson - Endangering Persons; Nobody in Bldg. and no B.I.
Assault by Prisoner
Burglary - Home: Person Present
Murder Inchoate - Conspiracy - no S.B.I.
Murder of The Third Degree
Murder Inchoate - Conspiracy with S.B.I.
Murder Inchoate - Attempt with S.B.I.
Murder Inchoate - Attempt - no S.B.I.
Murder Inchoate - Solicitation - no S.B.I.
Terroristic Threats
Robbery - Inflicts S.B.I.
Robbery - Threatens S.B.I.
Robbery - Commit/Threaten any F1 or F2
Robbery - Inflicts or Threatens B.I.
Unlawful Restraint-Victim <18 yrs old
Robbery of Motor Vehicle - with S.B.I.
Robbery of Motor Vehicle - without S.B.I.
Discharge of Firearm into an Occupied Structure
Voluntary Manslaughter
Endangering Welfare of Children - Course of conduct
Suicide, Causing or Aiding as Homicide - Murder 3
Intimidation of Witness/Victim-Listed Factors
Intimidation of Witness/Victim-Listed Factors/F-1 or Murder
Kidnapping
Murder of an Unborn Child - Murder 3
Simple Assault
Drug Offenses

Acquisition of C.S. by Fraud: Sch II Prescr. pills (Narcotic) (51 - 100 pills)
Acquisition of C.S. by Fraud: Sch II Prescr. pills (Narcotic) (> 100 pills)
Acquisition of C.S.by Fraud: Sch II Prescr.pills(Coc/Meth/PCP)(51-100 pills)
Acquisition of C.S. by Fraud: Sch II Prescr. pills (Any Other) (51 - 100 pills)
Acquisition of C.S. by Fraud: Sch II Prescr. pills (Any Other) (> 100 pills)
Acquisition of Controlled Substance by Fraud: Cocaine (100 - 1000 g)
Contraband - Provide controlled substance to inmate
Contraband - Possession of controlled substance by inmate (8/25/97)
Drug Delivery Resulting in Death
Possession With Intent to Deliver: Schedule IV
Possession With Intent to Deliver: Narcotic (10 - < 50 g)
Possession With Intent to Deliver: Narcotic (50 - < 100 g)
Possession With Intent to Deliver: Narcotic (100 - 1000 g)
Possession With Intent to Deliver: Narcotic (> 1000 g)
Possession With Intent to Deliver: Methamphetamine (2.5 - < 10 g)
Possession With Intent to Deliver: Methamphetamine (10 - < 50 g)
Possession With Intent to Deliver: Methamphetamine (50 - < 100 g)
Possession With Intent to Deliver: Methamphetamine (100 - 1000 g)
Possession With Intent to Deliver: Methamphetamine (> 1000 g)
Possession With Intent to Deliver: PCP (2.5 - < 10 g)
Possession With Intent to Deliver: PCP (10 - < 50 g)
Possession With Intent to Deliver: PCP (50 - < 100 g)
Possession With Intent to Deliver: PCP (100 - < 1000 g)
Possession With Intent to Deliver: Cocaine (< 2.5 g)
Possession With Intent to Deliver: Cocaine (2.5 - < 10 g)
Possession With Intent to Deliver: Cocaine (10 - < 50 g)
Possession With Intent to Deliver: Cocaine (50 - < 100 g)
Possession With Intent to Deliver: Cocaine (100 - 1000 g)
Possession With Intent to Deliver: Cocaine (> 1000 g)
Possession With Intent to Deliver: Drug Unknown
Acquisition of Controlled Substance by Fraud: Drug Unknown
Acquisition of Controlled Substance by Fraud: Heroin (50 - < 100 g)
Acquisition of Controlled Substance by Fraud: Heroin (100 - 1000 g)
Acquisition of Controlled Substance by Fraud: Narcotic (10 - < 50 g)
Acquisition of Controlled Substance by Fraud: PCP (100 - 1000 g)
Possession w/ Intent to Deliv.: Marijuana (10 - < 50lbs.)
Possession w/ Intent to Deliv.: Marijuana (21 - <50 plants)
Possession With Intent to Deliver: Heroin (< 1 g)
Possession With Intent to Deliver: Heroin (1 - < 10 g)
Possession With Intent to Deliver: Heroin (10 - < 50 g)
Possession With Intent to Deliver: Heroin (50 - < 100 g)
Possession With Intent to Deliver: Heroin (100 - 1000 g)
Delivery by practitioner: Cocaine (50 - < 100 g)
Delivery by practitioner: Drug Unknown
Delivery by practitioner: Heroin (> 1000 g)
Delivery by practitioner: Narcotic (100 - 1000 g)
Delivery by practitioner: Narcotic (> 1000 g)

Sex Offenses
Statutory Sexual Assault
Involuntary Deviate Sexual Intercourse
IDSI with a Child < 13 yrs.
Incest
Rape
Rape of a Child < 13 yrs.
Sexual Abuse of Children - Taking Photo
Sexual Assault
Aggravated Indecent Assault
Aggravated Indecent Assault of a Child

Property Offenses
Access Device Fraud - Att./obtain $500>
Burglary - Home: No One Present
Catastrophe - Intentionally Causing Failure to Remit Sales Tax
Theft - Unlawful Taking; > $100,000
Theft - Unlawful Taking - During Disaster
Theft - Unlawful Taking; Firearm
Theft - Unlawful Taking; > $2,000 - $25,000/Auto - etc.
Tax Violations
Theft - Deception; Firearm
Firearms; Sale or Transfer - Subsequent Offense
Owning/Operating a Chop Shop

Miscellaneous Offenses
Homicide by Vehicle While DUI
Aggravated Assault by Vehicle While DUI
Criminal Attempt-Unspecified
Corrupt Organizations
Criminal Conspiracy-Unspecified
Firearms; Possessed by Former Convict (eff. 2/14/00)
Firearms-Loaded; Persons Not To Possess, Use, etc.
Firearms-Unloaded; Persons Not To Possess, Use, etc.
Solid Waste: Management of Hazardous Waste
Escape - Other Escapes; this Subsection
Failure to Register (lifetime)
Firearms; Altering I.D.
Failure to Verify (lifetime)
Failure to Verify
False Identification to Law Enforcement Authorities
Homicide by Vehicle (w/DUI Conviction)
Involuntary Manslaughter
Involuntary Manslaughter - victim < 12 yrs.
Appendix D: Prior Offense Coding

Prior Violent Convictions
Murder and inchoate murder
Voluntary manslaughter
Kidnapping
Arson (F-1)
Robbery SBI
Robbery/motor vehicle/SBI
Aggravated assault (SBI)
Burglary (house and person)
Ethnic intimidation
Arson (F-1/no person)
Robbery
Robbery/motor vehicle/no SBI
Prior aggravated assault (att. SBI)
Simple assault
Murder and inchoate murder
Voluntary manslaughter
Kidnapping
Arson (F-1)
Robbery SBI
Robbery/motor vehicle/SBI
Aggravated assault (SBI)
Burglary (house and person)
Intimidation of witness
Assault by life prisoner
Ethnic intimidation
Simple assault (<12 years of age)

Prior Drug Convictions
Drug delivery causing death
Felony drug (>50 g)
Other felony drug
Drug delivery causing death
Felony drug (>50 g)
Other felony drug
Prior Sex Convictions
- Rape
- Involuntary deviant sexual intercourse
- Aggravated indecent assault
- Sexual assault
- Luring child into vehicle
- Indecent assault
- Indecent exposure
- Corruption of minor
- Rape
- Involuntary deviant sexual intercourse
- Aggravated indecent assault
- Incest
- Sexual assault
- Luring child into vehicle
- Indecent assault (<13 age)
- Indecent exposure (<16)
- Corruption of minor
- Unlawful contact with minor

Prior Property Convictions
- Burglary
- Burglary

Prior Miscellaneous Convictions
- Inchoate to 4 point F-1
- Felony-1
- Felony-2
- Felony -3
- Involuntary manslaughter
- Homicide by vehicle
- Possessing an instrument of crime
- Prohibited offensive weapon
- Possession of weapon on school property
- Possession of weapon on court property
- Endangering welfare of children
- Dealing infant children
- Violation of uniform firearms act
M-1 DUI
Weapons of mass destruction
Aggravated jury tampering
Other 4 point offenses
Inchoate to 4 point F-1
Other felony-1
Felony-2
Felony -3
Involuntary manslaughter
Accidents involving death
M-1 involving death
Possessing an instrument of crime
Prohibited offensive weapon
Electronic incapacitation
Possession of weapon on school property
Possession of weapons on court property
Violation of uniform firearms act
M-1 involving weapons
Endangering welfare of children
Dealing infant children
M-1 involving children
Unclassified DUI
M-2 DUI
M-1 DUI
Appendix E: Correlation Matrix for Covariance in Risk Assessment Analysis

Full Sample (N=7,935)

|          | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1. Race  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 2. Male  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 3. Age   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 4. White |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 5. Hisp |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 6. Age   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 7. Age   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 8. Age 4+|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 9. All     |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 10. Phila |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 11. Urban |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 12. Rural |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 13. MHC  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 14. Rent  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 15. Arrest |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 16. Arrest |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 17. Sex   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 18. Prior |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 19. Prior |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 20. Drug  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 21. Prop  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 22. Other |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 23. Spec  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 24. V    |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 25. Org   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 26. Prop  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 27. V    |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 28. Sex  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 29. V    |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 30. Other |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 31. Multi |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 32. V    |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 33. Inc  |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

179
### Appendix F: Pennsylvania Sentencing Matrix - 2015

#### §303.16(a). Basic Sentencing Matrix.


<table>
<thead>
<tr>
<th>Level</th>
<th>OGS</th>
<th>Example Offenses</th>
<th>Prior Record Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>LEVEL</td>
<td>5</td>
<td>State Incar</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Murder 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inchoate Murder (SBI)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rape (victim &lt;13 yrs)</td>
<td></td>
</tr>
<tr>
<td>LEVEL</td>
<td>11</td>
<td>State Incar/ RIP trade</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rape-Forcible Compulsion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDSI-Forcible Compulsion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robbery-Inflicts SBI</td>
<td></td>
</tr>
<tr>
<td>LEVEL</td>
<td>10</td>
<td>State Incar</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kidnapping</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agg Assault-Cause SBI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Voluntary Manslaughter</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sexual Assault</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PWID Cocaine (100-1,000 g)</td>
<td></td>
</tr>
<tr>
<td>LEVEL</td>
<td>9</td>
<td>State Incar/ RIP trade</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sexual Exploitation of Children</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robbery-Commit/Threat F1/F2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burglary/Home/Person Present</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arson-No Person in Building</td>
<td></td>
</tr>
<tr>
<td>LEVEL</td>
<td>8</td>
<td>State Incar/ RIP trade</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Agg Assault - Cause BI w/DW Theft (Firearm)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identity theft (3rd+/ &amp; Vic&gt;=60 yrs) Hom by Veh-DUI or Work Zone Theft (&gt;100,000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PWID Cocaine (10-&lt;50 g)</td>
<td></td>
</tr>
<tr>
<td>LEVEL</td>
<td>7</td>
<td>State Incar/ RIP trade</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Statutory Sexual Assault Theft (&gt;100,000-50,000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identity Theft (3rd/subq) PWID Cocaine (5-&lt;10 g)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Designated areas of the matrix indicate restrictive intermediate punishments may be imposed as a substitute for incarceration.

2. When restrictive intermediate punishments are appropriate, the duration of the restrictive intermediate punishment programs are recommended not to exceed the guideline ranges.

3. When the range is RS through a number of months (e.g. RS-6), RIP may be appropriate.

4. All numbers in sentence recommendations suggest months of minimum confinement pursuant to 42 Pa.C.S. 9755(b) and 9756(b).

5. Statutory classification (e.g., F1, F2, etc.) in brackets reflect the omnibus OGS assignment for the given grade.

Key:

- **RIP** = restrictive intermediate
- **RS** = punishments restorative sanctions
- **SBI** = serious bodily injury statutory limit
- **SL** = (longest minimum sentence) no
- **~** = recommendation (aggravated
  - sentence would exceed statutory
  - limit) less than; greater than

<table>
<thead>
<tr>
<th>Cnty Incar RIP</th>
<th>RS 1 (M3)</th>
<th>Most Misd. 3's; Theft (&lt;$50)</th>
<th>RS-1</th>
<th>RS-2</th>
<th>RS-3</th>
<th>RS-4</th>
<th>RS-6</th>
<th>3-6</th>
<th>NA</th>
<th>+/- 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEVEL 1 RS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theft ($50-$&lt;200)</td>
<td>RS</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>1-9</td>
<td>6-&lt;12</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Retail Theft (1st/2nd Offense)</td>
<td>Bad Checks ($500-$&lt;1,000)</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>1-9</td>
<td>6-&lt;12</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theft ($&lt;50)</td>
<td>RS</td>
<td>RS-1</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>3-6</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEVEL 2 Cnty Incar RIP RS</td>
<td>2 (M2)</td>
<td>Theft ($50-$&lt;200)</td>
<td>RS</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>1-9</td>
<td>6-&lt;12</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Retail Theft (1st/2nd Offense)</td>
<td>Bad Checks ($500-$&lt;1,000)</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>1-9</td>
<td>6-&lt;12</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theft ($&lt;50)</td>
<td>RS</td>
<td>RS-1</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>3-6</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEVEL 3 Cnty Incar RIP RS</td>
<td>3 (M1)</td>
<td>Simple Assault-Attempt/Cause BI</td>
<td>RS-1</td>
<td>RS-6</td>
<td>RS-9</td>
<td>RS-&lt;12</td>
<td>3-14 BC</td>
<td>6-16 BC</td>
<td>9-16 BC</td>
<td>12-18 BC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theft ($200-$2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carrying Explosives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Simple Possession</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEVEL 2 Cnty Incar RIP RS</td>
<td>4</td>
<td>Indecent Assault M2</td>
<td>RS-3</td>
<td>RS-9</td>
<td>RS-&lt;12</td>
<td>3-14 BC</td>
<td>6-16 BC</td>
<td>9-16 BC</td>
<td>21-30 BC</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forgery (Money, Stocks)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weapon on School Property</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEVEL 2 Cnty Incar RIP RS</td>
<td>5 (F3)</td>
<td>Burglary F2</td>
<td>RS-9</td>
<td>1-12 BC</td>
<td>3-14 BC</td>
<td>6-16 BC</td>
<td>9-16 BC</td>
<td>12-18 BC</td>
<td>24-36 BC</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Theft (&gt;25,000-$25,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bribery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PWID Marij (1&lt;10 lbs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Homicide by Vehicle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Burglary-Not a Home/Person Prsnt Theft (&gt;25,000-$50,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Endanger Property</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PWID Cocaine (2&lt;5 g)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

181
BC = boot camp
CNTY = county
INCAR = incarceration
PWID = possession with intent to deliver
REVOC = repeat violent offender category
RFEL = repeat felony 1 and felony 2 offender category
### Appendix G: Sentencing Matrix - 1997

#### §303.16. Basic Sentencing Matrix

**Edition (6/13/97)**

<table>
<thead>
<tr>
<th>Level</th>
<th>OGS</th>
<th>Example Offenses</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>RFEL</th>
<th>REVOC</th>
<th>AGG/MIT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LEVEL 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>Inchoate Murder/no SBi Drug Del. Result in Death PWID Cocaine, etc. (&gt;1,000 gms)</td>
<td>60-78</td>
<td>66-84</td>
<td>72-90</td>
<td>78-96</td>
<td>84-102</td>
<td>96-114</td>
<td>108-126</td>
<td>240</td>
<td>+/- 12</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>Rape (DSI) Robbery (SBi) Robbery/car (SBi)</td>
<td>48-66</td>
<td>54-72</td>
<td>60-78</td>
<td>66-84</td>
<td>72-90</td>
<td>84-102</td>
<td>96-114</td>
<td>120</td>
<td>+/- 12</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>Agg Asslt (SBi) Voluntary Manslaughter Sexual Assault PWID Cocaine, etc. (&lt;100-1,000 gms)</td>
<td>36-54 BC</td>
<td>42-60</td>
<td>48-66</td>
<td>54-72</td>
<td>60-78</td>
<td>72-90</td>
<td>84-102</td>
<td>120</td>
<td>+/- 12</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Kidnapping Arson (person inside) Agg Asslt (att. SBi) Robbery (threat. SBi) Agg. Indecent. Asslt Causing Catastrophe (F1) PWID Cocaine, etc. (50-&lt;100 gms)</td>
<td>22-36 BC</td>
<td>30-42 BC</td>
<td>36-48 BC</td>
<td>42-54</td>
<td>48-60</td>
<td>60-72</td>
<td>72-84</td>
<td>120</td>
<td>+/- 12</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>Robbery/car (no SBi) Robbery (F1/F2) Burglary (home/person) Arson (no person)</td>
<td>12-24 BC</td>
<td>18-30 BC</td>
<td>24-36 BC</td>
<td>30-42 BC</td>
<td>36-48 BC</td>
<td>48-60</td>
<td>60-72</td>
<td>120</td>
<td>+/- 12</td>
</tr>
<tr>
<td><strong>LEVEL 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LEVEL 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Robbery (inflcts/threatens BI) Burglary (home/ no person) Statutory Sexual Assault Theft (&gt;=$50,000&lt;$100,000) Sexual Abuse/Child (take photo) PWID Cocaine, etc. (5.5-&lt;10 gms)</td>
<td>6-14 BC</td>
<td>9-16 BC</td>
<td>12-18 BC</td>
<td>15-21 BC</td>
<td>18-24 BC</td>
<td>24-30 BC</td>
<td>35-45 BC</td>
<td>NA</td>
<td>+/- 6</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Invol. Mansl. (when no DUI) Hom. by Vehicle (when no DUI) Burglary (not home/person) Theft (&gt;=$25,000-$50,000) Arson (property) PWID Cocaine, etc. (&lt;2.5 gms)</td>
<td>3-12 BC</td>
<td>6-14 BC</td>
<td>9-16 BC</td>
<td>12-18 BC</td>
<td>15-21 BC</td>
<td>21-27 BC</td>
<td>27-40 BC</td>
<td>NA</td>
<td>+/- 6</td>
</tr>
<tr>
<td><strong>LEVEL 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Burglary (not home/no person) Corruption of Minors Robbery (prop by force) Firearms (loaded) Theft (&gt;=$2000-$25,000) PWID (1-&lt;10 lb of mari)</td>
<td>RS-9</td>
<td>1-12 BC</td>
<td>3-14 BC</td>
<td>6-16 BC</td>
<td>9-16 BC</td>
<td>12-18 BC</td>
<td>24-36 BC</td>
<td>NA</td>
<td>+/- 3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Indecent assault Forgery (money, stock, etc.) Firearms (unloaded) Crim Trespass (breaks in)</td>
<td>RS-3</td>
<td>RS-9</td>
<td>RS&lt;12</td>
<td>3-14 BC</td>
<td>6-16 BC</td>
<td>9-16 BC</td>
<td>21-30 BC</td>
<td>NA</td>
<td>+/- 3</td>
</tr>
<tr>
<td><strong>LEVEL 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Theft ($50-$200) Retail Theft (1st, 2nd ) DUI (M2)</td>
<td>RS</td>
<td>RS-2</td>
<td>RS-3</td>
<td>RS-4</td>
<td>RS-6</td>
<td>1-9</td>
<td>6-&lt;12 BC</td>
<td>NA</td>
<td>+/- 3</td>
</tr>
</tbody>
</table>

Note: RS = References for Sentencing.
<table>
<thead>
<tr>
<th>1</th>
<th>RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS-1</td>
<td>RS-2</td>
</tr>
</tbody>
</table>

**Key:**
- Level 1 = Purple
- Level 2 = White
- Level 3 = Blue
- Level 4 = Yellow
- Level 5 = Green
- AGG/MIT = Tan

1. Yellow (Level 4) and Blue (Level 3) shaded areas of the matrix indicate restrictive intermediate punishments may be imposed as a substitute for incarceration.

2. When restrictive intermediate punishments are appropriate, the duration of the restrictive intermediate punishment program shall not exceed the guideline ranges.

3. When the range is RS through a number of months (e.g. RS-6), RIP may be appropriate.

4. All numbers in sentence recommendations suggest months of minimum confinement pursuant to 42 PA.C.S. §9755(b) and §9756(b).

5. Statutory grades in brackets correspond with the omnibus OGS for the grade.

- BC = boot camp
- PWID = possession with intent to deliver
- CNTY = county
- RIP = restrictive intermediate punishments
- INCAR = incarceration
- RS = restorative sanctions
- Italics = Three Strikes Offense
- RFEL = repeat felony 1 and felony 2 offender category
- < ; > = less than;greater than
- REVOC = repeat violent offender category

---

184
Appendix H: Sample Distribution by Risk Level (Black/White)

Sample distribution by risk level (Black/White sample) N=7,022

<table>
<thead>
<tr>
<th>N Distribution</th>
<th>All</th>
<th>White (%)</th>
<th>Black (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (0-3)</td>
<td>1,315</td>
<td>889 (68%)</td>
<td>426 (32%)</td>
</tr>
<tr>
<td>Medium (4-5)</td>
<td>2,193</td>
<td>1,187 (54%)</td>
<td>1,006 (46%)</td>
</tr>
<tr>
<td>High (6-7)</td>
<td>2,580</td>
<td>1,169 (45%)</td>
<td>1,411 (55%)</td>
</tr>
<tr>
<td>Very High (8-10)</td>
<td>934</td>
<td>372 (40%)</td>
<td>562 (60%)</td>
</tr>
</tbody>
</table>
Julia A. Laskorunsky
Condensed Vita

EDUCATION

August 2017 (exp.)  Ph.D., Criminology, The Pennsylvania State University

2013  M.A., Criminology, The Pennsylvania State University

2007  B.A., Sociology (cum laude), University of California, Berkeley

TEACHING

2016  Introduction to Criminal Justice – Online (Instructor of Record)
2015  American Corrections (Instructor of Record)
2014  Introduction to Criminal Justice (Instructor of Record)

REFEREED JOURNAL ARTICLES


BOOK CHAPTERS


PROJECT REPORTS
