ABSTRACT

Title of Dissertation: THE IMPACT OF RETURNING TECHNICAL PAROLE VIOLATORS TO PRISON: A DETERRENT, NULL, OR CRIMINOGENIC EFFECT

Kristofer Bret Bucklen, Doctor of Philosophy, 2014

Dissertation directed by: Professor Raymond Paternoster, Department of Criminology and Criminal Justice

As a result of the significant U.S. prison population build-up over the past several decades, a large number of inmates are now being released from prison and returned to the community. One mechanism for facilitating this transition to the community is for inmates to be conditionally released under parole supervision. Once on parole, a parolee is subject to certain rules and conditions that, if violated, can result in a return to prison, even if not a criminal act. These types of non-criminal parole violations are typically referred to as Technical Parole Violations (TPVs). Many states return a large number of TPVs to prison each year, and TPVs contribute significantly to the prison population in many states. However, there is virtually no existing research examining what impact returning TPVs to imprisonment has on their subsequent rates of re-offending. While a large body of literature examining the overall impact of incarceration on recidivism has mostly concluded that imprisonment has a null or even slightly criminogenic effect, this overall finding is not necessarily generalizable to all sub-populations within the prison

population. Strong theoretical cases can be made each way, for the impact on recidivism of incarcerating TPVs.

This dissertation examines the impact on recidivism of sanctioning TPVs to imprisonment versus an alternative sanction, and also examines the dose-response impact on recidivism of varying lengths of stay in prison for a TPV, using a large sample of TPVs in one state (Pennsylvania). The bulk of the evidence supports the conclusion that recidivism rates are mostly lowered by using incarceration in response to first TPV violations. However, the evidence also suggests that the specific mechanism for lowering recidivism rates among incarcerated TPVs is largely attributable to aging and exposure time rather than to deterrence.

The findings on the dose-response impact of differential lengths of stay in prison for TPVs who are sanctioned to imprisonment are more mixed. Generally the evidence suggests somewhat lowered recidivism rates attributable to longer lengths of stay in prison for a TPV violation, yet the effect sizes are generally smaller and in some cases statistically insignificant. It again appears that the particular mechanism for reduced recidivism rates associated with longer lengths of stay in prison is associated with aging and exposure time rather than with traditionally formulated deterrence mechanisms.

A few contingencies of these findings are noted. First, the effect of imprisonment on recidivism among TPVs is likely highly contingent upon the swiftness, certainty, and perceived fairness of sanctioning, yet measures of these factors were not available for this study. Second, this dissertation only focuses on the first TPV violation instance after release from prison, and also is mostly limited to higher risk TPVs. Third, lower overall recidivism rates for TPVs sanctioned to imprisonment, and sanctioned for longer periods of time in prison, were influenced heavily by lower re-incarceration rates, whereas re-arrest rates did not significantly differ in any of the models. Since re-incarceration rates not only include new criminal activity but also new technical violations, it is unclear whether imprisonment for a first TPV reduces serious criminal behavior or rather mostly reduces additional technical violations and minor crimes. Future research must address these contingencies.

THE IMPACT OF RETURNING TECHNICAL PAROLE VIOLATORS TO PRISON: A DETERRENT, NULL, OR CRIMINOGENIC EFFECT

By

Kristofer Bret Bucklen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:

Professor Raymond Paternoster, Chair
Professor Brian Johnson
Professor Thomas Loughran
Professor Kiminori Nakamura
Professor Peter Reuter

ACKNOWLEDGEMENTS

It is difficult to know where to begin in offering my thanks to the many people who supported me, provided advice, and helped me along the way on this long journey. With a full-time career, a family to take care of, and a dissertation to write, I often felt like the best I could hope for was to succeed at two out of three at any one time. I certainly could not have made it to this point without the help of many people.

I want to first thank those who inspired me early on, and served as early mentors in my academic journey. During my master's degree studies at Carnegie Mellon, I was fortunate enough to work as a graduate assistant for Al Blumstein, who has been a continual mentor and an inspiration. I could not ask for a better mentor. Early on in my doctoral studies at University of Maryland, I was greatly inspired by John Laub, who served as my first advisor and dissertation chair. John provided much guidance and advice to get me started in this process.

I owe a special debt of gratitude to Ray Paternoster, who took over as my dissertation chair after John left for NIJ. Thanks Ray for your patience in sticking with me for the long haul, giving me just the push that I needed at just the right time, and for always providing encouragement and helpful advice. I am also grateful to Kiminori Nakamura, Tom Loughran, Brian Johnson, and Peter Reuter for taking the time to serve on my committee and provide thoughtful feedback on my dissertation. Thanks especially to Tom for the technical assistance on my propensity score models. And an extra special thanks to Kiminori for the many, many hours of conversation and discussion on my dissertation and on so many other topics. Kiminori invested significantly in me, and I am incredibly indebted to him for all of his advice and help.

Too many people to mention served as intellectual inspiration on the topic of my dissertation and on the related policy issues, but one person I'd particularly like to thank is Mark Kleiman. Mark sparked much of my thinking and intellectual curiosity on issues surrounding community corrections and probation/parole supervision. I appreciate the many email exchanges and dialogue with him on these topics.

Thanks goes to my friends and colleagues at the Pennsylvania Department of Corrections, especially to past and present staff in the Bureau of Planning, Research, and Statistics. Again, there are too many people to mention by name. To name just a few, thanks to Gary Zajac, Kathy Gnall, Jim Schaefer, Nikki Bell, and Bob Flaherty. It has been an immense pleasure to work with the most dedicated group of people in state government, and to lead the smartest team of researchers across any state correctional agency. Thanks also to current Secretary of Corrections John Wetzel for his leadership and inspiration, and to former Secretary of Corrections Jeff Beard for investing in my career early on. I owe a debt of gratitude to both Secretary Wetzel and Secretary Beard for making a significant investment in research at the Pennsylvania Department of Corrections, and for fostering an agency culture that values using research to guide our business. Both Secretary Wetzel and Secretary Beard supported me in my academic pursuit too.

I want to also thank the Pennsylvania Board of Probation and Parole, and particularly Fred Klunk and his staff, for working with me to provide the necessary data for my dissertation. Fred and his staff not only provided data, but also worked with me to understand it.

None of this would have been possible without an incredible support system of family and friends who came alongside me, helped me to stay sane, and helped me to keep my life together throughout this journey. I owe a special debt of gratitude to my mom and dad for supporting me in so many different ways. Thanks also to all of my extended family, including all of my brothers, sisters, and in-laws. From babysitting, to financial support, to emotional support, I am extraordinarily blessed to have you as my family. Thanks also to my New Covenant Fellowship church family who provided the same extended support network.

An extra special acknowledgment goes to the loves of my life, my three precious daughters, Kayla, Ashley, and Emma, who stuck with daddy throughout this long process. I love you so much. I'm sorry that you had to share me on so many occasions with my work. Here's to looking forward to more time together. You make daddy so very proud.

Finally, above all, I want to thank God. All credit and praise truly goes to Him. Soli Deo Gloria

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

TECHNICAL PAROLE VIOLATORS IN PRISON

Over 600,000 inmates are released from U.S. prisons each year, of which approximately 80% are conditionally released under some sort of parole supervision. This annual number of prison releases is more than quadruple the number of annual prison releases just 25 years ago, spawning immense criminal justice interest in a policy area commonly referred to as "prisoner reentry." One of the primary concerns of policymakers and criminologists studying the impact of prisoner reentry is the threat to public safety and the contribution to community crime rates posed by increasing numbers of returning prisoners (Hipp & Yates, 2009; Petersilia, 2003; Rosenfeld, Wallman, & Fornango, 2005). Parole supervision is purported to serve as an important mechanism for protecting the public from the potential threat posed by returning inmates (Piehl & LoBuglio, 2005). Specifically, parolees are given conditions of supervision by which they are expected to abide. Violations of these technical conditions of parole supervision are considered in many cases to be precursors to impending criminal behavior. Largely at the discretion of the parole agent (but also sometimes based on other factors such as agency policy and formal risk assessment guidelines), a parolee may be remanded to prison on a technical parole violation if it is deemed that the violation is an indication of imminent criminal behavior. The use of prison as a policy response to parole violations is reflected in the fact that the number of parole violator admissions to prison has increased seven-fold over the past two decades, currently representing over one-third of all state prison admissions (Travis & Lawrence, 2002). Indeed, research on correctional trends has recently pointed to parole violators as an increasing contribution to state prison populations (Blumstein & Beck, 2005).

However, states vary widely in their use of prison for parole violators (see Figure 1.1). For example states such as Florida, North Carolina, Virginia, Idaho, and Massachusetts admit less than 10% of prisoners as parole violators, whereas in California and Montana approximately 67% of all prison admissions are parole violators (Bureau of Justice Statistics, 2008). Pennsylvania is among the group of states with a relatively high proportion and number of parole violator prison admissions, with approximately one-third of its nearly 18,000 annual prison admissions being for parole violations. This places Pennsylvania in the top half of all states in terms of parole violators as a fraction of total state prison admissions. While the monthly rate of parole violators remanded to prison per supervised parole population has actually remained relatively stable in Pennsylvania over the past decade at approximately 1.7%, an increase of approximately 58% in the number of prison releases to parole supervision over the past decade means that this relatively stable rate has nonetheless translated into increasing numbers of parole violators being returned to prison. This increasing number of parole violator admissions to prison comes at a significant cost to Pennsylvania taxpayers. According to a recent Pew Report (2009), it costs Pennsylvania approximately $98 per day to incarcerate one offender in prison, whereas the cost of supervising that same offender for one day in the community under parole supervision is approximately $8 (Pew Center on the States, 2009). Thus, Pennsylvania can afford to keep a parole violator under parole supervision for approximately 12 days at the same cost as only one day in prison.

A subset of parole violators who are of particular interest from a policy perspective are those who are labeled technical parole violators (TPVs). What makes this group interesting is that a technical parole violation does not require the occurrence of a new crime. A technical parole violation typically occurs when a stipulated condition of parole supervision has been violated. Examples of technical violations can include failing to refrain from alcohol use, failing to report routinely to a parole agent, failing to obtain employment, failing to comply with required treatment programming, or changing an address without providing notification to the supervising parole authority. While not necessarily a criminal violation, a technical violation is nevertheless often considered to be a precursor to relapse into criminal behavior. As a result, parole agents are usually granted a certain degree of discretion in deciding to return a TPV to prison for a period of incarceration if criminal behavior is considered imminent.

This discretion to re-incarcerate a technical violator is viewed by parole authorities as an important mechanism for protecting the public. However, this also places a sub-population of inmates in prison who have not necessarily committed a new crime at all. It should be noted that sometimes technical parole violations do in fact include underlying criminal behavior but are processed as a technical violation instead of being prosecuted as a new crime in cases where evidentiary or procedural problems are expected to result in a non-conviction if prosecuted by the court (e.g., a likely plea agreement, etc.).[1] Due to a lack of research in this area, it remains unclear as to the degree of overlap between criminal behavior and technical parole violations. Further, the

[1] This would seem to raise an issue of fairness/legitimacy, since a lack of good evidence might otherwise normally preclude a period of incarceration. Perceived fairness/legitimacy of the technical violation sanctioning process may condition the impact of imprisonment.

claim that the use of prison for sanctioning TPVs prevents or deters criminal behavior remains largely untested.

PAROLE SUPERVISION IN PENNSYLVANIA

In Pennsylvania, approximately three-fifths of all parole violator prison admissions are for technical parole violations. The number of TPV prison admissions has in fact increased by approximately 41% over the past 10 years. The Pennsylvania Board of Probation and Parole (PBPP) stipulates six standard supervision conditions for all parolees, as well as any number of special supervision conditions for certain parolees, all of which can result in a technical parole violation if broken. The six standard conditions are:

1. Report in person or in writing within 48 hours of release from prison to the district office or sub-office and do not leave that district without prior written permission of the parole supervision staff.
2. Do not change residence from the approved residence without the written permission of the parole supervision staff.
3. Maintain regular contact with the parole supervision staff by: a) reporting regularly as instructed and following any written instructions of the Board or parole supervision staff, b) notifying the parole supervision staff within 72 hours of an arrest or a receipt of a summons/citation for an offense punishable by imprisonment upon conviction, and c) notifying the parole supervision staff within 72 hours of any change in status, including, but not limited to, employment, on-the-job training, and education.

4. Comply with all municipal, county, state and Federal criminal laws, as well as the provisions of the Vehicle Code and the Liquor Code.
5. Abstain from a) the unlawful possession or sale of narcotics and dangerous drugs or any drugs without a valid prescription, b) owning or possessing any firearms or other weapons, and c) any assaultive behavior.
6. Pay fines, costs, and restitution imposed by the sentencing court, which includes establishing within thirty days of release from prison a payment schedule for the fines, costs and restitution owed.

Special parole conditions may include: following treatment referrals, submitting to urinalysis testing, refraining from alcohol consumption, taking prescribed psychotropic medication, refraining from contacting or associating with specific individuals, paying a supervision fee of at least $25 per month, refraining from entering certain establishments such as bars or those that sell or dispense alcohol, providing dependent support, maintaining employment or educational/vocational training, and/or engaging in an active job search during a period of unemployment.

Technical parole violators sent to prison in Pennsylvania serve an average of 14 months (median of 12 months) in prison. According to PBPP's most recent statistics (Alibrio & Findley, 2005), the single most prevalent category of technical violations among those returned to prison is for an unapproved change of residence (20%). The broader category of drug and alcohol substance use violations (which is covered by multiple standard/special conditions) accounts for approximately 24% of all TPVs returned to prison. Other prevalent reasons for technical violations resulting in a prison term include failure to report for supervision or to report arrests (16%) and failure to comply with the rules of a residential halfway house (17%). The average number of conditions violated for those returned to prison is 2.1 per case. At least some preliminary research in Pennsylvania suggests that approximately 40% of technical parole violator cases returned to prison also involve criminal behavior (Kramer, Silver, Van Eseltine, Ortega, & Rutkowski, 2008; Bucklen, 2005).

One other important note about the parole violation process in Pennsylvania is that not all technical parole violators are returned to prison. Indeed, there are a number of alternative or diversionary options in the community that are available for sanctioning TPVs. These options include written warnings, travel restrictions, increased reporting requirements or urinalysis testing, imposed curfews, placement in in-patient or outpatient treatment, placement in a Day Reporting Center, imposition of electronic monitoring, placement in a Community Corrections Center (CCC) under a “halfway back” status, or placement in a secure violation center. While PBPP maintains a Violation Sanctioning Grid for guiding violation sanctioning decisions and generally views sanctioning along a continuum of seriousness, significant discretion is still granted to the parole agent for determining the ultimate sanctioning response for any given violation.

THEORETICAL FRAMEWORK

From a theoretical standpoint, there is a case to be made that sending TPVs to prison can serve: a) a deterrent effect by preventing future criminal behavior, b) a null effect by simply removing the violator from the community for a “time out” period, or c) a criminogenic effect by actually increasing the probability of future criminal behavior. From a deterrence point of view, the “broken windows” perspective (Wilson & Kelling,

1982) is particularly germane. Specifically, “broken windows” policing purports that serious crime is best prevented by devoting criminal justice resources to addressing minor crimes or community disorder. Applying this perspective to TPVs, it may be reasonable to expect a drop in serious criminal behavior by implementing a “zero tolerance” policy response to minor parole violations through the use of prison, with the expectation that minor technical violations are essentially equivalent to the type of “broken windows” disorder that can lead to more serious crime if left unaddressed.

Deterrence theory also purports that punishment must be certain, severe, and swift in order to be effective (Beccaria, 1764). More recent deterrence research has found that the certainty and swiftness of punishment appear to be more important factors than the severity of punishment (Kleiman, 2009). One might argue that a broad policy of returning TPVs to prison increases the certainty and swiftness of punishment in comparison to the certainty and swiftness of punishment typically delivered through the court system in traditional criminal cases. Thus, any deterrent impact of prison may be more effective for TPVs than for new court commitments.

However, those who would hypothesize a null effect from sending parole violators to prison will likely point to the general conclusion from recent reviews of the literature that the overall deterrent effect of prison is modest at best (Nagin, Cullen, & Jonson, 2009). More evidence appears to support an overall incapacitative effect of prison, where inmates are simply restricted from criminal behavior during their time of incarceration but largely return to their same level of criminal offending after prison. Best estimates to date suggest that a 10% increase in the prison population can lead to approximately a 4% decrease in the crime rate primarily through the incapacitation of additional criminals (Stemen, 2007). Bhati and Piquero (2008) examined the degree to which incarceration has a deterrent, criminogenic, or null effect on subsequent criminal offending, and concluded that for the largest proportion of prison releases incarceration has a null effect (they observed a null effect on subsequent criminal offending for 56% of the prison releases in their sample). Sending TPVs to prison thus may only temporarily delay any impending criminal behavior.

Recent research suggests that the benefits of prison are facing diminishing marginal returns due to expansive prison build-up over the past several decades (Liedka, Piehl, & Useem, 2006). It may thus be the case that the system has reached a "tipping point" and that sending parole violators to prison no longer has a deterrent or null effect (if it ever did), but instead may actually be criminogenic. From a labeling theory perspective, incarceration is indeed expected to lead to an increase in future criminal behavior through what is referred to as secondary deviance (Lemert, 1951). Sampson and Laub (1997), in their classic longitudinal follow-up of a sample of Boston boys who came of age during the 1950s, found that those boys who served time in prison were generally at an increased risk of future criminal behavior through a process they refer to as cumulative disadvantage (Sampson & Laub, 1997). Under this process of cumulative disadvantage, the adverse effect of prison is indirect, leading to disruption of social ties such as employment and family, which in turn leads to an increased risk of criminal activity. We do know that the recidivism rates for parole violators who are returned to prison and subsequently re-released are significantly higher than the recidivism rates for first time parole releases (Blumstein & Beck, 2005), which would on the surface suggest a criminogenic effect of returning parole violators to prison. Yet it is unclear as to whether this simply represents a selection effect, where those who are returned to prison on a parole violation were already at an increased risk of subsequent re-offending due to a higher criminal propensity as evidenced by their parole violation. However, research by Petersilia and Turner (1993) found little evidence that technical parole violations are proxies for criminal behavior, which suggests that a selection effect might not necessarily fully explain higher recidivism rates among TPVs. Thus, there is theoretical ground for expecting that a policy of returning TPVs to prison may lead to a criminogenic effect, actually increasing future criminal behavior.

CONCLUSION

The following dissertation will examine whether and to what extent a policy of returning technical parole violators to prison will reduce criminal behavior. More specifically, I will examine the use of prison for technical parole violators in Pennsylvania and attempt to disaggregate the degree to which prison serves a deterrent, null, or criminogenic effect.

An outline for this dissertation is as follows: Chapter 2 will review the literature on parole supervision practices and parole violation sanctioning as well as the general theoretical and empirical literature on the deterrent, null, and criminogenic effects of prison. Chapter 3 will review the data and methods utilized. This dissertation will investigate two different aspects of the impact of incarceration among TPVs: 1) the impact of the decision to incarcerate compared to delivering an alternative sanction for technical violations, and 2) the dose-response impact of length of stay in prison contingent upon being returned to prison for a technical violation. Datasets from the

Pennsylvania Department of Corrections (PA DOC), the Pennsylvania Board of Probation and Parole (PBPP), and the Pennsylvania State Police (PSP) are utilized. I make use of propensity score matching methods to examine both the impact of the decision to incarcerate a TPV and the dose-response impact of length of stay in prison for TPVs who are incarcerated.

Chapter 4 reports on results from models examining the relationship between the decision to incarcerate and subsequent recidivism among TPVs. Specifically, propensity score matching is used to compare a group of TPVs who are returned to prison for their first recorded violation sanction while on parole supervision to a group of TPVs who are not returned to prison for their first recorded sanction but instead receive an alternative community-based sanction. The sample will consist of 12,705 parolees who were initially released from prison onto parole between October 2006 and December 2009, and were either returned to prison or sanctioned to a community alternative for a TPV. Six-month, one-year, and three-year recidivism rates will be examined and compared for the two groups. Covariates in these models will be used to examine sub-populations of TPVs for which prison is more or less effective.

Chapter 5 will examine the issue of the dose-response impact of varying lengths of incarceration for technical violations on subsequent offending. Specifically, this analysis is limited to the 1,758 TPVs from the full sample for whom their first recorded post-release sanction for a technical violation was imprisonment. Again using propensity score techniques, varying lengths of incarceration among TPVs will be examined in order to estimate whether length of stay in prison for a technical violation demonstrates a positive, negative or null relationship to subsequent offending. Covariates in these models will be used to examine varying dose-response impacts among sub-populations of TPVs who are sent to prison.

Lastly, Chapter 6 is a concluding chapter, which summarizes the study findings and outlines specific policy implications as well as a future direction for research.

CHAPTER 2: THE IMPACT OF IMPRISONMENT

INTRODUCTION

Research on the impact of prison on individual criminal behavior is growing, but significantly limited in its ability to lead to firm policy conclusions. For a variety of reasons outlined in this chapter, estimates of the impact of imprisonment on reoffending are fragile. The extant research is even more limited in its ability to disaggregate the impact of imprisonment on subsequent criminal behavior among various subgroups of offenders. For example, only one known study specifically examines the impact of prison on subsequent criminal behavior among parole violators returned to prison. This is striking, given that parole violators comprise a relatively large percentage of state prison admissions nationwide. The following chapter will proceed by examining the general theoretical and empirical literature on the effects of imprisonment on criminal behavior and then conclude with a review of the theoretical and empirical literature relevant to parole violator sanctioning and the effectiveness of parole supervision.

Theories on the impact of imprisonment on subsequent criminal behavior can generally be divided into one of three categories: 1) prison as a specific deterrent, 2) prison as criminogenic, or 3) prison as a null impact.

PRISON AS DETERRENT

The earliest school of thought, prison as a specific deterrent, purports that prison discourages offenders from future criminal behavior by imposing a significant cost of offending. It should be noted at the outset that criminologists have long distinguished between general deterrence and specific deterrence. General deterrence considers the broad impact of the threat of punishment on would-be criminals within society at large, whereas specific deterrence is concerned with the impact of punishment on the individual who is actually punished or threatened with punishment. I focus on specific deterrence here, since I am most concerned with the impact of imprisonment on individual trajectories among known criminals rather than the impact of imprisonment on society at large.

Deterrence theory originated from the writings of 18th century political philosophers, most notably that of Cesare Beccaria and Jeremy Bentham, who proposed that punishment must be certain, severe, and swift in order to be effective in preventing criminal behavior (Beccaria, 1764; Bentham, 1781). The mechanics behind deterrence, as outlined by these early writers, were based on a rational actor view of human behavior in which humans weigh costs and benefits and make choices favoring actions in which benefits are perceived to outweigh costs. This early work in deterrence theory came to be known as the "classical school" of criminology.

Deterrence thought dwindled by around the mid-19th century and did not resurface until the late 1960s among economists and a select group of criminologists who sought to more fully elucidate the mechanisms behind how punishment might deter criminal behavior. Most notably among this group of scholars was the seminal work of Gary Becker (1968). Becker, an economist, argued that an expected utility model (or economic model of choice) was best suited to explain criminal behavior. The key elements of this model were that: 1) offenders hold expected rewards from alternate courses of legal and illegal actions, 2) offenders also hold expected costs for these actions, 3) the expectations of rewards and costs are subjective, and 4) if the subjectively perceived expected utility (i.e., rewards minus costs) for a criminal act is greater that the subjectively perceived expected utility for a non-criminal act then the individual will engage in the criminal act. Becker's work (1968) opened up the possibility that deterrence was not only realized through the formal properties of incarceration, but also through indirect effects such as psychological effects, loss of income while incarcerated, and the stigmatization of serving time in prison. Becker's work (1968) also made room for the fact that humans are not perfectly rational actors but that they have what Herb Simon (1957) referred to as a "bounded rationality" (or differential utility function). Beccaria (1764) had hinted around this potential for differential motivation but did very little to describe how it might operate. These contingencies of the deterrent impact of punishment formed what became known as rational choice theory, which is really an extension of classical school deterrence.

The first theoretical contingency of the deterrent effect of incarceration is the degree of importance and the relative presence of the three traditional components of deterrence: certainty, severity, and celerity. Most of the work to date has focused primarily on the certainty and severity of punishment, with evidence suggesting that the certainty of punishment serves more of a deterrent impact than the severity of punishment (Nagin, 1998; Paternoster, 1987; Pratt, Cullen, Blevins, Daigle, & Madensen, 2006). Further, in an early paper by Tittle and Rowe (1974), the authors hypothesized that there is a "tipping point" threshold for the certainty of punishment, with little deterrent impact when the probability of punishment falls below a "tipping point" threshold but a significant deterrent effect when the probability of punishment crosses above this threshold. Thus it may be that prison will serve little deterrent impact until the probability of arrest, conviction, and incarceration can be raised above a threshold of certainty. Similarly, Geerken and Gove (1977) proposed that the criminal justice system can reach a point of "overload" in which crime rates are too high, thus making it virtually impossible to capture, convict, and imprison offenders in a manner that would maintain a necessary threshold of punishment certainty needed in order for punishment to deter.

Relatively little attention in the literature has been given to the celerity (or swiftness) of punishment (Blumstein, 2011). A few noted exceptions are Nagin and Pogarsky (2001), Howe and Loftus (1996), Legge and Park (1994), and Yu (1994). What these authors point out is a relevant concept in economics known as discount rates (used to account for consequences realized at different times). It is well known that offenders are more impulsive and present-oriented (Wilson & Herrnstein, 1985; Gottfredson & Hirschi, 1990). They also tend more towards deliberately discounting future costs. Thus, the swiftness of punishment would seem to be all the more important to focus on for a group of individuals with an already established criminal history, such as parole violators. Focusing on increasing the swiftness of sanctioning for parole violators would also seem to be more feasible than doing so through a court system, since typically there is less due process for returning a technical parole violator to prison than there is for incarcerating an individual for the commission of a new crime. Conversely, some have hypothesized that deterrence may be stronger through a delayed punishment, since the offender may wish to simply "get it over with", and the anticipation of the punishment (as long as it is of sufficient certainty) might present an additional cost to the cost-benefit calculation of offending (O'Donoghue & Rabin, 2000). Thus, there is likely an interaction between the certainty, severity, and swiftness of punishment, but theoretical work in this area is fairly new in development.

The second theoretical contingency of the deterrent effect of incarceration is the relative congruence between the actual properties of punishment and the perception of punishment. In early work on perceptual deterrence, Geerkin and Gove (1977) pointed out that deterrence is really a social psychological theory of threat communication. Thus, even if the reality is that the certainty, severity, and swiftness of punishment are sufficiently high, if the general perception among offenders is that these properties are low then the perception may drive reality and lead to no deterrent impact from punishment. Other scholars have laid out a significant body of literature on perceptual deterrence (see Jensen, Erickson, & Gibbs, 1978; Kleck, Sever, Li, & Gertz, 2005; Paternoster, 1987). Results from subsequent empirical work have been mixed on whether the formal properties of punishment coincide with perceptions of those formal properties. It has been shown, for example, that the types of committed offenders who are more likely to end up in prison may also tend to perceive that punishment is unlikely, and therefore prison may not serve as a deterrent for these types of committed offenders (Pogarsky & Piquero, 2003).

Based on social learning variables, offenders may also change their perceptions of the properties of deterrence, which reflects the idea that perceptions are not static but instead dynamic. In economics, this has been referred to as “Bayesian updating”. Emerging research has examined how offenders go through this process of updating their perceptions of risk over time based on signals that they receive in their offending experience (Anwar & Loughran, 2011; Hjalmarsson, 2009; Horney & Marshall, 1992;

Lochner, 2007; Matsueda, Kreager, & Huizinga, 2006). Stafford and Warr (1993) outlined a learning theory of perceptual deterrence in which four factors updated risk perceptions: 1) personal punishment- being personally punished, 2) vicarious punishment- witnessing the punishment of someone else, 3) personal punishment avoidance- personally escaping punishment, and 4) vicarious punishment avoidance- witnessing someone else escape punishment. It is thus likely that parolees witness parole revocation policies and update their perceptions of the risk of being sent to prison for a parole violation based on their own experience as well as the experience of other parolees.

A third contingency of the deterrent impact of prison is what has been described as the "resetting effect" of punishment (or "the gambler's fallacy"), first put forth by Pogarsky and Piquero (2003). Under this notion, offenders who have already been subjected to punishment (e.g., incarceration) may find it exceedingly unlikely that they will again be subjected to punishment and thus may be undeterred by punishment. Under this proposition, offenders not yet incarcerated would be more likely deterred by the threat of imprisonment than would offenders who have already been incarcerated at least once in the past. Thus, there is reason to believe that parole violators, who by their very definition have already served time in prison, may be less deterred by incarceration than would other types of "newer" offenders.

A fourth contingency of the deterrent impact of prison is the oft-neglected element of the rational actor calculus- the benefits (or rewards) from crime. The three traditional deterrence elements – certainty, severity, and celerity – are about the cost side of the equation. A few scholars have pointed out that the net effect of the benefits of crime should be equally as important to consider (Piliavin, Gartner, Thornton, & Matsueda, 1986). There are of course the financial rewards for some types of crimes like drug selling, burglary, and theft. But in addition there are non-financial benefits, which Katz (1988) describes as the "seductions of crime." Offenders tend to be thrill-seekers and thus generate some benefit from their criminal lifestyle simply by the criminal act itself. So to the extent that offenders find more or less reward from their criminal endeavors, they may or may not be deterred by incarceration or the threat thereof. Even technical parole violators who commit no crime may find reward from a certain lifestyle (e.g., spending time in a bar, etc.) which outweighs the calculated cost of punishment for their technical violations.

A fifth contingency of the deterrent impact of prison involves a consideration of the heterogeneity of different types of offenders. Certain types of offenders may be found to be more or less deterred by actual or threatened incarceration. For example Pogarsky (2002) outlines three hypothetical types of offenders who vary in their susceptibility to deterrence: 1) "acute conformists" who are likely to comply in the future regardless of punishment, 2) "deterrables" who are the most susceptible to being deterred by punishment, and 3) "incorrigibles" who are likely so committed to the criminal enterprise that they will continue criminal behavior regardless of punishment. As was previously mentioned, there is also some debate around whether more impulsive types of offenders are more or less deterrable (Nagin & Paternoster, 1993). Several scholars have described and tested persistent individual differences that make offenders more or less prone to deterrence (Nagin & Paternoster, 1993; Wright, Caspi, Moffitt, & Paternoster, 2004).

Certain crime types might be more or less deterrable also. For example, Exum (2002) proposed that intoxicated and violent offenders may not be deterred because their crimes frequently are crimes of passion which involve less of a rational calculus. Similarly, Bouffard (2002) described sex offenders as being more impacted by current emotional states rather than by rational calculations of the costs and benefits of crime. Bachman, Paternoster, and Ward (1992) suggested that moral beliefs rather than rational assessments of costs and benefits have more of an impact on sex offenders. Similarly, Paternoster and Simpson (1993) found that costs and benefits of crime were irrelevant for corporate offenders when moral inhibitions were high. Grasmick and Bursik (1990) found that deterrence worked better for theft crimes, while shame was more important for drunk drivers.

A sixth contingency of the deterrent impact of prison is the degree to which an offender participates in rehabilitative programming during incarceration. A significant body of literature demonstrates that in-prison treatment programming can be effective in reducing future criminal behavior (Andrews & Bonta, 1998). Also, evidence suggests that offenders may benefit from educational or vocational training while incarcerated as well (MacKenzie, 2002). In this regard, in-prison treatment programming is part of the mechanics of deterrence. Note that there are two further contingencies on in-prison treatment programming though. The first is that it must be assumed or demonstrated that offenders are being appropriately treated at the individual level, with effective interventions. Second, it must be recognized that there is a cost imposed by prison programming in that quality programs typically take a significant amount of time to complete and thus may delay an offender's prison release date, especially if there are large waiting lists to get into programs. Thus, the trade-off must be considered of whether it is worth it to keep an offender incarcerated for perhaps a longer period of time in order to potentially generate a rehabilitative benefit from programming. The broader point, however, is that typically the literature has distinguished between rehabilitation and deterrence as separate purposes of prison, when in fact they serve very similar purposes, and rehabilitation, with its goal of restraining or correcting future criminal behavior, can actually be viewed as a subset of deterrence.

PRISON AS CRIMINOGENIC

An equally compelling theoretical case can be made that prison actually plays a criminogenic role, increasing the future criminal behavior for those who experience it. The theoretical tradition from which this school of thought primarily derives is labeling theory. Labeling theories came to prominence in criminology during the early 1970s. Paternoster and Iovanni (1989) describe two primary components of labeling theory: 1) the "status characteristic" hypothesis derived from conflict theory, and 2) the "secondary deviance" hypothesis derived from symbolic interaction theory. Of primary relevance to the relationship between punishment and future criminal behavior is the "secondary deviance" hypothesis, which generally holds that those who are punished subsequently experience some sort of negative social reaction or societal label which is internalized as a core identity and actually serves to reinforce or increase the propensity for future criminal behavior (Lemert, 1951). In a similar vein, Braithwaite (1989) has suggested that a stigmatizing label of imprisonment impacts socially relevant factors after an offender is released from prison, such as denying employment opportunities to exoffenders and eroding ties to family and the community, which in turn serve to increase the propensity for subsequent criminal behavior. Braithwaite (1989) contrasts this kind of stigmatizing label (or "shaming") with what he refers to as "reintegrative shaming," in which the act of the offender is condemned by the community, but at the same time the community works to embrace the offender back as a member (the old adage of "hate the sin but love the sinner"). Thus, the criminogenic impact of prison, according to Braithwaite (1989), would be contingent upon whether the offender experiences reintegrative shaming rather than stigmatizing shaming after release from prison.

Sampson and Laub's (1993) age-graded theory of informal social control has also incorporated elements of labeling theory in order to explain the impact of incarceration on subsequent criminal behavior within a developmental/life course perspective. In a process which they refer to as "cumulative continuity", imprisonment can lead to increases in future criminal behavior by altering the offender's identity, by excluding the offender from normal routines or conventional opportunities such as employment, and by increasing contact with deviant others (Sampson & Laub, 1997).

Matsueda's (1992) version of labeling theory, rooted in the symbolic interactionist tradition of sociological theory (Mead, 1934), views the impact of punishment leading to a criminogenic effect through what he refers to as "reflected appraisal," which involves an individual's perception of how others view the self. When offenders are more likely to perceive that others view them in a negative light, this perception takes on a dynamic of its own and in essence becomes a self-fulfilling prophecy for predicting increased future criminal behavior. This is in essence the labeling equivalent of perceptual deterrence in the deterrence literature; a negative label may or may not exist, but if through reflection the individual perceives one to exist then it essentially has the same impact.

Another body of literature that has been closely tied to labeling theory and speaks to the potential for a criminogenic impact of imprisonment is the theoretical work on procedural justice and the legitimacy of punishment. Tom Tyler, in his book entitled "Why People Obey the Law" (1990), outlines how process (i.e., procedural justice) is just as important as outcome (i.e., distributive justice) in determining the impact of criminal justice intervention on future offending. Offenders, like the rest of us, desire to be treated fairly, and if they perceive that there is legitimacy in the way that justice is handed out by the system, then they will be more likely to comply even under circumstances of unfavorable outcomes of punishment (e.g., a sentence of imprisonment or a longer sentence length). Tests of Tyler's (1990) theory have mostly confirmed its validity among different groups of offenders (see Paternoster, Brame, Bachman, & Sherman, 1997 for a test among domestic violence offenders). Recently Franke, Bierie, and MacKenzie (2009) have also examined the issue of legitimacy specifically within a correctional setting. It may be that prison is found to serve a criminogenic purpose among parole violators if they largely view the process of their revocation to prison to be arbitrary or unfair, and that prison may actually serve a deterrent impact by improving the legitimacy and transparency of the process for revoking parole (see Kleiman, 2009 for this basic argument).

An integrated perspective on the potential for a criminogenic impact of imprisonment is also found in defiance theory, which sets out to describe the conditions under which criminal sanctions reduce, increase, or have no impact on future crime

(Sherman, 1993). Defiance theory purports that sanctioning such as incarceration may lead to deterrence, defiance, or irrelevance depending on four primary factors: 1) the perceived legitimacy of the punishment, 2) the degree of social bonding, 3) the extent of recognized shame, and 4) the nature of pride. Specifically, if the offender experiences imprisonment as illegitimate, is weakly bonded to the sanctioning community, denies shame resulting from the sanction, and takes pride in isolation, then imprisonment is likely to have a criminogenic impact and increase future defiance of the law. The reverse of these factors lead to a deterrent impact of incarceration. If these four factors are evenly counterbalanced, then incarceration would likely have no impact on future criminal behavior (i.e., "irrelevance"). Defiance theory integrates Braithwaite's (1989) criminological theory on reintegrative shaming, Tyler's (1990) political science theory of procedural justice, and Scheff and Retzinger's (1991) sociological theory of the "master emotions" of pride and shame.

Prison is also hypothesized to increase future criminal behavior through social learning mechanisms. In early sociological studies on prison culture, it was found that an oppositional subculture, characterized by values supportive of crime, was typically present in prisons (Sykes, 1958). The origins of this deviant prison subculture have been variously attributed, but generally fall into one of two categories. Deprivation theory suggests that the "pains of imprisonment" lead to embracing a deviant subculture of adaptation in order to cope with the deprivation of prison life (Sykes, 1958). Importation theory, on the other hand, suggests that the prison subculture is simply a continuation of the street subculture that offenders bring with them to prison. Regardless of the specific mechanism, prison is viewed as a school of crime, in which criminal behavior is learned and reinforced during incarceration.

PRISON AS A NULL EFFECT

Writers on the topic of the impact of incarceration recognize that it may be that prison simply has a null impact on future criminal behavior, neither raising nor lowering the probability of a subsequent return to crime. Specifically, psychologists have tended to take the "minimalist" view of the effect of imprisonment (Gendreau, Cullen, & Goggin, 1999). Drawing upon the literature on learning and behavior modification, the social psychological literature on persuasion/coercion, and the personality literature, these psychologists find convincing theoretical reasons to believe that a term of imprisonment simply serves as a "psychological deep freeze" in which offenders maintain the same level of criminal propensity before, during, and after prison (Zamble & Porporino, 1988). In terms of the learning and behavioral modification literature, psychologists point out that it is unlikely that many of the previously outlined necessary contingencies of deterrence (e.g., immediacy and predictability of punishment, etc.) are present under the current system of imprisonment. In terms of the persuasion/coercion literature, it is pointed out that many of the prerequisites of human persuasion, such as credibility and empathy on the part of the messenger, are also highly unlikely under our current system of imprisonment. It is purported that repeated threat communication, such as that which operates through a system of mass imprisonment, leads to psychological "inoculation" in which individuals devise reasons to resist change (Eagly & Chaiken, 1993). This is very similar to the theoretical mechanisms previously outlined in the literature on legitimacy, procedural justice, and defiance theory. Finally, in terms of the literature on personality, it is suggested that imprisonment is not well-suited to serve as an effective change of behavior given the personality type of the typical criminal, which is characterized as antagonistic, egocentric, manipulative, and impulsive (Andrews & Bonta, 1998). Together, these various literatures form the minimalist view, hypothesizing a largely null impact of imprisonment.

One note of importance is that a null impact of imprisonment on subsequent criminal behavior does not necessarily mean that prison has no impact on restraining criminal behavior during the period of incarceration. Prison may serve a mere incapacitative effect, taking a slice out of the criminal career during the period of incarceration but having no impact on future criminal behavior. Thus, I introduce the incapacitation literature below, even though a null and incapacitative effect of imprisonment need not necessarily go together. Indeed prison can potentially simultaneously serve both an incapacitative and a deterrent/criminogenic effect.

Incapacitation research grew out of early criminological work examining criminal careers. Wolfgang, Figlio, and Sellin's groundbreaking work in 1972 found that approximately 6% of a Philadelphia birth cohort accounted for 52% of the arrests attributable to the same cohort. This spawned the notion of "selective incapacitation", in which it was suggested that it may be possible to identify a small group of high frequency offenders and incarcerate them for an extended period of time in order to prevent a large proportional share of criminal activity through incapacitation. Following the work of Wolfgang and colleagues, the concept of "lambda" was introduced into the discussion on incapacitation (Blumstein, Cohen & Nagin, 1978; Blumstein, Cohen, Roth, & Visher, 1986). Lambda is a value that represents the frequency of individual criminal behavior. The criminal career paradigm suggests that if lambda values, as well as lengths of criminal careers, could be reliably estimated, then it could be possible to generate estimates of how much crime is prevented through incapacitation by taking a slice out of the criminal career trajectory through imprisonment (Blumstein et al., 1986). Empirical research on the incapacitative effects of imprisonment blossomed between the mid-1970s and mid-1990s. This literature is reviewed in the next section of this chapter.

EMPIRICAL EVIDENCE ON THE IMPACT OF IMPRISONMENT

Three seminal reviews of the literature on the impact of imprisonment on subsequent criminal behavior have been completed to date. The first review, a meta-analysis by Gendreau et al. (1999), examined 50 studies dating to 1958 involving 336,052 offenders and producing a total of 325 correlations. Results were broken down by studies examining: 1) the impact on recidivism of serving a prison sentence versus receiving an alternative community-based sanction, and 2) the dose-response impact on recidivism of serving differential lengths of time in prison. The overall conclusion of this review was that prison produced a slight increase in subsequent recidivism rates. In three out of every four of the outcomes examined, recidivism rates were higher for those sentenced to prison or serving longer periods of time in prison. The average weighted effect size just among those studies which examined the difference between imprisonment and community sanctions was zero, meaning that prison was found to have no impact one way or the other on subsequent recidivism when compared to community-based sanctions. The average effect size for just those studies which examined the impact of length of time in prison was .03, meaning that a one unit increase in the length of time served in prison was associated with a 3% increase in recidivism. Low risk offenders were found to have particularly larger increases in recidivism associated with longer periods of time served in prison. The impact of imprisonment versus community-based sanctions was not found to vary by offender risk level, however. Reviewed studies which used more methodologically rigorous methods were more likely on average to find an increase in recidivism for those who went to prison compared to those who did not, but were not more likely to find differences in recidivism based on the length of time served in prison.

The second review of the literature on the impact of imprisonment on subsequent re-offending was a Campbell Collaboration systematic review conducted by Villettaz, Killias, and Zoder (2006) of over 3,000 abstracts. Only 23 studies met the criteria for inclusion in the final review. To be included, the study had to use: a) a randomized or natural experiment design, or b) a quasi-experimental design in which more than three potentially relevant independent variables were controlled for. Only studies conducted between 1961 and 2002 were included. Of the final 23 studies, only five studies used a randomized or natural experiment design. The final set of 23 included studies allowed for 27 statistical comparisons between custodial (e.g., prison) and non-custodial groups. Thirteen of these 27 comparisons produced statistically significant differences; re-offending rates were lower for non-custodial offenders in eleven comparisons, and were lower for custodial offenders in two comparisons. The remaining fourteen comparisons showed no statistically significant differences.

The authors of this review point out several limitations to their findings, however. First, randomized controlled experiments provide the strongest design for determining with a high degree of confidence the impact of custodial sanctions compared to non-custodial sanctions, but such experiments are exceedingly rare. Second, follow-up periods for recidivism rates rarely extended past two years in the included studies. Third, most studies used re-arrest or re-conviction as the measure of recidivism, instead of alternative measures such as self-reports. Fourth, in most of the studies only the occurrence of recidivism was examined, and not the frequency with which it occurred. Fifth, other outcomes such as on employment, health, family, and social networks were rarely examined. Sixth, no study considered placebo (or Hawthorne) effects where, for instance, offenders randomly assigned to non-custodial sanctions may have felt more fairly treated and adjusted their behavior accordingly. One final contingency that the authors of this review point out is that there is a lot of variation in the types of offenses examined in each of the included studies, and that prison might very well have a different impact for different types of offenses or offenders. Thus, for example, a null to slightly criminogenic effect of imprisonment might be found overall but a significant deterrent effect might be found if just examining parole violators.

In the last section of their review, Villettaz et al. (2006) perform a separate meta-analysis on just the five experimental design studies included in their review. From this meta-analysis, they find no overall significant difference in recidivism between custodial and non-custodial sanctions, and thus conclude based on these studies that prison has a null effect on subsequent re-offending. Given the methodological rigor of the five experimental studies, they warrant describing individually. The first study (Bergman,

1976) reports on the random assignment of “second felony” offenders in Michigan who would have otherwise received a prison sentence, to either probation or prison. A one-year follow-up period was used. The authors of this study found the probation group to have a lower recidivism rate (14% for probation vs. 33% for prison).

The second experimental study (Schneider, 1986) was a juvenile intervention implemented in Idaho, in which juvenile delinquents were randomly assigned to either probation or detention. After a 22-month follow-up, the probation group generally fared better (53% of the probation group versus 59% of the detention group reported further contact with the court), although the differences between groups in both incidence and prevalence of re-offending were not found to be statistically significant.

The third experimental study (Barton & Butts, 1990) reports on an intervention for male juvenile delinquents in Michigan. Participants were randomly assigned to either intensive community supervision or incarceration. A two year follow-up period was reported. Overall the results were mixed. The incarceration group indicated less frequent subsequent charges, but more self-reported delinquency.

The fourth experimental study (Killias, Aebi, & Ribeaud, 2000) reports on an interesting experiment in Switzerland in which a group of offenders were randomly assigned to either community service or jail. The incarcerated group served only up to 14 days in jail. The formula for community service was 8 hours for every potential day in jail for which the offender would have otherwise served. After a two year follow-up, no statistically significant differences were found, although the jail group reported a slightly higher incidence and prevalence of re-arrest.

The fifth experimental study (Van der Werff, 1979) was a natural experiment in which a royal pardon in honor of the wedding of Princess Beatrix of the Netherlands led to an automatic suspended sentence of offenders who committed their offense before a fixed date. These pardoned offenders were compared to offenders whose offense fell after the fixed cutoff date and were thus ineligible for the suspended sentence and served time in jail. After a six year follow-up period, no significant differences were found for traffic and property offenders, but significantly lower recidivism rates were found for violent offenders who received a suspended sentence and were not required to serve time in jail.

The third major review of the literature on the impact of imprisonment on reoffending was conducted by Nagin et al. (2009). These authors begin their review with a lengthy discussion of the various methodological difficulties faced by researchers attempting to study the impact of imprisonment on criminal behavior. These studies face a basic inference problem, in which the establishment of a causal (not merely correlational) link between imprisonment and subsequent criminal behavior is attempted. Offenders are non-randomly assigned to imprisonment in the vast majority of studies, which raises the immediate question as to whether pre-existing conditions explain any differences in observed recidivism rates between custodial and non-custodial groups. Nagin and colleagues (2009) lay out three aspects of the basic inference problem which are important to address in studies of the impact of imprisonment on re-offending: 1) the target population from which inferences are being drawn, 2) specification of treatment and control conditions, and 3) randomization. The target population of the study is important because many studies on the impact of imprisonment are conducted on lower risk samples of offenders, yet the results may be mistakenly conferred to the larger population of offenders who are sentenced to prison. Alternatively, focusing on a widely heterogeneous population can make inference of the results misleading or inaccurate when applied to any one subset of the total group. Therefore, there is a tradeoff between the cost of narrowness and the benefits of reduced heterogeneity which is important to consider.

Also important to consider, according to Nagin et al. (2009), is the specifics of the treatment and control conditions. These conditions can vary widely between studies and can greatly affect the results. It is thus important, for example, to specify the length and conditions of confinement for the custodial group as well as the length and conditions of treatment for the noncustodial group.

Finally, a third consideration is randomization of assignment to the treatment or control conditions. The benefits of randomization for determining causality have been clearly elucidated in the program evaluation literature (Shadish, Cook, & Campbell, 2002). The goal of randomization is to ensure that the subjects assigned to treatment and control groups (e.g., custodial versus noncustodial sanctions) differ in only one way (i.e., their treatment status). In the real world, most studies on the impact of imprisonment are not able to meet the so-called "gold standard" or randomization, however. In such cases, it becomes imperative on the researcher to use methods which approximate a random assignment process, or to use statistical techniques in order to understand and control for other significant differences between the treatment and control groups which may impact outcomes.

Nagin et al. (2009) break down their review of the literature into four types of studies: 1) experimental or quasi-experimental studies, 2) matching studies based on observational data, 3) regression studies based on observational data, and 4) studies using unique or sophisticated methodological/statistical techniques. Under the category of experimental or quasi-experimental studies, they identify the same five experimental studies as were identified in the previous review by Villettaz and colleagues (2006). Their conclusion from these five studies are that the studies point to a criminogenic effect of imprisonment, but that this conclusion is weak based on a small number of studies, few statistically significant relationships, and sample groups limited to juveniles or to shorter stays in prison.

The second category of studies, matching studies, included studies in which members in the custodial and noncustodial groups were matched either on a variable-by-variable case or through propensity score matching techniques. Eleven studies under this category were identified. Studies using a variable-by-variable matching approach overwhelmingly pointed towards a criminogenic effect of imprisonment for juveniles, but produced inconsistent results among adults. The propensity score matching studies also pointed to mixed results, but tended towards a criminogenic effect of imprisonment. On the whole, although the preponderance of the point estimates were not statistically significant, the matching studies provided even more evidence for a statistically significant criminogenic effect than did the experimental studies.

By far the largest category of studies was regression-based studies. Thirty-one studies were identified in this category, with 22 studies predominately favoring noncustodial sanctions, seven studies favoring custodial sanctions, and two studies producing mixed results. While more studies tended to favor a criminogenic effect of imprisonment, Nagin and colleagues (2009) were careful to point out that the sample characteristics of these studies were too varied to lead to an overall generic conclusion on the effect of imprisonment.

Seven studies fell under the final category of studies using unique or sophisticated methodological/statistical techniques. These studies are as much interesting for their methodology as they are for their results. The first study in this category (Drago, Galbiati, & Vertova, 2009) reports on a clemency bill in Italy in which more than 20,000 inmates were granted early conditional release from prison. Using a natural experiment framework, the authors found that each additional month that the offender would have otherwise served in prison but were instead allowed to serve in the community (i.e., their residual sentence length) was associated with a 1.24% reduction in reoffending, suggesting a criminogenic impact of imprisonment. The second study (Helland & Tabarrok, 2007) examined California's "three strikes" law by comparing reoffending rates for offenders who had one strike versus offenders who had two strikes. Those with two strikes were found to have about a 20% lower re-arrest rate, suggesting a deterrent impact of imprisonment based on the lengthier sentence which would have been served by this group had they been arrested for a third strike. The third study (Bhati & Piquero, 2008) uses an "information theoretic" hazard model to compare actual post-imprisonment criminal trajectories with estimated post-imprisonment criminal trajectories based on pre-imprisonment offending patterns. It was concluded from this study that imprisonment led to a very large reduction in reoffending due to a combination of deterrence and incapacitation. The fourth study (Wimer, Sampson, & Laub, 2008) applied recently developed advanced statistical methods for examining causal inference with non-experimental data, using the Glueck's famous longitudinal sample. They concluded that imprisonment was associated with higher re-arrest rates, but that these results were fragile when subjected to their more advanced methods of sorting out causal inference. The fifth study (Manski & Nagin, 1998) examined a sample of juvenile delinquents in Utah using a bounding approach rather than a traditional approach which produces a point estimate. Overall, results leaned towards a criminogenic impact of imprisonment, although assigning the highest rate offenders to custodial confinement tended to result in more of a deterrent effect. The final two studies (Berube & Green, 2007; Green & Winik, 2010) were both examinations of the same sample, in which the random assignment of cases to judges in the U.S. federal court system (who had different rates of sentencing offenders to prison) was exploited in order to approximate a natural experiment. Neither study found evidence that imprisonment affected recidivism rates one way or another, with point estimates equally divided between positive and negative results.

Nagin et al.'s (2009) review also examined 19 studies on the dose-response impact of imprisonment (i.e., the impact of the length of imprisonment). Only two of these studies used an experimental design in which offenders were randomly sentenced to longer versus shorter sentences in prison (Deschenes, Turner, & Petersilia, 1995; Berecochea & Jaman, 1981). Findings from the two experimental studies generally produced non-significant differences based on imprisonment length, although the Berecochea and Jaman (1981) study tended to favor longer sentences. It was concluded that the 17 remaining non-experimental studies on the dose-response issue were inconclusive and all suffered from a methodological weakness of failing to properly account for the impact of aging on recidivism.

Overall, Nagin et al.'s (2009) review is the most comprehensive to date. Their final assessment is that imprisonment has a null to slightly criminogenic impact on future criminal behavior, but that the existing evidence is not solid enough to lead to firm policy decisions. This assessment is closely aligned to the final assessment of the previous two reviews of the literature, and represents the state of the evidence on what we tentatively know about the impact of imprisonment on subsequent criminal behavior.

While it appears that on the whole imprisonment has a null to slightly criminogenic rather than a deterrent impact on future criminal behavior, this speaks nothing of the possible incapacitative impact of imprisonment. Empirical research on the incapacitative impact of incarceration is largely divided into two bodies of work. At the macro-level, economists have predominately investigated the relationship between incarceration rates and crime rates over time and between jurisdictions. At the micro-level, criminologists have used individual level data to generate estimates of "lambda" (or the frequency of offending), which are then used to simulate estimates of the amount of crime prevented through imprisonment.

Macro level studies of the relationship between imprisonment and crime rates face several methodological difficulties. First, such studies must account for a simultaneous relationship between prison and crime rates in which each theoretically influences one another. For example, we might reasonably expect that the use of imprisonment will go up in response to rising crime rates, but conversely if imprisonment serves any incapacitative effect at all then we would expect high imprisonment rates to be associated with lower crime rates. Sorting this issue of simultaneity out is a difficult task. The second methodological issue faced by macro-level studies is deciding what level of aggregation to use. Results differ by whether the association between crime rates and imprisonment rates are examined at the county level, state level, or national level. Generally, studies to date have found a smaller impact of incarceration rate on crime rates at lower levels of aggregation (Steman, 2007). The third methodological issue faced by macro-level studies is properly isolating the impact of imprisonment rates on crime rates net the wide variety of other factors that can also affect crime rates. Good models must account for these other factors. Finally, the last difficulty faced by macro-level studies is separating the incapacitative effects of imprisonment from the deterrent effects of imprisonment. By just examining incarceration rates at an aggregate level, there is no good way to date of isolating the incapacitative effects from the deterrent effects of imprisonment.

Several macro-level studies have been published, with widely varying estimates of the impact of incarceration rates on crime rates. Based on existing studies, one could conclude that a 10% increase in imprisonment rates could lead to anywhere from a 22% decrease in crime rates (Devine, Sheley, & Smith, 1988) to having no impact or actually even slightly increasing crime rates (Liedka et al., 2006). These varying estimates might be a function of the time period investigated too, since estimates are also likely dependent on the marginal incarceration rate and at a certain point incapacitation of larger numbers of offenders is likely to produce diminishing returns (Liedka et al., 2006). Three of the existing macro-level studies to date have dealt seriously with the major methodological issues outlined above (Levitt, 1996; Spelman, 2000; Spelman, 2005). Interestingly, these studies come to remarkably similar conclusions on the impact of incarceration rates on crime rates. These three studies generally conclude that a 10% increase in incarceration rates leads to somewhere between a 2 and 4% decrease in crime rates. The one factor that these three studies were still not able to account for was separating the incapacitative from the deterrent effects of imprisonment, which means that not all of the 2 to 4% decrease in crime rates is necessarily attributable to incapacitation. However, given the previously discussed literature on the deterrent, null, and criminogenic effects of imprisonment, it seems unlikely that much of this 2 to 4% decrease in crime rates is due to deterrence (or at least to specific deterrence).

Micro-level incapacitation studies by criminologists face their own set of difficulties (Piquero & Blumstein, 2007). Most of the problems relate to estimating the key parameter for simulating the incapacitative impact of prison, namely "lambda." First, lambda estimates typically rely on self-reports of the number of crimes committed within a given timeframe. Self-report measures are potentially susceptible to known problems of reporting and recall biases. Second, lambda estimates must account for, although often neglect, heterogeneity in offending levels across groups of offenders and types of offenses. Third, lambda estimates often assume a constant rate of offending at different ages, although there is strong evidence that offending declines with age. Fourth, estimates must also consider the length of the criminal career, and most estimates to date have assumed that the criminal career length is exponentially distributed. This assumption may or may not be accurate across different settings or types of offenders or crimes, however. Fifth, lambda estimate must contend with selection bias effects, in that most of the estimates of lambda are generated from among inmate populations which are known to offend at higher rates than other offenders who are not caught, arrested, and imprisoned. Sixth, most simulations assume that periods of incarceration do not change the rate of offending or length of a criminal career for a given offender. Seventh, most models do not account for crimes which involve co-offending, and as such assume that co-offending is irrelevant.

All estimates of lambda and subsequent micro-level simulation models are based on the early model of Avi-Itzhak and Shinnar (1973; see also Shinnar & Shinnar, 1975). This model is called a "steady-state" model, and proceeds by estimating the following five inputs: 1) the rate of offending (lambda), 2) the likelihood of an offender being caught and convicted, 3) the likelihood of receiving a prison sentence if convicted, 4) the average time served in prison, and 5) the average length of the offender's criminal career. Based on this model, widely varying estimates of lambda (and thus of crime prevented by incapacitation) have been generated (Piquero & Blumstein, 2007). Estimates of lambda vary from anywhere between 3 crimes to 187 crimes per person per year. Due to the complications of estimating lambda as well as the other parameters for these models, and due to the now dated work upon which much of these models/estimates are based, micro-level research by criminologists has not led to anything near a consensus on the average incapacitative effect of imprisonment.

PAROLE SUPERVISION AND SANCTIONING

Now that the research and theoretical work on the overall impact of imprisonment on criminal behavior has been examined, I turn my attention to the impact of imprisonment specifically on technical parole violators. Unfortunately empirical insight is extremely limited in this area. I begin by providing a brief history of parole supervision in America, then discuss some theoretical work which has been used to explain the potential mechanisms by which imprisonment of technical parole violators may impact their subsequent criminal behavior, and finally discuss the limited empirical work on effective sanctioning of parole violators.

The term “parole” derives from the French word “parol”, which literally means “word”, as in giving one’s word or promise (Petersilia, 2003). Thus, parole was first designed as a system in which offenders were granted early release from prison in exchange for their promise that they would abide by certain rules or conditions and obey the law. The earliest parole supervision system operated in an English penal colony off the coast of Australia, as implemented by Alexander Maconochie in the early 1800s. The practice was later adopted by the Irish penal system under the leadership of Sir Walter Crofton during the mid-1800s. Under Crofton’s system, parolees submitted monthly reports to the police, and were supervised by a civilian inspector (the precursor of the modern-day parole agent). Later in the 1800s, American penal reformers began taking notice of this system, which they referred to as the “Irish system”. Zebulon Brockway, a Michigan penologist, is largely credited with first adopting the parole system in the United States. The most complete workings of Brockway’s parole system were first implemented in New York’s Elmira Reformatory in 1876. The New York system held all of the elements of a modern day parole system: 1) an indeterminate sentencing structure, 2) a parole release mechanism, 3) post-release supervision, and 4) conditions for parole revocation for violation of rules. New York’s parole system spread so rapidly that by 1942, all of the states plus the federal government had a parole system. For a variety of reasons beginning in the mid-1970s, many states began to rethink the indeterminate sentencing structure as well as the parole release mechanism associated with the indeterminate structure, and took steps to repeal it. However, even among states that moved to repeal indeterminate sentencing and parole release, the one function of parole which has largely remained intact is using a period of post-release supervision after prison (Petersilia, 2003).

Primary questions about the functioning of the modern day parole system, which have remained largely unanswered, revolve around what the purpose of, and most effective sanctioning response to, technical parole violations should be. The thinking of those who support returning technical parole violators to prison is that technical violations are precursors to a return to criminal behavior, and that by sending technical violators back to prison we prevent or deter further criminal behavior (Piehl & LoBuglio, 2005). Critics point out that sending technical violators back to prison is costly and is a significant contributor to increased prison populations, with little evidence that such a policy actually prevents criminal behavior or is any more effective than less restrictive intermediate sanctions for technical violators (Burke, 1997).

While many of the theoretical frameworks previously discussed under the section on the general impact of incarceration may specifically apply to explaining the impact of imprisonment among technical parole violators, perhaps the most relevant theoretical perspective to this issue is the "broken windows" perspective. The "broken windows" perspective was first articulated by political scientists James Q. Wilson and George Kelling (1982). The basic idea behind the "broken windows" perspective is that there is a causal link between disorder and more serious criminal behavior, in which unaddressed disorder and minor nuisance offenses will eventually lead to more serious crimes if left unaddressed (Kelling & Coles, 1996; Wilson & Kelling, 1982). The primary policy implication extending from broken windows theorizing is that a strategy based on the criminal justice system closely responding to minor offenses and disorder will pay dividends in increased public safety.

The broken windows strategy, as has been utilized by the police, is also variously referred to as "zero-tolerance policing" or "order maintenance policing." Both labels are variants on the idea that there is some relationship between disorder/minor crimes and more serious crimes. It is important to note that the broken windows theory was originally formulated within the policing literature to explain crime rates at the aggregate level, rather than individual criminal behavior. According to the original authors, letting disorder and nuisance crimes go unattended in neighborhoods sends a signals to would-be criminal offenders that these neighborhoods lack investment and social controls. This in turn is theorized to lead to social decay and community lawlessness in which more serious crime can flourish. The original formulation of the "broken windows" approach therefore is seemingly closer related to the concept of general deterrence rather than specific deterrence. General deterrence considers the broad impact of the threat of punishment on would-be criminals within society at large, whereas specific deterrence is concerned with the impact of punishment on the individual who is actually punished or threatened with punishment. Given that this study is primarily focused on the specific deterrent impact of imprisonment among technical parole violators, the "broken windows" analogy is not perfectly applied, and may rather be more suitable for explaining how technical parole violation rates at the aggregate level could generally deter would-be criminal offenders from within the entire parole supervised population.

To date, the "broken windows" framework has primarily been adopted as a policing innovation and used by police forces in several jurisdictions, most notably New York City during the mid-1990s. While a natural connection would also seem to exist between the broken windows framework and community supervision (e.g., probation or parole supervision), "broken windows" has not specifically been raised as a relevant perspective in the community supervision literature, with a few noted exceptions (Farabee, 2005; Kleiman, 2009; Piehl & LoBuglio, 2005; Reinventing Probation Council, 1999). Technical violations of the conditions of probation/parole supervision largely do not involve criminal behavior, and would thus be analogous to the types of minor nuisance infractions or disorder which is thought to lead to more serious crime under the broken windows framework. Therefore, an increased focus on sanctioning of technical parole violators may results in substantial gains in terms of reducing serious crime rates.

Recent work by Mark Kleiman (2009) provides primary support for a deterrence-based, broken windows approach to sanctioning technical parole violators. The central premise of Kleiman's (2009) work is that certain and swift (but not necessarily severe) sanctioning of even minor technical infractions is most effective in reducing criminal behavior among probationers and parolees. The problem with the current operation of probation/parole, in Kleiman's (2009) view, is that sanctioning for technical infractions is done in almost a random manner, where technical violators are given many breaks before being sent back to prison. Kleiman's (2009) proposed strategy is to provide frequent and close monitoring of the behavior of probationers/parolees, coupled with quick and consistent enforcement of even minor infractions. Kleiman (2009) describes such a deterrence-based approach as holding particular promise among probationers/parolees because they are already subjected to monitoring and supervision (which can increase the certainty of punishment due to increased detection) and are also not typically subjected to the same level of due process for infractions as is a crime in a criminal court case (which can facilitate increased swiftness of punishment).

While Kleiman (2009) most clearly articulates a broken windows approach for sanctioning technical parole violators, he was not the first to draw a link between broken windows policing and probation/parole supervision.[2] In an early report by the Manhattan Institute on reforming probation supervision practices, the authors laid out a new model for probation supervision which they referred to as "Broken Windows Probation" (Reinventing Probation Council, 1999). As described under this approach, violations of probation conditions are to be enforced quickly and strongly. David Farabee (2005), in his monograph describing what he sees as disappointing results favoring reforming criminals through in-prison rehabilitation programs, lays out an alternative model for reforming criminals which is based directly on broken windows policing and is similar in nature to Kleiman's (2009) proposed model of focusing on increase certainty and swiftness of sanctioning among probation/parole violators (Farabee, 2005). In a review of the literature on supervision, Piehl and LoBuglio (2005) acknowledge the relevance of the broken windows perspective to the sanctioning of probation/parole violators, stating that in light of broken windows theory, "the revocation of the conditional terms of release

[2] Although Kleiman's (2009) proposed approach to the sanctioning of technical probation/parole violators conforms closely to a broken windows model, and while Kleiman himself draws a link between broken windows policing and his deterrence-based sanctioning approach for technical violators, he does not specifically refer to his approach as a "broken windows" approach.

for a large number of recent inmates, who increasingly comprise a larger percentage of new commitments to prison, may be a desired outcome."

Hawken and Kleiman (2009) provide the most convincing empirical evidence in support of a broken windows type of approach for sanctioning technical violators. They report on a randomized experiment of the Hawaii HOPE initiative, in which a group of repeat probation violators were randomly assigned to receive either probation as normal or a monitoring and sanctioning type of deterrence strategy. HOPE participants were told that from there forward they would be consistently and immediately sanctioned with short stays in jail for each and every violation of a technical condition of their supervision. The HOPE probationers, mostly meth addicts, were required to randomly receive drug testing at least on a weekly basis. The results of the HOPE experiment were impressive. The re-arrest rate for HOPE probationers was 21%, compared to 47% for the comparison group. The positive drug test rate for HOPE probationers was 13%, compared to a 46% positive drug test rate for the comparison group. Based primarily on the results from the Hawaii HOPE pilot, several recent policy papers on best practices for responding to probation/parole violations have advocated employing swift, certain, and consistently applied responses to technical violations in order to reinforce a deterrent effect (Pew, 2007; Urban Institute, 2008).

Evidence contrary to the broken windows style of enhanced monitoring and sanctioning for technical violators comes from the famous RAND study of Intensive Supervision Programs (ISP) conducted in the late 1980s to early 1990s (Petersilia & Turner, 1993). In this randomized field experiment, ISP programs across 14 sites (9 states) were compared to standard supervision practices. The authors concluded that ISP increased recidivism rates in most incidents. After a one-year follow-up, 37% of ISP participants had been re-arrested compared to 33% of the control group, and 27% of the ISP participants were returned to prison compared to 19% of the control group. The most significant difference was in the rate of technical violations, with 65% of the ISP group receiving a technical violation compared to 38% of the control group. In the end, the authors questioned whether the increased recidivism rates for ISP participants were due to increased monitoring and thus increased detection of violations, or whether the results simply spoke to a criminogenic effect of ISP supervision. Their preliminary evidence suggested that the differences were primarily due to increased detection. The next question then became whether increased detection and sanctioning of technical violations led to decreased criminal outcomes. Preliminary evidence from one of their evaluation sites (Washington state) suggested that technical violations were not a proxy for criminal behavior, although their assessment in this area was far from conclusive.

It is important to note that while theoretical and empirical work drawing on broken windows responses to technical violations might explain why imprisonment should be expected to exert a deterrent impact on technical parole violators, it does not necessarily speak directly to the issue of the relationship between imprisonment and subsequent criminal behavior among technical parole violators. The "broken windows" approach does not require prison as a necessary sanctioning response. Indeed, a parallel movement in the best practices literature on probation/parole supervision promotes "graduated sanctions" for technical violations (Pew, 2007; Urban Institute, 2008). Under a graduated sanctioning approach, imprisonment would largely not represent the most appropriate response for most types of technical violations. Interest in graduated sanctioning for technical (or administrative) violations began within the drug court movement (Harrell & Roman, 2001; Taxman, Soule, & Gelb, 1999). A graduated sanctioning matrix provides a structure of increasingly severe penalties for violations, with imprisonment serving as a last resort. Importantly, graduated sanctioning approaches such as those used in early drug court models can still rely on swift and certain sanctioning, but without coupling such sanctioning with an immediate return to incarceration as was done in the Hawaii HOPE model.

The empirical work on graduated sanctioning among probation/parole violators has been extremely limited to date. Two existing studies which have evaluated the impact of a graduated sanctioning protocol on probationer/parolee outcomes have found that graduated sanctioning reduced prison/jail time, but have found little support for any significant decrease in recidivism rates. Martin and Van Dine (2008) examined Ohio's progressive sanctioning grid for parole violators, and found that while moving to graduated sanctioning increased the progressiveness of responses to technical violations and decreased reliance on imprisonment, it had no statistically significant impact on reducing recidivism rates.[3] An earlier study in Pennsylvania examined outcomes among probationers under a "zero tolerance" model of more immediate returns to jail compared to probationers operating under a graduated sanctioning schema, and found generally lower recidivism rates for probation violators returned more immediately to jail (Civic Institute, 2005).

[3] In their bivariate analysis, recidivism rates were actually found to be higher among violators sanctioned under the progressive sanctioning grid, but these results diminished to non-significance in their multivariate analysis.

One type of intermediate sanctioning option for technical violators is residential halfway house centers. This option is often referred to as "halfway back," indicating its status as a sanction which is halfway back to prison. Offenders in "halfway back" programs typically reside in a center for a short period of time, during which they are allowed out during daytime hours to maintain employment and other social ties. There has only been one published study to date of a "halfway back" program (White, Mellow, Englander, & Ruffinengo, 2011). This study examined a "halfway back" program for New Jersey technical parole violators. Using propensity score matching techniques, the authors compared "halfway back" completers to a matched sample of technical parole violators who were returned to prison. No significant differences were found in re-arrest rates, although "halfway back" participants were found to have a slightly lower number of total arrests. Findings were further obscured by the fact that only "halfway back" completers were included, rather than following an intent-to-treat model where both program failures and completers were included. Since generally program completers are found to be more motivated to succeed, it may be that technical violators who were returned to prison would have demonstrated lower re-arrest rates if program failures and completers were included in the treatment group, although the completion rate for the "halfway back" program was found to be quite high.

While empirical evidence on the impact of graduated sanctioning grids for TPVs is far from conclusive, one hypothesized benefit of such an approach over a "zero tolerance" approach is that it enhances transparency and procedural justice, thereby enhancing the perceived legitimacy of the supervising authority (Taxman et al., 1999). This may in turn reduce re-offending rates among TPVs, and suggests that conversely an over-reliance on imprisonment for TPVs may in fact generate a criminogenic effect. Similar to the theoretical mechanisms previously discussed under the section on the overall criminogenic potential of imprisonment, sending TPVs to prison may actually lead to defiance (Sherman, 1993) if they perceive that the process of revocation to prison is not fair and transparent. This becomes particularly germane since most TPVs in prison have not committed a new crime and thus may find their punishment unfair.

One recent empirical study which is most germane to drawing a conclusion that incarcerating TPVs may actually have a criminogenic effect is a study conducted by the Washington State Institute for Public Policy (Drake & Aos, 2012). This study is in fact the only known study that comes close to the precise purpose of the current study reported on in the subsequent chapters here. In their evaluation using data on parolees in Washington state, Drake and Aos (2012) assess the impact of confinement for a TPV violation on felony recidivism rates using an instrumental variable (IV) approach based on a "natural experiment." Drake and Aos (2012) take advantage of the discovery that Community Corrections Officers (CCOs) who supervise parolees use confinement as a sanction for a TPV violation at differing rates, and that the Washington State Department of Corrections attempts to evenly distribute offenders to CCO caseloads by risk for re-offense in a way that mimics random assignment. Essentially, this random assignment of parolees to caseloads with differing rates of returning TPVs to imprisonment provides methodological advantages similar to the "gold standard" of a randomized controlled experiment for estimating the causal impact of the use of imprisonment for TPVs on subsequent recidivism rates. This approach is a very similar methodological approach to previous studies that use a similar "natural experiment" situation to examine the overall impact of imprisonment on incarceration by taking advantage of the observation that court cases are randomly assigned to judges with differing rates of sentencing offenders to prison (see Berube & Green, 2007; Green & Winik, 2010; Nagin & Snodgrass, 2013).

Drake and Aos's (2012) study began with all offenders in Washington state who were at risk for recidivism in the community between July 1, 2001 and June 30, 2008. They focus on 1,273 parolees during this timeframe with at least one violation and who were supervised by only one single CCO during the observation period. Their main finding is that felony recidivism is not lowered by using confinement for offenders who violate the technical conditions of their community supervision. In fact, in all of the models that they estimated, confinement for a violation was actually associated with increased recidivism. They offer two possible explanations for this finding of a criminogenic effect of incarceration for a TPV violation. First, they suggest that confinement may actually have a deleterious effect on the offender by leading to increased difficulties in the parolee reintegrating back home, such as increased difficulties in reentering the labor market. This explanation suggests a causal criminogenic impact of incarceration on recidivism for TPVs. Their second explanation is that CCOs had the ability to observe a parolee's risk for recidivism beyond what is measured in the department's risk classification system used for assigning parolees to caseloads, thus suggesting that the increased recidivism for those TPVs returned to prison may simply imply that some CCOs are routinely better at assessing higher risk offenders and using confinement accordingly. This second explanation would suggest a selection effect, in which it is unclear as to whether the criminogenic impact of incarceration for TPVs is causal.

Drake and Aos (2012) point out several limitations to their study. First, they do not estimate the number of potential crimes avoided during the confinement period for TPVs, which would examine the incapacitation effect of imprisonment. Second, the strength of their research design rests on the assumption that the assignment of parolees to CCO caseloads is essentially random. While they find evidence of this, there is always some uncertainty absent a true randomized controlled trial. Third, they point out that the generalizability of their results may be limited given that they restrict their analysis to only two percent of all parolees who had a violation event during the follow-up period. In addition to the limitations noted by the authors, another limitation is that the study does not examine the dose-response issue of the length of incarceration for a TPV on subsequent recidivism. These limitations aside, however, this study is the closest to the current study reported on in the chapters that follow. Based on this being the only existing study close to the current study, the tentative conclusion is that incarceration for a TPV violation has a slightly criminogenic effect. This is a far from conclusive result, however, since there remain questions by the authors of this study on the causal nature of the impact noted, the generalizability and external validity of the results both in Washington State and in other states, and the impact of relevant contingencies that may matter in determining differing impacts of incarceration for TPVs such as the certainty and swiftness of the response to the violation.

Taking one step back, some may ask whether supervision/sanctioning of released ex-prisoners even matters at all. For example, California has recently responded to the large number of technical violators filling up its prison system by enacting a parole status entitled “non-revokable parole”, in which parolees are not under supervision requirements and thus can only be revoked to prison for a new crime. One study to date which has examined whether supervision matters in terms of recidivism outcomes is a study by the Urban Institute based on a sample of offenders released from prison across 15 states during 1994 (Solomon, Kachnowski, & Bhati, 2005). This study examined recidivism rates for three groups: 1) unconditional releases from prison, 2) conditional-mandatory releases from prison, and 3) conditional-discretionary releases from prison. Both of the conditional release groups received community supervision, with the only difference being that the "conditional-mandatory" group was not released based on a parole board decision whereas the "conditional-discretionary" release group was. In their bivariate analysis, the authors find slightly lower two-year recidivism rates for the two supervised groups (61% for the conditional-mandatory group and 54% for the conditional-discretionary group) compared to the non-supervised group (62%). However, in the multivariate analysis, where several controls were included, these differences became insignificant, suggesting no added benefit in terms of lower recidivism rates for supervised versus non-supervised offenders.

Other than this Urban Institute study, only one additional study published to date has examined the impact of community supervision versus no supervision. This recently published study compared a group of parole releases in New Jersey to two groups of non-supervised max-out releases (Ostermann, 2011). After controlling for various differences between the groups, the author found no difference in terms of re-arrest rates, again suggesting no added benefit in terms of lower recidivism rates for supervised versus non-supervised offenders. The primary weakness faced by these studies, however, is that offenders were not randomly assigned to supervision versus non-supervision, and thus it becomes difficult to determine the true impact on recidivism rates attributable to supervision alone.

SENTENCING, CORRECTIONS, AND PAROLE IN PENNSYLVANIA

Before concluding this chapter, a brief summary of the context of sentencing, corrections, and parole in Pennsylvania is provided, since this particular study uses Pennsylvania specific data to examine the primary questions posed. It is important to understand the context of the study, since the results may be impacted by this particular context.

Pennsylvania is an indeterminate sentencing state with presumptive sentencing guidelines. Convicted offenders receive both a minimum and a maximum sentence date, with the maximum sentence length required to be at least double the minimum sentence length. Judicial sentencing discretion is limited by presumptive sentencing guidelines, with a judge having to justify in writing any sentence given outside of the guideline range. Guideline ranges are based on ‘offense gravity score’, ‘prior record score’, and aggravating and mitigating circumstances. Before November 2011 (which includes the timeframe for which all participants in this study would have been initially sentenced), offenders receiving a maximum sentence length of five years or more are automatically sentenced to the PA DOC. Judges are given discretion to sentence offenders receiving a maximum sentence length of between two years and less than five years to either PA DOC custody or to a county jail. All sentences with a maximum sentence length under two years are sent to a county jail.

Offenders sentenced to PA DOC custody are required to serve at least up to their minimum sentence date in state prison before even being considered for release. At the minimum sentence date, an offender then becomes eligible to be considered for parole release. The PBPP consists of nine parole board members who hear parole eligible cases and determine whether the offender may be released under parole supervision. If an offender is rejected for parole at the minimum sentence date, he or she may be re-considered for parole release at any time in between his or her minimum and maximum sentence date. An offender may serve no longer than his or her maximum sentence date in state prison, at which time the offender is unconditionally released with no parole supervision (referred to as a "max-out"). If an offender is in fact paroled somewhere between his or her minimum sentence and maximum sentence, the offender is released onto parole supervision and must be supervised on parole until the maximum sentence date is reached.

While under parole supervision, a parolee is subject to standard and special conditions of parole (as were outlined in Chapter 1). Any violation of these conditions may be considered a technical parole violation and can be sanctioned to a range of sanctioning options up to and including return to state prison. If returned to prison, a first level hearing is held within 14 days as a preliminary hearing for determining guilt for the technical violation charge(s). At this point the parolee may waive his or her right to a second level hearing and admit guilt. If essentially pleading not guilty, the parolee receives a second level hearing generally within three months of being re-incarcerated. If found guilty at the second level hearing, the parolee will serve a standard length of time in prison (referred to as a "parole hit"), and eventually will be reconsidered by the parole board for re-release back to parole supervision. If found not guilty at the second level hearing, the parolee is immediately re-released back to parole supervision under what is referred to as a "no recommit action" (NRA). Parolees who are sanctioned to anything less than imprisonment for a technical violation charge are actually not formally revoked and removed from active parole supervision, but rather are continued under parole supervision during the time of the sanction.

SUMMARY

Our knowledge of the impact of imprisonment on subsequent criminal behavior is growing, but remains far from lending to firm policy conclusions. One under-investigated area is the impact of imprisonment among specific subsets of the overall prison population. For example, prison may well serve a deterrent impact among one subset of offenders while at the same time serving a criminogenic impact among another subset of offenders. Given the large number of technical parole violators who are sent to prison each year in many states, we know remarkably little about what impact imprisonment serves for this subset of the prison population. Theoretical insight further complicates matters, since there are solid theoretical reasons to believe that the imprisonment of technical parole violators could serve either a deterrent, criminogenic, incapacitative, or null effect.

This chapter has reviewed the theoretical and empirical literature on the overall impact of imprisonment on criminal behavior, as well as reviewed the more limited theoretical and empirical literature on best parole supervision/sanctioning practices and on the impact of sending technical parole violators to prison. The tentative conclusion to date is that imprisonment serves a null to slightly criminogenic effect on subsequent criminal behavior, while simultaneously serving a marginal but diminishing incapacitative effect during the time of imprisonment. Among TPVs, it is not yet clear whether supervision or imprisonment matter at all in terms of reducing criminal behavior. Further, even if supervision matters, it is not clear as to what the most effective sanctioning strategy should be. Some theoretical and recent empirical evidence suggests that a "zero tolerance" type of approach in which technical violators are returned swiftly and consistently to prison may serve a significant deterrent effect. On the other hand, some evidence suggests that technical violators could be sanctioned certainly and swiftly but using less costly intermediate sanctioning options rather than relying on an immediate return to incarceration, with at least equal rates of reoffending. Given the large number of technical parole violators returned to prison each year, understanding the impact of imprisonment on criminal behavior among this subset of the prison population would seem to warrant a high priority for policymakers.

CHAPTER 3: DATA AND METHODS

INTRODUCTION

This study uses a quasi-experimental design, with propensity score matching techniques, to explore the impact of both the use of incarceration and the length of incarceration in response to TPVs on recidivism. Ideally a randomized controlled experiment would be the "gold standard" for exploring these questions, where TPVs are randomly assigned to either imprisonment or some other diversionary option, and where those TPVs who are assigned to imprisonment are further randomly assigned to differential lengths of incarceration. This type of experimental design would account for both observed and unobserved heterogeneity between TPVs who are sanctioned to imprisonment versus another option, and between imprisoned TPVs with varying lengths of stay in prison, thus addressing any potential selection bias in these sanctioning decisions. It is likely that there are important differences between TPVs who are sanctioned to imprisonment versus an alternative option and who are sanctioned to imprisonment for varying lengths of time, which affect the outcome of recidivism independently from the impact of imprisonment itself.

To explore the impact of imprisonment among TPVs, this study examines data from one state (Pennsylvania). Unfortunately, a randomized controlled experiment is not possible in this situation, since decision-makers in Pennsylvania are not willing to set aside current practice and assign TPVs to imprisonment and to varying lengths of imprisonment on a random basis. Fortunately one possible alternative technique for reducing selection bias is the use of propensity score matching to retrospectively assign TPVs to either a treatment or control group based on their propensity to receive the treatment condition. The impact of two separate treatment conditions are explored: 1) the impact of imprisonment for TPVs, and 2) the dose-response impact of varying lengths of stay in prison given that a TVP is sanctioned to prison.

DATA

This study is based on a primary sample of 12,705 parole releases from the PA DOC between October 2006 and December 2009 who were subsequently sanctioned for a TPV to either imprisonment or to an alternative sanctioning option. Important to note is that the primary sample is not a set of unique individuals since some individuals were released to parole multiple times within the sampling timeframe. The number of unique individuals in the sample is 12,242. The treatment group consists of 1,758 parole releases (1,593 unique individuals) whose sanction for their first detected TPV violation after their parole release from prison was re-imprisonment, and the control group consists of 10,947 parole releases (10,649 unique individuals) whose sanction for their first detected TPV violation after their parole release was any other alternative option other than re-imprisonment.

The primary dataset originated from staff in the PA DOC's Bureau of Planning, Research, and Statistics, who created a file from the PA DOC's inmate records database system of all parole releases between October 2006 and December 2009. PA DOC staff also attached various inmate demographics to the dataset, to be used as covariates and propensity score predictors. PA DOC staff also provided re-incarceration recidivism data and dates to calculate periods of incarceration. Based on the primary dataset provided from the PA DOC, the PBPP provided data on the subsequent TPV violations and sanctions received after each parole release date, as well as an indication of which geographic parole district office each case was primarily supervised in, and at what level of supervision the parolee was supervised upon release from prison. Full arrest history RAP sheet data for the primary dataset was provided by the Pennsylvania State Police (PSP).

SAMPLE SELECTION

The sampling timeframe of October 2006 to December 2009 was selected for two primary reasons. First, the beginning of the sampling timeframe (October 2006) was selected based on feedback from PBPP staff that complete and accurate parole violation and sanctioning data was only available starting in the fourth quarter of 2006. Second, the ending of the sampling timeframe (December 2009) was selected in order to allow enough time for recidivism follow-up, particularly for those in the treatment group since they were sanctioned to a period of re-incarceration and had to be re-released before being tracked for recidivism.

During the sampling timeframe (Oct. 2006 – Dec. 2009), a total of 31,561 parole releases from PA DOC prison custody were recorded. This includes 24,569 first-time parole releases on the given sentence (77.8%), and 6,992 reparole releases (22.2%). As noted before, these are not unique individuals but rather parole release incidents. The total parole releases during the sampling timeframe (n = 31,561) is comprised of 29,826 unique individuals.

For several reasons, certain cases from the universe of 31,561 parole releases during the sampling timeframe were removed in order to establish the primary analysis sample of 12,705 parole releases. These various reasons for removal are summarized and tabulated in Table 3.1.

First, 7,739 cases were removed because the parolee never received a TPV after the parole release date according to the PBPP and PA DOC data. These cases appear to be non-recidivists. There were 936 additional cases removed because they appeared to have a return to prison for a parole violation according to PA DOC data, but showed no sanctioning history for a return to prison from PBPP data. From further exploration of these 936 cases within PA DOC data, it is clear that the vast majority of these cases was returned to PA DOC for a police re-arrest and/or charge of a new crime rather than a TPV, and thus would not be included in the study since the study is limited to TPV violators. An additional 1,361 cases were removed because they could not be matched with PBPP violation/sanctioning data based on unique IDs provided by PA DOC. The unique ID provided by PA DOC for matching with PBPP data was the Parole Board Number, which was the only feasible field for matching cases between PA DOC and PBPP datasets. In all of these 1,361 unmatched cases, the only Parole Board Number that PA DOC databases had on record was actually the DOC Inmate Number, which is an altogether completely different number which PBPP staff do not have access to.

Of the remaining 21,525 cases, additional cases were removed for a variety of other reasons. A total of 5,521 cases were removed because there was an indication that the first violation (and associated sanction) was actually for a new crime and not a technical violation. All of the analysis in this study is based only on outcomes after the <u>first</u> recorded technical violation after parole release from prison, and is restricted only to cases where the first recorded violation was for a technical reason and not for an indication of a new crime being committed.[4] Of the 5,521 cases removed because of an indication that the first violation involved new criminal activity, 4,130 showed "Pending Criminal Charges" as the reason (or one of multiple reasons) in the PBPP violation data for their first violation after release from prison. Another 1,186 had an arrest record present between their release from prison and their first recorded sanction for a parole violation. An additional 205 cases showed up in PA DOC databases as being adjudicated for the first violation as a Convicted Parole Violator (CPV) or Technical Convicted Violator (TCV), indicating criminal activity.

One of the difficulties with the PBPP data is that the violation and sanctioning data are not directly connected. In other words, a sanction or set of sanctions is not explicitly assigned by PBPP in their data as being associated with a specific violation incident. To connect sanctions to violations some inference based on violation/sanction dates have to be made. In many cases violations and sanctions are recorded as happening on the same date, which is the simplest case to interpret. However sometimes a sanction date is actually before the next closest violation date, or is a significant period of time after the next closest violation date. For this study, only violations on the first violation date after parole release from prison were used, and only sanctions on the first sanctioning date after parole release from prison were used. To match the first violation(s) with the first sanction(s), the dates were compared. For the primary sample dataset a conservative approach was used, which means 178 cases were removed because the first sanctioning incident was before the first violation incident, and another 1,034

[4] While drug relapse is technically a new crime, in most cases drug relapse is included as a TPV violation here for this analysis, as long as it was processed as a TPV by PBPP staff and not prosecuted as a new crime. In addition, some low level misdemeanors and summary offenses are treated as TPVs if processed as TPVs by PBPP staff rather than as a new crime.

cases were removed because the first sanctioning incident was more than one month after the first violation incident. In addition, 253 cases were removed because another violation date was recorded between the first violation date and the first sanction date, making it unclear as to which violation incident the sanction was in relation to. In later analysis, these rules will be relaxed in order to test the sensitivity of the results to the rules set in place for connecting violations to sanctions.

Another 1,637 total cases were removed because of a disconnect among the treatment group, between the sanction to prison date as recorded in the PBPP sanctioning data and the incarceration date as recorded in the PA DOC data. Among these 1,637 cases, a total of 582 cases were removed because the PA DOC re-incarceration date was before the PBPP sanction to prison date, and another 793 cases were removed because the PA DOC re-incarceration date was more than one month after the PBPP sanction to prison date. Another 55 control group cases were removed because their first TPV sanction was something other than imprisonment based on the PBPP data, yet the PA DOC data showed them returning to prison within five days of the first violation date, so it was unclear as to whether these cases should have been in the treatment or control group. Finally, 207 cases were removed because the first TPV sanction in the PBPP data was a sanction to incarceration, but there was no record in the PA DOC dataset of a return to incarceration. Again, as with the connection of the PBPP violation to sanctioning data, later sensitivity analysis will be conducted in order to test the sensitivity of model results to the rules set in place for matching the PBPP arrest date with the PA DOC re-incarceration date.

Lastly, 105 treatment cases were removed because the PA DOC data showed no prison re-release date after the TPV re-incarceration date, which means that these cases cannot be followed for recidivism since they have not been subsequently re-released after their TPV incarceration. Approximately 75% of these 105 cases were re-incarcerated before 2010, meaning that they have been in PA DOC custody for a parole violation for as long as 4.5 years. An inspection of PA DOC records indicates that these cases are mostly parole violators who are pending adjudication for criminal charges, and thus would not be included in the study anyways because they are not TPVs. Also, 92 treatment cases were removed because they received some other sanction in addition to incarceration for their first TPV violation, so the impact of the treatment cannot be determined in these cases since they received both the treatment and control condition.

To summarize these reasons for cases being removed, the above reasons can be grouped into five basic categories (see Table 3.1). A total of 7,739 cases (41%) were removed because no recidivism incident was recorded after release from prison, 6,457 cases (34.2%) were removed because the first violation incident appeared to be for new criminal offense charge(s), 1,465 cases (7.8%) were removed because of difficulties in connecting the first violation incident to the first sanction, 2,943 cases (15.6%) were removed due to difficulties in matching PA DOC records to PBPP records, and 252 cases (1.3%) were removed for other miscellaneous reasons noted above.

Since approximately 27% of the total sample of cases with a first TPV violation (minus those cases showing no parole violation or a violation for a new crime rather than a TPV; n = 17,365) were removed for the reasons outlined above, one important factor to consider is how comparable the final sample is to the sample before removing this 27%

of cases, based on various relevant covariates. Table 3.2 provides a comparison of the primary analysis sample (n = 12,705) to the excluded cases (n = 4,660) across all of the covariates described in the next section. Table 3.2 shows several statistically significant differences between the primary analysis sample and the original sample before removing cases. Specifically, the final analysis sample shows statistically significant differences across age, race, criminal risk score, prior treatment programs, sentencing county, parole district office where supervised, supervision level, severity of first TPV violation, and whether the first violation occurred in a community corrections center. The final sample appears to be a slightly lower risk population. A particularly large difference is observed in the severity of the first TPV violation, with 40.4% of the final sample having a high severity first TPV violation, and 68.4% of the original sample before removing cases having a high severity first TPV violation. As such, a sensitivity analysis will be performed in later analysis by relaxing rules for matching violations to sanctions and for matching PA DOC imprisonment data to PBPP violation data, in order to add back in cases from the original sample and examine this impact on results.

MEASURES

Treatment Indicator Variables

There are two primary treatment indicator variables for this study. The first variable is a binary indicator of whether the case received a sanction for the first TPV violation of either imprisonment (n = 1,758; 13.8% of the total sample) or any other alternative, lower-level sanctioning option (n = 10,947; 86.2% of the total sample). One important note is that cases in the control group can receive multiple different sanctions for their first TPV violation(s). Sanctioning options are based on PBPP's standard Violation Sanctioning Grid (see Figure 3.1 - "Con I – Con II Arrest Worksheet") and vary across a continuum of seriousness. Sanction options are grouped on PBPP's Violation Sanctioning Grid into one of three levels of severity – Low, Medium, and High. Table 3.3 provides a tabulation of the various sanctions received for the first TPV among the control group. Again, the sum of this tabulation does not equal the control group sample size because each case may receive more than one sanction for the first violation(s).

On average, the control group received 1.6 sanctions for their first TPV violation incident (median=1 sanction). The number of sanctions for the first violation incident ranged from 1 to 9 sanctions. Approximately 63.9% of the control group only had one sanction for the first violation incident, and approximately 86.2% had either 1 or 2 sanctions for the first violation incident. The single most prevalent sanction within the control group was a written warning (32.4% of all sanctions). Approximately 18.0% of sanctions within the control group were to either drug/alcohol treatment (14.2%) or to increased urinalysis drug testing (3.8%). Another 8.4% of sanctions within the control group were to increased reporting requirements to parole staff, 12.4% to a curfew or increased curfew, and 1.4% to travel restrictions. Approximately 3.0% of sanctions were to electronic monitoring or GPS monitoring. Approximately 9.8% of sanctions were to a residential placement like a "Halfway Back" home or community corrections center. Finally, about 11.1% received some sort of "other" undetermined sanctioned, classified as either an "other-low" sanction (5.4%), "other-medium" sanction (4.1%), or "other-high" sanction (1.6%).

The second treatment variable is a continuous variable representing the calculated length of stay in prison (in months) for those who are in the first treatment condition of being returned to prison for a first TPV violation after release from prison. This variable is used to answer the second primary question of this study, which is the dose-response impact of varying lengths of stay in prison for a TPV. In order to use the length of stay variable in later propensity score modeling, a second categorical variable for length of stay is created, which represents the length of stay categorized into quintiles.

The average length of stay in prison for the treatment group cases who were returned to prison for a TPV is 12.3 months, and ranges from a low of 4 days to a high of 78.5 months. Figure 3.2 shows a histogram of the distribution of time served in prison for a TPV violation for the treatment group sample. Quintiles for the prison length of stay distribution are 0 to 3.6 months, 3.7 to 7.9 months, 8.0 to 12.1 months, 12.2 to 18.3 months, and 18.4 to 78.5 months.

Outcome Variables

The outcome examined in this study is recidivism after the first TPV. Three different measures of recidivism are used. The first measure of recidivism is 're-incarceration', defined as the first instance of return to PA DOC custody after either being sanctioned to prison for a TPV and subsequently re-released, or after being sanctioned to a control group condition of any other type of sanctioning option. Re-incarceration data is provided by the PA DOC.

The second measure of recidivism is 're-arrest,' defined as the first instance of police arrest after either being sanctioned to prison for a TPV and subsequently re-

released, or after being sanctioned to a control group condition of any other type of sanctioning option. Re-arrest data comes from police RAP sheets, provided by the PSP.

The third measure of recidivism is 'overall recidivism,' defined as the first instance of either return to PA DOC custody or police arrest, after either being sanctioned to prison for a TPV and subsequently re-released, or after being sanctioned to a control group condition of any other type of sanctioning option.

Given different dates for sample participants of either being sanctioned to alternative options or being sanctioned to prison and re-released, times of recidivism exposure varies. Based on recidivism data being pulled by PA DOC and PSP staff during mid-August 2014, recidivism exposure times for the sample varies from 2.5 months to 7.8 years. This study examines three recidivism follow-up periods for all three measures of recidivism: 6-month recidivism rates, 1-year recidivism rates, and 3-year recidivism rates. For the 6-month recidivism follow-up period, approximately 99.9% of the sample have six months or more recidivism exposure time (n = 1,754 for treatment group; n = 10,933 for control group). For the 1-year recidivism follow-up period, approximately 99.5% of the sample have one year or more recidivism exposure time (n = 1,748 for treatment group; n = 10,906 for control group). For the 3-year recidivism follow-up period, approximately 94.8% of the sample have three years or more recidivism exposure time (n = 1,547 for treatment group; n = 10,495 for control group). The vast majority of the sample thus clearly has three years or more of exposure for recidivism follow-up.

<u>Covariates</u>

Several covariates are considered, which may predict recidivism rates and also are likely to predict whether the sanction for a first TPV violation is to prison (treatment condition) or to another alternative (control condition). Thus, these covariates will be used in later propensity score matching models to predict the propensity for assignment to the treatment condition. The covariates considered are age at time of the first TPV violation, race, gender, LSI-R criminal risk assessment score, offense type, the number of prior treatment programs participated in, sentencing county, prison release type, parole district office reporting to under supervision, level of supervision, and the severity of the most serious violation for the first TPV violation. Table 3.4 provides summary descriptive statistics of these covariates for the overall sample, and comparisons between the treatment and control group samples.

On basic demographic statistics, approximately 42% of the sample is white, 92% of the sample is male, and the average age at the time of the first violation among the sample is 35. The treatment group is slightly younger than the control group. No statistically significant differences on race were identified between the treatment and control groups. There is a statistically significant difference between the treatment and control groups by gender, with males representing approximately 94.6% of the treatment group and 91.4% of the control group.

One factor that should influence the decision of parole staff of whether to sanction a TPV to prison or to another option, is the parolee's criminal risk assessment score. PBPP utilizes the Level of Service Inventory-Revised (LSI-R) as its primary criminal risk assessment tool. The LSI-R is a widely used, 54-item assessment instrument which measures an individual's propensity to criminally recidivate. Each item on the LSI-R is scored as a 0 or 1, with all of the 54 items summing up to generate an overall test score. Higher scores indicate a higher risk of criminal recidivism. Potential scores range from 0 to 54. PBPP staff administer the LSI-R to all parolees just prior to their release from prison onto parole, and then again at least once per year for every year under parole supervision in the community. This study considers the LSI-R score for the last LSI-R assessment date on record before the first TPV violation associated with the sanction that triggers treatment assignment. LSI-R scores in the sample range from 2 to 54, with an average score of 22.4. The treatment group indicates a significantly higher average LSI-R score than the control group (25.1 vs. 21.9), suggesting that those in the treatment group are at a higher pre-violation risk of criminal recidivism than the control group.

A breakdown of the primary offense type for the sample's originally sentenced offense reveals that approximately 27% of the total sample had a violent offense, 21% had a property offense, 33% a drug offense, and 18% a public order or other type of offense. Group differences reveal that the treatment group is significantly more likely to have a violent or property offense, while the control group is more likely to have a drug offense.

The decision to sanction a first TPV violation either to prison or to another alternative is likely to be influenced by the number of prior treatment programs the parolee participated in. Parole agents will theoretically be more willing to sanction TPVs to prison for those already given multiple chances to improve behavior through prior treatment program participation. One covariate considers a count of the total number of prior in-prison treatment programs participated in before the first TPV violation. On average, the total sample participated in 2.85 treatment programs prior to receiving their TPV violation. Note that treatment participation includes any program that the parolee participated in during prison, regardless of whether the program was successfully completed. Standard programs include cognitive behavioral therapy, violence prevention, batterer intervention, and various alcohol and other drug (AOD) treatments like a Therapeutic Community (TC) or outpatient treatment. As theorized, the treatment group participated in a statistically significant higher number of prior treatment programs than the control group (3.2 vs. 2.8).

One statistically significant difference is observed between the treatment and control groups in the county that the parolee was originally sentenced from. Sentencing county is measured in two ways. The first variable is a binary indicator of whether the county of sentencing was Philadelphia or Allegheny County (the two most populous counties in Pennsylvania) versus any other county. The treatment and control groups do not significantly differ on this county indicator. The second variable is a binary indicator of whether the sentencing county was a Class 3 or higher county based on definition in Pennsylvania statute. By statute all of Pennsylvania's 67 counties are classified into one of nine classes based on population size. A Class 1 county has a population of 1.5 million or more. Philadelphia is the only Class 1 county. A Class 2 county has a population of between 800,000 and 1,499,999. Allegheny County (which contains the city of Pittsburgh) is the only Class 2 county. A Class 2-A county has a population of 500,000 to 799,999, and a Class 3 county has a population of 210,000 to 499,999. Any county with less than a 210,000 population is a Class 4 county or lower. The control group is significantly more likely to have been originally sentence from a Class 3 or higher county than the treatment group. Approximately 28% of the total sample was originally sentenced from either Philadelphia or Allegheny County, and approximately 71% of the total sample was originally sentenced from a Class 3 county or higher.

The decision to sanction a first TPV to either imprisonment or to an alternative option may also be influenced by whether the parolee is released from prison for a first parole release on the sentence or alternatively is being re-released after already serving time in prison for a parole violation. Approximately 76% of the total sample was initial parole releases, and 24% were re-parole releases. The treatment group was significantly more likely than the control group to be re-parole releases (32% of the treatment group versus 22% of the control group). This makes sense intuitively, since parole agents are probably more willing to sanction a TPV to imprisonment if the parolee has already served time in prison for a parole violation.

A sanction to imprisonment for a TPV may also be influenced by length of time the parolee is to be under parole supervision. The length of parole supervision is defined as the difference between the parolee's release from prison and the parolee's maximum sentence expiration date. Overall, the parole supervision term for the sample averages 47.3 months. The control group shows a significantly longer total supervision period than the treatment group (48.1 months versus 42.1 months respectively).

Due to the discretionary nature of parole sanctioning, there is also likely to be differences in TPV sanctioning severity for a first TPV violation across different parole district offices. In Pennsylvania there are ten geographic parole district offices covering state parolees throughout the state, and an eleventh parole district office category (Central Office) which supervises special cases. The covariate utilized here records the parole district office that the parolee is supervised in at the time of the first TPV sanction. The largest parole district offices for the total sample are the Philadelphia District Office (supervising 24% of the cases), Harrisburg District Office (supervising 17% of the cases), Pittsburgh District Office (supervising 15% of the cases), and Allentown District Office (supervising 15% of the cases). A significantly higher proportion of cases sanctioned to prison for a first TPV (i.e., the treatment group) were supervised in the Allentown, Pittsburgh, and Scranton district offices. A significantly higher proportion of cases sanctioned to an alternative option for a first TPV (i.e., the control group) were supervised in the Chester, Erie, Harrisburg, Mercer, and Williamsport district offices.

Another factor which is likely to affect whether a first TPV is sanctioned to imprisonment or to an alternative option is at what level of intensity the parolee is being supervised. There are four primary levels of state parole supervision in Pennsylvania, ranging from the least intense being "minimum" supervision to the most intense being "enhanced" supervision. Minimum supervision level requires one face-to-face parolee contact per three (3) months, with at least every other face-to-face being at the parolee's approved residence, and one face-to-face collateral contact per three (3) months with an employer, spouse, treatment provider, relative, etc. Medium supervision level requires three face-to-face contacts per three (3) months, one of which being at the parolee's approved residence, and one face-to-face collateral contact per three (3) months. Maximum supervision level requires six face-to-face contacts per three (3) months, two of which must be at the parolee's approved residence, and one face-to-face collateral contact per month. Enhanced supervision level requires four face-to-face contacts per month, one of which must be at the approved residence, and two collateral contacts per month, one of which must be face-to-face. Supervision level is primarily determined by the results of the LSI-R risk assessment score, which is administered to parolees just prior to their release from prison, and then again at least once per year during their supervision period. The variable used here for supervision level represents the supervision level at the time of the parolee's first TPV sanction. Approximately 20% of the total sample was released to a minimum supervision level, 23% to a medium supervision level, 53% to a maximum supervision level, and 3% to an enhanced/special supervision level. As expected, those in the treatment group (sanctioned to prison for a first TPV) were significantly more likely to be supervised at a higher supervision level.

Perhaps the most important factor that should predict the decision to sanction a first TPV to either imprisonment or to an alternative sanction is the severity of the associated violation(s). Table 3.5 provides a breakdown of the TPV violations associated with the first TPV sanctioning incident for the total sample, as well as a comparison of the breakdown of the TPV violations between the treatment and comparison groups. Note that a TPV violation incident may be associated with multiple violation reasons. On average, there were 1.5 TPV violations associated with the first TPV sanctioning incident (an average of 2.1 violations for the treatment group and 1.4 violations for the control group). The range of number of TPV violations associated with the first sanctioning incident was between 1 and 18 (between 1 and 10 for the treatment group, and between 1 and 18 for the control group). Overall, the most prevalent violation was a positive urinalysis for alcohol or other drugs (33% of the total violations). Yet a violation for a positive urinalysis was nearly three times more likely in the control group as in the treatment group (38.0% vs. 12.7%). Other overall prevalent violations for the entire sample were:

- failure to pay supervision fees (6.4% of all violations; 0.3% of treatment group violations; 7.9% of control group violations),

- changing residence without permission (5.9% of all violations; 14.4% of treatment group violations; 3.8% of control group violations),
- removal from treatment or a community corrections center (5.8% of all violations; 14.5% of treatment group violations; 3.7% of control group violations),
- failure to abide by Board imposed special conditions (7.0% of all violations; 8.7% of treatment group violations; 6.5% of control group violations),
- absconding or failure to report as instructed (8.2% of all violations; 22.0% of treatment group violations; 4.8% of control group violations)

According to PBPP's Violation Sanctioning Matrix (see Figure 3.1), all violations are assigned a severity level of low, medium, or high. The descriptive statistics table (Table 3.4) and later models examine the severity level of the most serious violation associated with the first TPV violation incident. Approximately 16% of the total sample was first violated for a low severity violation, 45% for a medium severity violation, and 39% for a high severity violation. As expected, the treatment group was significantly more likely than the control group to be sanctioned for a high severity violation for their TPV associated with their first sanction (90% of the treatment group versus 31% of the control group). The control group was significantly more likely than the treatment group to be sanctioned for a low severity first TPV violation (19% of the control group versus 0.4% of the treatment group).

One final covariate indicates whether or not the parolee was housed in a community corrections center (i.e., a "halfway house") at the time of the first TPV violation. Approximately 16% of the total sample was housed in a community corrections center at the time of their first TPV violation. The treatment group indicates a significantly higher percent of TPVs housed in a community corrections center at the time of the first violation when compared to the control group (25% of the treatment group versus 14% of the control group).

The above description of the various covariates highlights significant baseline differences across several relevant factors between the treatment and control groups, which likely influence recidivism rates through group assignment, thus confounding an evaluation of the impact of the treatment condition itself on recidivism rates. Those in the treatment group (assigned to prison for a first TPV) are younger males who are more at risk of criminal recidivism, more likely to have been to prison already for a parole violation, more likely to have been originally sentenced for a violent or property offense, more likely to have participated in a larger number of prior treatments, supervised at a higher supervision level, and whose first TPV violation was for something more serious. The treatment group is thus at an overall higher risk for recidivism across a number of different dimensions. These covariates will therefore be used in later propensity score models to generate a predicted score of the propensity to be in either the treatment or control condition, which will later then be used in the models to reduce the confounding effects of these individual differences on the outcomes.

METHODS

Propensity Score Matching

When there are significant differences between treatment and control group conditions on a significant number of observable covariates (as is seen in Table 3.4 for the covariates examined in this study) this presents difficulties in determining whether the observed outcomes (in this case recidivism rates) are due to the treatment itself or to the confounding effect of pre-existing differences in other factors. In the case of this study, it is thus difficult to isolate the causal effect of imprisonment for a technical violation on recidivism rates. A randomized controlled trial (RCT) would be an ideal (a "gold standard") for answering this question, as random assignment sanctioning decisions to either imprisonment or to an alternative sanctioning option would maximize comparability between these two conditions on observable and unobservable characteristics. In an RCT, assignment to a treatment or control condition is based on pure chance ($p = .50$), so that the only differences between treatment and control groups on all observed and unobserved covariates are due to chance alone. This is often referred to as "conditional independence." The results of an RCT experiment are thus said to produce unbiased and consistent estimates of the average treatment effect on the subjects in the experiment.

Unfortunately a prospective RCT is not feasible for this study. Parole agents in Pennsylvania are not willing to relax discretion in assigning sanctions for technical violations so that a sanction to imprisonment versus an alternative option is assigned randomly. We are thus left with observational data for attempting to explore the impact of imprisonment for a TPV on subsequent recidivism. The apparent selection bias from observed covariates must be addressed in order to try and draw causal inference, however.

Several quasi-experimental methods exist for attempting to deal with the problem of selection bias within observational data, in attempting to draw causal inferences. One particularly strong quasi-experimental method for attempting to deal with this problem is propensity score matching techniques, developed by Rosenbaum and Rubin (1983). In propensity score techniques, a score of the propensity for assignment to a treatment group condition, based on a set of observed predictors, is used to balance observed characteristics so that they are independent of the treatment assignment. For this study, two types of propensity score techniques will be used in order to answer the primary research questions. Traditional propensity score matching techniques will be used primarily to examine differences between those who are sanctioned to prison versus an alternative sanction for a first TPV. Based on common support issues (discussed later), a few extensions of propensity score modeling (i.e., propensity score weighting and propensity score stratification) may be further used to examine the question of the impact of TPV imprisonment assignment on recidivism. The second type of model used in this study is an extension of propensity score modeling for measuring treatment in doses, at different levels (Lu, Zanutto, Hornik, & Rosenbaum, 2001; Zanutto, Lu, & Hornik, 2005). This is used to examine the "dose-response" question of the impact of differential lengths of incarceration for those who are sent to prison for a first TPV.

Propensity score matching models are advantageous to other non-experimental methods in addressing selection bias for several reasons. First, traditional regression models assume a linear relationship between independent and dependent variables. Propensity score matching models make no such assumption since the outcome is not used in the matching procedure. Second, regression-based models ignore whether there is sufficient overlap in the distribution of the covariates between the treatment and control groups. If sufficient overlap does not exist, regression models draw conclusions outside of the range of the treatment and control group overlap thus leading to less meaningful comparisons. Third, propensity score models allow for matching on one score rather than controlling for many variables. One extension of this benefit is that it leads to an ease in understanding and interpretation of results for a non-technical audience. Fourth, propensity score models facilitate balance over many observed factors that impact the treatment assignment decision, thus more closely replicating the benefits of an RCT, at least for observed and measurable covariates. The goal here, as is the goal with an RCT, is to develop a convincing counterfactual (i.e., what would have happened to subjects in the treatment condition had they in fact been assigned to the control condition instead). While the counterfactual can never actually be observed (i.e., individual are always only either in the treatment or control condition at a given point in time), propensity score models can build a strong case for making conclusions about the counterfactual.

In propensity score matching, the propensity score is defined as "the conditional probability of assignment to a particular treatment given a vector of observed covariates" (Rosenbaum & Rubin, 1983). The propensity score is represented by the following formula:

$$e(x) = pr(Z=1|x)$$

where Z is a binary indicator that is one if the TPV sanction is to imprisonment and zero if the TPV sanction is to any other alternative, and x is a vector of individual observed characteristics. The estimated propensity score, $e(x)$, ranges from 0 to 1, with a higher score representing a greater likelihood of being assigned to the treatment group (in this case sanctioned to prison).

The propensity score can be estimated with either a logit or probit model. I use a logit model in this study for estimating the propensity score. To maximize the benefit of propensity score modeling, it helps if there is a strong understanding of, and ability to measure, the decisional rules that affect the assignment to the treatment condition. In this case, that would mean it is important to have a strong understanding of the predictors of sanctioning to prison for a first TPV. Propensity score models can only balance on factors that can be observed and measured. RCTs have the added advantage of also balancing unobserved/unmeasurable factors. Fortunately the dataset for this study demonstrates strong predictors for the treatment assignment, indicating a decent understanding of what factors are considered by PBPP staff in deciding whether to sanction a first TPV violation to prison. Specifically, the supervision level of the parolee, the severity of the violation, and the actuarial criminal risk of the parolee are all strong predictors of whether or not the parolee is sanctioned to prison for a first TPV.

After generating a propensity score from the logit model, it is important to assess whether there is sufficient overlap in the propensity score distribution of the treatment and control groups. This is referred to as 'common support.' Figure 3.3 and 3.4 provide graphical representations of the distributions and overlap of the propensity scores between the treatment and control groups used in this study. These are useful graphical representation of common support. As is seen, there is very little overlap at the high and low ends of the propensity score range between the treatment and control groups. There are very few treatment group cases with a low propensity for assignment to treatment, and virtually no control group cases with a very high propensity for assignment to treatment. While on the one hand this situation of relatively low common support makes modeling more challenging, on the other hand it highlights the benefit of propensity score modeling over traditional regression-based modeling, since regression models simply ignore this problem of low common support.

One important distinction to make at this point is the difference between the Average Treatment Effect (ATE) and the Average Treatment Effect on the Treated (ATT). The ATE is the expected effect of the treatment on a randomly drawn case from the population and is represented by the following formula:

$$E(y_1 - y_0)$$

In cases where there is low common support as is demonstrated here, estimating the ATE is not practical because it averages across the entire population and there are likely cases that are simply not eligible for treatment. In this example, low common support at high values of the propensity score distribution due to virtually no control group cases with a propensity score above .6 means that there clearly seems to be situations where a first TPV violation would simply never be eligible to be sanctioned to imprisonment for the violation.

An alternative to estimating the ATE is to estimate the ATT. The ATT is the average treatment effect for those who actually are "treated" (or in this case sanctioned to imprisonment for a TPV). Estimating the ATT is a more appealing quantity to estimate here given the observed low common support and the evidence that there are certain TPV violators who simply do not appear eligible to be sanctioned to imprisonment for their violation. The formula for the ATT is:

$$E(y_1 - y_0) \mid Z = 1)$$

After determining common support, the next step in propensity score modeling is to determine the matching technique. The matching technique that can be used is in part constrained by the identified degree of common support. Since relatively low common support is identified in this study, there are primarily three matching options to consider. The first option is to consider an inverse probability of treatment weighting, which in effect weighs cases with lower propensity scores for treatment assignment as a higher value, with stronger emphasis in later models for calculating the effect size. The second option is to consider stratifying the propensity score, generating estimates of effect sizes within each stratum, and averaging over strata to generate an overall estimated effect size. The third option is to use traditional nearest-neighbor matching, but with replacement of appropriate control cases to be used multiple times, and with a caliper of an accepted range of nearest neighbors to match.

Nearest neighbor matching with replacement and a caliper will primarily be used for this study. In nearest neighbor matching, each treated unit *i* is matched to a non-treated unit *j* as such:

$$|e_i - e_j| = \min_{k \in \{Z=0\}}\{|e_i - e_j|\}$$

This selects the control group case which looks the most similar to each treatment group case. Using replacement allows for each matched control group case to be re-used by being returned to the reservoir after being matched with a treatment group case. By doing so, it maximizes utilization of the control group cases that are able to be matched, given that there are unmatchable control group cases at the far right end of the propensity score distribution.

Selecting a caliper addresses the potential situation with low common support where the "nearest neighbor" may actually be far away and therefore may not look similar enough to be useful in developing the counterfactual. The caliper sets a maximum distance between nearest neighbors for matching. The caliper range is the number of standard deviations of the propensity score within which to select control cases. While difficult to determine a priori what the appropriate distance should be for setting a caliper, Rosenbaum and Rubin (1985) suggest a caliper size of .25 times the standard deviation of the logit for the estimated propensity score estimation. This generally translates into a caliper of around .05.

After selecting a matching technique and performing the match between treatment and control group cases, there are a few ways to assess that the selected matching technique has adequately balanced the treatment and control groups on the observed covariates. First, t-tests can be used to assess the mean difference in covariates between treated and non-treated cases after matching. Rosenbaum and Rubin (1985) also recommend using the standardized difference in means between the treatment and control group for each observed covariate. This standardized bias measure can be calculated as such:

$$\frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2 + s_c^2}{2}}}$$

where $\bar{x}_t$ and $\bar{x}_c$ represent the mean of the covariate for the treatment and control group and s_t^2 and s_c^2 represent the variance of the covariate for the treatment and control group.

If the standardized bias is an absolute value of less than 0.25, the covariate is considered balanced. Graphical summaries, such as quantile-quantile plots, can also be used to compare the distribution of treatment and control group cases after matching. If a number of covariates outside of random chance remain statistically unbalanced, alternative matching strategies must be attempted until an acceptable balance between the treatment and control groups is achieved.

Once balance is achieved through a successful matching technique, the treatment effect size can be estimated (in this case the ATT effect size). The ATT effect size resulting from an appropriately balanced model allows for an inference of the counterfactual to be drawn. In other words, the effect size indicates the causal impact of the treatment condition on the outcome for those who receive the treatment. In this case, the effect size would indicate the causal impact of imprisonment for a first TPV among those who are imprisoned for a first TPV.

Dose-Response Model

The second main question examined in this study is the impact of differential lengths of incarceration on recidivism, among the treatment group of those who are sanctioned to prison for a first TPV violation. This type of effect is often referred to as the dose-response impact. As with the first main research question of the effect of incarceration itself on recidivism, propensity score modeling can again be used here to examine the dose-response impact of differential lengths of stay in prison for a TPV on recidivism, while addressing potential selection bias in the assignment of differential lengths of stay in prison. It is not likely that the length of incarceration for TPV is assigned at random by PBPP staff, but rather affected by factors such as the seriousness of the violation preceding the TPV and the general riskiness of the parolee for recidivating, which themselves should predict recidivism and thus confound the ability to isolate the impact of the length of imprisonment itself on recidivism.

One issue with modeling a dose-response impact like differential lengths of incarceration is that the length of incarceration is not a binary variable. Fortunately recent extensions of propensity score matching techniques have been utilized for addressing this problem. Specifically, two approaches have been recently developed. Lu et al. (2001) use an expanded version of the original Rosenbaum and Rubin (1983) model to accommodate a categorical treatment effect of multiple dosage categories, where the propensity score is calculated using an ordinal logic model. In this model, the distribution of doses for an individual, i, D_i, given a vector of observed covariates, x_i is model as:

$$\log\left[\frac{P(D_i \geq d)}{P(D_i < d)}\right] = \propto_d + \beta' x_i \text{ for d=2,3,4,\ldots}$$

The distribution of doses given covariates thus depends on the observed covariates only through $b(x_i) = \beta x_i$, such that the observed covariates, x, and the doses, D, are conditionally independent, given the propensity score, $b(x_i)$. After this, the propensity score from the above model can then be used like in traditional propensity score models to balance the distribution of a large number of covariates among individuals with various treatment doses simultaneously. Lu et al. (2001) suggest matching individuals of different dose levels.

Zanutto et al. (2005) use a slightly different approach, suggesting that the propensity score identified from the above ordinal logit model be sub-classified instead of matched. This sub-classification approach is the one employed here in this study. Sub-classification on the propensity score is accomplished by stratifying the propensity score into equally sized groups, usually five groups based on quintiles. Once the propensity score is sub-classified, covariates should no longer predict treatment dose level. In other words, the covariates for cases in the same sub-class should look as if the cases were randomly assigned to one of the dosage levels.

Table 3.6 shows descriptive statistics for the quintiles of the prison length of stay dosage groups, across all of the considered covariates, at baseline before a propensity score is estimated. ANOVAs and logit regressions are used to test the statistical significance of baseline differences across the various covariates. There are clearly several factors that show statistically significant differences across the treatment dosage groups. In particular, 12 covariates (age, race, LSI-R risk score, offense type, whether sentenced from Philadelphia/Pittsburgh, parole supervision length, whether the supervising district office is Philadelphia or Scranton, and the severity of the first TPV violation) all significantly differ across the treatment dosages. This represents significantly more differences than would be expected from chance alone, thus suggesting imbalance across the treatment dosages on these covariates before sub-classification.

Once the propensity score is sub-classified, balance of the covariates can be assessed through a two-way ANOVA, where the dependent variable is the covariate, and the two factors are the propensity score sub-class and the treatment dosage level. If the main effect of the treatment dose level, or the interaction of the propensity score sub-class and the treatment dose level, are statistically significant, a given covariate is considered unbalanced. If covariates remain unbalanced after sub-classification, the procedure to estimate the propensity scores can be repeated using quadratic and interaction terms to improve the model.

After generating a propensity score model that demonstrates balance on covariates after sub-classification, the next step is to use the sub-classifications in order to identify the average dose-response effect of differential lengths of the treatment on the outcome. This can be done by estimating stratum-weighted averages of the outcome, which in this case is recidivism. The stratum-weighted mean for the outcome, conditional on receiving dose, *d*, is given as:

$$\bar{Y}_d = \sum_{i=1}^{5} \frac{1}{5} \bar{Y}_{d,i}$$

where $\bar{Y}_{d,i}$ is the observed mean outcome among individuals receiving dosage level, *d*, in balancing score quintile, *i*. The standard error is calculated as:

$$SE(\bar{y}_{di}) = \frac{1}{5} \sqrt{\sum_{i=1}^{5} \frac{s_{di}^2}{n_{di}}}$$

where s_{di}^2 and n_{di} are the sample variance and frequency, respectively, among individuals in treatment dose level *d* in propensity score quintile *i*. One note, while the standard error estimate has been found to be a reasonable approximation, as noted by Zanutto et al. (2005), it is not totally unbiased due to the fact that the sub-classification is based on propensity scores, which are estimated from the data, and the outcomes, both between and within each sub-class, are not independent.

The averages from each propensity score quintile can then be plotted to create a dose-response curve to determine whether a change in dosage (in this case incarceration length for a TPV) has an impact on the outcome (in this case post-release recidivism).

CONCLUSION

In order to determine the impact of incarceration for a TPV on recidivism, ideally a randomized controlled trial would be utilized. In many cases, including this study, an RCT design is not feasible. Fortunately several strong alternatives exist for approximating an RCT design in order to make causal conclusions about the impact of a treatment (in this case imprisonment for a TPV). This study makes use of quasi-experimental propensity score models using observational data for estimating the impact of incarceration for a TPV on recidivism. Models are derived from a large primary sample of 12,705 parolees released from prison onto parole in Pennsylvania between October 2006 and December 2009. The primary sample contains a treatment group of 1,758 cases which subsequent to their release from prison receive a first TPV violation which is sanctioned to imprisonment, and a control group of 10,947 cases which subsequent to their release from prison receive a first TPV violation which is sanctioned to any other alternative other that imprisonment. From this sample, a propensity score matching model is used in order to generate estimates of the impact of imprisonment itself for a TPV on subsequent recidivism. A sub-classification model is used, also based on generating a propensity score, in order to further determine the dose-response impact within the treatment group of differential lengths of imprisonment for a TPV. In both cases, a large number of relevant covariates are utilized for generating the propensity score later used for either matching or sub-classification.

The next chapter (Chapter 4) presents the results of the propensity score matching models for estimating the impact of imprisonment itself for a first TPV on various measures of recidivism. Chapter 5 explores the dose-response models for determining the impact of different lengths of stay in prison for a TPV on various measures of recidivism. Finally, Chapter 6 provides a summary of findings, a discussion of policy implications, and future directions for further understanding the impact of the use of incarceration among TPVs on recidivism.

CHAPTER 4: PROPENSITY SCORE MATCHING MODEL OF TREATMENT EFFECT

INTRODUCTION

This chapter details results from the propensity score models used to examine the first of the two primary questions explored in this study – the treatment effect on recidivism of using imprisonment in response to a first TPV violation after release from prison compared to using another alternative sanctioning option. Before matching, baseline comparisons of differences between the treatment and control groups across various relevant covariates are first examined. Baseline recidivism rates prior to matching are also first examined. Then the results of propensity score matching models, using different matching parameters and different measures of recidivism outcomes, are detailed. Following the details of the propensity score matching outputs, a profile of the matched cases is next provided in order to examine a profile of the relevant (on-support) group of TPVs for which policy conclusions from these models can be drawn. Finally, this chapter concludes with several post-estimation sensitivity analyses to examine the robustness of the findings generated from the propensity score matching models.

BASELINE COMPARISONS

Comparisons of Covariate Differences Between Treatment and Control Group

Prior to the utilization of propensity score matching, it is first important to determine whether there are significant differences between the treatment and control groups across covariates that may impact both the treatment assignment and the recidivism outcomes. If significant differences do exist, this can lead to selection bias that must be addressed (in this case through propensity score matching) prior to making a causal inference of the impact of the treatment itself on the outcome. Making the treatment and control groups similar on all important/relevant variables prior to estimating the effect size of the treatment on the outcome is referred to as "balancing" the two groups.

As was detailed in Chapter 3 (see Table 3.4), there are indeed important covariate differences that exist between the treatment and control groups. These differences are further highlighted in Table 4.1. As can be seen from Table 4.1, a large number of covariates differ significantly between the treatment and control groups. Using a T-test to test for statistically significant differences, 26 of the 36 covariates (72% of the covariates) show statistically significant differences between the treatment and control groups at $p<=.01$. This is well above the number of covariates that might be expected to be statistically different by chance alone.

Some have argued that hypothesis tests such as T-tests are not most appropriate for assessing balance because they are impacted in part by other factors than balance alone (see Ho, Imai, King, & Stuart, 2007). Thus, another way to inspect pre-matching balance between the treatment and control groups on the covariates is to examine the standardized bias statistic (Rosenbaum & Rubin, 1985) for each covariate.[5] This statistic is the standardized difference in means between the treatment and the control group on an

---

[5] The standardized bias statistic is calculated as follows:

$$\frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2 + s_c^2}{2}}}$$

where $\bar{x}_t$ and $\bar{x}_c$ represent the mean of the covariate for the treatment and control group and s_t^2 and s_c^2 represent the variance of the covariate for the treatment and control group.

observed covariate. The covariate is considered out of balance if the associated standardized bias has an absolute value of greater than 0.2. As shown in Table 4.1, nine of the 36 covariates (25% of the covariates) have an absolute standardized bias score above 0.2.

Overall, it is clear that there are some fairly large and significant differences between those who are sanctioned to imprisonment for a first TPV violation versus those who are sanctioned to an alternative sanction for a first TPV violation, across several different covariates. These differences are in the direction of what might be expected, with TPVs sanctioned to imprisonment mostly being higher risk and more serious TPV violators. The largest selection effect appears to come from the severity of the first violation, the parole supervision level, and the LSI-R criminal risk assessment score. TPVs with a higher severity TPV violation, who are supervised at a more intense level, and who have a higher LSI-R criminal risk score are much more likely to be sanctioned to imprisonment rather than an alternative sanction for a first violation.

Comparisons of Recidivism Rates Before Matching

Three measures of recidivism are used in the subsequent analysis. The re-incarceration rate represents the first instance of return to PA DOC imprisonment. A parolee can be re-incarcerated either as a parole violator (charged with a technical and/or new crime violation), or through police arrest, prosecution, and then sentencing to incarceration by a court. The re-arrest rate represents the first instance of a police arrest. The overall recidivism rate is a combination of re-arrest and re-incarceration, and represents the first instance of either of those two events. A re-incarceration can happen without being preceded by a re-arrest instance as is the case with technical parole violators, and a re-arrest may not always result in a re-incarceration. Thus the most inclusive definition of recidivism is to consider the overall recidivism rate by combining re-arrest and re-incarceration. It is still useful, however, to examine the re-arrest and re-incarceration rates separately in order to better understand the source of what is driving observed recidivism as measured by the overall recidivism rate combining re-arrest and re-incarceration.

Using multiple follow-up time periods is also useful when examining recidivism rates. Patterns of short-term change in recidivism may differ from patterns of longer-term change in recidivism. Three follow-up periods are used in this analysis: 6-month recidivism rates, 1-year recidivism rates, and 3-year recidivism rates.

For framing purposes, first presented here are the recidivism rates for the full population of all parolees released from PA DOC custody during the time period examined (parole releases from PA DOC between October 2006 and December 2009). The 6-month re-arrest rate is 11.1%, the 1-year re-arrest rate is 22.0%, and the 3-year re-arrest rate is 46.0%. The 6-month re-incarceration rate is 12.6%, the 1-year re-incarceration rate is 25.5%, and the 3-year re-incarceration rate is 47.7%. The 6-month overall recidivism rate is 19.2%, the 1-year overall recidivism rate is 35.8%, and the 3-year overall recidivism rate is 61.7%.

Next examined are the observed recidivism rates before matching among the primary analysis sample of those sanctioned for a TPV after release from prison. Observed recidivism rates are compared between the treatment group (those sanctioned to imprisonment for a first TPV) and the control group (those sanctioned to any other alternative sanction for a first TPV). Table 4.2 presents these observed recidivism rates.

For the 6-month follow-up recidivism rates, TPVs sanctioned to imprisonment (the treatment group) are 13.9% more likely to be re-arrested (x^2 = 4.8, p = 0.028), and have an 11.7 % higher overall recidivism rate (x^2 = 6.9, p = 0.008). While the observed re-incarceration rate is 1.5 percentage points lower for the treatment group than the control group at the 6-month follow-up, the difference is not statistically significant (x^2 = 2.1, p = 0.143). At the 6-month follow-up mark, approximately 52% of the first recidivism events are re-arrests and 48% are re-incarcerations.

For the 1-year follow-up recidivism rates, the treatment group of those TPVs sanctioned to imprisonment have a 17.1% higher re-arrest rate (x^2 = 14.4, p < 0.001) and a 9.4% higher overall recidivism rate (x^2 = 10.2, p = 0.001), but a 15.1% lower re-incarceration rate (x^2 = 16.7, p < 0.001). At the 1-year follow-up mark, approximately 59% of the first recidivism events are re-arrests and 41% are re-incarcerations.

For the 3-year follow-up recidivism rates, the treatment group of those TPVs sanctioned to imprisonment have a 9.1% higher re-arrest rate (x^2 = 12.0, p = 0.001) and a 4.3% higher overall recidivism rate (x^2 = 6.0, p = 0.014), but an 11.6% lower re-incarceration rate (x^2 = 21.1, p < 0.001). At the 3-year follow-up mark, approximately 71% of the first recidivism events are re-arrests and 29% are re-incarcerations.

These observed recidivism rates thus generally show statistically significant higher recidivism rates for the treatment group of TPVs sanctioned to imprisonment for a first violation, especially as measured by re-arrest rates and overall recidivism rates at all follow-up periods. At least at the 1-year and 3-year follow-up, those TPVs sanctioned to imprisonment have significantly lower re-incarceration rates, but these lower re-incarceration rates do not have enough impact on overall recidivism to lead to lower (or non-significant) overall recidivism rates. Re-incarceration has a diminishing impact on overall recidivism rates with longer follow-up periods, as re-incarceration events compose less and less of the first recidivism events with longer follow-up periods. The important question to now turn to is whether these generally higher recidivism rates for TPVs sanctioned to imprisonment hold up after matching on relevant covariates that affect both the type of sanctioning response (i.e., the treatment assignment) and the recidivism outcomes.

PROPENSITY SCORE MATCHING MODELS

Naïve Logistic Regression Model

As a first step before presenting propensity score matching results, Table 4.3 presents the results from logistic regression models using all covariates plus the treatment assignment indicator to predict recidivism outcomes. As detailed in Chapter 3, traditional regression models have several disadvantages to propensity score matching models for adequately drawing causal inference about the treatment effect. The logistic regression models might be thought of as naïve models to compare to the later propensity score matching models, however. The results of these models (see Table 4.3) generally show a change in the sign of the relationship between treatment and recidivism rates compared to the raw recidivism rates observed above. Specifically, re-incarceration rates and overall recidivism rates at all three follow-up periods now show significantly lower recidivism for TPVs who are first sanctioned to imprisonment versus an alternative sanction. For TPVs first sanctioned to imprisonment, the odds of re-incarceration are 40% lower at 6 months, 48% lower at 1 year, and 50% lower at 3 years. The odds of overall recidivism are 17% lower at 6 months, 18% lower at 1 year, and 27% lower at 3 years. Re-arrest rates generally show non-significant differences between the treatment and control groups, except for at the 1 year follow-up where the odds of re-arrest are 18% higher for TPVs first sanctioned to imprisonment.

Calculating Propensity Score

The first step in developing propensity score matching models is to calculate the propensity score itself. The propensity score is the likelihood, based on the distribution of the covariates, that an individual will be assigned to the treatment group, regardless of actually observed treatment group assignment. The propensity score may be generated through either a logit or probit regression model. For ease of interpretation, the propensity score here is estimated using a logit regression model. It is important that the covariates used to generate the propensity score are measured prior to treatment assignment and are at least theoretically relevant to the assignment decision. All of the covariates meet both criteria, and thus are all used in the logit regression in order to generate the propensity score. Table 4.4 presents the logit regression output used to generate the propensity score.

Common Support

Once a propensity score has been generated for each individual in the sample, it is next important to consider the common support between the treatment and control groups across the distribution of the propensity scores. Common support for the propensity score distribution based on the propensity scores generated for the primary sample was previously discussed in Chapter 3 (see discussion in Chapter 3, as well as Figure 3.3 and Figure 3.4). To review, an examination of the common support reveals that there is very little overlap at the high and low ends of the propensity score distributions between the treatment and control groups. There are very few treatment group cases with a low propensity for assignment to treatment, and virtually no control group cases with a very high propensity for assignment to treatment. There is an apparent region of common support in the middle of the distributions, however, where the propensity score distribution of the treated group (i.e., those sanctioned to imprisonment for a first TPV violation) overlaps with the propensity score distribution for the control group (i.e., those sanctioned to an alternative option for a first TPV violation). Since I only examine the ATT in this analysis, it will be important to profile and understand which group falls in the overlap region of the propensity score distributions between the treatment and control groups, since this will be the policy relevant group for which the results will be able to be generalized to in the end. On the surface, from an examination of the common support graphs (Figure 3.3 and 3.4), it does appear that there may be a sizeable enough region of common support to continue with matching procedures.

<u>Selecting Matching Algorithm</u>

Choosing an appropriate matching procedure requires optimizing several objectives. The first objective is to keep the largest number of cases in the total sample for matching. In order to match it requires that the cases are "on-support" and chosen for matching by the particular matching algorithm parameters selected. A larger matched sample is better, all else equal. The second objective is to maximize the balance between the treatment and control groups on all observed covariates after matching. The third objective is to minimize the distance between the individual treatment and control group cases that are matched. The “distance” is the difference in the propensity score value for each matched pair in the matching sample.

Based on the above three objectives, several nearest neighbor matching strategies were considered here. In nearest neighbor matching, the treatment group cases are randomly sorted and then matched, one at a time, to control group cases with the closest propensity score. Options to consider within nearest neighbor matching are whether are not to allow replacement so that control group cases can be matched more than once, and whether or not to set a caliper of the maximum distance that will be allowed between the propensity scores of treatment and control cases. Fundamentally, these options are constructed to deal with the reality that distance between “nearest neighbors” can become quite large over successive matches as matching proceeds. On the surface, both setting a caliper level and allowing for the replacement of control group cases to be matched more than once would seem appropriate here since it has already been illustrated that there is fairly low common support on the two tails of the overlap in distributions between the treatment and control group propensity scores.

To run the propensity score matching models, the ‘psmatch2’ procedure within Stata 13.0 was used. The nearest neighbor propensity score matching models that were attempted were one-to-one nearest neighbor matching with replacement, without replacement, with a caliper of .01, with a caliper of .05, and with the “common” option selected within ‘psmatch2’. The “common” option imposes a common support by dropping treatment group observations whose propensity score is higher than the maximum or less than the minimum propensity score of the control group cases. All combinations of the above one-to-one nearest neighbor matching options were attempted, also including models with no caliper or common support trim selected. The model that consistently produced the best results was one-to-one nearest neighbor matching with replacement and with a caliper of .01 imposed. Models using matching with replacement were clearly advantageous, with the average distance in propensity scores between matched cases being on average 25 times higher for the models without replacement versus the models with replacement. A caliper or .01 produced the lowest average distance between matched cases without suffering a significant drop in the sample size of matched cases. For the large dataset of n=11,934 cases for which a propensity score could be estimated and for which a full three years of follow-up was available for calculating recidivism rates, the number of cases retained for matching using one-to-one nearest neighbor matching with replacement and a .01 caliper was 2,645 (1,542 treatment group cases and 1,103 control group cases). The average distance between propensity scores for matched cases was .00016, with a range of a minimum distance of 0 and a maximum distance of .0068. Only approximately 26% of the control group cases were matched more than once, and only 8% of the control group cases were matched more than twice (maximum number of matches of a control group case was 8; only one control case was matched 8 times). Finally, the average standardized bias score for the covariates after matching was only 1.9%, for the one-to-one nearest neighbor matching with replacement and a .01 caliper. As such, all models below use one-to-one nearest neighbor matching with a caliper of .01 imposed.

Assessing Balance Pre- and Post-Matching

Once a matching algorithm has been chosen, the next important step before examining the outcomes is to examine the improvement in balance on all of the covariates between the treatment and control groups before versus after matching. The goal is to maximize balance on all observed covariates. Several steps can be taken in order to examine balance after matching. First, T-tests are conducted in order to determine whether there are remaining statistically significant differences between the means of the covariates between the treatment and control groups. Table 4.5 presents measures of balance between the treatment and control groups for all of the covariates after matching, and the improvement in balance from before matching. As can be seen from Table 4.5, none of the mean differences between the treatment and control groups on any of the covariates remain statistically significant at $p<=.05$ after matching, suggesting that conditional independence has been met.

As was previously discussed, however, some have argued that conventional hypothesis testing is not adequate for assessing balance (see Ho et. al., 2007). For this reason, the standardized bias statistic (Rosenbaum & Rubin, 1985) is suggested as a better indicator of balance. A standardized bias statistic for a given covariate of less than an absolute value of 0.2 indicates a balanced covariate. Table 4.5 shows the standardized bias statistic for all of the covariates before versus after matching as well. Before matching, the average standardized bias score for the covariates is 0.22, above the critical 0.2 value. All standardized bias scores after matching are clearly well below the critical 0.2 value, with the absolute value of the highest standardized bias score of only 0.063, and an average standardized bias score of 0.02 for all of the covariates. Further, the average reduction in bias from before versus after matching for all of the covariates is 84%, and is above 90% for nearly half of the covariates. This clearly confirms the adequacy of the matching. In fact, this balance is better than might be expected through randomization given that two out of 31 covariates might be expected to show statistically significant differences between the treatment and control groups by chance alone even under randomization.

The balance after matching between the treatment and control groups can also be visualized graphically. Figure 4.1 shows a jitter dot-plot of the propensity score distributions of the matched treatment and control cases. As can be seen from Figure 4.1, the distributions after matching are now closely aligned and very similar. Figure 4.2 also provides a histogram comparison of the propensity score distributions between treatment and control group cases after matching. When compared to the histogram of the distributions before matching (see Figure 3.3 from previous chapter), the selected matching parameters clearly significantly improve the overlap in distributions between the treatment and control groups after matching. These figures provide strong further confirmation that balance has been achieved.

Effect Sizes from Propensity Score Matching Results

Now that a propensity score has been generated, a suitable matching strategy has been selected, and balance on all observed covariates after matching has been demonstrated, the results of the selected matching model can be simply presented as the ATT. Table 4.6 shows the recidivism rates and effect sizes (i.e. the ATT) for the 6-month, 1-year, and 3-year recidivism rates based on all three measures of recidivism (i.e., re-arrest rate, re-incarceration rate, and overall recidivism rate). After matching, re-incarceration rates at all three follow-up points are significantly lower for the treatment group of those TPVs first sanctioned to imprisonment versus the control group of those TPVs sanctioned to an alternative sanction. The 6-month re-incarceration rate is 7.2

percentage points lower for the treatment group (18.3% vs. 25.5%), the 1-year re-incarceration rate is 14 percentage points lower for the treatment group (29.2% vs. 43.2%, and the 3-year re-incarceration rate is 13.1 percentage points lower for the treatment group (48.1% vs. 61.2%). All of these differences are highly statistically significant at $p < .01$.

Re-arrest rates mostly show no statistically significant differences between the treatment and control groups. Only the 1-year re-arrest rate is statistically significantly higher for the treatment group than for the control group. At the 1-year re-arrest rate mark, the re-arrest rate is 4.5 percentage points higher for the treatment group than for the control group (30.1% vs. 25.6%), and this difference is statistically significant at $p < .05$.

Overall recidivism rates are lower for the treatment group than for the control group at all three follow-up points, and the effect size grows with longer follow-ups. The 6-month overall recidivism rate is 2.1 percentage points lower for the treatment group (30.9% vs. 33%), but this difference does not quite reach a level of statistical significance at $p < .05$. The 1-year overall recidivism rate is 5.2 percentage points lower for the treatment group (49.2% vs. 54.4%). The 3-year overall recidivism rate is 6 percentage points lower for the treatment group (72.5% vs. 78.5%).

Interestingly, these results after propensity score matching generally confirm the "naïve" logistic regression results with covariate controls performed before matching (see Table 4.3), and both tell a somewhat different story than the raw observed recidivism differences before covariate controls or matching is considered (see Table 4.2). Before matching, and without consideration of relevant covariates, recidivism rates were generally higher for the treatment group of TPVs first sanctioned to imprisonment compared to the control group of TPVs first sanctioned to another alternative. After matching, recidivism rates (specifically overall recidivism and re-incarceration) are significantly lower for the treatment group of TPVs first sanctioned to imprisonment. Since overall recidivism is the most comprehensive definition of recidivism, it is especially notable that overall recidivism rates are lower for the treatment group than for the control group. While re-arrest rates appear slightly higher for the treatment group, they are generally not statistically significant differences. Clearly lower re-incarceration rates are heavily driving lower overall recidivism rates. It thus appears that after propensity score matching the general evidence supports a deterrent effect from sanctioning TPVs for a first violation to imprisonment rather than to an alternative option.

DESCRIPTION OF MATCHED CASES

In the process of matching, only a select group of comparable cases which are on-support and able to be matched are retained in order to generate the recidivism outcomes presented above. As a result of matching, 9,397 cases are discarded – 0.3% (n=5) of the treatment group and 89.5% (n=9,392) of the control group. Very few of the treatment group cases are discarded, but a significant proportion of the control group cases are discarded. It is thus important to especially examine a profile of the control group cases which are retained for matching, so that the results previously presented may be placed in context. From the previous examination of common support before matching, it was clear that a large proportion of the control group cases showed a very low probability of receiving treatment. Since the effect estimated here is the ATT, the matched group for which the ATT effect sizes can be generalized should be profiled. This is also important since the control group is a quite heterogeneous group, containing a continuum of various sanctioning alternatives for a first TPV.

Table 4.7 shows a comparison in mean differences between the matched and unmatched groups across all of the covariates. In addition, Table 4.8 shows a comparison between the matched and unmatched groups of the individual types of most serious TPV violations associated with the first TPV instance. Also, Table 4.9 shows a comparison of the distribution of different sanctions among the control group using the most serious sanction, between the matched and unmatched groups.

From Table 4.7, several differences are observed between the matched and unmatched groups. Among the most important differences, the matched group appears to be at a higher risk of criminal re-offending (i.e., a higher LSI-R score), are much more likely to be supervised at the maximum supervision level, and are much more likely to have violated for a high severity technical violation. The matched group is also more likely to have been supervised in a halfway house at the time of their first TPV violation, and is also more likely to have been re-released from prison after already previously serving time in prison for a parole violation at the time of their first TPV violation. Altogether, this paints a picture of the matched sample as a group of higher risk TPVs who commit a more serious technical violation.

Table 4.8 shows several differences between the matched and unmatched groups on the individual types of TPV violations for the most serious violation associated with the first TPV. Among the largest differences, the matched group is significantly more likely to receive a violation for changing residences without permission, failure to abide by Board imposed special conditions, absconding, and removal from or failure in treatment. On the other hand, the matched group is significantly less likely to receive a violation for failure to pay supervision fees and a positive urinalysis or use of drugs among those with no previous history of positive urinalysis and/or use of drugs. The most prevalent violations among the matched group are changing residences without permission (21.4%), a positive urinalysis for use of alcohol among those with a previous history of positive urinalysis for alcohol use (21.3%), failure to abide by Board imposed special conditions (15.8%), and removal from or failure in treatment (15.8%).

An examination of sanctioning differences between the matched and unmatched samples among the control group (in Table 4.9) also shows some differences. The most frequent sanction among the control group cases retained for matching is placement in a "Halfway Back" community corrections center (24.9% of matched control group cases). A "Halfway Back" community corrections center is a non-secure residential halfway house which is used as an intermediate sanction, where residents are still provided a large degree of latitude to leave the center for various reasons such as employment. The second most frequent sanction among the matched control group cases is a written warning (16.9% of matched control group cases), although the prevalence of written warnings among all sanctions for matched control group cases is only about half of the prevalence of written warnings among the sanctions for unmatched control group cases (16.9% vs. 33.3%). So while written warnings are quite prevalent amongst matched control group cases, they are significantly less prevalent than among unmatched control group cases. The third most frequent sanction among the matched control group cases is placement in an inpatient drug and alcohol facility (12.1% of matched control group cases). Overall, it appears that Halfway Back placements and inpatient drug and alcohol placements are significantly more prevalent among the matched group sample than among the unmatched group, whereas less severe sanctions such as written warnings (although still quite prevalent), travel restrictions, and increased reporting requirements are much less prevalent among the matched group sample than among the unmatched group. It thus appears that the treatment effect being compared in the previously described propensity score matching models is largely between the imposition of imprisonment versus the imposition of a less secure residential placement for a first TPV violation.

SENSITIVITY ANALYSIS

The propensity score matching results presented above appear to mostly show a deterrent effect (i.e., mostly lower recidivism rates) for TPVs who are first sanctioned to imprisonment versus an alternative sanction, especially when the trade-off is between imprisonment versus a non-secure residential placement such as in-patient treatment or a halfway house. The question to now to turn to is how consistent and robust this finding is. Several sensitivity analyses can be performed in order to examine the robustness of these propensity score model results.

Scenario 1: Relaxing Data Exclusion Criteria

The first sensitivity analysis performed is to relax some of the exclusion criteria applied to the original dataset. In particular, as was detailed in Chapter 3, cases were removed from analysis because the first TPV sanction date was either before the first TPV violation date or was more than 30 days after the first TPV violation date. Also, cases were removed because the PBPP violation/sanctioning data indicated a first TPV sanction to imprisonment but PA DOC data did not indicate the presence of a re-incarceration date, or the PA DOC re-incarceration date was more than 30 days after the first violation date. These exclusion rules were placed in order to address difficulties in directly connecting PBPP violation incidents to sanctioning incidents, and also in connecting PBPP sanctioning records to PA DOC imprisonment records for those sanctioned to imprisonment.

In order to examine the potential impact of these data restrictions on the results, some of the eliminated cases are added back in. Specifically, cases are added back in where the first TPV sanction date is more than 30 days after the first TPV violation date but no other violations or police arrests are recorded in between them. Also, cases are added back in where the PBPP data indicates that the TPV is a treatment group case (i.e., the first TPV sanction is to imprisonment) and yet there is no matching indication of imprisonment in the PA DOC data records, or the next available record in the PA DOC data is an incarceration date more than 30 days afterward with no other PBPP violation or police arrest in between. By adding these cases back in which were previously removed, an additional 1,070 cases are included in the dataset before matching (n=13,004). Using the same matching parameters (i.e., one-to-one nearest neighbor matching with replacement and a caliper of 0.01), an additional 668 treatment group cases are retained for matching (a 43% increase in matched treatment group cases), and an additional 365 control group cases are retained for matching (a 33% increase in matched control group cases). A total of 2,210 treatment group cases are matched with a total of 1,468 control group cases. Post-matching balance shows good balance on all covariates, with none of the covariates showing statistically significant differences between the treatment and control groups, and none showing a standardized bias score above 0.20.

Scenario #1 in Table 4.10 shows the resulting estimated effect sizes from the propensity score matching model after relaxing the data exclusion criteria. The results show generally larger effect sizes than the original propensity score matching model results. Specifically, the overall recidivism rate differences between the treatment and control groups under this scenario of relaxing the data exclusion rules are generally around twice as large as the recidivism rate differences from the original propensity score matching results. The 6-month overall recidivism rate is 7.0 percentage points lower for the treatment group than for the control group (27.8% versus 34.8%), compared to the original estimate of a 2.1 percentage point lower overall recidivism rate for the treatment group than the control group (30.9% versus 33.0%). The 1-year overall recidivism rate is 8.3 percentage points lower for the treatment group than for the control group (44.6%% versus 52.9%%), compared to the original estimate of a 5.2 percentage point lower overall recidivism rate for the treatment group than the control group (49.2% versus 54.4%). The 3-year overall recidivism rate is 12.3 percentage points lower for the treatment group than for the control group (65.0% versus 77.3%), compared to the original estimate of a 6.0 percentage point lower overall recidivism rate for the treatment group than the control group (72.5% versus 78.5%). Similar to the original model, re-incarceration rate differences appear to drive the lower overall recidivism rates for the treatment group. Also similar to the original model, re-arrest rates generally show non-significant differences. So at least under this sensitivity analysis, it not only confirms a deterrent impact of incarceration for a first TPV sanction, but it also generally shows a larger deterrent impact than previously estimated using the primary sample group.

Scenario 2: Only High Severity TPV Violations

Another type of sensitivity analysis that can be performed is to limit the analysis sample to only those TPVs who have a high severity violation for their first TPV violation. As has been previously demonstrated, the estimated ATT effect is largely limited to high severity violations given that high severity violations are much more prevalent in the matched group than in the unmatched group from the original sample. Scenario #2 in Table 4.10 also shows estimated effect sizes from a propensity score matching model limited to only those TPV cases that had a high severity TPV violation for their first violation. One-to-one nearest neighbor matching with replacement and a caliper of 0.01 is once again used. Good balance after matching is once again achieved. A total of 937 control group cases are matched with 1,376 treatment group cases. The results are similar to the primary results from the original propensity score matching model. Overall recidivism rates are significantly lower for the treatment group than for the control group at the 6-month, 1-year, and 3-year follow-up, with slightly larger effect size estimates when compared to the original propensity score matching estimates. The 6-month overall recidivism rate is 7.4 percentage points lower for the treatment group than for the control group, the 1-year overall recidivism rate is 8.4 percentage points lower, and the 3-year overall recidivism rate is 8.7 percentage points lower. Once again, it appears that significantly lower re-incarceration rates for the treatment group are primarily driving the lower overall recidivism rates, and re-arrest rates are not significantly different between the treatment and control group.

Scenario 3: Limiting TPV Sanction Types Among Control Group

One other way to examine the robustness of these findings is to create a more homogenous group within the control group by limiting the control group to only certain sanctioning options for which to compare to incarceration among the treatment group. Since it was previously demonstrated that a high percent of the control group cases retained for matching in the primary analysis were sanctioned to either a “Halfway Back” non-secure residential community corrections center placement or to inpatient treatment, it may be instructive to limit the control group sample to only those who receive one of these two sanctions. A propensity score matching model is thus generated, comparing the treatment of imprisonment for a first TPV to a comparison group consisting of either “Halfway Back” placement or inpatient treatment placement for a first TPV. One-to-one nearest neighbor matching with replacement and a caliper of 0.01 is once again utilized, and post-matching balance is once again achieved. A total of 755 control group cases and 1,536 treatment group cases are retained for matching. The results are presented in Scenario #3 in Table 4.10. Results once again show lower overall recidivism rates for the treatment group than for the control group, with a 4.6 percentage points lower 6-month recidivism rate, a 6.6 percentage points lower 1-year recidivism rate, and a 7.1 percentage points lower 3-year recidivism rate. Once again, significantly lower re-incarceration rates also appear to drive the lower overall recidivism rates. One different result from this model compared to the primary analysis model, however, is that the re-arrest rate at the 6-month mark is significantly higher for the treatment group than for the control group. The re-arrest rate at the 1-year mark is also significantly higher for the treatment group than for the control group, but this was similarly found in the original primary analysis model. Only the 3-year re-arrest rate shows no statistically significant difference between the treatment and control group.

Rosenbaum Bounds

Another way to look at a sensitivity analysis is to ask what type of impact an unobserved variable or set of variables would need to have in order to alter the previously outlined findings. While propensity score matching can do very well at creating suitable comparisons between the treatment and control groups on observed and measurable covariates, unobserved factors may still impact both the treatment assignment and the outcome, and thus lead to remaining hidden selection bias. A randomized controlled experiment has the advantage of balancing on all group differences, both observed and unobserved. Propensity score matching models are limited to balancing only on observed covariates. In situations such as the one in this study where there is a strong understanding of the factors affecting treatment assignment, remaining hidden bias is theoretically less of a concern. Nonetheless, it is informative to consider the potential impact of remaining hidden selection bias.

Existing statistical procedures for performing sensitivity analysis after propensity score matching allow the user to simulate what impact potentially remaining hidden bias would need to have in order to alter findings. One available procedure within Stata for doing this is the ‘mhbounds’ procedure (Becker & Caliendo, 2007). The ‘mhbounds’ procedure is essentially an extension of the Rosenbaum bounds calculation (Rosenbaum, 2002). Essentially, Rosenbaum bounds evaluate the sensitivity of the observed effects under different scenarios where the magnitude of an unobserved confounder varies. A bound estimate of $\Gamma=1$ is equivalent to no hidden bias. Bound estimates higher than 1 represent varying degrees to which the treatment effect may be under-estimated or over-estimated based on unobserved confounders. As an example, a Γ=1.2 represents a situation where there is hidden bias which would increase the odds of receiving the treatment by 20% compared to the control group. An associated *p* value with each value of Γ indicates the probability of accepting the null hypothesis, which is that the effect size remains statistically significantly different between the treatment and control group. At the point where Γ has an associated *p* value above the critical level of p=.05, this would indicate the critical level of the associated bias which would render the effect size no longer significantly different if present.

Table 4.11 presents Rosenbaum bounds to show the robustness of the primary propensity score matching results in the presence of remaining hidden bias. Overall recidivism rates are fairly sensitive to hidden bias. A hidden bias that increases the odds of being sanctioned to imprisonment for a first TPV by between 5% and 10% would render the 1-year and 3-year overall recidivism treatment effects no longer significant at $p < .05$, meaning that it could no longer be assumed that TPVs sanctioned to imprisonment have a lower overall recidivism rate. A hidden bias increasing the odds of treatment by between 35% and 50% could lead to a reverse in the sign of the effect size for the 1-year and 3-year overall recidivism rates, where it might be concluded that imprisonment for a first TPV actually leads to significantly higher overall recidivism rates compared to an alternative sanction for a first TPV. Since the baseline estimated effect size for the 6-month overall recidivism rate was not statistically significant to begin with, Table 4.11 shows that a hidden bias increasing the odds of differential treatment assignment by 10% could lead to a significantly lower 6-month overall recidivism rate for the treatment group, and a hidden bias increasing the odds of differential treatment assignment by 30% could lead to a significantly higher 6-month overall recidivism rate for the treatment group.

Re-incarceration rates, on the other hand, are fairly robust and insensitive to hidden bias based on Rosenbaum bounds. It would generally take a hidden bias that increases the odds of treatment by between 30% and 60% to render the estimated lower re-incarceration rates for the treatment group no longer statistically significant. It would take a hidden bias that increases the odds of treatment by between 75% and 110% in order to potentially reverse the sign of the effect size for the re-incarceration rates, and lead to a conclusion that re-incarceration rates are in fact significantly higher for the treatment group of those imprisoned for a first TVP versus an alternative sanction.

Re-arrest rates at baseline, based on the original propensity score matching model results, were significantly higher for the treatment group at the 6-month and 1-year follow-up mark, and did not significantly differ at the 3-year follow-up mark. The Rosenbaum bound results show that these estimates are fairly sensitive. For the 6-month and 1-year re-arrest rates, a hidden bias increasing the odds of differential treatment assignment by between 5% and 10% would lead to the conclusion that these re-arrest rates no longer significantly differ between the treatment and control groups. A hidden bias increasing the odds of differential treatment assignment by 50% would reverse the sign of the effect size and lead to the conclusion that 6-month and 1-year re-arrest rates are actually significantly lower for the treatment group. For the 3-year re-arrest rates, where the baseline re-arrest rates did not significantly differ between the treatment and control groups, a 10% increase in the odds of differential treatment assignment could lead to the conclusion that 3-year re-arrest rates are significantly higher for the treatment group, and a 25% increase in the odds of differential treatment assignment could lead to the conclusion that 3-year re-arrest rates are significantly lower for the treatment group.

To summarize the Rosenbaum bound estimates, overall recidivism rates and re-arrest rates are generally fairly sensitive to hidden bias, whereas re-incarceration rates are generally fairly robust and insensitive to hidden bias. An examination of the odds ratios for the individual covariates from Table 4.4 (which is the logistic regression model used to generate the original propensity scores by predicting the probability of treatment assignment using the observed covariates) shows that around two-thirds of the observed covariates have odds ratios of similar size to those for which the Rosenbaum bounds show that an unobserved bias would need to have in order to render the significant re-arrest rate and overall recidivism rate differences insignificant. Therefore it is certainly possible that an unobserved covariate with a moderately sized impact on treatment assignment could render null differences in recidivism for re-arrest rates and overall recidivism rates. On the other hand, it has already been demonstrated that re-incarceration rates appear to contribute heavily to the lower estimated overall recidivism rates for the treatment group, and re-incarceration rates are found here to be fairly robust and insensitive to hidden bias. Further, as previously discussed, given that there appears to be a fairly good understanding of the factors affecting treatment assignment based on the strength of the relationship between some of the observed covariates and the probability of treatment assignment, it is unlikely that there exists a set of unobserved factors that are significantly related to treatment assignment which could thus alter the effect size estimates.

Further Exploring The Treatment Effect: Deterrence, Aging, or Exposure Time?

Based on all of the evidence presented so far, it appears that sanctioning a first TPV violation to imprisonment has a slight deterrent effect on overall recidivism when compared to an alternative sanction such as treatment or a non-secure residential placement, and that these results are fairly robust and primarily driven by significantly lower re-incarceration rates for the treatment group. The last type of sensitivity analysis that is examined here is to explore deeper this apparent deterrent effect on overall recidivism of sanctioning first TPVs to imprisonment. One plausible theory is that the lower overall recidivism rates for sanctioning to imprisonment do not represent a deterrent effect but rather an aging or exposure time effect. It is well established in the criminological literature that age is inversely related to crime (and recidivism). Recidivism follow-up for the control group of those sanctioned to an alternative to imprisonment begins immediately at the time of the sanction, given that all control group sanctions involve sanctions where the parolee is not securely confined and thus is at risk for potentially re-offending immediately. On the other hand, recidivism follow-up for the treatment group only begins after the parolee is re-released from imprisonment since the parolee is confined for the TPV sanction. The average confinement time for the TPVs in the treatment group who are imprisoned for a first TPV is 12.3 months. Thus, on average, the TPVs in the treatment group are one year older at the time of the start of recidivism follow-up. Since this aging occurs after treatment assignment, it is not considered in the balancing of observed covariates between the treatment and control groups.[6] Therefore, those in the treatment group may show lower overall recidivism rates

[6] The age variable used among the list of covariates for predicting treatment assignment represents the age at time of the first TPV violation, not the age at the time of the start of recidivism follow-up.

simply because they have aged by a year on average before being followed for recidivism. In a sense, this aging might be considered as part of the treatment itself. If this aging fully accounts for the lower overall recidivism rates for the treatment group, however, it might not be fair to call the treatment effect a deterrent effect. Rather it might be more accurate to characterize the treatment effect as an aging effect. Or the treatment effect might represent some combination of an aging effect and a deterrent effect, where overall recidivism rates are lower for the treatment group both because they have aged in prison and because something about imprisonment has changed their individual cost-benefit assessment to where it deters from future recidivism.

Another plausible explanation for why overall recidivism rates are lower for the treatment group than for the control group is that the treatment group is closer to reaching the maximum term of their sentence when recidivism follow-up begins. Parolees are no longer under parole supervision when they reach their maximum sentence date. Since re-incarceration rates are largely driven by technical parole violations, and technical violations are no longer possible once a parolee is no longer under parole supervision, a shorter period of remaining supervision among the treatment group compared to the control group might explain the significantly lower re-incarceration rates and overall recidivism rates for the treatment group. Since it has been previously shown that the lower re-incarceration rates for the treatment group show large effect sizes, are strongly robust, and are the driving force behind the lower overall recidivism rates, it is plausible that overall recidivism rates might be lower for the treatment group due to less remaining time on parole for which to be subject to potential re-incarceration as a result of parole violations. The observed remaining time on parole between the treatment and control groups at the time of the start of recidivism follow-up shows that the treatment group has on average 17.1 fewer remaining months under parole supervision. This differential impact may be referred to as exposure time, since recidivism follow-up begins right away for the control group yet begins on average more than one year later for the treatment group.

One way to examine whether aging and exposure time could in part or fully explain the lower overall recidivism rates for the treatment group is to run a logistic regression model among only the sample retained in the propensity score matching model for matching, in which treatment assignment is used to predict recidivism rates controlling for age and residual time on parole at the beginning of recidivism follow-up. Table 4.12 shows the results of logistic regression models predicting 6-month, 1-year, and 3-year overall recidivism rates, controlling for the two aging factors (i.e., age and residual time on parole at the beginning of recidivism follow-up). The results show that after controlling for age and residual time remaining on parole at the beginning of recidivism follow, the treatment group of those imprisoned for a first TPV no longer shows significantly lower 6-month, 1-year, and 3-year overall recidivism rates at a *p*-critical value of p=.05. Further, the aging and exposure time variables show a significant relationship with overall recidivism rates in the expected direction. An increase in age is significantly related to lower overall recidivism rates, and a shorter remaining time on parole is also significantly related to lower overall recidivism rates. So we find some evidence here that a large part (perhaps all) of the treatment effect observed from the propensity score matching results represents an aging and exposure time effect among the treatment group, rather than a deterrent effect.

CONCLUSION

This chapter presented results of propensity score matching models to examine the treatment effect of sanctioning first TPV violations to imprisonment versus an alternative sanction. An examination of observed recidivism rates before matching generally showed higher overall recidivism rates among the treatment group of those sanctioned to imprisonment. After matching, however, it appears that overall recidivism rates and re-incarceration rates were generally significantly lower for TPVs first sanctioned to imprisonment, suggesting a deterrent effect of imprisonment. Best estimates were that overall recidivism rates after at least one year of recidivism follow-up were around 5 to 6 percentage points lower for the treatment group than for the control group. Lower overall recidivism rates for the treatment group were largely driven by significantly lower re-incarceration rates. Re-arrest rates did not generally show statistically significant differences. These average treatment effects are only generalizable to the treated (i.e., the Average Treatment Effect on the Treated, or ATT), which largely represents a comparison among high risk parolees who are under higher levels of parole supervision and are sanctioned for a high severity TPV violation. Also, the relevant control group types of sanctions for comparing to the treatment group are largely sanctions to a non-secure residential halfway house or to inpatient treatment (although a significant proportion of written warnings are also present within the matched control group too). Propensity score matching was successful at producing a high degree of post-matching balance across all observed covariates, and was generally successful at retaining a significant number of treatment and control group cases in the final matched sample, without resorting to a heavy reliance on replacement among the control group cases for matching. One-to-one nearest neighbor matching with replacement and a caliper of 0.01 generally produced the best propensity score matching models.

A number of different types of sensitivity analyses were undertaken, in order to examine the robustness of these findings. One type of sensitivity analysis involved adding back in cases that were removed based on decision rules for connecting violations to sanctions and for connecting PBPP data to PA DOC data. These results showed similar (and in some instances larger) effect sizes confirming the conclusion of a deterrent effect on recidivism for sanctioning first TPV violations to imprisonment. A second sensitivity analysis limited the matched group to a comparison among only high severity first TPV violations, and also to a comparison of only control group cases who were sanctioned to a halfway house or inpatient treatment. These results again confirmed generally robust findings, meaning lower overall recidivism rates for TPVs first sanctioned to imprisonment. A third sensitivity analysis used Rosenbaum bounds to estimate the robustness of these findings in the presence of remaining hidden bias. This analysis generally showed more sensitive results for overall recidivism rates, yet re-incarceration rates remained fairly insensitive to the existence of remaining hidden bias. Given the fairly robust re-incarceration rate findings, the fact that re-incarceration rates tended to drive overall recidivism rates, and evidence that treatment assignment is fairly well understood here, overall recidivism rates were likely not to be impacted by remaining hidden bias, although this can never fully be ruled out in the absence of a true experimental evaluation design using random assignment to treatment. Overall, lower recidivism rates among the treatment group showed a moderate to high degree of robustness.

Finally, the treatment effect of lower overall recidivism rates for the treatment group was examined in further detail. Specifically, some evidence was present to suggest that the lower overall recidivism rates among the treatment group were largely explained by an aging and exposure time effect, where the treatment group was on average one year older and around 17 months closer to being released from parole supervision at the time of the start of recidivism follow-up. The treatment effect may thus likely be explained in part or in full by an aging and exposure time effect rather than by a deterrent effect, as deterrence is traditionally defined. Therefore, setting aside the aging and exposure time differences between the treatment and control group, sanctioning TPVs to imprisonment may otherwise have a null impact on overall recidivism.

The next chapter will examine the dose-response impact of differential lengths of imprisonment among those TPVs who are in the treatment group, meaning that they were sanctioned to imprisonment for a first TPV violation.

CHAPTER 5: STRATIFICATION MODELS OF DOSE-RESPONSE EFFECTS

INTRODUCTION

This chapter details results from propensity score stratification models intended to answer the second of the two primary questions examined in this study – the dose-response effect on recidivism of longer versus shorter times incarcerated for a TPV violation. The chapter proceeds first by examining baseline differences on all examined covariates. Second, the results of "naïve" regression-based models are presented, predicting differences in recidivism by length of stay in prison for a TPV, while controlling for all covariates. Following the results from the naïve regression-based models, results are presented from propensity score stratification (or sub-classification) models in order to examine recidivism outcomes across different dosages of lengths of stay in prison once more fully accounting for covariate differences that may predict both assignment to differential lengths of stay in prison and recidivism. Propensity score stratification models are also presented with control group TPVs included, in order to examine recidivism rates between different lengths of stay in prison compared to recidivism rates for an alternative sanction to imprisonment for a TPV violation. By comparing dosages of the treatment condition to the control condition, this ties together the results from the previous chapter with the results presented below. Finally, some extended analysis is presented, in order to examine the robustness of the propensity score stratification model results under different scenarios, and also in comparison to the control group of TPVs not sanctioned to imprisonment for a first violation.

BASELINE BIVARIATE COMPARISONS

Prior to examining recidivism outcomes by different lengths of stay in prison among TPVs, it is first important to determine whether there is significant variation across the distribution of length of stay in prison among important covariates that may predict both the length of stay in prison and recidivism itself. If important differences exist, this may lead to selection bias confounding the examination of the causal impact of the length of stay in prison itself on recidivism. Table 5.1 presents simple correlations between each of the full list of covariates examined in the previous chapter, and the continuously measured length of stay in prison (in months) among the treatment group of TPVs sanctioned to imprisonment for a first violation. All of the same covariates previously examined in Chapter 4 for considering the assignment of TPVs to imprisonment versus an alternative sanction are also theoretically relevant predictors of assignment to differential length of stay in prison given a sanction to imprisonment for a TPV violation.

In addition to all of the previously examined covariates, one additional covariate is included in the analysis in this chapter, which is relevant as a predictor of "assignment" to differential lengths of stay in prison. This new covariate measures whether the TPV violation was adjudicated as a "No Recommit Action" (NRA). For TPVs who are sanctioned to confinement in prison, they are entitled to an adjudication hearing for their violation within three months of being sent back to prison. At this hearing, some TPVs will essentially be found "not guilty" and will be released from prison as an NRA. NRAs are disproportionately represented among the shortest lengths of stay in prison, as more fully described below, thus relevant as a predictor of dosage to be included for examination with the other covariates. NRA status only occurs after assignment to prison for a TPV charge and a subsequent adjudication hearing. Thus the NRA status *did not precede* treatment assignment to imprisonment versus an alternative sanction and was not included as a covariate in the previous chapter for examining the impact of imprisonment versus an alternative sanction for a TPV.

Table 5.1 also presents simple correlations between each of the full set of covariates and the three measures of recidivism (re-arrest, re-incarceration, and overall recidivism).

An examination of Table 5.1 shows that nearly half of the covariates (16 covariates) are significantly correlated with time served in prison at $p < .05$. Significant correlations with length of stay in prison include: offense type for the original sentence, the LSI-R criminal risk score, the number of prior programs participated in, whether the parolee was under first time parole versus having been re-paroled for a previous violation, the total parole supervision length, the severity of the instant technical parole violation, the parole district where under supervision, the supervision level, and whether or not the TPV charge was adjudicated as an NRA. Parolees originally sentenced for a violent offense, who have a higher LSI-R criminal risk score, who have participated in more prior treatment programs, who have a longer parole supervision period, who have a high severity TPV violation, who are under a maximum supervision level, and who are supervised in either Allentown or Harrisburg parole district, spend significantly longer lengths of time in prison for a TPV violation. Parolees originally sentenced for a drug offense, who are under first time parole, who have a medium severity TPV violation, who are under a minimum supervision level, who are supervised in either Altoona or Scranton parole district, and who are adjudicated as an NRA for the TPV charge, spend significantly shorter lengths of time in prison for a TPV violation. There are clearly some significant differences across a number of relevant covariates in the length of stay spent in prison for a TPV violation, with longer lengths of stay generally associated with higher risk parolees who are supervised more intensely and are sanctioned for a more serious TPV violation.

Simple bivariate correlations between all of the covariates and the various recidivism measures also show some significant differences, although generally less so than between the covariates and the length of stay in prison examined above. Table 5.1 shows that between 4 and 11 of the covariates are significantly correlated with recidivism at $p < .05$, depending on the recidivism definition. Higher overall recidivism rates are generally associated with younger age, higher LSI-R criminal risk scores, and being under higher levels of parole supervision intensity. Being supervised in the Williamsport parole district appears to be associated with lower overall recidivism rates.

Similar factors are associated with re-arrest rates too. Higher re-arrest rates are generally associated with younger age, being original sentenced in a Class 3 or higher population county, having a shorter parole supervision length, being supervised in Philadelphia, being under higher levels of parole supervision intensity, and being in a community corrections center at the time of the parole violation. Some indication is also present that re-arrest rates are also associated with race. Higher re-arrest rates are generally observed among black parolees, and lower re-arrest rates generally observed among white parolees. Being supervised in the Scranton parole district appears to be associated with lower re-arrest rates.

Higher re-incarceration rates are generally associated with younger age, higher LSI-R criminal risk scores, longer parole supervision lengths, lower severity TPV violations, being on first time parole supervision, and being adjudicated as an NRA for the TPV charge. Being supervised in the Williamsport parole district is generally associated with lower re-incarceration rates.

NAÏVE REGRESSION MODEL PREDICTING RECIDIVISM

Before moving on to present the propensity score stratification models for recidivism rates by dosages of time served in prison, Table 5.2 shows results from "naïve" logistic regression models to predict recidivism, using length of stay in prison in its original form as a continuously measured variable, and controlling for all covariates. As was detailed in Chapter 3, regression-based models hold several disadvantages to propensity score models for adequately drawing causal inference about the treatment effect. The logistic regression models presented in Table 5.2 may thus be thought of as naïve baseline recidivism models to compare to later propensity score models in order to examine if results differ once selection bias is more fully accounted for.

The results of Table 5.2 show that overall recidivism is only associated with the length of time TPVs serve in prison for the 3-year follow-up rate. A one month increase in the length of time a TPV is incarcerated is associated with a 2% percent reduction in the odds of 3-year overall recidivism ($p < .05$). Neither the 1-year nor the 6-month overall recidivism rates are significantly associated with length of stay in prison for a TPV.

Re-arrest rates are not significantly associated with length of stay in prison for a TPV at any of the follow-up time periods. The coefficients are in a positive direction (i.e., longer lengths of stay associated with higher re-arrest rates), but the coefficients do not come close to statistical significance at a conventional p-level of .05.

Re-incarceration rates at all follow-up time periods are significantly associated with length of stay in prison for a TPV. Longer lengths of stay in prison are significantly associated with lower re-incarceration rates at 6-month, 1-year, and 3-year follow-ups. For a 3-year follow-up, a one month increase in length of stay in prison is associated with a 3.8% reduction in the odds of re-incarceration. For a 1-year follow-up, a one month increase in length of stay in prison is associated with a 2.5% reduction in the odds of re-incarceration. For a 6-month follow-up, a one month increase in length of stay in prison is also associated with a 2.5% reduction in the odds of re-incarceration.

DEFINING AND COMPARING TREATMENT DOSAGES

Defining Treatment Dosages

While the length of stay in prison variable is measured continuously (in months) in the original dataset, dosage categories need to be discretized in order to proceed with the propensity score stratification approach used later on. Figure 3.1 shows a histogram of the distribution of lengths of stay in prison in months. Lengths of stay vary from less than one month to just over 45 months. When creating dosage categories, the actual categorization of doses is somewhat arbitrary, with no set *a priori* rule for either the number of dosage categories or the range of data to include within each dosage category.

Considerations for choosing dosage categories include sample size within each dosage bin and practical/policy implications for the selection of dosage cut-offs.

Also, when selecting dosage categories, consideration should be given to the ability to balance across all relevant covariates after propensity score calculation in the later stratification models. The selection of both the number of dosage categories and the number of cases within each category may have an impact on the later ability to adequately balance groups, which is critical to the approach taken here for estimating the casual impact of dosage on recidivism. In this regard, as is detailed later on in this chapter, a number of different dosage categorizations were attempted in order to assist with difficulties that occurred in achieving balance across the covariates as the analysis proceeded. None of the dosage categorization approaches successfully aided in the range of strategies attempted in order to achieve balance. What became clear was that two covariates were essentially impossible to achieve adequate balance on, and therefore, for reasons later described below, the treatment group sample needed to be limited on these two covariates before proceeding with creating dosage categories and moving on to the propensity score stratification analysis. These two covariates were: 1) whether or not the TPV was adjudicated as an NRA, and 2) seriousness level of the TPV violation(s). Before proceeding with creating dosage categories, the treatment sample was limited to: 1) those who were not adjudicated as an NRA for their TPV violation(s), and 2) only TPVs who received high severity violations. Because nearly all of the NRA cases were at the very low end of the distribution of length of stay, and most of the treatment group were sanctioned to imprisonment for high severity TPV violations, the implication that these two limiters had on the treatment group sample size across the full distribution of lengths of stay in prison was minimal. Again, this decision of limiting the treatment group to non-NRAs and to high severity TPV violations is further discussed later on in the discussion of the propensity score stratification models. By removing NRAs and limiting to high severity TPV violations, the total treatment group size decreased by 403 (26% reduction), from n = 1,547 to n = 1,144.

Several dosage categorizations were considered after removing NRAs and limiting the sample to high severity TPV violations. Ultimately the best categorization for achieving the highest degree of balance across all of the covariates in later analysis, while also maintaining adequate sample size within each dosage bin, was to categorize the length of stay into quintiles of five approximately equal size dosage categories.[7] The five dosage categories are defined as 1) 0 to 5.1 months, 2) 5.2 to 8.4 months, 3) 8.5 to 11.8 months, 4) 11.9 to 16.3 months, and 5) 16.4 to 45.3 months. Each dosage category contains approximately 230 TPV cases.

<u>Comparing Covariates Between Dosages Before Propensity Score Stratification</u>

Once the continuously measured dosage variable (length of stay in prison) has been discretized into categories, it is next important to examine differences between the dosage categories on all of the covariates in order to determine the level of balance between dosage categories before proceeding with the propensity score stratification approach used below. This is similar to the pre-matching balance tests which were used in the previous chapter before proceeding with the propensity score matching models.

---

[7] Different dosage categorizations considered included: 1) four dosage categories with an equal number of cases in each, 2) five dosage categories of 0-3 months, 3-6 months, 6-9 months, 9-12 months, and more than 12 months, 3) four dosage categories of 0-6 months, 6-12 months, 12-18 months, and more than 18 months, and 4) six dosage categories with an equal number of cases in each.

Table 5.3 shows comparisons across the quintiles of the length of stay dosages for all of the covariates. One-way ANOVAs are used to measure statistically significant differences between the dosage bins for the covariates. Similar to the differences identified at the beginning of this chapter when looking at the bivariate correlations between the continuously measured time-served variable and all of the covariates (Table 5.1), Table 5.3 shows a significant number of differences between the lengths of stay dosage bins for several of the covariates. Specifically, 10 covariates (more than one-third of the covariates) show statistically significant differences ($p < .05$) between the discrete categories of lengths of stay in prison. Longer lengths of stay in prison are associated with parolees who were originally sentenced for a violent offense and who are serving longer periods of total time under parole supervision. Shorter lengths of stay in prison are associated with parolees who were originally sentenced for a drug offense. Several of the out of balance covariates do not vary progressively or linearly across dosage categories. Interestingly, for several of the covariates, the first dosage category seems to differ substantially from the other four dosage categories, with the other four dosage categories looking fairly similar among one another. For example, approximately 59% of those parolees in the lowest dosage bin are black, while the percent black in the remaining four higher dosage bins only varies between 41% and 45%. Similarly, 36% of those in the lowest dosage bin were in a community corrections center (halfway house) at the time of their first TPV violation, whereas only between 21% and 27% of the parolees in the remaining four higher dosage bins were in a community corrections center at the time of their first TPV violation. Also, approximately 37% of those in the lowest dosage bin were originally sentenced from Philadelphia or Pittsburgh, whereas only between

20% and 29% in the remaining four dosage bins were originally sentenced from Philadelphia or Pittsburgh. While 27% of parolees in the lowest dosage bin were supervised in the Pittsburgh parole district, only between 13% and 19% of parolees in the remaining four dosage bins were supervised in the Pittsburgh parole district. While average LSI-R risk scores show a statistically significant difference between dosage bins, in absolute numerical terms the LSI-R risk scores show little variation, only varying between 25.8 and 27.1 across dosage bins.

Altogether, these differences between dosage levels among many of the covariates clearly indicate potential for selection bias, which may confound an assessment of the causal impact of differential lengths of stay in prison on recidivism. Thus, it is important to first achieve balance across these covariates before moving on to an assessment of the impact of imprisonment dosage on recidivism.

Comparing Observed Recidivism Rates by Dosage before Propensity Score Stratification

Before moving on to the propensity score stratification method for attempt to balance covariates across dosage levels, Table 5.4 presents baseline observed recidivism rates by each of the length of stay dosage categories. Generally it appears that lower overall recidivism rates and re-incarceration rates are associated with longer lengths of stay. The results are less clear for re-arrest rates. Even among overall recidivism rates and re-incarceration rates, differences are minimal across the five dosage bins. Interestingly, similar to the pattern observed previously where covariates significantly differed between the lowest dosage category and the other four dosage categories, observed recidivism rates show somewhat of a same pattern. For instance, the observed 3-year overall recidivism rate in the lowest dosage category is 75.7%, but the observed 3-

year overall recidivism rate in the four higher dosage categories were between 68.1% and 74.3%. Similarly, when looking at observed re-arrest rates, the 3-year re-arrest rate for the lowest dosage category is 61.3%, whereas the 3-year re-arrest rate for the remaining four dosage categories varies between 55.4% and 58.6%. Only 3-year and 1-year observed re-incarceration rates show a pattern resembling a linear association with length of stay in prison, with 3-year and 1-year re-incarceration rates showing generally a 1 to 2 percentage point drop in re-incarceration rates across each successively higher dosage category of length of stay in prison.

PROPENSITY SCORE STRATIFICATION MODELS

In order to estimate the effect of length of stay in prison for a TPV violation on recidivism rates while more fully accounting for potential selection bias, a recent extension of the propensity score methodology is utilized. This methodology, set forth by Zanutto et al. (2005), is referred to as a propensity score stratification (or sub-classification) approach. Details of the propensity score stratification approach were provided in Chapter 3.

Ordinal Logit Model for Generating the Propensity Score

Similar to the traditional propensity score methodology, a single balancing score (i.e., propensity score) is first generated in order to proceed with the propensity score stratification methodology. However, since there are more than two treatment categories (i.e., multiple dosage categories), the propensity score is modeled using an ordinal logistic regression. The propensity score is the likelihood, based on the distribution of covariates, that an individual will be in one of the treatment dosage categories. The goal is to identify an ordinal logit model which generates a propensity score distribution that best balances all of the covariates across the dosage categories. As such, the modeling of the propensity score is an iterative process.

During the iterative process of generating a propensity score, before removing the NRA parolees and limiting the cases to only parole violators who had a high severity TPV violation, a full ordinal logit model was first attempted using main effects for all 32 covariates (n = 1,547). Once a predicted propensity score was generated after running the ordinal logit model, the propensity score was then stratified into quintiles based on the propensity score distribution. Rosenbaum and Rubin (1984) suggest that over 90% of bias due to covariate imbalance can be eliminated by stratifying the propensity score distribution into five equal sub-classes. This is the core of the propensity score stratification approach.

Balance across the covariates was then assessed using two-way ANOVAs (for continuously measured covariates) and logit regressions (for binary covariates), with each covariate as the dependent variable and the dose category and propensity score category as the two predictor factors. A covariate is considered out of balance if there is a statistically significant main effect of the dose or a statistically significant interaction effect between the dose category and the propensity score quintile (Zanutto et. al., 2005). Balance was not achieved using main effects for the full set of 32 covariates in the ordinal logit model, without excluding NRAs or TPV violations other than high severity violations. Since both the main effect of the dose category and the interaction effect between the dose category and the propensity score quintile are tested for statistical significance when assessing post-stratification balance, there are two statistical significance tests associated with each covariate examined. As such, when examining 32 covariates for balance, there are in fact 64 statistical significance tests. We can reasonably assume that perhaps 5% of these statistical significance tests will show statistically significant differences ($p<.05$) by chance alone. Thus, out of 64 statistical significance tests, a tolerance of approximately 3.2 tests could be statistically significant without assuming that the full range of covariates remain out of balance post-stratification. Or in terms of the number of covariates, out of 32 covariates, we might expect 1.5 covariates to be out of balance by chance alone. However, using main effects for all 32 covariates in the ordinal logit model to predict the propensity score, and before removing NRAs and TPV violations that are not high severity violations, 21 of the 64 statistical significance tests showed statistically significant results (indicating lack of balance) post propensity score stratification.[8] This translates into 12 covariates remaining out of balance, which is significantly higher than would be expected by chance along, thus indicating that the dosage categories remain poorly balanced.

When dosage groups remain out of balance after initial propensity score stratification, several approaches may be utilized as part of an iterative modeling process in order to attempt to create a propensity score distribution that can achieve balance across the range of covariates after stratification. First, as recommended by Rosenbaum and Rubin (1984), the ordinal logit model can be improved by adding quadratic terms and interaction terms involving the out-of-balance covariates. Second, the dosage categories can be revisited, by going back and creating more or fewer dosage categories (e.g., four categories, six categories, etc.), and/or by changing the cut-off points for categories or

[8] For the dosage categories of this first model using all treatment group cases before removing NRAs and limiting to high severity TPV violations, a quintile of five equally sized dosage categories was utilized.

creating unequally sized categories rather than equally sized quintiles. Third, the ordinal logit model may be over-specified with too many covariates, so a pared down set of covariates may be attempted.[9]

All of the above strategies were attempted, in a systematic process, in order to improve the ordinal logit model so as to create a propensity score distribution that when stratified led to optimal balance across all covariates. First, quadratic and interaction terms were created and added back into the ordinal logit model for all combinations of the 12 covariates that remained out of balance after the initial model attempt. The best model after including various combinations of quadratic and interaction terms still left at least 5 covariates (15.6% of the covariates) out of balance. This is significantly more than would be expected by chance alone. Second, reformulation of various categorization of the dosage categories were attempted. Four equally sized dosage categories, six equally sized dosage categories, five dosage categories with an unequal number of cases (0-3 months, 3-6 months, 6-9 months, 9-12 months, and more than 12 months), and four dosage categories with an unequal number of cases (0-6 months, 6-12 months, 12-18 months, and more than 18 months) were attempted. Also, dropping cases in the sample with a length of stay of less than one month, of more than 18 months, and of both less than one month and more than 18 months was attempted, also using various combinations of quadratic and interaction terms for unbalanced covariates in addition to main effects for all of the covariates. The best model among all of these combinations

[9] It is important to note that not all of the covariates need to be used in the ordinal logit model for generating the propensity score, as long as later balance is assessed across all of the covariates, regardless of if they were used in the ordinal logit model or not. The goal is to specify the ordinal logit model which will generate a predicted propensity score that when stratified into quintiles will lead to optimal balance across all covariates.

still left four covariates out of balance (12.5%), which is still more than expected by chance alone. Third, the ordinal logit model was modified for possible over-specification. Specifically, the ordinal logit model was run with only the 12 covariates which remained out of balance after the initial model. The remaining covariates were dropped from the ordinal logit model. This model came close to generating balance after propensity score stratification, with only three covariates remaining out of balance (9.4%). This is still higher than expected by chance alone. The set of 12 covariates in the ordinal logit model were further adjusted, based on the three covariates which remained out of balance, to include quadratic and interaction terms. This still did not improve balance better than three covariates remaining out of balance after propensity score stratification. Further ordinal logit models were attempted, with more or less covariates added in or out of the model, and all sorts of different combinations of quadratic and interaction terms included. This was done iteratively, by running the ordinal logit model, stratifying the predicted propensity score, checking for balance across all of the covariates, and making adjustments back in the ordinal logit model to covariates which remained out of balance. At no point could a model be produced where less than three covariates remained out of balance, which is not of sufficient balance for adequately isolating and addressing potential selection bias before attempting to estimate the causal impact of dosage variation on recidivism.

After the above systematic attempts to improve balance, a further exploration of each of the individual covariates revealed some discoveries. First, parolees who were adjudicated for their TPV violation charge(s) as an NRA (i.e., essentially a "not guilty" finding) showed some interesting patterns. There were 286 NRAs in the full treatment group sample (18.5% of the full treatment group sample). The NRAs had very short lengths of stay in prison when compared against the full distribution of lengths of stay for the full sample of TPVs. Approximately 79% of the NRAs fell into the first quintile of the length of stay dosage, meaning that they spent 2.8 months or less incarcerated. This left only 21% of the NRAs in the remaining four quintiles of dosage categories, with most of the remaining NRAs (80%) in the next lowest quintile (i.e., between 2.9 and 6.2 months served in prison). Further, 77% of all of the cases in the lowest dosage category were NRAs, so NRAs made up the vast majority of the low dosage category. An interesting pattern was also discovered when looking at observed recidivism rates for the NRAs. Specifically, the observed re-incarceration rates for NRAs were significantly higher than for the remaining non-NRA TPVs. For instance the 3-year re-incarceration rate for NRAs was 61.5%, whereas the 3-year re-incarceration rate for non-NRAs was 45.0%. Recidivism differences were only evident for re-incarceration rates (i.e., not for overall recidivism rates or re-arrest rates), but were quite dramatically different for re-incarceration rates. Clearly the NRAs were disproportionately composing a large percent of the shortest dosage category and were also showing significantly higher re-incarceration rates. The re-incarceration rate difference between NRAs and non-NRAs sets up a perfect example of why it is important to balance covariates across the distribution of dosages before examining the causal impact of the dosage itself on recidivism. Obviously a fair amount of selection bias appears present among NRAs. Ideally this would be handled by balancing NRA cases across the full distribution of the dosage bins of lengths of stay in prison, but the problem is that there are so few NRA cases outside of the first dosage category, that it makes balancing on NRA status even after stratification extremely difficult. The NRA status was consistently one of the covariates which remained out of balance regardless of any of the approaches documented above for attempting to modify the ordinal logit model in order to achieve balance. Thus the decision was made to attempt balancing after excluding NRA cases. This essentially also had the impact of changing the range of the dosage bins, since NRAs disproportionately represented the cases in the lowest dosage category before their removal. Thus, the examination of the dose-response impact on recidivism rates is limited after removing NRAs, in what can be concluded about those cases at the very low end of the length of stay distribution. By removing the NRAs and creating new dosage quintiles, the lowest dosage category essentially changes from being between 0 and 2.8 months before removing NRAs, to between 0 and 5.1 months after removing NRAs. After removing NRAs, the same iterative process as was previously described was used in order to attempt to find an ordinal logit model which produced a predicted propensity score than when stratified maximized balance across all of the covariates. Unfortunately, after many iterations, even removing the NRAs could not produce a model leading to remaining imbalance on less than three covariates. The dosage categories thus still remained out of balance by more than what would be expected by chance alone.

The second observation that was made was that the set of variables capturing the severity of the TPV violation for which the parolee was sanctioned to prison were consistently remaining out of balance as well. Three binary variables indicate the severity of the TPV violation: ‘Low Severity Violation’ indicates whether the TPV was a low severity violation, ‘Medium Severity Violation’ indicates whether the TPV was a medium severity violation, and ‘High Severity Violation’ indicates whether the TPV violation was a high severity violation. In most of the attempted ordinal logit models, 'Low Severity Violation' was excluded as the reference category. What was identified was that 89% of all of those in the full treatment group were sanctioned to imprisonment for a high severity TPV violation. This was in fact previously discussed in Chapter 4 when comparing common support between the treatment group and the control group, since a disproportionate percent of TPVs with high severity TPV violations are found in the treatment group of those sanctioned to imprisonment. In fact, only 7 cases in the treatment group (less than 1% of the treatment group) were sanctioned for a low severity TPV violation. While observed recidivism rates between the degrees of TPV violation severity did not differ substantially, and the high violation severity cases were not clustered at one end or another of the dosage categories like the pattern with the NRAs, the very low number of cases without a high violation severity in the total treatment group likely makes it difficult to balance across dosages for the range of TPV violation severity. It was thus attempted to remove the 164 treatment group cases which were not for a high violation severity, thus limiting the treatment group to only high violation severity TPVs for the dose-response examination.

After removing NRAs and limiting the treatment group to only high severity TPV violations, it then became possible to generate an ordinal logit model which achieved maximal balance. In fact, balance was now achieved among all of the covariates in the model. Table 5.5 presents results from the final ordinal logit regression model used to predict the propensity score for stratification. The final ordinal logit model included main effects for 14 of the covariates, quadratic terms for two of the covariates, and six interaction terms between covariates.

Propensity Score Stratification

Based on the results of the final selected ordinal logit model, a predicted propensity score is generated for each case in the sample. As recommended by Zanutto et. al. (2005), the propensity score is then stratified (or sub-classified) into quintiles of five equally sized categories. Table 5.6 shows a tabulation of the number of cases within each dosage (exposure) category by each of the quintiles of the propensity score grouping. Also, Figure 5.1 shows histograms of the distribution of the propensity scores among the five dosage (exposure) categories. Table 5.6 and Figure 5.1 show generally high overlap and a decent number of cases in each cell of the propensity score quintile by dosage quintile tabulation (n in each cell varies from 15 to 88), suggesting that the data will support comparisons across the five dosage categories.

Post-Stratification Examination of Covariate Balance

As discussed previously, once the sample is divided into quintiles based on the stratification of the propensity score generated from the ordinal logit model, it is next important to examine post-stratification balance. This is evaluated using two-way ANOVAs (for continuously measured covariates) and logit regressions (for binary covariates), with each covariate as the dependent variable and the dosage category and propensity score quintile as the two predictors. A covariate is considered out of balance if there is a statistically significant main effect of the dosage categories or a statistically significant interaction effect between the dosage categories and the propensity score quintiles. Table 5.7 shows results from the ANOVAs and logit regressions to test post-stratification balance for each of the covariates, and also comparing to pre-stratification balance for each covariate. As is seen in Table 5.7, after propensity score stratification based on the final produced ordinal logit model (see above description for iterations of the ordinal logit model), none of the covariates remain out of balance. Thus, a level of balance at least as good as (if not better than) what would be expected by randomized dosage assignment is achieved, since by chance alone approximately 1 or 2 covariates would show imbalance at $p < .05$.

<u>Dose-Response Curves and Recidivism Rates</u>

Now that balance is achieved on all covariates through propensity score stratification, the effect of length of stay in prison on recidivism rates can be assessed by creating a dose-response curve. First, the average likelihood of recidivism (using the various definitions of recidivism; a separate dose-response curve is generated for each recidivism definition) is estimated. The average likelihood of recidivism estimate is simply the average recidivism rate for each dosage category, weighted by the size of the propensity score quintile. These estimates are calculated as:

$$\bar{Y}_d = \sum_{i=1}^{5} \frac{1}{5} \bar{Y}_{d,i}$$

where, $\bar{Y}_{d,i}$ is the observed mean outcome among individuals receiving dosage level, *d*, in balancing score quintile, *i*. A common estimate of the corresponding standard error (Zanutto et al., 2005) is calculated as:

$$SE(\bar{y}_{di}) = \frac{1}{5}\sqrt{\sum_{i=1}^{5} \frac{s_{di}^2}{n_{di}}}$$

where, s^2_{di} and n_{di} are the sample variance and frequency, respectively, among individuals in treatment dose level *d* in propensity score quintile *i*.[10]

Figure 5.2 shows separate dose-response curves for each of the recidivism definitions and follow-up time periods (i.e., 6-month, 1-year, and 3-year overall recidivism, re-arrest, and re-incarceration rates). Also, Table 5.8 provides the numerical outputs (average recidivism rates and standard errors) associated with the graphical presentation of the dose-response curves in Figure 5.2. Overall recidivism rates generally show a slight downward trend with successfully longer lengths of stay in prison but some of the point estimates for the overall recidivism rates have associated standard errors that overlap across dosage categories, suggesting that the downward trend in overall recidivism rates may not be statistically significant. For the 3-year overall recidivism rates, the lowest and highest dosage categories have associated standard errors which do not overlap, suggesting that the longest dosage category has a statistically significant lower overall recidivism rate than the shortest dosage category. The standard error is wide for the longest dosage category for the 3-year overall recidivism rate though, so it may be that the longest dosage category is being influenced by extremely long lengths of stay in prison which are outliers. This is further explored later on. For the 1-year overall recidivism rates, the standard errors for all of the dosages overlap except for the highest dosage category, suggesting little statistically significant difference in overall recidivism rates between successively longer dosages except for a drop in recidivism associated with the longest dosage. For 6-month overall recidivism rates, the higher four dosages appear

[10] One note, while the standard error estimate has been found to be a reasonable approximation, as noted by Zanutto et al. (2005), it is not totally unbiased due to the fact that the sub-classification is based on propensity scores, which are estimated from the data, and the outcomes, both between and within each sub-class, are not independent.

to have a significantly lower recidivism rate than the lowest dosage level, based on non-overlapping standard errors.

Re-arrest rates generally show a quite flat pattern across dosages of length of stay in prison, with no indication of statistically significant differences in re-arrest rates between dosage categories. Standard errors overlap for all re-arrest rates among all three follow-up periods (i.e., 3-year, 1-year, and 6-month re-arrest rates).

Re-incarceration rates generally show the clearest evidence of a downward trend across successfully higher dosages of lengths of stay in prison, suggesting a deterrent effect on re-incarceration rates for longer lengths of stay in prison. Important to note though is that the standard errors for most of the re-incarceration rates are quite wide, and are generally larger than the standard error rates for re-arrest rates and overall recidivism rates. Most of the standard errors overlap between the different dosages for the 3-year and 1-year re-incarceration rates. For 6-month re-incarceration rates, the re-incarceration rate in the lowest dosage category is between 6.5 and 12 percentage points higher than for the other successively higher dosage categories, and most differences are statistically significant (based on non-overlapping standard errors).

Another way to examine the impact of dosage length on recidivism rates post propensity score stratification is to run logistic regressions for each of the recidivism measures, predicting the likelihood of recidivism with treatment dose category (treated as continuous) and the propensity score quintile as predictors. In addition, a two-way ANOVA analysis can be conducted (treating the dosage categories as ordered categorical data rather than continuously measured), with both treatment dosage and propensity score quintiles again used to predict recidivism rates based on each of the definitions of recidivism. Table 5.9 shows the results from these logit regression and ANOVA models. For overall recidivism rates, the logit and ANOVA models generally confirm no statistically significant differences in overall recidivism rates after propensity score stratification. Only the 3-year overall recidivism rate based on the logit model (not for the ANOVA model) shows a statistically significant reduction in recidivism with progressively longer lengths of stay in prison at $p < .05$. No statistically significant results are found for the 1-year or 6-month overall recidivism rates, nor for the 3-year overall recidivism rate based on the ANOVA model. This generally confirms observations from the dose-response curves.

For re-arrest rates, none of the logit or ANOVA models for the 3-year, 1-year, or 6-month re-arrest rates show a statistically significant coefficient for dosage. The p-values for all of the re-arrest rate coefficients are far from reaching statistical significance, thus again confirming the visual inspection of the dose-response curves showing no difference in re-arrest rates by length of stay in prison.

For re-incarceration rates, all of the logit and ANOVA models show statistically significant differences across dosages at $p < .05$, except for the 6-month re-incarceration rate from the ANOVA model. These results confirm the findings from the dose-response curves that longer lengths of stay in prison appear to produce lower re-incarceration rates at all follow-up points.

In summary, these results suggest that the effects of length of stay in prison on recidivism after release, prior to propensity score stratification, are confounded by pre-existing differences between TPV violators. Once controlled for, there is little relationship between length of stay in prison and recidivism rates, except for re-incarceration rates which are generally reduced with longer lengths of stay in prison.

EXTENDED ANALYSIS

This chapter concludes with supplemental analyses of the propensity score stratification models examined above, in order to further understand and contextualize the findings described above. For economy of space, and because to this point results have not substantially varied by the recidivism follow-up period within each measure of recidivism, only 3-year recidivism rates (overall recidivism, re-arrest, and re-incarceration) are examined in the below analysis.

Comparison of Treatment Doses to the Control Group

The first extended analysis which is performed is to compare the recidivism rates of the various doses of lengths of stay in prison among the treatment group TPVs, to the control group of TPVs who are not sanctioned to imprisonment. This analysis essentially ties together the findings from this chapter and the previous chapter (Chapter 4), by contextualizing dosage differences in recidivism rates within the context of recidivism rates for the control group of TPVs who are not sanctioned to imprisonment for a first TPV violation. In order to do this, the propensity score stratification method will again be utilized. This time the control group is added as the first category within the exposure categories. Since the control group cases essentially represent a prison exposure time of 0 because they are not sanctioned to imprisonment and they are thus at exposure for recidivism immediately, they becomes the first "dosage" category. The treatment group is then stratified into quintiles exactly as performed above in the primary analysis. There are therefore now six exposure categories (1 = control group; 2 = more than 0 to 5.1 months in prison, 3 = 5.2 to 8.4 months in prison, 4 = 8.5 to 11.8 months in prison, 5 = 11.9 to 16.3 months in prison, and 6 = 16.4 or more months in prison).

As with the models examined above, NRAs are excluded, and only high severity TPV violations are included in both the control group and the treatment group doses. In addition, because the control group is so much larger than the treatment group (control group n = 3,269; more than two times the entire treatment group), a random sample of 10% of the control group cases were drawn for this analysis. Because conventional statistical significance tests are used to assess balance of the covariates across dosage categories after propensity score stratification, and because conventional statistical significance tests are influenced by sample size, it was virtually impossible to attain balance across enough of the covariates without limiting the control group sample to a random 10% draw, even though the absolute numerical differences across the dosage categories for the vast majority of the covariates were minimal. Smaller absolute differences show statistical significance as sample size increases.

The final ordinal logit model included main effects for all of the covariates, two quadratic terms, and five interaction terms. After the propensity score was stratified, ANOVA and logit regressions run to show post-stratification balance revealed that only one of the covariates remained out of balance (3.4% of the covariates), which is what might be expected by chance alone. The remaining out of balance covariate was the total length of time supervised on parole (in months). Even though the this covariate remained out of balance ($p < .05$), absolute numerical differences were generally less than 10 months between dosage categories within each propensity score quintile. Differences on this covariate were thus minimal, even though statistically significant. Again, this covariate may be out of balance by chance alone. It is thus safe to conclude that balance is achieved after the best propensity score stratification model.

Figure 5.3 presents dose-response curves comparing the control group to the five dosages of length of stay in prison among the treatment group. Table 5.10 also presents the numerical outputs of recidivism rates and standard errors associated with the dose-response curves. For 3-year overall recidivism, the control group recidivism rate is higher than the recidivism rate among all doses of lengths of stay in prison in the treatment group. The control group recidivism rate is 4.6 to 11.8 percentage points higher than the treatment dosages.

Re-arrest rates appear to be consistently flat between the control group and each of the five treatment group dosage categories. Thus there is no evidence here that the 3-year re-arrest rates significantly differ between the control group and any of the treatment dosages for a TPV violation. The absolute value of the 3-year re-arrest rates for three of the treatment group dosages shows higher re-arrest rates than the control group, but again these differences are not statistically significant.

Re-incarceration rates show a generally large downward trend among the treatment group across successively higher dosages when compared to the control group. All treatment group dosages show statistically significant lower 3-year re-incarceration rates than the control group, by between 12 and 20.5 percentage points. None of the treatment group dosages except for the longest dosage show significantly different 3-year re-incarceration rates among one another, however, even though all are significantly lower than the control group.

Controlling For Aging and Exposure Time

In the previous chapter, it was explored as to whether the reduced overall recidivism rates among the treatment group might be related more to aging and exposure time mechanisms rather than to deterrence mechanisms. Recidivism follow-up for the control group of those sanctioned to an alternative to imprisonment begins immediately at the time of the sanction, given that control group sanctions involve those where the parolee is not securely confined and thus is at risk for potential recidivism immediately. On the other hand, recidivism follow-up for the treatment group only begins after the parolee is re-released from imprisonment. Depending on how long the term of imprisonment for the TPV violation is, a parolee might age by a year or more before being released. It is well established that recidivism declines with age. Since this aging occurs after treatment assignment, it is not considered as a covariate to balance when predicting treatment assignment. Similarly, treatment group parolees are closer to reaching their maximum sentence date and thus being discharged from parole supervision. A large mechanism for recidivism (i.e., return to prison for a parole violation) is taken away once a parolee is no longer under parole supervision. These two explanations were referred to in the last chapter as aging and exposure time effects, in contrast to a deterrent effect as deterrence is typically conceptualized in the criminological literature. In the previous chapter, it was found that once controlling for the age at the time of the start of recidivism follow-up, and for residual supervision time left at the time of the start of recidivism follow-up, overall recidivism rates between the treatment and control groups were no longer statistically significant. This suggests that the "deterrent" impact of imprisonment might actually be largely due to an aging and exposure time effect.

This similar logic applies to the current examination of the dose-response impact of differential lengths of stay in prison for a TPV on recidivism. Longer lengths of stay in prison might show lower recidivism rates because of aging and exposure time rather than deterrence mechanisms. In order to explore this possibility, covariates measuring the age at time of release from prison and the residual time remaining under parole supervision are added as predictors to the logit model generated after the primary propensity score stratification model presented previously, in order to predict recidivism rates while controlling for the propensity score quintile. The only overall recidivism rate that differed significantly across dosages from the primary models presented before was the 3-year overall recidivism rate, based on the logit model (the 3-year overall recidivism rate did not show statistically significant differences based on the ANOVA model). After running the logit model to control for age at time of release from prison and residual time under parole supervision, the dosage no longer predicts 3-year overall recidivism ($p = .20$). Both the age at time of release ($p < .01$) and residual time left under parole supervision at the time of release ($p < .01$) were significant predictors of 3-year overall recidivism, with a shorter residual time under parole supervision and older age both significantly related to a lower 3-year overall recidivism rate. None of the re-arrest models showed statistically significant differences across dosages of length of stay in prison based on the primary models previously presented. The dosage remained consistently insignificant for all re-arrest rates after controlling for age at time of release from prison and residual time remaining under parole supervision. Based on the logit models, all three follow-up periods for re-incarceration rates showed significantly lower re-incarcerations rates for successively longer dosages of length of stay in prison, based on the primary models previously presented. After controlling for age at time of release and residual parole supervision length, all three measures of re-incarceration rates no longer reached statistical significance (p = .016 for 3-year re-incarceration; p = .032 for 1-year re-incarceration; p = .047 for 6-month re-incarceration). Older age and shorter residual lengths of time under parole supervision at time of release from prison were once again significantly related to lower re-incarceration rates. Altogether, this analysis once again suggests that aging and recidivism exposure time are the primary "treatment" mechanism at work rather than traditionally formulated deterrence. It appears that where longer times in prison are related to lower recidivism rates, aging and exposure time may largely explain these differences.

Censoring Outliers of Long Exposures

As a final sensitivity analysis, it is explored whether censoring the right tail of the distribution of length of stays in prison, by removing very long stays in prison for a TPV, will generate similar or different results. This may be another way to attempt to disentangle aging effects, since by removing extreme outliers of long dosages will reduce the extent of aging within the highest dosage category. Currently the highest dosage category in the main propensity score stratification model varies from 16.4 to 45.3 months. This is a quite wide range of lengths of stay in prison compared to the lower four dosage categories. Further, on the surface, parole violators sanctioned to imprisonment should likely not be spending years in prison simply for TPV violations. Extremely long dosages might represent outliers where criminal behavior is actually present along with the technical violation(s), and yet was not identified or detected in the arrest records and violation data for removal from the sample.

This analysis limits the maximum length of stay in prison to 18 months, and drops all cases with a length of stay of longer than 18 months in prison. Once cases with lengths of stay of longer than 18 months are dropped, the dosage covariate measured continuously (in months) is once again discretized into quintiles of five equally sized dosage categories. NRAs and low/medium severity TPV violations are once again dropped in order to achieve later balancing across covariates. The new dosage categories are: 1 = 0 to 4.5 months; 2 = 4.6 to 7.3 months; 3 = 7.4 to 10.4 months; 4 = 10.5 to 13.4 months; and 5 = 13.5 to 18 months. The best fitting ordinal logit model for predicting the propensity score in order to stratify and achieve balance was the same model as used in the primary analysis above (see Table 5.5 for the covariates included in the ordinal logit model), including 14 main effect covariates, 2 quadratic terms, and 6 interaction terms. The predicted propensity score from this ordinal logit model was then stratified, and post-stratification balance was assessed. Only one covariate remained out of balance at $p < .05$, which is still better than chance alone. The one remaining covariate out of balance was the total parole supervision length. While out of balance at $p < .05$, the absolute numerical differences across dosage categories in average parole supervision lengths varied minimally, with a maximum difference of 10 months in the average parole supervision lengths between dosage categories across propensity score quintiles. It is thus fair to assume that the dosage categories were balanced after propensity score stratification.

Figure 5.4 presents the dose-response curves for the 3-year recidivism rates for the five dosage categories after limiting the sample only to doses under 18 months. Table 5.11 also presents the numerical outputs of 3-year recidivism rates and standard errors associated with the dose-response curves. For 3-year overall recidivism, only the first dosage has non-overlapping standard errors with other dosages. Otherwise the 3-year overall recidivism rates are fairly flat across dosages, with overlapping standard errors. Logit regression and ANOVA models, controlling for the propensity score quintile, show non-significant differences between dosages (p = .268 for logit; p = .416 for ANOVA).

For 3-year re-arrest rates, no statistically differences are identified between dosages, and re-arrest rates remain remarkably flat across lengths of stay in prison To confirm this, logit regression and ANOVA models, controlling for the propensity score quintile, show non-significant differences between dosages (p = .771 for logit; p = .792 for ANOVA).

For 3-year re-incarceration rates, the re-incarceration rates are more flat across lengths of stay in prison by dropping cases with a length of stay above 18 months, when compared to the main models presented earlier, except for the longest dosage category (13.5 to 18 months). However, logit regression and ANOVA models, controlling for the propensity score quintile, still show statistically significant differences between dosages (p = .004 for logit; p = .026 for ANOVA). These statistically significant differences from the logit and ANOVA equations appear to be influenced by the highest dosage category (13.5 to 18 months). The re-incarceration rate for the highest dosage category is 35%, whereas the re-incarceration rates for the lower four dosage categories range from 45.1% to 51.7%. Other than this drop in re-incarceration rate from the fourth to the fifth dosage category, the dose-response curve in Figure 5.4 shows a fairly flat re-incarceration rate between doses.

To summarize, results do not substantially differ by removing cases with more than 18 months in prison for a TPV. There appears to be some influence on re-incarceration rates, leading to slightly flatter re-incarceration rates than were seen in the main models previously presented, but otherwise very little difference.

CONCLUSION

This chapter examined the dose-response impact of differential lengths of stay in prison for a first TPV violation on recidivism rates post-release. In order to address selection bias impacting the assignment to differing lengths of stay in prison for a TPV violation, a recent extension of the propensity score modeling approach was utilized – the propensity score stratification approach (Zanutto et al., 2005). Before propensity score stratification, balance was assessed among a large set of covariates that are theoretically relevant for impacting the assignment of differing lengths of stay in prison for a TPV violation. By examining simple bivariate correlations between each of the covariates and the continuously measured length of stay in prison (measured in months), approximately half of the covariates showed statistically significant correlations with length of stay in prison. After creating discrete dosage categories (quintiles) of length of stay in prison, and running logit and ANOVA models to examine pre-propensity score stratification balance, approximately one-third of the covariates were out of balance. Clearly there exists a potentially large degree of selection bias in assignment to differential lengths of stay in prison for a TPV violation, which may confound examination of the causal impact of exposure time (i.e., length of stay in prison) on recidivism.

Naïve logit regression models predicting recidivism rates using continuously measured length of stay in prison while controlling for all the other covariates, showed that the 3-year overall recidivism rate was significantly lower for longer lengths of stay in prison. No statistically significant association between length of stay in prison and re-arrest rates were identified from the naïve regression models. Contrarily, lower re-incarceration rates at all follow-up time periods were statistically significantly associated with longer lengths of stay in prison.

When proceeding to propensity score stratification, it became extremely difficult to balance the set of covariates across dosage categories. It was observed that a very large percent of the full treatment group were sanctioned for high severity TPV violations, and a very large percent of those cases in the lowest dosage category were re-released under a 'No Recommit Action" (NRA) status. Nearly all of the NRAs were in the lowest dosage category. Further, NRA's had significantly higher re-incarceration rates. Propensity score stratification was thus attempted after removing NRAs and low/medium severity TPV violations from the full treatment group. After removing NRAs and low/medium TPV violations, balance was achieve on all of the covariates. After removing NRAs and low/medium severity TPV violations, five dosage categories of lengths of stay in prison were created, which roughly equated to 0-5 months, 5-8.5 months, 8.5-12 months, 12-16 months, and more than 16 months.

After propensity score stratification, overall recidivism rates appeared to drop slightly with progressively longer lengths of stay in prison, but these differences were generally not statistically significant. Re-arrest rates were remarkably flat across progressive doses of length of stay in prison, showing no statistical differences. Re-incarceration rates trended downwards with progressively longer lengths of stay in prison, with mixed but stronger evidence that this downward trend was statistically significant. Crossover in standard errors for some of the doses in the dose-response curves for re-incarceration rates showed no statistically significant differences, but logit and ANOVA models predicting re-incarceration rates controlling for propensity score stratification showed statistically significant differences, with lower re-incarceration rates resulting from longer lengths of stay in prison. It is important to note that these results can only be generalized to TPVs who were sanctioned to imprisonment for their first high severity TPV violation(s), and who were not adjudicated as a NRA upon being sent to prison.

This chapter concluded with several extension analyses. One such extension was to compare imprisonment dosages among the treatment group, to the control group of those who were not sanctioned to imprisonment for a first TPV. Propensity score stratification was again utilized, with the control group representing the lowest "dosage" category. Results showed mixed evidence of differences between the control group and the treatment group dosages for overall recidivism rates, with some evidence that short and medium term lengths of stay in imprisonment for a TPV produced lower 3-year overall recidivism rates than the control group of those not sanctioned to imprisonment. No differences were identified between the control group and the treatment group doses based on re-arrest rates. Re-incarceration rates were generally 12 to 20.5 percentage points lower for all of the treatment group doses when compared to the control group of those not sanctioned to imprisonment.

Another extended analysis was conducted to examine the degree to which aging and exposure time mechanisms might explain some of the apparent lower recidivism rates for longer lengths of stay in prison. From all of the previous models, it appeared that longer lengths of stay in prison deterred TPV violators from recidivating, at least when considering re-incarceration as a measure of recidivism. No evidence of deterrence was present based on re-arrest rates though. One factor related to re-incarceration rates is that it is heavily influenced by the residual length of time remaining under parole supervision. Revocation for a parole violation is a major contributor to re-incarceration rates, which is not possible once a parolee finishes the period of parole supervision. After controlling for age at time of release from prison and the residual parole supervision length at time of release from prison, re-incarceration rates were no longer statistically significantly related to length of stay in prison. This suggests that the mixed evidence possibly suggesting a deterrent effect on re-incarceration rates for longer lengths of stay in prison from the primary models, may in fact not represent deterrence mechanisms as traditionally formulated, but may instead reflect aging and exposure time mechanisms.

Finally, a sensitivity analysis was performed to see if the primary recidivism results held after removing cases with longer lengths of stay in prison (lengths of stay in prison longer than 18 months). The conclusions from these models confirmed all of the findings from the primary analysis, with very little differences identified. The downward trend in re-incarceration rates for longer lengths of stay in prison flattened out somewhat after removing doses higher than 18 months, perhaps providing more confirming evidence that aging mechanisms might be at work. Overall recidivism rates and re-arrest rates remained unchanged.

Overall, there is little evidence that longer lengths of stay in prison for a TPV violation lead to appreciably lower recidivism rates. Especially for re-arrest rates, it appears that lengths of stay in prison for less than 5 months have the same impact on recidivism as lengths of stay in prion of more than 15 months.

Chapter six concludes with a further discussion of findings and policy implications from the results of both this chapter and the previous chapter. Chapter six also outlines limitations of the current study, and provides recommendations for future research examining the impact of imprisonment on recidivism rates for TPV violators.

CHAPTER 6: SUMMARY AND DISCUSSION

REVISITING STUDY MOTIVATION

A natural result of nearly four decades of build-up of the U.S prison population is that an increasingly large number of inmates are now finishing their prison term and being released back to the community. One mechanism for managing this transition of a large number of inmates from prison to the community is through conditional release to parole supervision. In addition to being monitored for a return to criminal behavior, those under parole supervision are also subject to various supervision conditions which do not constitute criminal behavior if broken, but can nonetheless result in sanctioning up to a return to prison when violated. Parole violators who break these supervision conditions are typically referred to as technical parole violators (TPVs).

In many states, TPVs represent a significant percent of state prison admissions, and a significant contributor to prison population. As states look for ways to contain prison costs and reduce prison population, they are increasingly revisiting the use of imprisonment for TPVs. Several states have greatly reduced the number of TPVs returned to prison, or eliminated the use of imprisonment as an option for TPV sanctioning altogether. One concern with this approach is that virtually no research exists to inform whether specifically sanctioning TPVs to prison has a deterrent, null, or criminogenic effect on subsequent recidivism. While a larger body of literature examining the overall impact of incarceration on recidivism has mostly concluded that imprisonment has a null or even slightly criminogenic effect, this overall finding is not necessarily generalizable to all sub-populations within the prison population. Strong theoretical cases can be made each way, for the impact on recidivism of incarcerating

TPVs. Reducing the use of imprisonment for TPVs might be penny wise but pound foolish if indeed imprisonment deters recidivism among TPVs. Conversely, if imprisonment of TPVs has a null or criminogenic effect on recidivism, reducing the use of incarceration in response to TPVs can reduce prison spending and prison population while at the same time not jeopardizing (and perhaps even enhancing) public safety.

SUMMARY OF FINDINGS

This dissertation examined two primary questions in response to the above set of issues. The first question examined was whether sanctioning TPVs to imprisonment versus any other alternative sanction for a first violation has a deterrent, null, or criminogenic effect on subsequent recidivism. The bulk of the evidence supports the conclusion that recidivism rates are mostly lowered by using incarceration in response to TPV violations. Overall recidivism rates appear to be reduced by 5 to 6 percentage points when TPVs are sanctioned to imprisonment versus another alternative. Lower overall recidivism rates for TPVs sanctioned to imprisonment are especially influenced by lower re-incarceration rates for TPVs sanctioned to imprisonment. In fact re-arrest rates are mostly unaffected by the mode of sanctioning (prison vs. an alternative sanction) for TPVs. Some possible reasons for this difference in findings between re-incarceration rates and re-arrest rates are discussed further below. It should be noted too that this finding is not generalizable to all TPVs. The study focused on sanctioning in response to the first instance of a TPV violation after initial release from prison onto parole. Parolees who committed a new crime after initial release from prison were not included, nor were repeat TPV violators. Further, findings were mostly limited in comparability to higher risk TPVs who were supervised more intensely on parole and who committed a technical violation that was considered more serious. Finally, while overall recidivism rates appear to be lowered among first time high risk TPVs sanctioned to imprisonment, evidence was found to suggest that the specific mechanism for lowering recidivism rates among incarcerated TPVs is largely attributable to aging and exposure time rather than to deterrence. Further discussion of this finding is also provided below.

The second question examined in this study was the impact on recidivism of differential lengths of stay in prison among those TPVs sentenced to imprisonment for a first violation. This type of investigation is often referred to as an investigation into the "dose-response" effect of imprisonment on recidivism. The findings here are more mixed, but suggest somewhat lowered recidivism rates attributable to longer lengths of stay in prison for a TPV violation, yet with contingencies. The effect sizes are generally smaller and in some cases statistically insignificant for the impact of longer lengths of stay in prison on recidivism, when compared to the effect sizes for the impact of imprisonment itself versus an alternative sanction. When recidivism is measured by re-arrest rates, the evidence suggests no impact at all of differential lengths of stay in prison. When directly comparing TPVs sanctioned to an alternative sanction to TPVs sanctioned to varying lengths of stay in prison, recidivism rates look more similar among different lengths of stay in prison than between any time in prison versus an alternative to prison. Finally, where evidence existed that recidivism rates were lowered by longer periods of incarceration for a TPV, it again appears that the particular mechanism for this recidivism reduction is largely aging and a reduced recidivism exposure time rather than traditionally formulated deterrence mechanisms. It should again be noted that these findings are limited in their generalizability to first-time, high severity TPV violations. Further, TPVs sent to prison who were adjudicated "not guilty" of the TPV and re-released (i.e., NRAs) were excluded from the dose-response analysis. Practically speaking, this means that the analysis can say very little about the impact on recidivism of extremely short lengths of stay in prison for a TPV violation.

METHODOLOGICAL ADVANTAGES OF THE STUDY

One of the primary difficulties that this study had to wrestle with was attempting to draw causal conclusions about the impact of incarceration on recidivism among TPVs. Chapter 4 and 5 presented evidence that there were significant pre-existing differences across a large set of relevant covariates between TPVs sanctioned to imprisonment versus an alternative sanction, and between TPVs sanctioned to longer versus shorter periods of confinement in prison for a TPV violation. This evidence shows that TPVs sanctioned to imprisonment, and for longer periods of imprisonment, tend to be higher risk parolees who are supervised more intensely and who are sanctioned for a more severe technical violation. These pre-existing differences can affect both the sanctioning assignment and the recidivism outcome, leading to significant selection bias which confounds the ability to directly tie the impact of the "treatment" itself (i.e., imprisonment vs. an alternative sanction; varying lengths of imprisonment) to the outcome (i.e., recidivism). Ideally a prospective randomized controlled trial would be used to examine the questions in this study, where TPVs are randomly assigned to imprisonment versus an alternative sanction, and, conditional upon being sanctioned to imprisonment, randomly assigned to varying lengths of imprisonment. This was unfortunately not possible here (or likely anywhere), so this study was left with observational data in order to attempt to draw causal conclusions about the impact of imprisonment among TPVs.

This study made use of a relatively robust quasi-experimental approach for attempting to draw a causal inference about the impact of imprisonment among TPVs. Propensity score methods provide several advantages to non-experimental regression-based methods in addressing selection bias. Most important for this study, propensity score methods facilitate creating a more convincing counterfactual by: a) focusing analysis only on the subset of cases within both the treatment and control groups that demonstrate sufficient overlap in the distribution of the covariates that predict treatment assignment, and b) facilitating balance through the use of a propensity score between the matched treatment and control groups over many observed covariates affecting treatment assignment. By focusing only on comparable cases within the treatment and control groups which have sufficient overlap in the distribution of relevant covariates predicting treatment assignment, propensity score models estimate a quantity known as the Average Treatment Effect on the Treated (ATT). By contrast, regression-based models estimate an overall Population Average Treatment Effect (PATE). Regression-based models ignore the concern of whether there is sufficient overlap between the treatment and control groups, and thus can lead to conclusions being drawn which are outside of the range of comparable cases. This produces less convincing counterfactuals, and is also less useful (and possibly misleading) for understanding the policy-relevant group for which results can be accurately generalized to.

In this regard, this study benefited from propensity score modeling by generating a better understanding of the group of TPVs who can accurately be compared to TPVs who are sanctioned to imprisonment, an understanding that regression-based models overlook. It was found that there was a relatively large group of first time TPVs who were sanctioned to an alternative sanction with very little probability of ever receiving a sanction to imprisonment based on observed covariates predicting sanctioning assignment, and that these control group cases did not have similar cases within the treatment group (i.e., cases with a low probability of being sanctioned to imprisonment even though they were in fact sanctioned to imprisonment). Further, there was a smaller group of cases who were sanctioned to imprisonment and were in fact almost guaranteed to be sanctioned to imprisonment based on their very high predicted propensity for receiving an imprisonment sanction, and this group had no comparable cases within the control group. Propensity score matching thus aided a causal inference by focusing the analysis on cases demonstrating "common support", so that there is a clearer understanding of exactly what group of TPVs the findings on the causal impact of imprisonment may be generalizable to. The conclusion here was essentially that the results of this study can be generalized primarily to high risk TPVs who are supervised at a more intense level and who commit a relatively more serious TPV violation. Further, the results are most applicable to comparisons primarily between imprisonment and non-secure residential sanctioning options like inpatient treatment or a non-secure halfway house. Lower level sanctions are less comparable to imprisonment. This is important context for the conclusions of this study, which was facilitated by the use of the propensity score methodology.

One flip side to this, however, is that the population for which the ATT effect size can be generalized may become so small and limited in scope that it loses policy relevance. Some might consider the population average effect size to be the more interesting and policy relevant effect size to estimate. Unfortunately there is a trade-off here between methodological rigor and generalizability. It might be desirable to be able to generalize the results of this study to the full population of TPVs facing a sanction. Notwithstanding, the ATT effect size seems quite policy relevant from a specific deterrence standpoint here, since the treated population in this study represents a significantly large share of parolees, and since it focuses in on the group of TPVs for which imprisonment would actually be seriously considered as a sanctioning response while ignoring TPVs with no real likelihood of being sanctioned to imprisonment.

This study also benefitted from the propensity score methodology by making use of a relatively new and innovative extension of the propensity score framework in order to examine the question of the dose-response impact of imprisonment. The propensity score stratification approach (Zanutto et al., 2005) is relatively new to criminology, but proved useful to this study for building upon the previously mentioned benefits of propensity score matching while also accommodating a situation which traditional propensity score matching is not designed to handle. Specifically, traditional propensity score matching is only designed for examining a binary treatment assignment situation where cases are assigned to either a treatment or control group, yet the second question of this study (i.e., the dose-response question) involved a continuously measured treatment variable of the length of stay in prison (in months). The propensity score stratification approach is an extension of the propensity score methodology, for situations with a continuously measured or multi-categorical measured treatment variable. It is hoped that this study will encourage further use of the propensity score stratification approach within criminological research.

One final interesting observation from the methodology used to answer the questions in this study is that regression-based models essentially produced the same results as the propensity score models. In the analysis used to answer each of the two primary questions in this study, the analysis first started with what was termed a "naïve" regression model, and then proceeded to the primary propensity score models. In both cases, the naïve regression-based models generated similar findings to the propensity score models, although in the case of the dose-response question the evidence of lowered recidivism rates associated with longer imprisonment terms was weaker and less consistent in the propensity score models than in the naïve regression models. It might be tempting to conclude from this that there was no added benefit to using the propensity score models given that they largely came to the same conclusion as the regression-based models. However the discussion above about the generalizability of the results is important to keep in mind. By focusing on estimating the ATT effect, the propensity score models had the added benefit of clarifying exactly for whom the results of the study can be generalized to. This benefit should not be overlooked, and is important for moving to policy discussions of what exactly may be done with the findings from this study.

DISCUSSION

Aging and Exposure Time

This study set out to determine whether there is a deterrent, null, or criminogenic effect of sanctioning TPVs to imprisonment. In a sense, the conclusion of the study is "none of the above." While the direction of the effect sizes tended to favor a conclusion of a deterrent impact of imprisonment, further analysis suggested that the "treatment" effect was probably less due to deterrence as has been traditionally formulated, and more due to an aging and a recidivism exposure time effect. One inherent limitation of studies on the impact of incarceration, which has been previously noted in research on the dose-response impact of incarceration in general (see Nagin et al., 2009; Snodgrass et al., 2011), is that it is not possible to completely disentangle the elements of the "treatment effect" of imprisonment in order to separate the contribution of deterrence from aging. Aging is of course perfectly correlated with the length of incarceration, but both deterrence and aging may be thought of as components of the "treatment" of incarceration itself. While impossible to completely disentangle aging from deterrence, this is an important issue that should not simply be overlooked. Unfortunately this issue has been mostly overlooked in previous research on the general dose-response impact of imprisonment. As Nagin et al. (2009) note in their review of the existing studies on the general dose-response impact of incarceration, all of the 19 studies that they reviewed suffered from a methodological weakness of failing to account for the impact of aging on recidivism.

While again impossible to fully address, this study has made an effort to overcome the limitation of previous dose-response studies by at least raising the issue of aging and attempting as best as possible to separate the impact of aging versus deterrence. In addition to aging, differential recidivism exposure time was also explored in this study. In terms of aging, since exposure time for recidivism did not begin until re-release from imprisonment, those sanctioned to imprisonment for a TPV, and those sanctioned to longer terms of imprisonment for a TPV, were thus older in biological age at the time of re-release from prison. In terms of recidivism exposure time, TPVs sanctioned to imprisonment and sanctioned to longer terms of imprisonment had less residual time left to serve under parole supervision at the beginning of their recidivism exposure (i.e., at the time of their re-release from prison). This is important because those under parole supervision are subject to a major and prevalent mechanism of recidivism that those no longer under parole supervision are no longer subject to, which is re-incarceration for a technical parole violation. The prevalence of re-incarceration for technical parole violations was in fact a major motivating factor for this study to begin with. Once parolees finish their parole supervision term, they can only be returned to imprisonment through the court system for a new crime, and can no longer be returned to imprisonment for a technical violation. Thus those serving longer lengths of stay in prison should have lower re-incarceration rates (especially at longer follow-up periods) simply because they have less of a chance to be returned for a technical violation since they are ending their supervision terms earlier. This is indeed what this study found. Recall that lower recidivism rates for imprisonment (and for longer terms of imprisonment) were found when measuring recidivism by re-incarceration but not when measuring recidivism by police re-arrest. Further, after propensity score modeling, both chapters used regression models among the matched (or stratified) samples, in order to examine the impact of the treatment assignment on overall recidivism after controlling for the aging and recidivism exposure time (i.e., the age at time of re-release from prison and the residual time left under parole supervision at the time of re-release from prison). In both instances it was found that imprisonment and longer lengths of imprisonment were no longer correlated with overall recidivism when controlling for aging and exposure time. Both aging and exposure time were statistically significant predictors of overall recidivism, however. Together this suggests that aging and exposure time were the primary mechanism at work for the treatment effect of imprisonment for TPVs.

It is sometimes said in corrections circles that the most effective treatment program that we have is age. There is of course a tradition in criminological theory which would support this assertion as well (e.g., Gottfredson & Hirschi, 1990). Using imprisonment to generate an aging effect will look very different than using imprisonment for a deterrent effect. Treatment through aging (and reducing recidivism exposure time) is mostly pessimistic about the criminal justice system's ability to meaningfully and directly affect offender behavioral change, whereas treatment through deterrence is more optimistic in this regard. Treatment through aging basically holds that the passage of time itself will take care of reducing future recidivism, and that younger parolees need simply to "catch up" in age in order for their propensity for recidivism to be reduced. For an aging impact, there is very little left for the correctional system to do but to wait.

On the other hand, specific deterrence is a process by which offenders are expected to learn from the punishment itself in a way that will alter their future calculation of the cost of recidivism such that any benefit of re-offending is perceived as no longer worth the cost. Deterrence may even come in the positive form of rehabilitation programming, whereby the weighing of costs and benefits of returning to offending is altered by impacting fundamental internal change within the offender through programming.

If aging and reduced recidivism exposure time are the primary mechanisms generating lower recidivism rates among TPVs who are incarcerated, then this would suggest that there is little else for a correctional system to do but to incapacitate/house TPVs in prison for an appropriate amount of time until they have “aged” to a point of a lowered recidivism probability. However, at this point the cost of incarceration must be considered. Incarceration is a very expensive option for simply providing a constraining environment for violators to pass time until reaching an appropriate age. It should be considered that it may be possible to incapacitate TPVs in an alternative environment until “aging” occurs, so that similar recidivism outcomes are observed once aging has occurred. In other words, say that a TPV violator is placed under house arrest with a GPS ankle bracelet for 6 months, instead of being incarcerated for 6 months. At the end of the 6 month period, if the violator is effectively incapacitated from re-offending during the 6 month period via house arrest, does the recidivism probability after house arrest then look similar to that of the violator who spent that same 6 month period incarcerated in prison? The findings from this study suggest that this may indeed be the case, since recidivism rates were no longer significantly different between treatment and control group conditions once controlling for aging and recidivism exposure time. If so, there is a potentially large cost advantage of using a cheaper setting than imprisonment (e.g., house arrest with GPS monitoring). The ability of an alternative sanction other than incarceration to effectively incapacitate is an important contingency though, one that will be discussed further below.

Deterrence

While the conclusion of this study was that deterrence was largely not responsible for lower recidivism rates among TPVs sanctioned to imprisonment, it is important to note that this does not provide evidence that deterrence through the use of imprisonment in response to TPVs cannot work. Rather, it suggests that deterrence is not currently a *primary* mechanism at work under the TPV sanctioning regime currently in operation within the Pennsylvania parole system. As is well noted in the existing research literature, there are three primary components of deterrence: 1) the certainty of detection for a violation and of sanctioning given detection, 2) the swiftness of sanctioning for a violation, and 3) the severity of sanctioning for a violation. The general literature on deterrence has found that swiftness and certainty matter much more than the severity of the sanction. It may well be that sanctioning for a TPV in Pennsylvania's currently operating parole system is not delivered with a high degree of certainty or swiftness, and that if the system could be changed to increase the certainty and swiftness of sanctioning then a significant deterrent effect of imprisonment might be detected. This study was not able to provide any measure of the degree of certainty and swiftness of sanctioning in response to TPV violations. If Pennsylvania's parole system operates like most other parole systems around the country, however, it is likely not set up to produce a high degree of certainty or swiftness in sanctioning. The typical parole system is often under immense pressure to resist sanctioning parole violators to imprisonment due to prison bed constraints, and parole systems often lack the resources to respond certainly and quickly to all violations as well. Some evidence of this can be seen in the data for this current study, given that the modal type of sanction for a TPV violation is by far a written warning. While a written warning is considered a TPV sanction within the Pennsylvania parole violator sanctioning continuum, a case could be made that in fact a written warning is not a sanction at all, but instead is a warning of a sanction promised in the future. So the observation that a written warning is the most frequent type of sanction in response to a TPV suggests that violators may be warned repeatedly without a credible sanction being delivered. This may serve to minimize the consistency (the certainty) and swiftness of sanctioning for TPV violations. There is emerging evidence from supervision models like the Hawaii HOPE model (Kleiman, 2009) that when the certainty and swiftness of sanctioning in response to technical violations are increased, short periods of confinement can lead to a rather large deterrent effect on recidivism.

Another contingency to the conclusion of little to no deterrent effect of incarceration here is that procedural justice research suggests that sanctioning must be perceived as fair in order to deter. It may also be the case that in this particular example (i.e., parole supervision in Pennsylvania), the parole sanctioning system is not perceived by parolees as being fair. This in fact goes along with the certainty and swiftness of sanctioning, in that if sanctioning is not consistent it may ultimately be viewed as arbitrary and thus less procedurally fair. This study provides no evidence of the degree of perceived legitimacy of the TPV sanctioning process in Pennsylvania.

In summary, when a particular treatment is found not to work, the important next question to ask is whether it can’t work at all or rather is simply not currently designed to work. What policy decisions to make will differ based on the answer to this question. In the case of this study it is concluded that imprisonment in response to a TPV violation in Pennsylvania is not generating a large deterrent effect, so one question for future exploration is why. If a deterrent effect is not being detected because it can't exist, then policy makers should stop using imprisonment in response to TPVs based on the justification of deterrence alone. If a deterrent effect is not being detected because the system isn't currently set up to facilitate one, then the policy response for consideration is how to re-engineer the supervision and sanctioning regiment so that a deterrent effect can be realized. A cost-benefit calculation should also accompany this. The goal should be to provide the maximum optimal deterrent effect from the sanction, while minimizing the costs resulting from the sanction. A deterrent effect might be possible through the use of incarceration, but the benefit might not be enough to outweigh the significant cost of incarceration in comparison to a less restrictive option. These are the policy relevant factors that should be considered.

Recidivism Measures

Another important point of discussion from the findings of this study is around the differences in outcomes found when using different measures of recidivism. Recall that imprisonment for a first TPV was found to produce significantly lower re-incarceration rates, but no differences in re-arrest rates. Since re-incarceration largely reflects a return to prison for technical violations, which generally are not new crimes, it may be tempting to conclude that imprisonment for a first TPV reduces subsequent technical violations but does not reduce new crime, since re-arrest rates are unaffected. To be sure, this may indeed be the case. If this were the case, it would seem that public safety is not significantly enhanced by the use of incarceration in response to TPVs. If new criminal activity is not reduced through the use of incarceration, then it might be concluded that a higher rate of technical violations coming from using alternative sanctions to imprisonment is worth the cost in order to benefit from the cheaper and less restrictive sanctioning environment presented by alternatives to imprisonment.

It cannot necessarily be concluded from these findings, however, that lower re-incarceration rates but re-arrest rates that are not lower means no reduction in criminal activity by returning TPVs to imprisonment. The reason this cannot be concluded is two-fold. First, some proportion of cases which are re-incarcerated are indeed returned to prison for charges of new criminal behavior. Not all of re-incarceration is for technical violations. In some instances this happens when a parolee is first re-arrested for a criminal charge and then re-incarcerated. Such a case would be counted in both the re-arrest measure and the re-incarceration measure. In fact nearly half of all of the re-incarceration instances in this study (48%) were first preceded by an arrest incident, suggesting that nearly half of those re-incarcerated were first re-arrested for a new criminal charge. Another scenario is that a parolee is caught by his or her parole agent with a charge of criminal activity, but the parole department decides to handle the criminal charges internally or directly with the court rather than turning over to the police. In this situation the parolee would not be counted in the re-arrest measure, but would be counted in the re-incarceration measure with a charge of a new crime. In either situation, re-incarceration can certainly capture new criminal behavior.

The second reason that lower re-incarceration rates but not lower re-arrest rates cannot necessarily be equated with a null impact of imprisonment on criminal behavior is that even if re-incarceration rates did completely represent recidivism for technical parole violations, it may be the case that re-incarcerating technical violators anticipates and prevents criminal recidivism. This explanation fits within a “broken windows” type of model, where minor violations and disorder are precursors to more serious crime. Thus, while lower re-incarceration rates may primarily (or only) directly reduce recidivism for technical violations, there may be an indirect benefit of incapacitating technical violators before they have the chance to return to more serious criminal behavior.

POLICY IMPLICATIONS

Direct policy implications are a bit difficult to draw from this study given the contingencies and limitations in generalizability of the results found here. In general, the primary policy conclusion appears to be that once adjusting for aging and recidivism exposure time, the use of imprisonment (and longer periods of imprisonment) in response to first-time, high risk TPVs mostly has a null impact on subsequent recidivism. The empirical basis for using imprisonment in response to TPVs in order to specifically deter future recidivism is thus weak. Therefore, within the given context of this study, the use of imprisonment in response to TPV violations should not be justified on specific deterrence grounds. Arguments might be made for the use of imprisonment as a response to TPV violations on the basis of incapacitation or general deterrence grounds, but this study did not directly examine either of these two types of impacts and thus cannot speak to their hypothetical benefits. Also important to note is that it should not be implied from this study that imprisonment in response to TPVs cannot possibly serve a specific deterrent role. Based on the contingencies outlined in this chapter, it is certainly possible that under different circumstances (e.g., increased swiftness and certainty in responses to violations) the use of imprisonment may serve a specific deterrent role among TPVs. It can only be said from this study that within the specific context of the study it does not appear that a specific deterrent impact currently exists. It is thus imperative for those policymakers working within the policy environment where this study was conducted (i.e., within parole supervision in Pennsylvania as it currently exists) that the full range of the costs of imprisonment in response to TPVs be considered. Since the use of imprisonment cannot be justified on specific deterrence grounds, the question is whether the cost of using imprisonment in response to TPVs can be justified on other grounds such as based on general deterrence or incapacitation. Also, Pennsylvania policymakers should take stock of the current supervision approach in order to see how important contingencies such as the swiftness, certainty, and perceived fairness of sanctioning might be modified in order to realize a specific deterrent effect from the use of imprisonment among TPVs.

LIMITATIONS

A few limitations of this study should be noted. The primary limitation is the generalizability of the results. Several aspects limit the generalizability of these findings. First, this study is only of one state (Pennsylvania). It is unclear as to how the results of sanctioning TPVs in Pennsylvania generalize to other states with parole systems that may operate differently. For example, as previously discussed, other states may have parole systems with a higher degree of certainty, swiftness, and perceived fairness in sanctioning, which may in turn lead to a larger deterrent effect on recidivism produced by incarceration of TPVs. Or another possibility, other states may have more effective alternative sanctioning options for TPVs (e.g., effective rehabilitation programming, etc.), which may produce lower recidivism rates for TPVs who are not sanctioned to imprisonment. Also, other states may have varying averages and distributions in the length of stay in prison for TPVs sanctioned to imprisonment, which may affect the outcomes of the dose-response impact. For example in Pennsylvania very few TPVs are sanctioned to a short stay of imprisonment (one month or less). Very short stays of imprisonment may be more or less effective, yet this study is unable to generalize findings to a comparison between very short stays in prison and alternatives to imprisonment on the one hand, and between very short stays versus longer stays of imprisonment on the other hand.

The generalizability of these results is also limited to only a sub-set of TPV violators within Pennsylvania. Results were only examined here for sanctioning outcomes associated with the first TPV violation incident after initial release from prison. Results may differ based on the sanction used in response to second and subsequent TPV violations. Also, as previously discussed, one advantage of using the propensity score methodology is that it forces the analysis to focus on only comparable cases based on the propensity to receive the treatment. The limitation of this approach, however, is that it has the potential to narrow the generalizability of the findings to only a sub-set of the population of interest. In this case, results were mostly narrowed to high risk TPVs who were supervised at a higher intensity level, who committed a more serious technical violation associated with their first TPV sanction, and who faced a higher level sanctioning option for the technical violation.

Another limitation of this study is that it cannot easily separate the aging and exposure time effects of incarceration from the deterrent effect of incarceration. An attempt has been made here to at least raise the issue of the difference between aging, exposure time, and deterrent effects, and some attempt has been made to separate these. The conclusion is that incarceration for a TPV appears to mostly serve an aging and exposure time effect. It would be ideal for future studies to look at comparisons of parolees in treatment and control group conditions who are similar in age at time of recidivism exposure and who have a similar amount of recidivism exposure time left at the beginning of recidivism exposure. Essentially that was what was attempted in this study by controlling for age and exposure time at the beginning of recidivism exposure within the final matched sample group. Future studies might make this comparison more explicitly. However, it is impossible to completely separate out the contribution of the aging effect from the exposure time effect from the deterrent effect, since they together comprise the treatment effect here. In a real sense, this problem is an intractable problem which cannot be completely overcome.

This study is also limited by relying on a quasi-experimental approach for estimating a treatment effect, rather than being able to benefit from a true experimental design. The underlying concern is that while a strong quasi-experimental design has been carefully used here (i.e., propensity score modeling), it can only account for observed factors known to impact the treatment assignment decision. The added benefit of a true experimental design is that it can also account for unobserved factors that may influence the treatment assignment decision. If important unobserved factors exist which influence whether TPVs are sanctioned to imprisonment or not, it might change the size or direction of the effect sizes found in this study. Based on the Rosenbaum Bounds sensitivity analysis performed in Chapter 4, this is certainly a real possibility since most of the effect sizes appeared relatively sensitive to the simulated impact of possible unobserved factors. Any relevant missing or unobserved covariate(s) would thus not need to have a substantial impact on the treatment assignment in order to alter the findings made in the study.

A final limitation of this study is that there are no measures of the swiftness, certainty, or perceived fairness of sanctioning among TPVs in the study sample. Unfortunately no measures of these dimensions were available. It will be important in future research to look for ways to measure these dimensions.

FUTURE RESEARCH

There are several important avenues that future research should focus on for extending the findings of this study in order to better understand the impact of imprisonment in response to technical parole violations. One important area for future research is to explore the potential role that imprisonment serves in incapacitating technical parole violators from committing criminal activity during the period of incarceration. This study only explored the impact of imprisonment on subsequent re-offending after re-release from imprisonment. Even if imprisonment has a null impact on re-offending after release, it may still serve an incapacitation role by preventing re-offending during the period of confinement. Ideally both the deterrent and incapacitation effects of incarceration for TPVs would be understood so that the full benefits and costs of incarceration for TPVs could be weighed. Incapacitation is difficult to measure and untangle, but it is important for future research to attempt to do so.

In addition to investigating the incapacitation effect of imprisonment for TPVs, it is also important that future research investigates the general deterrent impact of the use of imprisonment in response to TPV violations. In addition to any specific deterrent and incapacitation effects of imprisonment, the use of imprisonment in response to TPVs might also serve as a general deterrent within the broader population of parolees under supervision. Would-be parole violators may be deterred from recidivating by the presence and use of re-imprisonment as a sanctioning response used among parole violators. This is where the broken windows framework described earlier on in Chapter 2 is particularly germane. Under broken windows policing, addressing community disorder is theorized to serve a general deterrent impact on preventing more serious criminal behavior within neighborhoods. Similarly, a policy of addressing minor technical violations through the use of imprisonment may generally prevent more serious criminal behavior by deterring would-be offenders under parole supervision. It will be important for future research to attempt to separate out estimates of the specific deterrent versus the general deterrent impact of the use of incarceration in response to TPV violations.

A second important area for future research is to explore potentially heterogeneous effects of incarceration of TPVs. It may be found that incarceration has more or less of a deterrent effect among sub-populations of TPVs. As one example, this study only focused on sanctioning in response to the first TPV violation incident after release from prison to parole. A natural question then is what impact sanctioning to imprisonment has among second and subsequent TPV incidences given that the first incident was not sanctioned to imprisonment. Also of interest would be to understand what impact sanctioning to imprisonment has among second and subsequent TPV incidences given that the first incident was sanctioned to imprisonment.

Another important contingency to consider in future research is the swiftness, certainty, and perceived fairness of sanctioning for TPVs. Future studies should attempt to find ways to measure these dimensions so that it can be understood how the deterrent effect of incarceration among TPVs varies based on these dimensions.

Also of importance for future research is to further explore the impact of incarcerating TPVs based on different types of recidivism. As was previously discussed, the re-incarceration measure used in this study is a measure which mixes recidivism for technical violations with recidivism for new crimes. It would be interesting to separate these, so that it can be understood whether lower re-incarceration rates produced by sanctioning first TPVs to imprisonment is reflective primarily of lower rates of return to prison for technical violations versus lower rates of return to prison for new crimes. Also, for both re-incarceration for new crimes and for police re-arrest, it would be important to understand whether imprisonment for a first TPV is more or less effective at deterring different crime types.

Another important area for future research is to explore whether the apparently observed aging effect of incarceration can be actualized within a different setting. The objective of parole should be to optimally reduce recidivism while also delivering the minimal level of sanctioning necessary to do so. One example for further exploration is whether the same aging effect can be observed from incarcerating TPVs in a prison versus a secure halfway house. The PA DOC operates secure halfway houses called

Parole Violator Centers (PVCs). Violators housed in a PVC are included in the PA DOC population count, and a PVC is considered a secure detention where violators are detained and not allowed out of the Center for any reason. For this study, TPVs housed in a PVC were counted in the treatment group (i.e., TPVs imprisoned for a first violation). TPVs sanctioned to a PVC represented a small fraction of the treatment group since PVCs were only recently introduced in Pennsylvania. It would be interesting to know whether the same aging effect is observed for TPVs sanctioned to a PVC versus to a prison. The reason this is of interest is that housing in a PVC can have a different cost than housing in a prison, depending on the group size. PVC beds are contracted out to private contractors and have a fixed per diem cost per resident, whereas a prison bed comes at a variable cost depending on the number of inmates. A PVC bed typically costs between $60 and $70 per day, whereas a prison bed when looking at a group size of 500 or more inmates is typically around $90 per day. When looking at a smaller group of inmates (100 inmates or less), the cost of a prison bed is only around $15 per day, however. The reason for different costs of a prison bed based on the size of the population is due to the difference between marginal cost and average cost. Moving one inmate from a prison bed to a PVC bed saves very little in prison spending since the same level of staffing is needed. On the other hand, if 500 or more inmates are moved from a prison to a PVC, larger costs can be saved since a whole prison unit can be de-staffed and closed. There are enough TPV violators currently sanctioned to imprisonment each year in Pennsylvania that it could save a significant amount of money by moving TPVs from prison to a secure PVC if it were found that secure PVCs generated the same aging/deterrence effect on recidivism. Future studies should explore whether there are differential impacts based on different types of sanctioning environments.

Finally, future studies are needed on this topic more generally, to build a larger body of knowledge about the impact of imprisonment specifically among TPVs. As was noted earlier in this study, this is only the second known study which has directly tested the impact of incarceration on subsequent offending specifically among TPVs. An accumulation of future studies will help to build a body of knowledge around this topic. It is hoped that at some point in the future, when enough studies have been conducted, a review can be conducted to synthesize the findings from multiple studies on this topic in order to build confidence in conclusions. Similar to meta-analyses and reviews of research on the general impact of imprisonment, it would be beneficial to one day have meta-analyses and reviews of research on the impact of imprisonment specifically among technical violators, given the significant policy focus on parole violators in many states.

TABLE 3.1
Summary of Data Exclusion Reasons

	#	%
No recidivism	7,739	41.0%
First violation is criminal offense	6,457	34.2%
Matching first violation to first sanction	1,465	7.8%
Matching PA DOC data to PBPP data	2,943	15.6%
Other/Miscellaneous	252	1.3%
TOTAL:	18,856	100.0%

TABLE 3.2
Descriptive Statistics for Total Sample, Primary Analysis Sample, and Missing Data

	Total Sample (N=17,365)		Primary Analysis Sample (N=12,705)		Missing Data (N=4,660)	
	Mean	SD	Mean	SD	Mean	SD
Age (at first violation)**	35.08	10.07	35.25	9.83	34.63	9.68
Race						
*White**	0.41	0.49	0.42	0.49	0.41	0.49
Black	0.46	0.50	0.45	0.50	0.46	0.50
Hispanic	0.12	0.33	0.12	0.33	0.13	0.33
Gender (*male*)	0.92	0.28	0.92	0.27	0.91	0.28
LSI-R criminal risk score**	22.82	7.83	22.44	7.81	23.94	7.78
Offense type (original sentence)						
Violent	0.28	0.45	0.28	0.45	0.28	0.45
Property	0.22	0.41	0.22	0.41	0.22	0.42
Drugs	0.33	0.47	0.33	0.47	0.32	0.47
Public Order/Other	0.18	0.37	0.18	0.38	0.17	0.38
Prior treatment programs**	2.77	2.29	2.86	2.27	2.53	2.32
Sentencing county						
Philadelphia/Pittsburgh	0.29	0.45	0.29	0.45	0.29	0.45
*Class 3+ county***	0.71	0.45	0.72	0.45	0.68	0.47
Release type						
Initial parole	0.77	0.42	0.76	0.43	0.77	0.42
Re-parole	0.23	0.42	0.24	0.43	0.23	0.42
Parole supervision length	46.84	56.87	46.91	60.47	46.64	44.49
Parole district office						
*Allentown***	0.14	0.35	0.15	0.36	0.11	0.31
*Altoona***	0.04	0.19	0.04	0.21	0.02	0.13
*Central Office***	0.00	0.04	0.00	0.03	0.00	0.06
Chester	0.05	0.22	0.05	0.22	0.05	0.22

	Total Sample (N=17,365)		Primary Analysis Sample (N=12,705)		Missing Data (N=4,660)	
	Mean	SD	Mean	SD	Mean	SD
Erie	0.06	0.24	0.06	0.24	0.07	0.25
*Harrisburg**	0.16	0.37	0.17	0.37	0.14	0.35
Mercer	0.02	0.15	0.02	0.16	0.02	0.15
*Philadelphia***	0.26	0.44	0.25	0.43	0.31	0.46
Pittsburgh	0.16	0.37	0.16	0.37	0.18	0.38
*Scranton**	0.07	0.25	0.06	0.24	0.08	0.27
*Williamsport**	0.03	0.16	0.03	0.16	0.02	0.13
Supervision level						
*Minimum***	0.10	0.30	0.10	0.30	0.08	0.27
*Medium**	0.13	0.34	0.14	0.34	0.12	0.33
*Maximum***	0.72	0.45	0.71	0.45	0.76	0.43
Enhanced/Special	0.05	0.21	0.05	0.21	0.04	0.20
First violation severity (most serious)						
*Low***	0.14	0.35	0.16	0.36	0.06	0.23
*Medium***	0.41	0.49	0.44	0.50	0.26	0.44
*High***	0.45	0.50	0.40	0.49	0.68	0.46
In Center at Time of Violation**	0.09	0.28	0.16	0.37	0.04	0.20

Note: Due to rounding, some categories may not equal 100%
* $p < .01$
** $p < .001$

TABLE 3.3

Tabulation of First TPV Sanctions for the Control Group (N=18,886)

SANCTION TYPE	#	%
Written Warning	6,128	32.4%
Placement in Outpatient D & A Treatment	1,813	9.6%
Increased Reporting Requirements	1,586	8.4%
Imposition of Curfew	1,539	8.1%
Community Parole Corrections Half Way Back	1,524	8.1%
Other Low-level Sanction	1,017	5.4%
Imposition of Increased Curfew	818	4.3%
Placement in Inpatient Drug and Alcohol Treatment	814	4.3%
Other Medium-level Sanction	779	4.1%
Imposition of Increased Urinalysis Testing	710	3.8%
Imposition of Electronic Monitoring	569	3.0%
Placement in Violation Center Contract Facility	314	1.7%
Other High-level Sanction	304	1.6%
Written Travel Restriction	266	1.4%
Documented Job Search	162	0.9%
Obtain treatment evaluation	157	0.8%
Refer to ASCRA Group	120	0.6%
Deadline for Securing Employment	74	0.4%
Imposition of Community Service	70	0.4%
Placement in Drug and Alcohol Detox Facility	60	0.3%
Placement in a Mental Health Facility	19	0.1%
Imposition of Mandatory Antabuse Use	17	0.1%
Refer to Violence Prevention Booster	14	0.1%
Placement in a Day Reporting Center	6	0.0%
Refer to Re-entry Court	3	0.0%
Community Parole Corrections Half Way Out	1	0.0%
Imposition of Global Positioning	1	0.0%
Imposition of Passive Global Positioning	1	0.0%
TOTAL	18,886	100.0%

TABLE 3.4
Descriptive Statistics for Combined Sample, Treatment Group, and Control Group

	Combined Sample (N=12,705)		Treatment Group (N=1,758)		Control Group (N=10,947)	
	Mean	SD	Mean	SD	Mean	SD
Age (at first violation)*	35.19	9.81	34.82	9.72	35.25	9.83
Race						
White	0.42	0.49	0.44	0.50	0.42	0.49
Black	0.45	0.50	0.44	0.50	0.46	0.50
Hispanic	0.12	0.32	0.11	0.32	0.12	0.33
Gender (*male*)**	0.92	0.27	0.95	0.22	0.91	0.28
LSI-R criminal risk score**	22.36	7.82	25.15	7.60	21.91	7.76
Offense type (original sentence)						
*Violent***	0.27	0.45	0.31	0.46	0.27	0.44
*Property**	0.21	0.41	0.24	0.43	0.21	0.41
*Drugs***	0.33	0.47	0.28	0.45	0.34	0.47
Public Order/Other	0.18	0.38	0.18	0.38	0.18	0.38
Prior treatment programs**	2.85	2.26	3.17	2.40	2.80	2.23
Sentencing county						
Philadelphia/Pittsburgh	0.28	0.45	0.27	0.44	0.29	0.45
*Class 3+ county***	0.71	0.45	0.68	0.47	0.72	0.45
Release type						
*Initial parole***	0.76	0.42	0.68	0.47	0.78	0.42
*Re-parole***	0.24	0.42	0.32	0.47	0.22	0.42
Parole supervision length**	47.25	62.11	42.17	48.76	48.07	63.97
Parole district office						
*Allentown***	0.15	0.35	0.19	0.39	0.14	0.35
Altoona	0.05	0.23	0.06	0.23	0.05	0.22
Central Office	0.00	0.06	0.00	0.05	0.00	0.06
*Chester***	0.05	0.22	0.02	0.14	0.06	0.24

	Combined Sample (N=12,705)		Treatment Group (N=1,758)		Control Group (N=10,947)	
	Mean	SD	Mean	SD	Mean	SD
*Erie**	0.06	0.23	0.04	0.20	0.06	0.23
*Harrisburg***	0.17	0.37	0.13	0.33	0.17	0.38
*Mercer***	0.03	0.17	0.01	0.09	0.04	0.18
Philadelphia	0.24	0.42	0.24	0.43	0.24	0.42
*Pittsburgh***	0.15	0.36	0.18	0.38	0.14	0.35
*Scranton***	0.06	0.25	0.11	0.32	0.06	0.23
*Williamsport***	0.03	0.18	0.02	0.13	0.04	0.18
Supervision level						
*Minimum***	0.20	0.40	0.10	0.30	0.22	0.41
*Medium***	0.23	0.42	0.13	0.34	0.24	0.43
*Maximum***	0.53	0.50	0.72	0.45	0.50	0.50
*Enhanced/Special**	0.03	0.18	0.04	0.20	0.03	0.18
First violation severity (most serious)						
*Low***	0.16	0.37	0.00	0.06	0.19	0.39
*Medium***	0.45	0.50	0.10	0.30	0.50	0.50
*High***	0.39	0.48	0.90	0.31	0.31	0.46
In Center at Time of Violation**	0.16	0.37	0.25	0.43	0.14	0.35

Note: Due to rounding, some categories may not equal 100%
* $p < .01$
** $p < .001$

TABLE 3.5
Tabulation of TPV Violations Associated with First TPV Sanction

	Control Group (N=17,435)		Treatment Group (N=4,348)		Total (N=21,783)	
	#	%	#	%	#	%
Positive Urinalysis/Use of Alcohol (Previous History)	3,404	19.5%	414	9.5%	3,818	17.5%
Positive Urinalysis/Use of Drugs (No History)	3,230	18.5%	137	3.2%	3,367	15.5%
Failure to Abide by Board Imposed Special Conditions	1,138	6.5%	378	8.7%	1,516	7.0%
Failure to Pay Supervision Fee	1,386	7.9%	13	0.3%	1,399	6.4%
Changing Residence without Permission	655	3.8%	624	14.4%	1,279	5.9%
Removal From Treatment/CCC Failure	640	3.7%	630	14.5%	1,270	5.8%
Failure to Report as Instructed	694	4.0%	277	6.4%	971	4.5%
Violating Curfew/Approved Schedule	842	4.8%	64	1.5%	906	4.2%
Failure to Abide by Field Imposed Special Conditions	763	4.4%	133	3.1%	896	4.1%
Failure to Abide by Written Instructions	727	4.2%	93	2.1%	820	3.8%
Absconding	131	0.8%	677	15.6%	808	3.7%
Travel Violation	360	2.1%	149	3.4%	509	2.3%
Failure to Participate/Attend Treatment	417	2.4%	30	0.7%	447	2.1%
Positive Urinalysis/Use of Drugs (Previous History)	319	1.8%	92	2.1%	411	1.9%
Assaultive Behavior	225	1.3%	180	4.1%	405	1.9%
Entering Prohibited Establishment	321	1.8%	52	1.2%	373	1.7%
Failure to Maintain Employment	353	2.0%	7	0.2%	360	1.7%
Conviction Summary Offense (No Court Record)	336	1.9%	17	0.4%	353	1.6%
Possession of Unauthorized Contraband, Cell Phone or Beeper	259	1.5%	37	0.9%	296	1.4%
Failure to Notify Agent of Changes of Status	241	1.4%	34	0.8%	275	1.3%
Failure to Complete Treatment	130	0.7%	73	1.7%	203	0.9%
Failure to Pay Restitution and/or Other Court Ordered Fee	173	1.0%	5	0.1%	178	0.8%
Associating with Known Felons, Gangs, Co-Defendant, etc.	124	0.7%	9	0.2%	133	0.6%
Changing Employment Without Agent Notification/Permission	102	0.6%	10	0.2%	112	0.5%
Possession of Offensive Weapon	62	0.4%	39	0.9%	101	0.5%

	Control Group (N=17,435)		Treatment Group (N=4,348)		Total (N=21,783)	
	#	%	#	%	#	%
Failure to Pay Urinalysis Fee	63	0.4%	0	0.0%	63	0.3%
Failure to Report Upon Release	17	0.1%	29	0.7%	46	0.2%
Failure to Provide Urine	40	0.2%	3	0.1%	43	0.2%
Electronic Monitoring Violation	36	0.2%	5	0.1%	41	0.2%
Failure to Take Prescribed Medication as Prescribed by MD	34	0.2%	7	0.2%	41	0.2%
Associating with Crime Victims	27	0.2%	11	0.3%	38	0.2%
Possession of Firearms	12	0.1%	18	0.4%	30	0.1%
Failure to Support Dependent	14	0.1%	1	0.0%	15	0.1%
Failure to Participate in Community Service	7	0.0%	2	0.0%	9	0.0%
Other/Unknown	153	0.8%	98	2.2%	251	1.1%
TOTAL	17,435	100.0%	4,348	100.0%	21,783	100.0%

TABLE 3.6
Descriptive Statistics for Treatment Group Prison Length of Stay Quintiles

	Quintile 1: 0 – 3 months (N=345)		Quintile 2: 3.1 – 6.8 months (N=351)		Quintile 3: 6.9 – 11.3 months (N=355)		Quintile 4: 11.4 – 16.8 months (N=356)		Quintile 5: 16.9 – 78.5 months (N=352)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (at first violation)*	34.43	8.71	35.40	10.00	34.80	9.42	34.14	9.68	36.56	9.74
Race (*white*)*	0.51	0.50	0.43	0.50	0.39	0.49	0.39	0.49	0.43	0.50
Gender (*male*)	0.95	0.22	0.93	0.26	0.92	0.27	0.97	0.17	0.96	0.19
LSI-R criminal risk score**	23.59	7.81	24.77	7.59	24.78	7.78	25.13	7.34	26.53	7.75
Offense type (original sentence)										
*Violent***	0.24	0.43	0.20	0.40	0.27	0.44	0.32	0.47	0.50	0.50
Property	0.26	0.44	0.24	0.43	0.25	0.43	0.22	0.41	0.23	0.42
*Drugs***	0.31	0.46	0.34	0.47	0.30	0.46	0.29	0.45	0.16	0.37
*Public Order/Other***	0.20	0.40	0.22	0.42	0.19	0.39	0.18	0.38	0.10	0.30
Prior treatment programs**	2.95	2.57	3.12	2.34	2.95	2.33	3.38	2.49	3.56	2.40
Sentencing county										
*Philadelphia/Pittsburgh**	0.22	0.42	0.28	0.45	0.31	0.46	0.33	0.47	0.31	0.46
Class 3+ county	0.65	0.48	0.74	0.44	0.73	0.45	0.76	0.43	0.67	0.47
Release type										
Initial parole	0.70	0.46	0.69	0.46	0.70	0.46	0.70	0.46	0.61	0.49
Re-parole	0.30	0.46	0.31	0.46	0.30	0.46	0.30	0.46	0.39	0.49
Parole supervision length**	33.53	32.83	28.43	28.12	39.19	37.13	44.33	43.90	63.85	71.08
Parole district office										
Allentown	0.15	0.36	0.21	0.41	0.22	0.42	0.19	0.39	0.15	0.35
Altoona	0.08	0.28	0.03	0.17	0.04	0.19	0.03	0.18	0.05	0.21
Central Office	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.00
Chester	0.02	0.13	0.03	0.17	0.02	0.14	0.02	0.15	0.02	0.13
Erie	0.07	0.25	0.03	0.18	0.05	0.22	0.05	0.22	0.05	0.23

	Quintile 1: 0 – 3 months (N=345)		Quintile 2: 3.1 – 6.8 months (N=351)		Quintile 3: 6.9 – 11.3 months (N=355)		Quintile 4: 11.4 – 16.8 months (N=356)		Quintile 5: 16.9 – 78.5 months (N=352)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Harrisburg	0.12	0.33	0.13	0.33	0.11	0.31	0.13	0.34	0.14	0.34
Mercer	0.01	0.09	0.00	0.07	0.01	0.08	0.01	0.11	0.01	0.10
*Philadelphia***	0.18	0.39	0.31	0.46	0.32	0.46	0.31	0.46	0.32	0.47
Pittsburgh	0.18	0.39	0.13	0.34	0.14	0.34	0.14	0.35	0.18	0.39
*Scranton***	0.18	0.38	0.10	0.31	0.09	0.28	0.09	0.29	0.07	0.26
Williamsport	0.01	0.11	0.02	0.14	0.01	0.11	0.01	0.10	0.01	0.11
Supervision level										
Minimum	0.08	0.26	0.07	0.26	0.06	0.25	0.04	0.20	0.05	0.21
Medium	0.12	0.33	0.11	0.31	0.09	0.29	0.12	0.33	0.09	0.28
Maximum	0.77	0.42	0.79	0.41	0.78	0.41	0.76	0.43	0.83	0.38
Enhanced/Special	0.03	0.17	0.03	0.17	0.06	0.24	0.07	0.26	0.04	0.19
First violation severity (most serious)										
Low	0.01	0.10	0.00	0.07	0.00	0.07	0.00	0.07	0.00	0.07
*Medium***	0.16	0.37	0.12	0.32	0.07	0.25	0.06	0.24	0.10	0.30
*High***	0.83	0.38	0.88	0.33	0.93	0.26	0.93	0.25	0.90	0.30

Note: Due to rounding, some categories may not equal 100%
* $p < .01$
** $p < .001$

TABLE 4.1
Bias Statistics for Covariates before Matching

	Treatment Group (N=1,758)	Control Group (N=10,947)	% Bias	T-Test	
				T	*p*
Age (at first violation)	34.82	35.25	-4.5%	-1.74	0.08
Race					
White	0.44	0.42	4.2%	1.64	0.10
Black	0.44	0.46	-2.6%	-1.01	0.31
Hispanic	0.11	0.12	-1.9%	-0.72	0.47
Other	0.00	0.00	-3.2%	-1.15	0.25
Gender (male)	0.95	0.91	13.2%	4.78	0.00
LSI-R criminal risk score	25.15	21.91	42.1%	16.26	0.00
Offense type (original sentence)					
Violent	0.31	0.27	8.4%	3.30	0.00
Property	0.24	0.21	6.8%	2.71	0.01
Drugs	0.28	0.34	-14.7%	-5.60	0.00
Public Order/Other	0.18	0.18	0.4%	0.14	0.89
Prior treatment programs	3.17	2.80	16.2%	6.47	0.00
Sentencing county					
Philadelphia/Pittsburgh	0.27	0.29	-4.3%	-1.67	0.10
Class 3+ county	0.68	0.72	-8.2%	-3.22	0.00
Initial parole	0.68	0.78	-21.4%	-8.70	0.00
Parole supervision length	42.17	48.07	-10.4%	-3.70	0.00
Parole district office					
Allentown	0.19	0.14	12.6%	5.10	0.00
Altoona	0.06	0.05	3.0%	1.18	0.24
Central Office	0.00	0.00	-0.8%	-0.30	0.76
Chester	0.02	0.06	-19.8%	-6.67	0.00
Erie	0.04	0.06	-6.9%	-2.55	0.01
Harrisburg	0.13	0.17	-13.2%	-4.90	0.00
Mercer	0.01	0.04	-18.8%	-6.07	0.00
Philadelphia	0.24	0.24	0.9%	0.34	0.73
Pittsburgh	0.18	0.14	9.7%	3.89	0.00
Scranton	0.11	0.06	19.7%	8.62	0.00
Williamsport	0.02	0.04	-12.1%	-4.20	0.00
Supervision level					
Administrative	0.00	0.01	-10.3%	-3.41	0.00
Minimum	0.10	0.22	-32.6%	-11.48	0.00
Medium	0.13	0.24	-28.5%	-10.28	0.00
Maximum	0.72	0.50	47.7%	17.88	0.00
Enhanced/Special	0.04	0.03	5.7%	2.36	0.02

	Treatment Group (N=1,758)	Control Group (N=10,947)	% Bias	T-Test	
				T	*p*
First violation severity (most serious)					
Low	0.00	0.19	-66.1%	-19.80	0.00
Medium	0.10	0.50	-97.6%	-32.83	0.00
High	0.90	0.31	150.3%	51.64	0.00
In Center at Time of Violation	0.25	0.14	27.8%	11.73	0.00

TABLE 4.2
Observed Recidivism Rates Before Matching

	Treatment Group (N=1,547)	Control Group (N=10,495)	Chi-Square	
			χ^2	p
Re-incarceration Rates				
6-Month	16.9%	18.4%	2.1	0.143
1-Year	27.6%	32.5%	16.7	0.000
3-Year	48.0%	54.3%	21.1	0.000
Re-arrest Rates				
6-Month	16.4%	14.4%	4.8	0.028
1-Year	29.5%	25.2%	14.4	0.000
3-Year	56.2%	51.5%	12.0	0.001
Overall Recidivism Rates				
6-Month	29.7%	26.6%	6.9	0.008
1-Year	47.6%	43.5%	10.2	0.001
3-Year	72.5%	69.5%	6.0	0.014

TABLE 4.3
"Naïve" Logistic Regression Models Predicting Recidivism (N=11,934)

	Re-incarceration					
	6-Month		1-Year		3-Year	
	OR	*p*	OR	*p*	OR	*p*
Sanctioned to Imprisonment (treatment)	0.60	0.00	0.52	0.00	0.50	0.00
Age (at first violation)	0.98	0.00	0.98	0.00	0.97	0.00
Race						
White	1.37	0.43	0.91	0.76	1.01	0.97
Black	1.28	0.55	0.82	0.50	1.10	0.75
Hispanic	1.37	0.44	0.85	0.59	1.11	0.72
Gender (male)	1.51	0.00	1.55	0.00	1.52	0.00
LSI-R criminal risk score	1.03	0.00	1.03	0.00	1.04	0.00
Offense type (original sentence)						
Violent	0.88	0.10	0.85	0.01	0.75	0.00
Property	1.08	0.31	1.03	0.62	1.08	0.23
Drugs	0.76	0.00	0.75	0.00	0.76	0.00
Prior treatment programs	1.01	0.33	1.01	0.22	1.02	0.01
Sentencing county						
Philadelphia/Pittsburgh	0.98	0.80	0.95	0.46	0.89	0.09
Class 3+ county	0.85	0.02	0.95	0.45	1.01	0.93
Initial parole	0.76	0.00	0.87	0.00	0.89	0.02
Parole supervision length	1.00	0.76	1.00	0.00	1.01	0.00
Parole district office						
Allentown	1.15	0.79	0.88	0.74	0.96	0.90
Altoona	1.34	0.57	1.06	0.87	1.10	0.80
Chester	0.86	0.77	0.73	0.43	0.87	0.71
Erie	1.18	0.75	0.78	0.53	0.79	0.52
Harrisburg	1.19	0.73	0.98	0.97	1.04	0.92
Mercer	0.76	0.60	0.79	0.56	0.76	0.48
Philadelphia	1.47	0.44	1.12	0.77	1.05	0.89
Pittsburgh	1.31	0.59	1.09	0.83	1.04	0.91
Scranton	2.06	0.15	1.63	0.20	1.38	0.39
Williamsport	1.23	0.68	0.95	0.91	0.86	0.70
Supervision level						
Administrative	0.30	0.05	0.23	0.00	0.25	0.00
Minimum	1.02	0.88	0.85	0.21	0.65	0.00
Medium	1.08	0.61	0.95	0.66	0.72	0.01
Maximum	1.30	0.06	1.08	0.49	0.85	0.15
First violation severity (most serious)						
Medium	1.78	0.00	1.53	0.00	1.61	0.00
High	2.11	0.00	1.93	0.00	1.94	0.00
In Center at Time of Violation	1.22	0.00	1.19	0.00	1.17	0.01

TABLE 4.3 (continued)
"Naïve" Logistic Regression Models Predicting Recidivism (N=11,934)

	Re-arrest					
	6-Month		1-Year		3-Year	
	OR	*p*	OR	*p*	OR	*p*
Sanctioned to Imprisonment (treatment)	1.09	0.32	1.18	0.02	1.04	0.51
Age (at first violation)	0.98	0.00	0.97	0.00	0.97	0.00
Race						
White	1.70	0.25	0.71	0.26	1.33	0.31
Black	1.78	0.21	0.77	0.38	1.47	0.17
Hispanic	1.48	0.40	0.62	0.11	1.19	0.54
Gender (male)	1.37	0.01	1.31	0.00	1.24	0.00
LSI-R criminal risk score	1.01	0.00	1.02	0.00	1.02	0.00
Offense type (original sentence)						
Violent	0.88	0.13	0.85	0.02	0.79	0.00
Property	1.12	0.17	1.11	0.14	1.20	0.00
Drugs	0.99	0.91	0.94	0.31	0.93	0.18
Prior treatment programs	1.01	0.35	1.02	0.06	1.02	0.08
Sentencing county						
Philadelphia/Pittsburgh	1.21	0.04	1.21	0.01	1.13	0.07
Class 3+ county	1.09	0.31	1.18	0.02	1.11	0.06
Initial parole	0.90	0.08	0.85	0.00	0.83	0.00
Parole supervision length	1.00	0.04	1.00	0.01	1.00	0.86
Parole district office						
Allentown	1.75	0.36	1.34	0.50	1.59	0.20
Altoona	1.85	0.32	1.53	0.34	1.67	0.17
Chester	2.03	0.25	1.50	0.36	2.22	0.03
Erie	1.41	0.58	0.94	0.89	1.21	0.61
Harrisburg	2.23	0.19	1.70	0.22	1.91	0.08
Mercer	2.65	0.12	1.94	0.14	2.11	0.05
Philadelphia	2.18	0.20	1.79	0.18	2.50	0.01
Pittsburgh	1.45	0.55	1.10	0.82	1.49	0.27
Scranton	1.21	0.76	1.02	0.97	1.33	0.44
Williamsport	1.50	0.52	1.26	0.61	1.30	0.49
Supervision level						
Administrative	0.76	0.48	0.63	0.15	0.73	0.20
Minimum	0.90	0.53	0.85	0.22	0.97	0.80
Medium	0.99	0.96	0.89	0.36	0.99	0.96
Maximum	1.05	0.72	0.92	0.48	1.00	0.97
First violation severity (most serious)						
Medium	1.09	0.30	1.18	0.01	1.32	0.00
High	1.18	0.07	1.17	0.03	1.40	0.00
In Center at Time of Violation	0.68	0.00	0.77	0.00	0.85	0.00

TABLE 4.3 (continued)
"Naïve" Logistic Regression Models Predicting Recidivism (N=11,934)

	Overall Recidivism					
	6-Month		1-Year		3-Year	
	OR	*p*	OR	*p*	OR	*p*
Sanctioned to Imprisonment (treatment)	0.83	0.01	0.82	0.00	0.73	0.00
Age (at first violation)	0.98	0.00	0.97	0.00	0.97	0.00
Race						
White	1.66	0.16	0.91	0.73	1.42	0.23
Black	1.71	0.14	0.91	0.75	1.60	0.11
Hispanic	1.66	0.17	0.84	0.55	1.38	0.28
Gender (male)	1.45	0.00	1.54	0.00	1.55	0.00
LSI-R criminal risk score	1.03	0.00	1.03	0.00	1.04	0.00
Offense type (original sentence)						
Violent	0.89	0.09	0.87	0.02	0.79	0.00
Property	1.12	0.08	1.09	0.15	1.15	0.04
Drugs	0.84	0.01	0.82	0.00	0.75	0.00
Prior treatment programs	1.01	0.18	1.03	0.00	1.03	0.00
Sentencing county						
Philadelphia/Pittsburgh	1.01	0.92	1.00	0.96	0.94	0.42
Class 3+ county	0.96	0.54	1.04	0.46	1.11	0.10
Initial parole	0.81	0.00	0.84	0.00	0.80	0.00
Parole supervision length	1.00	0.70	1.00	0.03	1.00	0.00
Parole district office						
Allentown	1.38	0.46	1.09	0.80	1.20	0.63
Altoona	1.49	0.37	1.30	0.48	1.37	0.42
Chester	1.29	0.57	1.05	0.90	1.37	0.41
Erie	1.27	0.59	0.89	0.76	0.94	0.87
Harrisburg	1.53	0.33	1.27	0.51	1.40	0.37
Mercer	1.33	0.53	1.16	0.69	1.24	0.58
Philadelphia	1.56	0.31	1.25	0.54	1.49	0.29
Pittsburgh	1.44	0.41	1.18	0.65	1.31	0.48
Scranton	1.91	0.14	1.58	0.21	1.47	0.32
Williamsport	1.32	0.54	1.07	0.87	1.04	0.93
Supervision level						
Administrative	0.52	0.08	0.39	0.00	0.45	0.00
Minimum	0.96	0.76	0.87	0.25	0.81	0.13
Medium	1.05	0.69	0.89	0.32	0.85	0.21
Maximum	1.21	0.11	1.01	0.92	0.94	0.61
First violation severity (most serious)						
Medium	1.40	0.00	1.47	0.00	1.62	0.00
High	1.67	0.00	1.76	0.00	2.05	0.00
In Center at Time of Violation	1.01	0.87	1.02	0.68	1.06	0.35

TABLE 4.4
Logit Regression Predicting Treatment Assignment (N=11,934)

	β	OR	Z	*p*
Age (at first violation)	-0.01	0.99	-3.50	0.00
Race				
White	0.19	1.21	0.33	0.74
Black	0.21	1.23	0.37	0.71
Hispanic	0.04	1.04	0.06	0.95
Gender (male)	0.43	1.54	3.31	0.00
LSI-R criminal risk score	0.02	1.02	3.58	0.00
Offense type (original sentence)				
Violent	-0.02	0.98	-0.23	0.82
Property	-0.05	0.95	-0.52	0.60
Drugs	-0.25	0.78	-2.57	0.01
Prior treatment programs	0.02	1.02	1.46	0.14
Sentencing county				
Philadelphia/Pittsburgh	-0.19	0.82	-1.82	0.07
Class 3+ county	-0.26	0.77	-2.86	0.00
Initial parole	-0.36	0.69	-5.10	0.00
Parole supervision length	0.00	1.00	-3.80	0.00
Parole district office				
Allentown	-0.01	0.99	-0.02	0.98
Altoona	0.10	1.10	0.17	0.87
Chester	-0.93	0.40	-1.59	0.11
Erie	-0.16	0.86	-0.28	0.78
Harrisburg	-0.87	0.42	-1.58	0.12
Mercer	-1.31	0.27	-2.07	0.04
Philadelphia	-0.20	0.82	-0.36	0.72
Pittsburgh	-0.11	0.90	-0.19	0.85
Scranton	0.45	1.57	0.82	0.41
Williamsport	-0.85	0.43	-1.43	0.15
Supervision level				
Administrative	-0.80	0.45	-1.23	0.22
Minimum	-0.53	0.59	-2.80	0.01
Medium	-0.64	0.53	-3.55	0.00
Maximum	-0.07	0.93	-0.44	0.66
First violation severity (most serious)				
Medium	1.98	7.26	5.11	0.00
High	4.61	100.02	12.08	0.00
In Center at Time of Violation	0.11	1.11	1.40	0.16
Constant	-4.94	0.01	-5.55	0.00

TABLE 4.5
Bias Statistics and Balance Improvement for Covariates After Matching

	Treatment Group (N=1,542)	Control Group (N=1,103)	% Bias After Matching	% Bias Reduction	T-Test T	T-Test *p*
Age (at first violation)	34.15	34.59	-4.6%	50.9%	-1.31	0.19
Race						
White	0.44	0.45	-2.0%	53.8%	-0.54	0.59
Black	0.45	0.43	2.7%	-23.5%	0.76	0.45
Hispanic	0.11	0.12	-1.2%	50.9%	-0.34	0.74
Gender (male)	0.94	0.95	-2.5%	77.5%	-0.80	0.43
LSI-R criminal risk score	25.33	25.40	-1.0%	97.5%	-0.28	0.78
Offense type (original sentence)						
Violent	0.29	0.26	5.5%	-9.5%	1.53	0.13
Property	0.24	0.25	-2.8%	61.2%	-0.75	0.45
Drugs	0.28	0.29	-1.4%	89.2%	-0.40	0.69
Prior treatment programs	3.13	3.12	0.4%	97.0%	0.11	0.92
Sentencing county			0.0%	0.0%		
Philadelphia/Pittsburgh	0.27	0.27	-1.2%	72.9%	-0.32	0.75
Class 3+ county	0.69	0.68	1.0%	86.7%	0.27	0.79
Initial parole	0.68	0.68	-0.1%	99.3%	-0.04	0.97
Parole supervision length	38.93	40.01	-2.3%	86.5%	-0.69	0.49
Parole district office						
Allentown	0.18	0.21	-6.3%	42.2%	-1.64	0.10
Altoona	0.06	0.06	0.6%	85.0%	0.15	0.88
Chester	0.02	0.01	3.1%	85.2%	1.32	0.19
Erie	0.04	0.03	4.5%	39.3%	1.43	0.15
Harrisburg	0.13	0.12	1.5%	89.2%	0.44	0.66
Mercer	0.01	0.01	0.5%	97.6%	0.21	0.83

	Treatment Group (N=1,542)	Control Group (N=1,103)	% Bias After Matching	% Bias Reduction	T-Test	
					T	*p*
Philadelphia	0.24	0.23	3.0%	-263.2%	0.85	0.39
Pittsburgh	0.18	0.19	-0.3%	96.5%	-0.09	0.93
Scranton	0.12	0.12	-2.8%	87.2%	-0.67	0.51
Williamsport	0.02	0.02	-1.6%	87.3%	-0.55	0.58
Supervision level						
Administrative	0.00	0.00	-0.9%	91.0%	-0.38	0.71
Minimum	0.09	0.10	-0.7%	97.6%	-0.25	0.81
Medium	0.13	0.14	-1.7%	94.3%	-0.53	0.60
Maximum	0.73	0.72	2.6%	94.4%	0.77	0.44
First violation severity (most serious)						
Medium	0.10	0.10	0.0%	100.0%	0.00	1.00
High	0.89	0.89	0.2%	99.9%	0.06	0.95
In Center at Time of Violation	0.26	0.25	0.6%	97.6%	0.16	0.87

TABLE 4.6

Recidivism Rates and Treatment Effect Sizes After Propensity Score Matching

	Treatment Group (N=1,542)	Control Group (N=1,103)	ATT	S.E.	T
Re-incarceration Rates					
6-Month	18.3%	25.5%	-7.2%	0.02	-3.99**
1-Year	29.2%	43.2%	-13.9%	0.02	-6.70**
3-Year	48.1%	61.2%	-13.1%	0.02	-6.16**
Re-arrest Rates					
6-Month	16.6%	13.6%	3.0%	0.02	1.94
1-Year	30.1%	25.6%	4.5%	0.02	2.33*
3-Year	56.2%	54.8%	1.4%	0.02	0.66
Overall Recidivism Rates					
6-Month	30.9%	33.0%	-2.1%	0.02	-1.03
1-Year	49.2%	54.4%	-5.3%	0.02	-2.43*
3-Year	72.5%	78.5%	-6.0%	0.02	-3.24**

* $p < .05$
** $p < .01$

TABLE 4.7
Comparison of Matched Group to Unmatched Group

	Matched Group (N=2,645)	Unmatched Group (N=9,289)	% Bias	T-Test T	T-Test *p*
Age (at first violation)	34.38	35.08	-7.4%	-3.31	0.00
Race					
White	0.44	0.42	4.2%	1.91	0.06
Black	0.45	0.46	-2.2%	-0.99	0.32
Hispanic	0.11	0.12	-2.4%	-1.08	0.28
Other	0.00	0.01	-3.8%	-1.60	0.11
Gender (male)	0.94	0.91	11.9%	5.15	0.00
LSI-R criminal risk score	25.12	21.95	42.0%	18.95	0.00
Offense type (original sentence)					
Violent	0.28	0.27	2.6%	1.21	0.23
Property	0.24	0.21	7.2%	3.33	0.00
Drugs	0.29	0.35	-11.4%	-5.09	0.00
Public Order/Other	0.19	0.18	2.8%	1.29	0.20
Prior treatment programs	3.12	2.76	15.6%	7.24	0.00
Sentencing county					
Philadelphia/Pittsburgh	0.27	0.28	-2.6%	-1.18	0.24
Class 3+ county	0.69	0.72	-7.2%	-3.32	0.00
Initial parole	0.69	0.78	-20.5%	-9.60	0.00
Parole supervision length	40.33	47.34	-15.2%	-6.80	0.00
Parole district office					
Allentown	0.18	0.14	12.4%	5.84	0.00
Altoona	0.06	0.05	3.2%	1.49	0.14
Central Office	0.00	0.00	0.9%	0.44	0.66
Chester	0.02	0.06	-23.3%	-9.24	0.00
Erie	0.04	0.06	-9.2%	-3.99	0.00
Harrisburg	0.13	0.18	-12.2%	-5.36	0.00
Mercer	0.01	0.04	-19.3%	-7.56	0.00
Philadelphia	0.24	0.23	1.2%	0.55	0.59
Pittsburgh	0.18	0.14	11.2%	5.26	0.00
Scranton	0.11	0.05	20.7%	10.38	0.00
Williamsport	0.02	0.04	-10.9%	-4.56	0.00
Supervision level					
Administrative	0.00	0.01	-9.1%	-3.60	0.00
Minimum	0.10	0.21	-30.4%	-12.79	0.00
Medium	0.14	0.25	-29.2%	-12.49	0.00
Maximum	0.71	0.50	46.2%	20.40	0.00
Enhanced/Special	0.04	0.03	5.8%	2.75	0.01

	Matched Group (N=2,645)	Unmatched Group (N=9,289)	% Bias	T-Test T	*p*
First violation severity (most serious)					
Low	0.01	0.20	-67.3%	-24.80	0.00
Medium	0.12	0.55	-104.5%	-42.76	0.00
High	0.88	0.25	164.6%	69.67	0.00
In Center at Time of Violation	0.25	0.14	27.4%	13.27	0.00

TABLE 4.8
Comparison of TPV Violations between Matched Group and Unmatched Group

VIOLATION	Matched Group (N=2,645)	Unmatched Group (N=9,289)	% Bias	T-Test T	T-Test *p*
Changing Residence without Permission	21.4%	3.7%	55.5%	31.60	0.00
Positive Urinalysis/Use of Alcohol (Previous History)	21.3%	18.0%	8.2%	3.78	0.00
Failure to Abide by Board Imposed Special Conditions	15.8%	5.7%	33.0%	17.17	0.00
Removal From Treatment /CCC Failure	15.8%	2.1%	49.5%	29.34	0.00
Absconding	6.4%	0.2%	35.2%	23.30	0.00
Assaultive Behavior	3.8%	1.0%	18.2%	10.01	0.00
Positive Urinalysis/Use of Drugs (No History)	3.6%	24.2%	-62.3%	-24.09	0.00
Positive Urinalysis/Use of Drugs (Previous History)	2.7%	1.6%	7.4%	3.60	0.00
Failure to Report as Instructed	1.6%	4.0%	-14.9%	-6.08	0.00
Failure to Abide by Field Imposed Special Conditions	1.2%	4.4%	-19.2%	-7.62	0.00
Possession of Offensive Weapon	0.9%	0.4%	6.5%	3.36	0.00
Failure to Complete Treatment	0.7%	0.5%	2.7%	1.27	0.20
Failure to Abide by Written Instructions	0.6%	3.9%	-21.9%	-8.41	0.00
Entering Prohibited Establishment	0.5%	1.6%	-11.5%	-4.57	0.00
Possession of Unauthorized Contraband, Cell Phone, or Beeper	0.5%	1.7%	-12.2%	-4.81	0.00
Associating with Crime Victims	0.4%	0.1%	5.0%	2.66	0.01
Failure to Maintain Employment	0.4%	2.2%	-16.1%	-6.20	0.00
Failure to Participate/Attend Treatment	0.4%	2.8%	-19.4%	-7.39	0.00
Travel Violation	0.4%	1.9%	-14.5%	-5.61	0.00
Violating Curfew/Approved Schedule	0.3%	3.8%	-24.4%	-9.17	0.00
Failure to Notify Agent of Changes of Status	0.3%	1.6%	-13.4%	-5.17	0.00
Failure to Report Upon Release	0.3%	0.1%	4.9%	2.71	0.01
Conviction Summary Offense (No Court Record)	0.3%	1.9%	-16.1%	-6.12	0.00
Failure to Pay Supervision Fee	0.2%	10.3%	-46.4%	-17.02	0.00
Associating with Known Felons, Gangs, Co-Defendents, etc.	0.1%	0.4%	-5.3%	-2.10	0.04
Possession of Firearms	0.1%	0.1%	0.9%	0.42	0.67

VIOLATION	Matched Group (N=2,645)	Unmatched Group (N=9,289)	% Bias	T-Test T	T-Test *p*
Changing Employment Without Agent Notification/Permission	0.0%	0.3%	-7.0%	-2.63	0.01
Failure to Provide Urine	0.0%	0.1%	-2.6%	-1.03	0.30
Failure to Take Prescription Medication as Prescribed by MD	0.0%	0.1%	-3.6%	-1.43	0.15
Electronic Monitoring Violation	0.0%	0.1%	-4.6%	-1.68	0.09
Failure to Participate in Communnity Service	0.0%	0.0%	-2.9%	-1.06	0.29
Failure to Pay Restitution and/or Other Court Ordered Fees	0.0%	0.5%	-10.2%	-3.72	0.00
Failure to Pay Urinalysis Fee	0.0%	0.2%	-6.4%	-2.31	0.02
Failure to Support Dependent	0.0%	0.1%	-4.1%	-1.50	0.13
Other	0.0%	0.0%	-2.1%	-0.75	0.45

TABLE 4.9
Comparison of Sanctions Received between Matched and Unmatched Control Group Cases

SANCTION TYPE	Matched Group (N=1,103)	Unmatched Group (N=9,287)	% Bias	T-Test T	*p*
Community Parole Corrections Halfway Back	24.9%	8.3%	45.7%	17.55	0.00
Written Warning	16.9%	33.3%	-38.6%	-11.16	0.00
Placement in Inpatient Drug and Alcohol Treatment	12.1%	5.1%	25.3%	9.48	0.00
Placement in Outpatient Drug and Alcohol Treatment	9.9%	11.6%	-5.5%	-1.67	0.09
Imposition of Curfew	5.6%	7.3%	-6.6%	-2.00	0.05
Other High-Level Sanction	5.3%	2.2%	16.8%	6.43	0.00
Electronic Monitoring	4.9%	4.0%	4.2%	1.36	0.17
Other Medium-Level Sanction	4.9%	4.9%	0.2%	0.06	0.95
Increased Curfew	4.3%	5.8%	-6.9%	-2.06	0.04
Other Low-Level Sanction	4.2%	6.4%	-9.9%	-2.89	0.00
Increased Reporting Requirements	2.0%	4.7%	-15.0%	-4.11	0.00
Increased Urinalysis Testing	1.5%	2.1%	-4.5%	-1.32	0.19
Placement in Drug and Alcohol Detox Facility	0.5%	0.4%	2.4%	0.80	0.42
Deadline for Securing Employment	0.5%	0.6%	-1.7%	-0.51	0.61
Documented Job Search	0.5%	0.6%	-2.0%	-0.59	0.56
Community Service	0.4%	0.6%	-3.1%	-0.90	0.37
Placement in a Mental Health Facility	0.4%	0.1%	5.0%	2.04	0.04
Mandatory Antabuse Use	0.3%	0.1%	4.1%	1.64	0.10
Obtain Treatment Evaluation	0.3%	0.5%	-3.4%	-0.96	0.34
Refer to ASCRA Group	0.3%	0.5%	-3.1%	-0.88	0.38
Refer to Violence Prevention Booster	0.3%	0.1%	4.7%	2.01	0.04
Written Travel Restrictions	0.2%	1.0%	-11.1%	-2.79	0.01
Imposition of Passive Global Positioning	0.0%	0.0%	-1.5%	-0.34	0.73
Placement in a Day Reporting Center	0.0%	0.0%	-2.5%	-0.59	0.55

TABLE 4.10
Sensitivity Analysis Results

Scenario # 1: Relaxing Data Exclusion Criteria

	Treatment Group (N=2,210)	Control Group (N=1,468)	ATT	S.E.	T
Re-incarceration Rates					
*6-Month***	16.8%	27.3%	-10.5%	0.02	-6.71
*1-Year***	27.1%	42.6%	-15.4%	0.02	-8.68
*3-Year***	43.8%	62.9%	-19.1%	0.02	-10.53
Re-arrest Rates					
6-Month	14.6%	13.9%	0.7%	0.01	0.53
1-Year	26.7%	25.2%	1.4%	0.02	0.90
3-Year	49.8%	53.2%	-3.3%	0.02	-1.80
Overall Recidivism Rates					
*6-Month***	27.8%	34.8%	-7.0%	0.02	-4.03
*1-Year***	44.6%	52.9%	-8.3%	0.02	-4.47
*3-Year***	65.0%	77.3%	-12.3%	0.02	-7.49

Scenario # 2: Only High Severity TPV Violations

	Treatment Group (N=1,376)	Control Group (N=937)	ATT	S.E.	T
Re-incarceration Rates					
*6-Month***	17.6%	29.4%	-11.8%	0.02	-5.95
*1-Year***	28.4%	46.6%	-18.2%	0.02	-8.11
*3-Year***	47.5%	69.0%	-21.4%	0.02	-9.61
Re-arrest Rates					
6-Month	16.4%	16.6%	-0.3%	0.02	-0.17
1-Year	30.6%	29.3%	1.3%	0.02	0.62
3-Year	57.0%	56.3%	0.7%	0.02	0.32
Overall Recidivism Rates					
*6-Month***	30.5%	37.9%	-7.5%	0.02	-3.41
*1-Year***	49.2%	57.6%	-8.4%	0.02	-3.64
*3-Year***	72.8%	81.5%	-8.7%	0.02	-4.53

TABLE 4.10 (continued)
Sensitivity Analysis Results

Scenario # 3: Only "Halfway Back and Inpatient TPV Sanctions in Control Group

	Treatment Group (N=1,536)	Control Group (N=755)	ATT	S.E.	T
Re-incarceration Rates					
*6-Month***	18.4%	30.7%	-12.3%	0.02	-5.15
*1-Year***	29.4%	48.4%	-19.0%	0.03	-7.15
*3-Year***	48.2%	69.1%	-20.9%	0.03	-8.17
Re-arrest Rates					
*6-Month***	16.7%	10.4%	6.3%	0.02	3.55
*1-Year***	30.1%	22.1%	8.1%	0.02	3.46
3-Year	56.3%	52.3%	4.0%	0.03	1.49
Overall Recidivism Rates					
6-Month	31.1%	35.7%	-4.6%	0.03	-1.80
*1-Year**	49.3%	55.9%	-6.6%	0.03	-2.44
*3-Year***	72.7%	79.8%	-7.2%	0.02	-3.21

* $p < .05$
** $p < .01$

TABLE 4.11
Rosenbaum Bounds for the Treatment Effects

	Bias for Non-significant Group Difference		Bias for Reversal of Treatment Effect Sign	
	Γ	*p* critical	Γ	*p* critical
Re-incarceration Rates				
6-Month	1.30	0.09	1.75	0.05
1-Year	1.60	0.08	2.10	0.04
3-Year	1.50	0.09	1.95	0.03
Re-arrest Rates				
6-Month	1.05	0.09	1.50	0.04
1-Year	1.10	0.06	1.50	0.03
*3-Year***	1.10	0.01	1.25	0.04
Overall Recidivism Rates				
*6-Month***	1.10	0.02	1.30	0.02
1-Year	1.05	0.09	1.35	0.04
3-Year	1.10	0.07	1.50	0.03

** Since these effect sizes were non-significant ($p < .05$) at baseline in the main Propensity Score Matching Model, these Rosenbaum Bounds indicate the size of the bias in order to move the effect size to a statistically significant level in each direction. Thus, for the 3-year re-arrest rate and 6-month overall recidivism rate, the bias under the "Bias for Non-Significant Group Difference" column is actually the bias needed to move these two effects to a significant group difference.

TABLE 4.12

Logit Regression Predicting Overall Recidivism Among Matched Sample, Controlling for Aging and Exposure Time Effects

	Overall Recidivism Rates (N=2,580)											
	6-Month				1-Year				3-Year			
	β	OR	Z	*p*	β	OR	Z	*p*	β	OR	Z	*p*
Sanctioned to Imprisonment (treatment)	-0.0493	0.95	-0.56	0.57	-0.12	0.89	-1.46	0.14	-0.13	0.88	-1.33	0.19
Age at Time of Recidivism Follow-up	-0.0164	0.98	-3.54	0.00	-0.02	0.98	-4.85	0.00	-0.03	0.97	-6.75	0.00
Residual Parole Time at Recidivism Follow-up	0.0021	1.00	1.81	0.07	0.00	1.00	2.93	0.00	0.01	1.01	4.86	0.00

TABLE 5.1
Bivariate Correlations (N = 1,547)

	Length of Stay		Overall Recidivism					
			3-Year		1-Year		6-Month	
	Corr.	*p*	Corr.	*p*	Corr.	*p*	Corr.	*p*
Age (at first violation)	0.01	0.70	-0.15	0.00	-0.13	0.00	-0.09	0.00
Race								
White	0.00	0.87	-0.01	0.65	-0.04	0.12	-0.01	0.69
Black	-0.03	0.27	0.01	0.75	0.03	0.18	0.02	0.50
Hispanic	0.03	0.20	0.00	0.99	0.00	0.88	-0.01	0.70
Gender (male)	-0.01	0.02	0.02	0.35	0.00	0.88	0.00	0.95
LSI-R criminal risk score	0.13	0.00	0.05	0.03	0.09	0.00	0.07	0.00
Offense type (original sentence)								
Violent	0.13	0.00	-0.01	0.83	-0.02	0.45	0.00	0.94
Property	-0.01	0.83	0.04	0.14	0.02	0.50	0.02	0.37
Drugs	-0.09	0.00	-0.02	0.34	0.00	0.88	-0.05	0.04
Prior treatment programs	0.05	0.04	0.02	0.37	0.00	0.98	-0.01	0.59
Sentencing county								
Philadelphia/Pittsburgh	0.00	0.98	-0.01	0.82	0.00	0.88	0.01	0.58
Class 3+ county	-0.01	0.66	0.01	0.69	0.05	0.07	0.02	0.49
Initial parole	-0.07	0.01	0.04	0.14	0.06	0.02	0.04	0.12
Parole supervision length	0.21	0.00	0.02	0.50	0.01	0.57	0.00	0.96
Parole district office								
Allentown	0.05	0.04	-0.01	0.63	-0.01	0.77	-0.02	0.51
Altoona	-0.06	0.01	0.04	0.12	0.02	0.37	0.04	0.11
Chester	-0.03	0.26	0.01	0.77	-0.02	0.50	-0.02	0.49
Erie	0.00	0.91	-0.02	0.54	0.01	0.80	0.05	0.06
Harrisburg	0.08	0.00	-0.03	0.22	0.00	0.95	-0.03	0.27
Mercer	0.00	0.99	0.00	0.85	-0.03	0.27	-0.01	0.65
Philadelphia	0.02	0.37	0.01	0.70	-0.01	0.77	0.01	0.69
Pittsburgh	-0.03	0.30	0.04	0.14	0.01	0.57	0.01	0.66
Scranton	-0.08	0.00	-0.02	0.53	0.01	0.70	0.00	0.90
Williamsport	0.02	0.45	-0.07	0.01	-0.05	0.06	-0.07	0.01
Supervision level								
Administrative	-0.04	0.08	-0.01	0.82	-0.01	0.58	0.00	0.93
Minimum	-0.09	0.00	-0.05	0.06	-0.06	0.02	-0.06	0.03
Medium	-0.05	0.06	0.04	0.16	0.02	0.53	0.02	0.54
Maximum	0.09	0.00	-0.02	0.39	0.02	0.44	0.01	0.62
First violation severity (most serious)								
Medium	-0.11	0.00	0.00	0.98	0.01	0.77	0.04	0.09
High	0.11	0.00	0.00	0.86	-0.01	0.70	-0.04	0.14
In Center at Time of Violation	0.01	0.64	-0.01	0.59	-0.02	0.50	0.02	0.52
No Recommit Action (NRA)	-0.48	0.00	0.04	0.12	0.04	0.11	0.02	0.44

TABLE 5.1 (continued)
Bivariate Correlations (N = 1,547)

	Re-Arrest Rates					
	3-Year		1-Year		6-Month	
	Corr.	*p*	Corr.	*p*	Corr.	*p*
Age (at first violation)	-0.13	0.00	-0.10	0.00	-0.09	0.00
Race						
White	-0.03	0.29	-0.06	0.03	-0.02	0.03
Black	0.04	0.10	0.06	0.03	0.03	0.32
Hispanic	-0.03	0.31	0.00	0.90	-0.01	0.66
Gender (male)	0.02	0.45	0.02	0.55	0.02	0.44
LSI-R criminal risk score	0.04	0.10	0.05	0.04	0.02	0.53
Offense type (original sentence)						
Violent	-0.03	0.18	-0.02	0.47	0.01	0.84
Property	0.04	0.14	-0.01	0.82	-0.01	0.74
Drugs	-0.01	0.66	0.00	0.88	-0.03	0.20
Prior treatment programs	0.05	0.07	0.03	0.19	-0.02	0.54
Sentencing county						
Philadelphia/Pittsburgh	0.04	0.15	0.05	0.03	0.04	0.14
Class 3+ county	0.05	0.04	0.09	0.00	0.03	0.19
Initial parole	0.00	0.90	0.01	0.64	0.01	0.60
Parole supervision length	-0.13	0.00	-0.07	0.00	-0.06	0.02
Parole district office						
Allentown	-0.03	0.23	0.01	0.69	0.01	0.60
Altoona	0.01	0.60	0.02	0.53	0.06	0.03
Chester	0.01	0.63	0.00	0.86	-0.01	0.74
Erie	-0.01	0.70	-0.03	0.21	-0.01	0.79
Harrisburg	-0.02	0.43	0.00	0.99	-0.02	0.46
Mercer	0.03	0.19	0.01	0.81	0.02	0.43
Philadelphia	0.08	0.00	0.07	0.01	0.05	0.07
Pittsburgh	0.00	0.85	-0.03	0.22	-0.02	0.36
Scranton	-0.06	0.02	-0.04	0.08	-0.05	0.06
Williamsport	-0.06	0.03	-0.02	0.43	-0.04	0.08
Supervision level						
Administrative	0.01	0.71	0.00	0.90	0.02	0.44
Minimum	-0.05	0.04	-0.06	0.01	-0.03	0.25
Medium	0.02	0.53	-0.01	0.56	0.00	0.94
Maximum	0.00	0.88	0.04	0.09	0.01	0.63
First violation severity (most serious)						
Medium	-0.02	0.38	-0.02	0.43	0.03	0.27
High	0.03	0.18	0.02	0.33	-0.02	0.41
In Center at Time of Violation	-0.05	0.07	-0.07	0.01	-0.06	0.02
No Recommit Action (NRA)	-0.06	0.02	-0.05	0.04	-0.02	0.33

TABLE 5.1 (continued)
Bivariate Correlations (N = 1,547)

	Re-Incarceration Rates					
	3-Year		1-Year		6-Month	
	Corr.	*p*	Corr.	*p*	Corr.	*p*
Age (at first violation)	-0.10	0.00	-0.05	0.07	-0.02	0.36
Race						
White	-0.02	0.46	0.00	0.98	-0.01	0.61
Black	-0.01	0.62	0.00	0.92	0.02	0.54
Hispanic	0.04	0.08	0.00	0.98	-0.01	0.83
Gender (male)	0.02	0.35	-0.01	0.76	0.00	0.98
LSI-R criminal risk score	0.06	0.02	0.06	0.02	0.06	0.02
Offense type (original sentence)						
Violent	-0.01	0.66	0.00	0.91	-0.01	0.83
Property	0.03	0.20	0.01	0.57	0.03	0.26
Drugs	0.00	0.87	-0.02	0.53	-0.04	0.14
Prior treatment programs	-0.02	0.51	-0.03	0.25	0.00	0.87
Sentencing county						
Philadelphia/Pittsburgh	-0.03	0.26	-0.03	0.24	-0.01	0.72
Class 3+ county	-0.02	0.44	-0.03	0.24	-0.01	0.79
Initial parole	0.11	0.00	0.09	0.00	0.05	0.07
Parole supervision length	0.21	0.00	0.14	0.00	0.08	0.00
Parole district office						
Allentown	0.01	0.69	-0.04	0.16	-0.04	0.10
Altoona	0.04	0.08	0.06	0.03	0.00	0.94
Chester	-0.02	0.35	-0.04	0.08	-0.03	0.30
Erie	0.01	0.84	0.03	0.26	0.06	0.02
Harrisburg	-0.04	0.16	-0.02	0.38	-0.02	0.45
Mercer	-0.04	0.11	-0.04	0.11	-0.02	0.37
Philadelphia	-0.02	0.46	-0.01	0.57	0.00	0.89
Pittsburgh	0.02	0.39	0.02	0.47	0.03	0.23
Scranton	0.02	0.38	0.05	0.05	0.03	0.21
Williamsport	-0.06	0.03	-0.06	0.01	-0.05	0.05
Supervision level						
Administrative	-0.01	0.61	0.00	0.88	0.02	0.50
Minimum	-0.05	0.08	-0.01	0.69	-0.03	0.33
Medium	0.04	0.11	0.02	0.38	0.03	0.31
Maximum	-0.03	0.20	-0.02	0.36	-0.02	0.55
First violation severity (most serious)						
Medium	0.04	0.15	0.05	0.04	0.06	0.01
High	-0.04	0.13	-0.06	0.02	-0.06	0.02
In Center at Time of Violation	-0.01	0.81	0.04	0.12	0.08	0.00
No Recommit Action (NRA)	0.13	0.00	0.15	0.00	0.07	0.01

TABLE 5.2
"Naïve" Logistic Regression Models Predicting Recidivism (N=1,547)

	Re-incarceration					
	6-Month		1-Year		3-Year	
	OR	*p*	OR	*p*	OR	*p*
Length of Stay in Prison (Dosage)	0.98	0.03	0.98	0.01	0.96	0.00
Age (at first violation)	0.99	0.19	0.98	0.01	0.97	0.00
Race						
White	0.56	0.63	0.44	0.44	0.36	0.42
Black	0.75	0.81	0.54	0.56	0.41	0.49
Hispanic	0.78	0.84	0.56	0.59	0.50	0.59
Gender (male)	0.85	0.60	0.77	0.33	1.08	0.76
LSI-R criminal risk score	1.03	0.01	1.03	0.00	1.03	0.00
Offense type (original sentence)						
Violent	0.79	0.27	0.89	0.53	0.74	0.08
Property	1.10	0.64	1.07	0.71	1.12	0.50
Drugs	0.72	0.13	0.89	0.51	0.87	0.42
Prior treatment programs	1.00	0.94	0.98	0.41	1.00	0.99
Sentencing county						
Philadelphia/Pittsburgh	0.78	0.26	0.72	0.10	0.79	0.21
Class 3+ county	0.96	0.84	0.91	0.58	0.90	0.47
Initial parole	1.19	0.27	1.39	0.02	1.32	0.03
Parole supervision length	1.01	0.00	1.01	0.00	1.02	0.00
Parole district office						
Allentown	0.35	0.27	0.28	0.20	0.30	0.30
Altoona	0.42	0.37	0.45	0.43	0.37	0.40
Chester	0.27	0.25	0.18	0.12	0.30	0.32
Erie	0.71	0.73	0.41	0.39	0.30	0.32
Harrisburg	0.36	0.29	0.28	0.20	0.24	0.22
Mercer	0.20	0.26	0.08	0.08	0.11	0.11
Philadelphia	0.51	0.49	0.41	0.37	0.34	0.36
Pittsburgh	0.53	0.50	0.39	0.34	0.36	0.38
Scranton	0.50	0.47	0.42	0.38	0.35	0.37
Williamsport	0.08	0.07	0.05	0.02	0.13	0.11
Supervision level						
Administrative	1.84	0.65	0.67	0.76	0.18	0.19
Minimum	0.85	0.70	0.76	0.46	0.34	0.00
Medium	1.04	0.93	0.80	0.50	0.47	0.03
Maximum	0.76	0.40	0.69	0.21	0.41	0.00
First violation severity (most serious)						
Medium	1.15	0.90	0.22	0.06	0.53	0.45
High	0.76	0.80	0.17	0.03	0.48	0.37
In Center at Time of Violation	1.50	0.01	1.15	0.32	0.85	0.24
No Recommit Action (NRA)	1.24	0.27	1.76	0.00	1.47	0.02

TABLE 5.2 (continued)
"Naïve" Logistic Regression Models Predicting Recidivism (N=1,547)

	Re-arrest					
	6-Month		1-Year		3-Year	
	OR	*p*	OR	*p*	OR	*p*
Length of Stay in Prison (Dosage)	1.00	0.80	1.01	0.46	1.00	0.63
Age (at first violation)	0.97	0.00	0.97	0.00	0.97	0.00
Race						
White	n.a.	n.a.	0.44	0.42	0.45	0.50
Black	n.a.	n.a.	0.51	0.51	0.50	0.55
Hispanic	n.a.	n.a.	0.41	0.39	0.41	0.45
Gender (male)	1.18	0.61	0.97	0.92	0.96	0.88
LSI-R criminal risk score	1.01	0.64	1.01	0.26	1.01	0.36
Offense type (original sentence)						
Violent	0.86	0.46	0.80	0.22	0.84	0.31
Property	0.78	0.23	0.90	0.58	1.11	0.55
Drugs	0.61	0.02	0.76	0.12	0.78	0.15
Prior treatment programs	0.98	0.50	1.03	0.21	1.04	0.08
Sentencing county						
Philadelphia/Pittsburgh	1.33	0.10	1.11	0.57	0.91	0.61
Class 3+ county	1.17	0.36	1.68	0.00	1.30	0.07
Initial parole	1.00	0.99	1.09	0.53	0.95	0.70
Parole supervision length	0.99	0.03	1.00	0.01	0.99	0.00
Parole district office						
Allentown	n.a.	n.a.	1.94	0.56	0.29	0.28
Altoona	n.a.	n.a.	3.06	0.34	0.43	0.46
Chester	n.a.	n.a.	1.50	0.74	0.34	0.38
Erie	n.a.	n.a.	1.36	0.79	0.28	0.27
Harrisburg	n.a.	n.a.	1.99	0.55	0.30	0.29
Mercer	n.a.	n.a.	3.13	0.38	0.94	0.96
Philadelphia	n.a.	n.a.	2.03	0.54	0.46	0.50
Pittsburgh	n.a.	n.a.	1.76	0.62	0.37	0.39
Scranton	n.a.	n.a.	1.54	0.71	0.24	0.22
Williamsport	n.a.	n.a.	1.92	0.60	0.15	0.12
Supervision level						
Administrative	3.07	0.39	2.32	0.52	2.80	0.44
Minimum	0.64	0.30	0.63	0.20	0.71	0.31
Medium	0.80	0.57	0.89	0.71	0.99	0.96
Maximum	0.83	0.58	0.91	0.74	0.75	0.33
First violation severity (most serious)						
Medium	n.a.	n.a.	1.53	0.70	4.20	0.20
High	n.a.	n.a.	1.61	0.67	4.51	0.18
In Center at Time of Violation	0.62	0.01	0.64	0.00	0.77	0.04
No Recommit Action (NRA)	0.90	0.63	0.85	0.35	0.80	0.18

TABLE 5.2 (continued)
"Naïve" Logistic Regression Models Predicting Recidivism (N=1,547)

	Overall Recidivism					
	6-Month		1-Year		3-Year	
	OR	*p*	OR	*p*	OR	*p*
Length of Stay in Prison (Dosage)	0.99	0.19	0.99	0.20	0.98	0.02
Age (at first violation)	0.98	0.00	0.97	0.00	0.96	0.00
Race						
White	1.56	0.71	0.34	0.35	0.00	0.99
Black	1.80	0.62	0.43	0.47	0.00	0.99
Hispanic	1.72	0.65	0.39	0.43	0.00	0.99
Gender (male)	0.83	0.47	0.82	0.40	1.01	0.98
LSI-R criminal risk score	1.02	0.01	1.03	0.00	1.02	0.01
Offense type (original sentence)						
Violent	0.83	0.29	0.84	0.31	0.89	0.51
Property	0.97	0.88	1.03	0.85	1.16	0.44
Drugs	0.65	0.02	0.81	0.20	0.77	0.16
Prior treatment programs	0.98	0.53	1.00	0.95	1.02	0.37
Sentencing county						
Philadelphia/Pittsburgh	1.07	0.74	0.94	0.74	0.77	0.19
Class 3+ county	1.13	0.44	1.31	0.06	1.13	0.44
Initial parole	1.07	0.59	1.17	0.18	1.05	0.70
Parole supervision length	1.00	0.42	1.00	0.13	1.00	0.06
Parole district office						
Allentown	0.76	0.77	0.29	0.28	0.00	0.99
Altoona	1.15	0.88	0.42	0.45	0.00	0.99
Chester	0.64	0.66	0.24	0.24	0.00	0.99
Erie	1.13	0.90	0.32	0.33	0.00	0.99
Harrisburg	0.69	0.69	0.30	0.29	0.00	0.99
Mercer	0.68	0.74	0.21	0.23	0.00	0.99
Philadelphia	0.81	0.82	0.29	0.28	0.00	0.99
Pittsburgh	0.82	0.84	0.35	0.35	0.00	0.99
Scranton	0.77	0.78	0.33	0.33	0.00	0.99
Williamsport	0.15	0.12	0.16	0.14	0.00	0.99
Supervision level						
Administrative	1.25	0.87	0.70	0.78	0.60	0.70
Minimum	0.65	0.23	0.69	0.26	0.43	0.05
Medium	0.92	0.79	0.92	0.79	0.61	0.24
Maximum	0.77	0.35	0.80	0.41	0.43	0.03
First violation severity (most serious)						
Medium	1.93	0.55	0.45	0.31	1.08	0.92
High	1.45	0.74	0.44	0.29	1.10	0.91
In Center at Time of Violation	1.01	0.95	0.83	0.13	0.89	0.40
No Recommit Action (NRA)	1.03	0.87	1.20	0.25	1.13	0.50

TABLE 5.3
Descriptive Statistics for Treatment Group Prison Length of Stay Quintiles

	Quintile 1: (0 – 5.1 months) (n=230)		Quintile 2: (5.2 – 8.4 months) (n=233)		Quintile 3: (8.5 – 11.8 months) (n=222)		Quintile 4: (11.9 – 16.3 months) (n=233)		Quintile 5: (16.4 – 45.3 months) (n=226)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (at first violation)	32.98	9.14	34.89	10.17	33.81	9.12	33.3	9.5	34.48	8.98
Race										
*White**	0.33	0.47	0.45	0.5	0.4	0.49	0.48	0.5	0.43	0.5
*Black***	0.59	0.49	0.41	0.49	0.42	0.49	0.42	0.49	0.45	0.5
Hispanic	0.09	0.28	0.13	0.34	0.17	0.38	0.1	0.3	0.11	0.31
Gender (*male*)	0.96	0.2	0.93	0.24	0.92	0.27	0.95	0.22	0.94	0.23
LSI-R criminal risk score*	25.79	6.77	25.6	7.31	25.15	7.65	25.69	7.23	27.16	7.56
Offense type (original sentence)										
*Violent***	0.24	0.43	0.2	0.4	0.28	0.45	0.33	0.47	0.41	0.49
Property	0.24	0.43	0.26	0.44	0.27	0.45	0.24	0.42	0.24	0.43
*Drugs***	0.34	0.48	0.33	0.47	0.27	0.45	0.25	0.43	0.21	0.41
Prior treatment programs	3.33	2.76	2.95	2.28	3.04	2.39	3.26	2.35	3.4	2.45
Sentencing county										
*Philadelphia/Pittsburgh***	0.37	0.48	0.2	0.4	0.24	0.43	0.28	0.45	0.29	0.45
Class 3+ county	0.74	0.44	0.67	0.47	0.69	0.46	0.72	0.45	0.65	0.48
Initial Parole	0.64	0.48	0.7	0.46	0.69	0.46	0.69	0.46	0.59	0.49
Parole supervision length**	30.96	40.44	27.14	23.18	37.9	34.56	43.36	51.31	57.96	84.6
Parole district office										
*Allentown***	0.14	0.35	0.22	0.42	0.25	0.43	0.18	0.39	0.18	0.39
Altoona	0.02	0.13	0.06	0.24	0.07	0.25	0.03	0.18	0.04	0.21
Chester	0.03	0.16	0.02	0.16	0.01	0.09	0.03	0.17	0	0.07
Erie	0.03	0.16	0.03	0.17	0.04	0.2	0.05	0.22	0.03	0.17

	Quintile 1: 0 – 5.1 months (N=230)		Quintile 2: 5.2 – 8.4 months (N=233)		Quintile 3: 8.5 – 11.8 months (N=222)		Quintile 4: 11.9 – 16.3 months (N=233)		Quintile 5: 16.4 – 45.3 months (N=226)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Harrisburg	0.1	0.3	0.13	0.34	0.14	0.35	0.15	0.36	0.16	0.37
Mercer	0.01	0.11	0.01	0.09	0	0.07	0.01	0.09	0.01	0.09
Philadelphia	0.33	0.47	0.24	0.43	0.22	0.42	0.25	0.43	0.28	0.45
*Pittsburgh***	0.26	0.44	0.13	0.34	0.15	0.36	0.17	0.38	0.19	0.39
Scranton	0.07	0.25	0.11	0.32	0.1	0.3	0.1	0.3	0.07	0.26
Williamsport	0.02	0.13	0.02	0.14	0	0.07	0.01	0.11	0.03	0.17
Supervision level										
Minimum	0.07	0.25	0.09	0.29	0.08	0.27	0.05	0.23	0.05	0.21
Medium	0.09	0.29	0.11	0.31	0.1	0.3	0.16	0.37	0.1	0.3
Maximum	0.79	0.41	0.77	0.42	0.76	0.43	0.72	0.45	0.82	0.39
First violation severity										
Medium	-	-	-	-	-	-	-	-	-	-
High	-	-	-	-	-	-	-	-	-	-
In Center at Time of Violation**	0.36	0.48	0.21	0.41	0.24	0.43	0.27	0.44	0.27	0.44
No Recommit Action (NRA)	-	-	-	-	-	-	-	-	-	-

TABLE 5.4
Observed Recidivism Rates Before Stratification

	Quintile 1: 0 - 5.1 months (n=230)	Quintile 2: 5.2 - 8.4 months (n=233)	Quintile 3: 8.5 - 11.8 months (n=222)	Quintile 4: 11.9 - 16.3 months (n=233)	Quintile 5: 16.4 - 45.3 months (n=226)
Re-incarceration Rates					
6-Month	20.9%	15.9%	14.9%	15.9%	15.9%
1-Year	27.8%	25.3%	24.8%	27.5%	23.5%
3-Year	47.0%	45.1%	47.3%	43.4%	41.6%
Re-arrest Rates					
6-Month	17.0%	15.5%	19.4%	15.9%	15.9%
1-Year	33.0%	30.0%	32.0%	30.5%	31.0%
3-Year	61.3%	55.4%	58.6%	58.4%	55.8%
Overall Recidivism Rates					
6-Month	33.9%	27.9%	30.2%	30.5%	27.9%
1-Year	52.2%	47.6%	46.9%	50.2%	45.6%
3-Year	75.7%	70.4%	71.6%	74.3%	68.1%

TABLE 5.5
Ordinal Logit Regression Predicting Dosage Assignment (N=1,144)

	coefficient	std. err.	Z	*p*
Age (at first violation)	0.02	0.04	0.41	0.68
% White	-1.34	1.19	-1.12	0.26
%Black	-1.67	1.19	-1.40	0.16
%Hispanic	-1.48	1.20	-1.24	0.22
LSI-R criminal risk score	0.00	0.01	-0.08	0.94
Violent (original offense)	0.55	0.19	2.92	0.00
Drugs (original offense)	-0.13	0.14	-0.95	0.35
Prior treatment programs	0.01	0.03	0.22	0.83
Philadelphia/Pittsburgh Sentencing County	-0.08	0.14	-0.58	0.56
Initial parole	-0.17	0.12	-1.40	0.16
Parole supervision length	0.01	0.01	1.63	0.10
Allentown Parole District Office	0.01	0.15	0.07	0.95
Pittsburgh Parole District Office	-0.73	0.54	-1.35	0.18
In Center at Time of Violation	-0.20	0.13	-1.55	0.12
Age^2	0.00	0.00	-0.37	0.72
Age*Pittsburgh	0.01	0.01	0.95	0.34
Age*ParoleTime	0.00	0.00	-0.94	0.35
$ParoleTime^2$	0.00	0.00	-3.58	0.00
ParoleTime*Violent	-0.01	0.00	-1.49	0.14
ParoleTime*PriorTreatmentPrograms	0.00	0.00	-0.09	0.93
ParoleTime*LSI-R	0.00	0.00	2.54	0.01
ParoleTime*Pittsburgh	0.00	0.00	-0.27	0.79

TABLE 5.6
Cross-tabulation of Number of Cases in
Each Dosage by Propensity Score Quintile

Propensity Score Quintile	Length of Stay Dosage				
	Dosage 1	Dosage 2	Dosage 3	Dosage 4	Dosage 5
Quintile 1	88	51	37	37	15
Quintile 2	50	61	50	41	26
Quintile 3	36	60	48	41	44
Quintile 4	27	36	39	63	63
Quintile 5	29	25	48	51	79

TABLE 5.7
ANOVAs and Logit Regressions of Pre- vs Post-Stratification Balance

	Pre-Stratification		Post-Stratification			
	Dosage		Dosage (Main Effect)		Dosage*Propensity Score (Interaction)	
	Z or F Score	*p*	Z or F Score	*p*	Z or F Score	*p*
Age (at first violation)	1.66	0.16	1.31	0.26	1.35	0.16
Race						
White	2.30	0.02	-1.48	0.14	1.41	0.16
Black	-2.61	0.01	1.20	0.23	-1.09	0.28
Hispanic	0.25	0.80	0.72	0.47	-0.86	0.39
Gender (male)	-0.31	0.76	0.38	0.70	-0.37	0.71
LSI-R criminal risk score	2.40	0.05	0.45	0.77	1.02	0.43
Offense type (original sentence)						
Violent	4.83	0.00	-1.42	0.16	1.36	0.17
Property	-0.27	0.79	-0.27	0.79	0.29	0.77
Drugs	-3.62	0.00	1.06	0.29	-0.90	0.37
Prior treatment programs	1.40	0.23	1.77	0.13	1.34	0.17
Sentencing county						
Philadelphia/Pittsburgh	-0.74	0.46	0.47	0.64	-0.45	0.66
Class 3+ county	-1.29	0.20	0.73	0.46	-1.14	0.26
Initial parole	-1.14	0.26	1.56	0.12	-1.65	0.10
Parole supervision length	12.72	0.00	2.35	0.05	1.43	0.12
Parole district office						
Allentown	0.56	0.58	1.35	0.18	-1.48	0.14
Altoona	0.54	0.59	0.60	0.55	-0.71	0.48
Chester	-1.33	0.18	-0.23	0.82	0.01	0.99
Erie	0.82	0.41	0.09	0.93	0.00	1.00
Harrisburg	1.89	0.06	0.11	0.91	0.42	0.67
Mercer	-0.43	0.67	0.86	0.39	-1.20	0.23

	Pre-Stratification		Post-Stratification			
	Dosage		Dosage (Main Effect)		Dosage*Propensity Score (Interaction)	
	Z or F Score	*p*	Z or F Score	*p*	Z or F Score	*p*
Philadelphia	-1.03	0.30	-0.62	0.54	0.27	0.79
Pittsburgh	-1.17	0.24	-0.89	0.38	1.23	0.22
Scranton	-0.16	0.87	0.08	0.94	-0.10	0.92
Williamsport	0.66	0.51	-0.27	0.79	0.55	0.58
Supervision level						
Minimum	-1.51	0.13	0.13	0.89	0.82	0.41
Medium	0.92	0.36	0.45	0.66	0.45	0.65
Maximum	0.01	0.99	-0.34	0.74	0.62	0.53
In Center at Time of Violation	-1.36	0.18	-0.64	0.52	0.92	0.36

TABLE 5.8
Recidivism Rates and Standard Errors from Primary Dose-Response Curve

	Quintile 1: 0 - 5.1 months (n=230)		Quintile 2: 5.2 - 8.4 months (n=233)		Quintile 3: 8.5 - 11.8 months (n=222)		Quintile 4: 11.9 - 16.3 months (n=233)		Quintile 5: 16.4 - 45.3 months (n=226)	
	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.
Re-incarceration Rates										
6-Month	23.2%	0.027	16.9%	0.015	14.6%	0.012	14.8%	0.028	11.2%	0.050
1-Year	32.1%	0.038	28.3%	0.043	24.5%	0.027	25.7%	0.043	16.3%	0.075
3-Year	52.2%	0.046	48.6%	0.058	46.4%	0.057	41.1%	0.059	33.3%	0.083
Re-arrest Rates										
6-Month	18.0%	0.021	14.8%	0.011	19.4%	0.026	16.2%	0.013	15.5%	0.026
1-Year	32.7%	0.037	28.8%	0.017	32.5%	0.029	31.2%	0.024	29.3%	0.032
3-Year	60.4%	0.036	52.2%	0.045	58.7%	0.026	56.2%	0.017	55.8%	0.022
Overall Recidivism Rates										
6-Month	35.8%	0.027	28.5%	0.014	30.4%	0.015	24.7%	0.014	23.9%	0.050
1-Year	52.4%	0.042	48.1%	0.016	47.0%	0.012	49.7%	0.016	39.5%	0.073
3-Year	77.1%	0.022	70.7%	0.039	71.1%	0.020	73.5%	0.026	64.7%	0.044

TABLE 5.9
Logistic Regressions and ANOVAs
Predicting Recidivism Using Dosage, Post Propensity Score Stratification

	Logit Models		ANOVAs	
	Z	*p*	F	*p*
Re-incarceration Rates				
6-Month	-2.61	0.009	2.14	0.074
1-Year	-3.07	0.002	2.83	0.024
3-Year	-4.29	0.000	4.84	0.001
Re-arrest Rates				
6-Month	0.18	0.855	0.45	0.769
1-Year	0.33	0.742	0.20	0.936
3-Year	0.17	0.862	0.59	0.673
Overall Recidivism Rates				
6-Month	-1.46	0.145	0.98	0.416
1-Year	-1.70	0.088	1.14	0.337
3-Year	-2.18	0.030	2.00	0.092

TABLE 5.10
Recidivism Rates and Standard Errors from Dose-Response Curve Comparing Dosages to Control Group

	Control Group (n=327)		Treatment Dose 1: 0 - 5.1 months (n=237)		Treatment Dose 2: 5.2 - 8.4 months (n=226)		Treatment Dose 3: 8.5 - 11.8 months (n=235)		Treatment Dose 4: 11.9 - 16.3 months (n=224)		Treatment Dose 5: 16.4 - 45.3 months (n=222)	
	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.
Re-incarceration Rates												
3-Year	58.9%	0.030	46.9%	0.034	45.0%	0.024	45.7%	0.036	42.6%	0.050	38.4%	0.035
Re-arrest Rates												
3-Year	58.2%	0.019	59.5%	0.031	55.0%	0.030	58.5%	0.026	59.3%	0.033	55.5%	0.035
Overall Recidivism Rates												
3-Year	78.7%	0.017	73.7%	0.025	70.3%	0.030	71.1%	0.023	74.1%	0.026	66.9%	0.020

TABLE 5.11
Recidivism Rates and Standard Errors from Dose-Response Curve (Dropping Doses > 18 Months)

	Quintile 1: 0 - 4.5 months (n=191)		Quintile 2: 4.6 - 7.3 months (n=187)		Quintile 3: 7.4 - 10.4 months (n=214)		Quintile 4: 10.5 - 13.4 months (n=188)		Quintile 5: 13.5 - 18 months (n=191)	
	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.	Rate	S.E.
Re-incarceration Rates										
3-Year	51.7%	0.043	47.1%	0.075	45.8%	0.043	45.1%	0.081	35.0%	0.075
Re-arrest Rates										
3-Year	58.3%	0.045	55.2%	0.045	56.5%	0.048	60.2%	0.031	58.4%	0.025
Overall Recidivism Rates										
3-Year	76.4%	0.022	70.5%	0.034	71.7%	0.025	72.2%	0.038	69.9%	0.033

FIGURE 1.1
Parole Violators as Percent of Prison Admissions (2007)

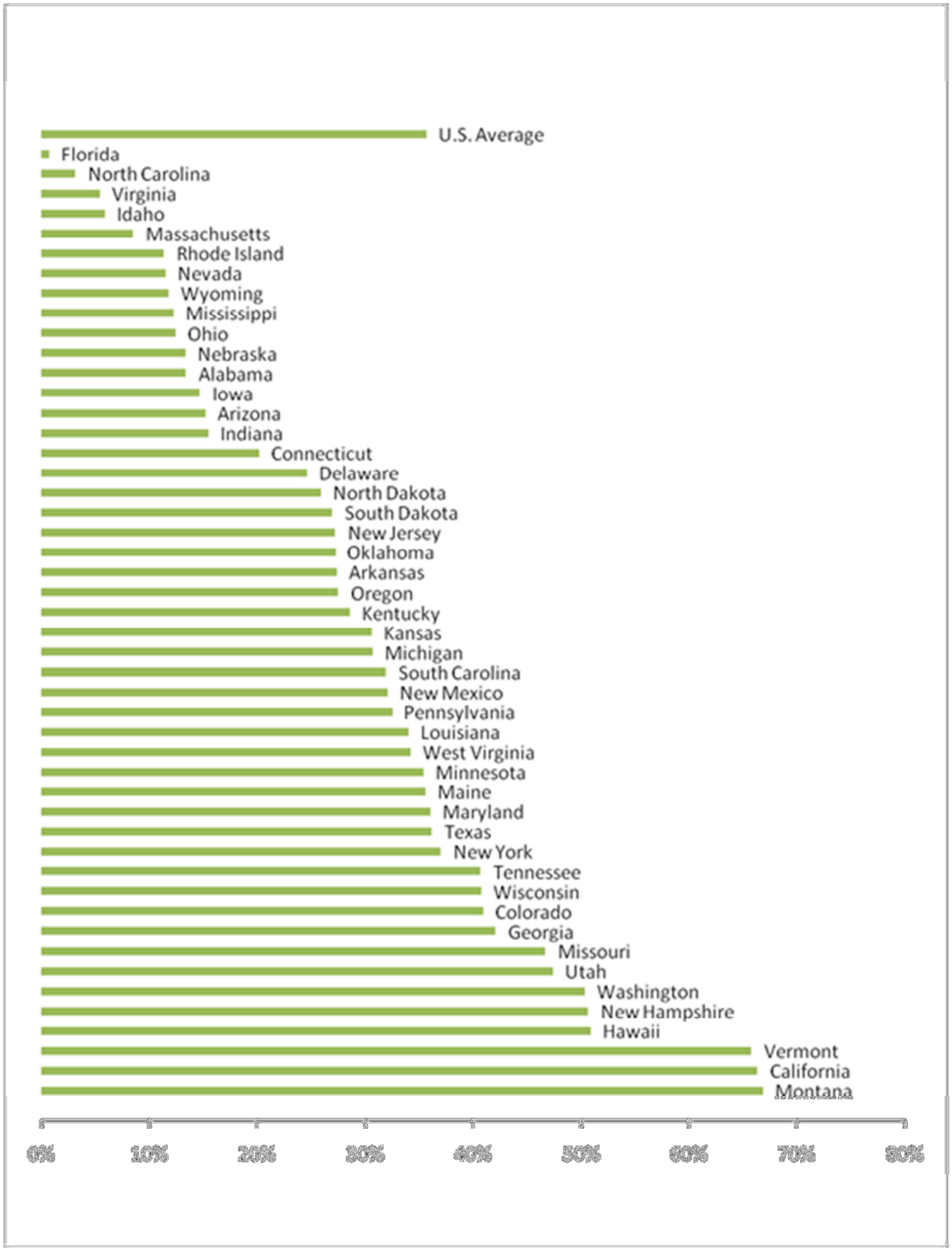


FIGURE 3.1
PBPP Violation Sanctioning Matrix

CON I - CON II
WORKSHEET

pennsylvania
BOARD OF PROBATION AND PAROLE

PBPP 347

OFFENDER		**DATE**		**AGENT/SUPV**			
PAROLE#		**TIME**		**UNIT#**		**OVRT**	

Violations - Severity

Tech	Conditions	Low	Med	High
	Condition 1			
	Failure to report upon release			H06
	Travel violations		M04	
	Condition 2			
	Changing residence without permission			H01
	Absconding			H09
	Condition 3			
	A - Failure to report as instructed		M02	
	B - Failure to notify agent of arrest or citation within 72 hrs		M19	
	C - Failure to notify agent of change in status/employment	L07	M01	
	Condition 4			
	Conviction of a summary offense		M13	
	Condition 5			
	A - Positive urine, drugs (No Hx)	L08	M03	H12
	B - Possession of offense weapon			H10
	B - Possession of firearm			H11
	C - Assaultive behavior			H08
	Condition 6			
	Failure to pay court ordered fees, restitution	L06		
	Condition 7			
	Failure to participate in community service	L01		
	Failure to pay supervision fees	L03		
	Failure to pay urinalysis fees	L04		
	Failure to support dependents	L05		
	Possession of contraband, cell phones, etc.		M05	
	Failure to take medications as prescribed		M06	
	Failure to maintain employment		M07	
	Failure to participate or maintain treatment		M08	
	Entering prohibited establishments		M09	
	Associating with gang members, co-defendants, etc		M10	
	Failure to abide by written instructions		M11	
	Failure to abide by field imposed special conditions		M12	
	Conviction of summary offense		M13	
	Positive urine, alcohol (Hx)	L02	M14	H03
	Violating curfew		M15	
	Violating electronic monitoring		M16	
	Failure to provide urine		M17	
	Failure to complete treatment		M18	
	Conviction of Misdemeanor Offense		M20	
	Associating with crime victims			H02
	Failure to abide by Board Imposed Special Conditions			H05
	Removal from Treatment/CCC Failure			H07
	Pending criminal charges (UCV) not detained			H04

	Low	Med	High
Severity of Violation			

	Low	Med	High
LSIR/Static 99 SCORE			

Risk/Severity Scale

SEVERITY OF VIOLATION				
	H	M	H	H
	M	L	M	H
	L	L	L	M
		L	M	H
		RISK (LSIR)		

	Low	Med	High
Baseline Sanctioning Range			

Code	Low Sanction Range
WTWR	Written Warning
DJBS	Documented Job Search
WTVR	Written Travel Restriction
IRPT	Increased Reporting Requirements
TXEV	Obtain Treatment/Evaluation
ACCG	Refer to ASCRA groups
CURF	Imposition of Curfew
LOTR	Other

Code	Medium Sanction Range
DFSE	Deadline for Securing Employment
URIN	Imposition of Increased Urinalysis Testing
ICRF	Imposition of Increased Curfew
GVPB	Refer to Violence Prevention Booster
COMS	Imposition of Community Service
OPAT	Placement in Out-Patient D&A Treatment
RECT	Refer to Re-entry Court
EMOS	Imposition of Electronic Monitoring
MOTR	Other

Code	High Sanction Range (Need Supv Approval)
IDOX	Placement in D&A Detox Facility
AGPS	Imposition of Global Positioning
CPCP	Placement in CCC Half Way Back
IPAT	Placement in In-Patient D&A Treatment
IPMH	Placement in Mental Health Facility
VCCF	Placement in Violation Center Contract Facility
HOTR	Other
ARR2	Incarceration

Criminogenic Needs

M/H	#	Needs	Currently a problem? Yes	No
	1	Criminal history		
	2	Education/Employment		
	3	Financial		
	4	Family/Marital		
	5	Accommodations		
	6	Leisure/Recreation		
	7	Companions		
	8	Substance abuse		
	9	Emotional/Personal		
	10	Anti-social attitudes		

Referral Information/Comments

Is there a departure from the Baseline Sanctioning Range; Yes or No - List reasons for change.

FIGURE 3.2
Histogram of TPV Length of Stay (in months) for Treatment Group

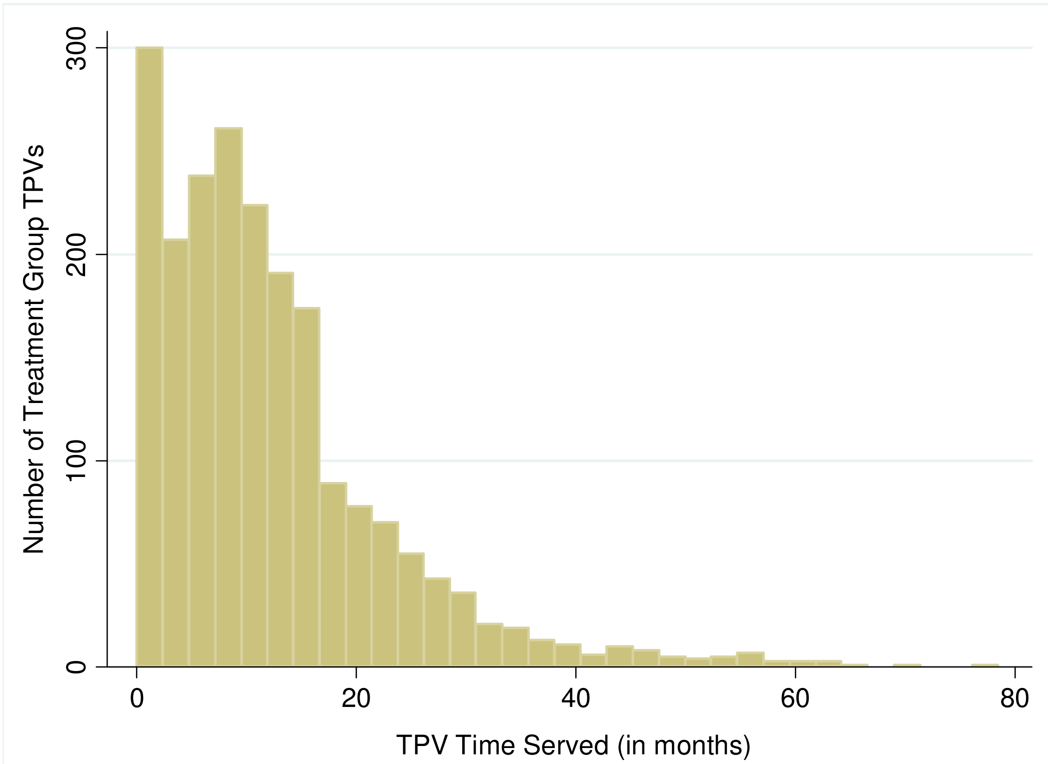


FIGURE 3.3

Box Plot of Propensity Score Distributions between Treatment and Control Groups

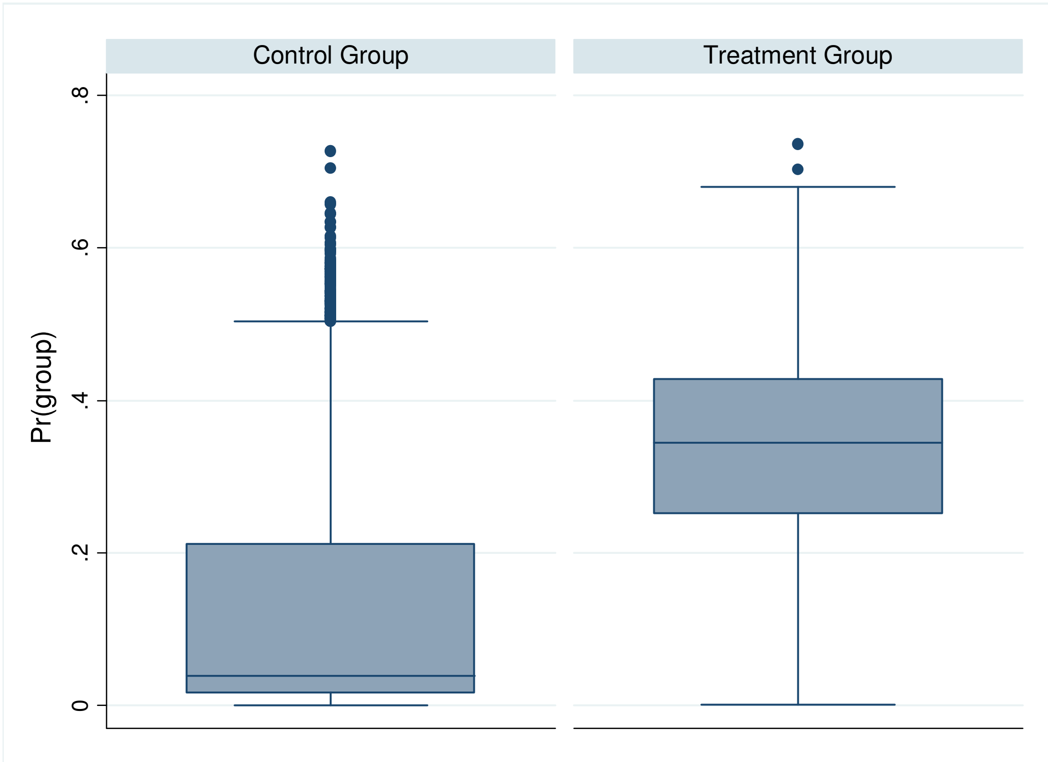


FIGURE 3.4
Distribution of Propensity Scores between Treatment and Control Groups

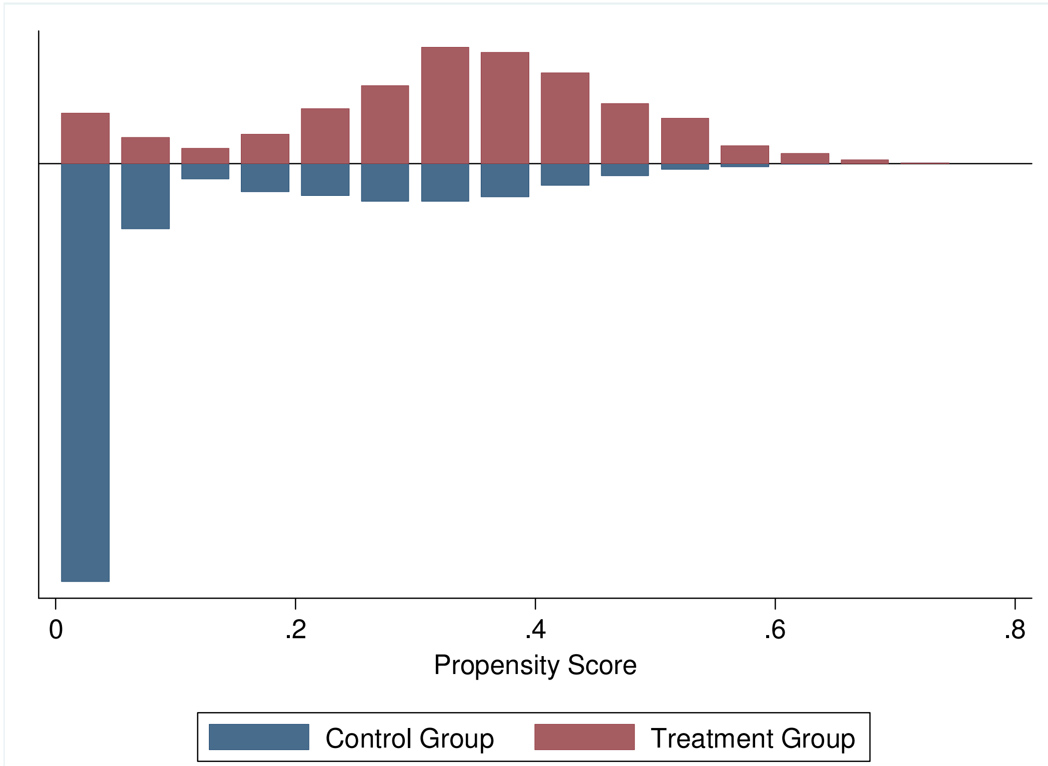


FIGURE 4.1
Jitter Dot-Plot of Propensity Score Distribution among Matched Sample

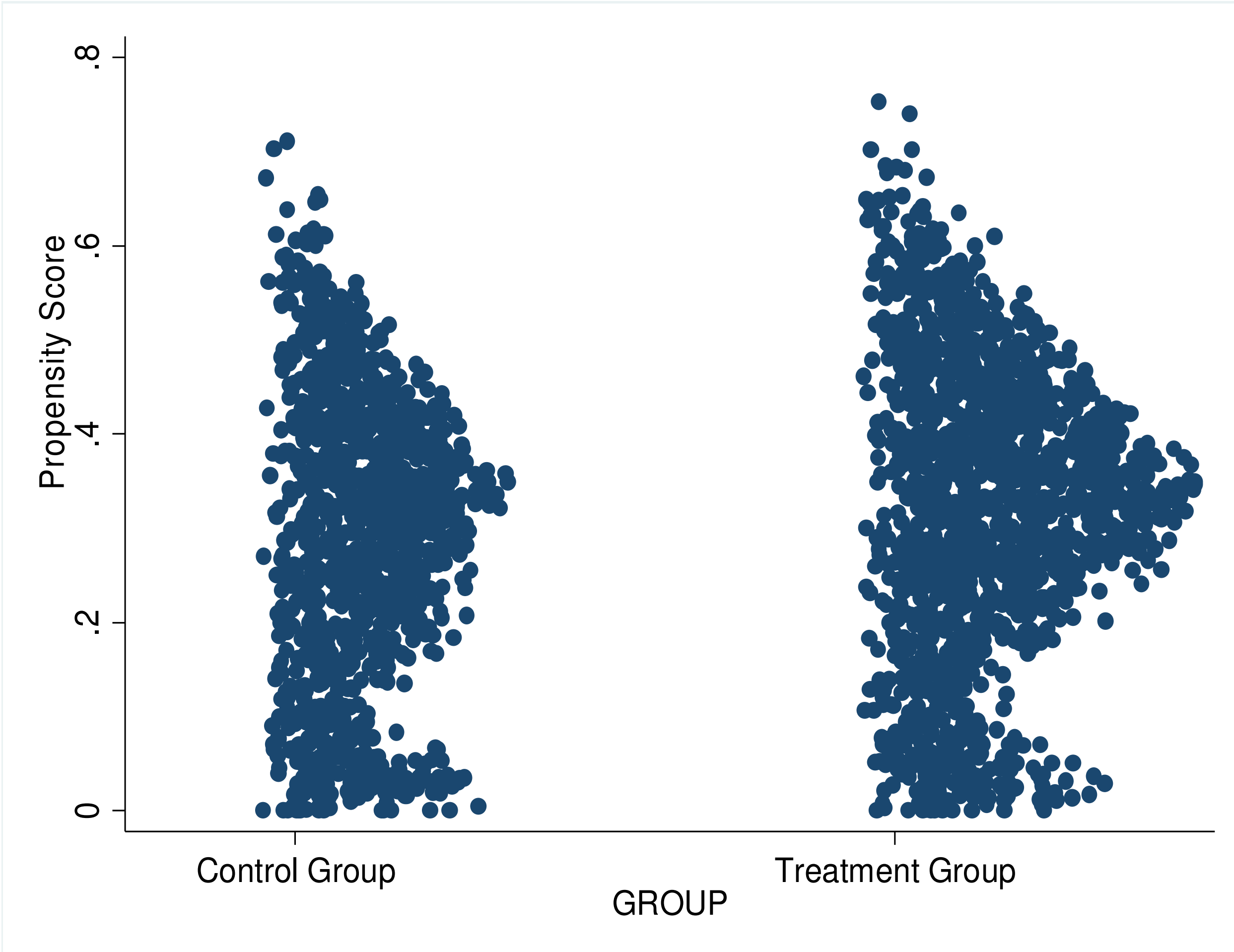


FIGURE 4.2
Histogram of Propensity Score Distribution between Groups after Matching

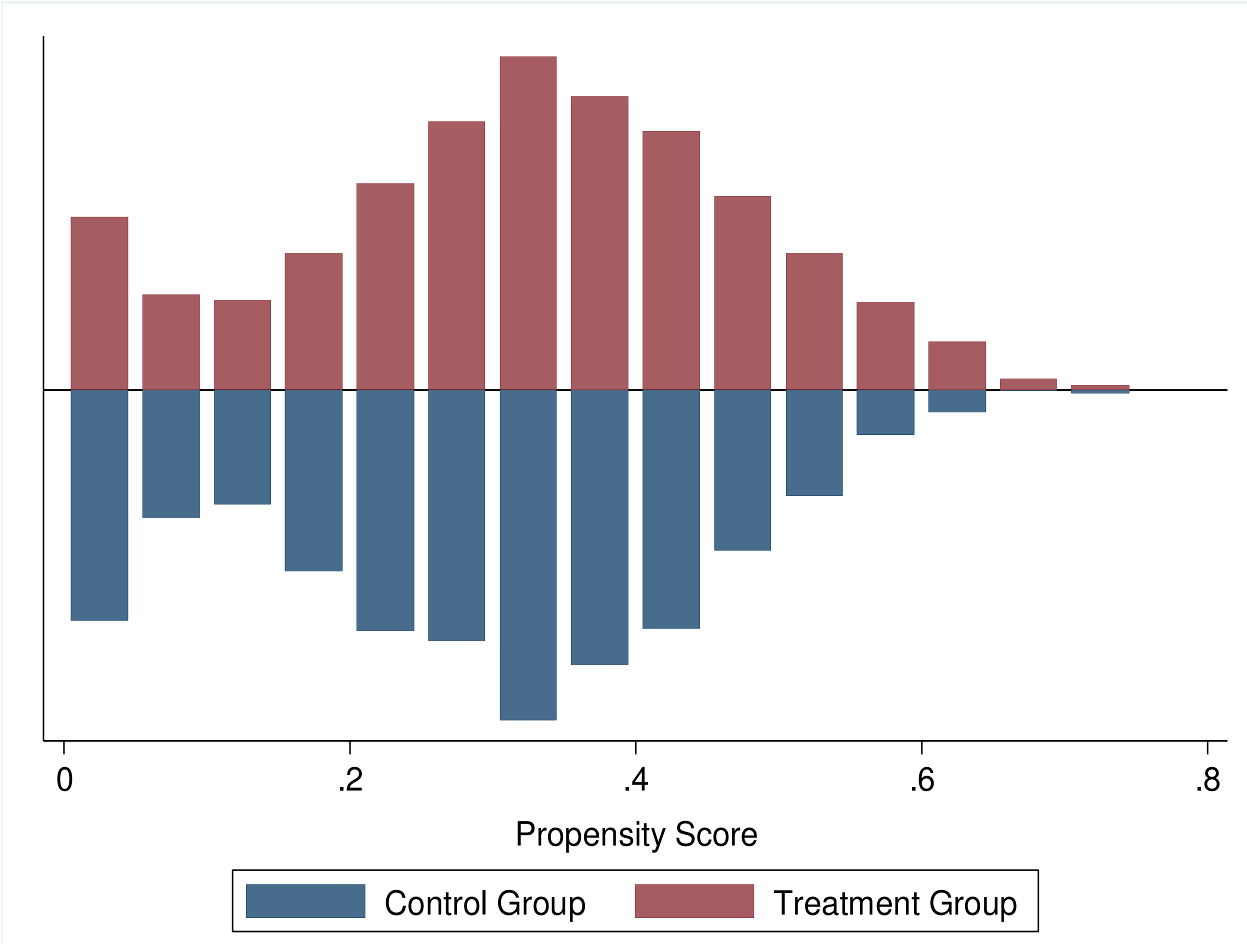
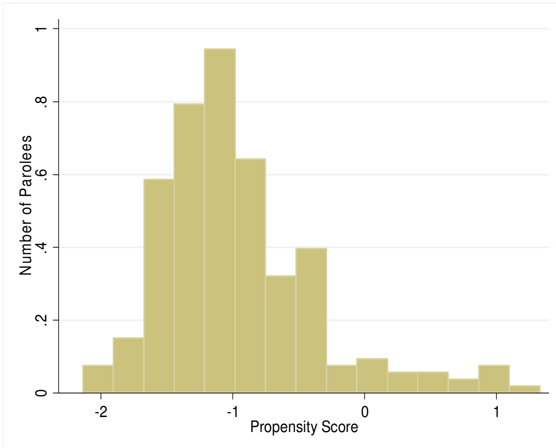
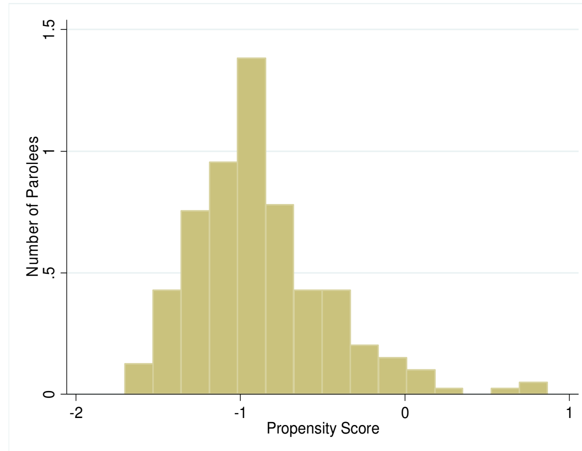


FIGURE 5.1
Histograms of the Estimated Propensity Score, in the Five Exposure Levels

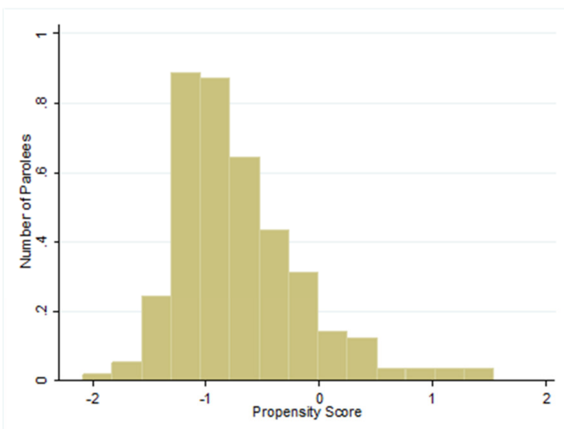
Dosage 1



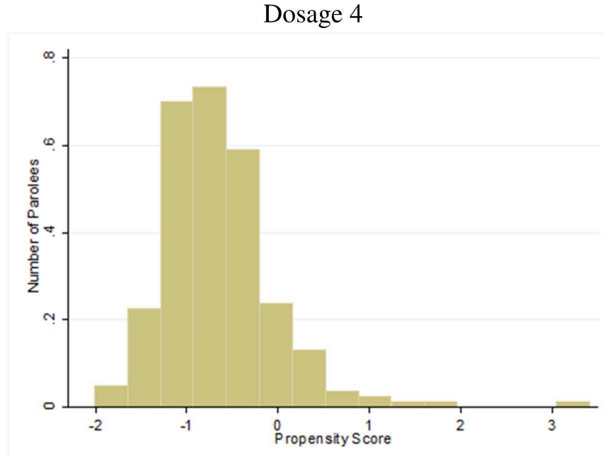
Dosage 2



Dosage 3



Dosage 4



Dosage 5

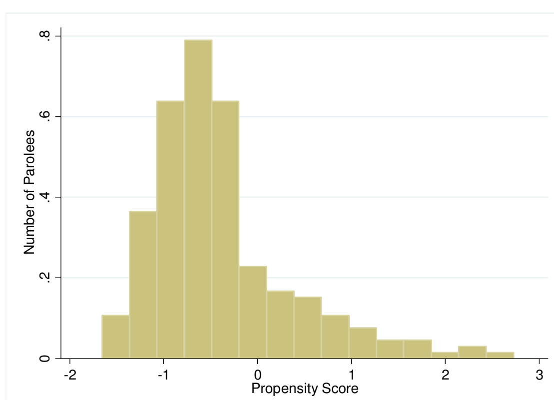


FIGURE 5.2
Dose-Response Curves for Main Models

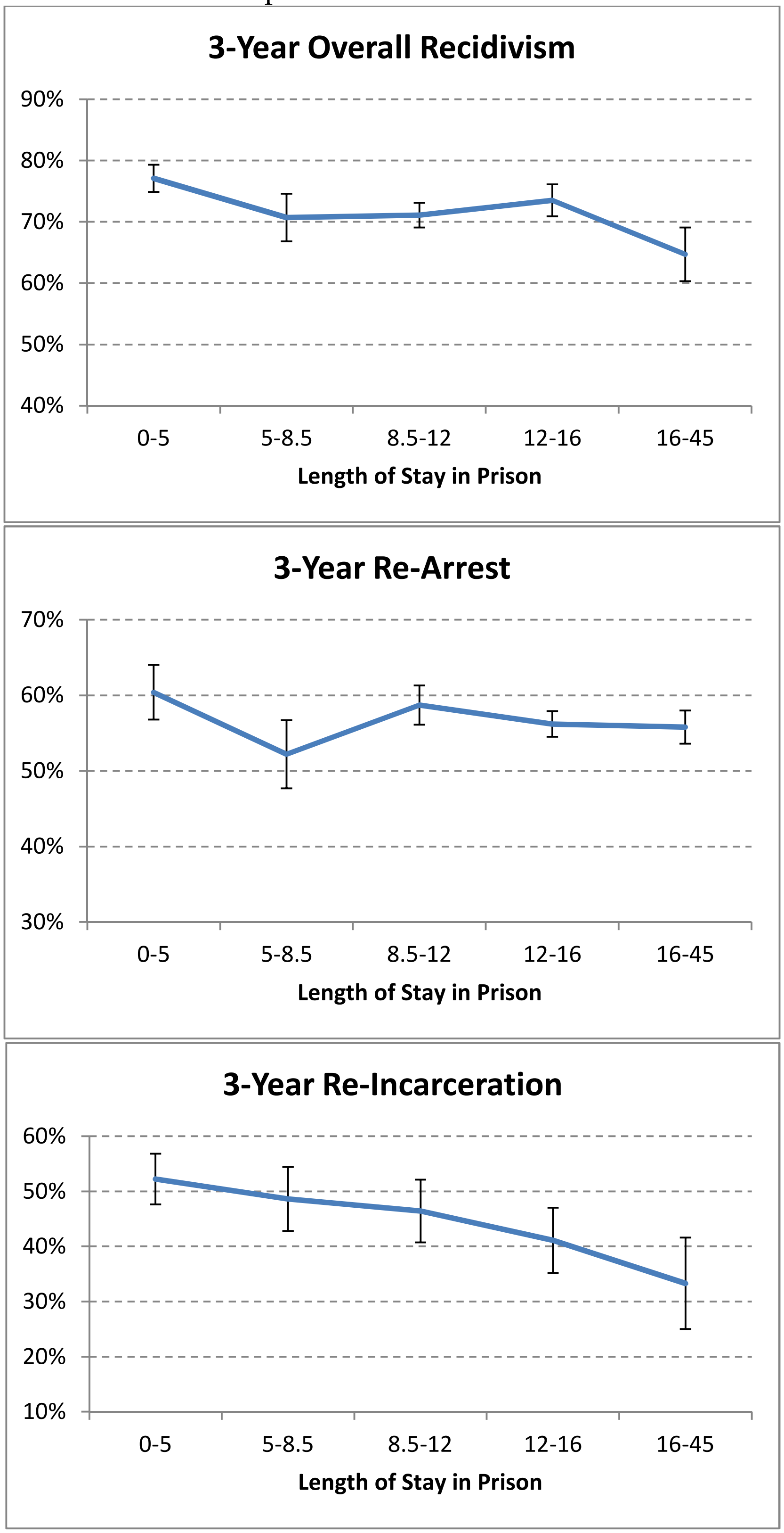


FIGURE 5.2 (continued)
Dose-Response Curves for Main Models

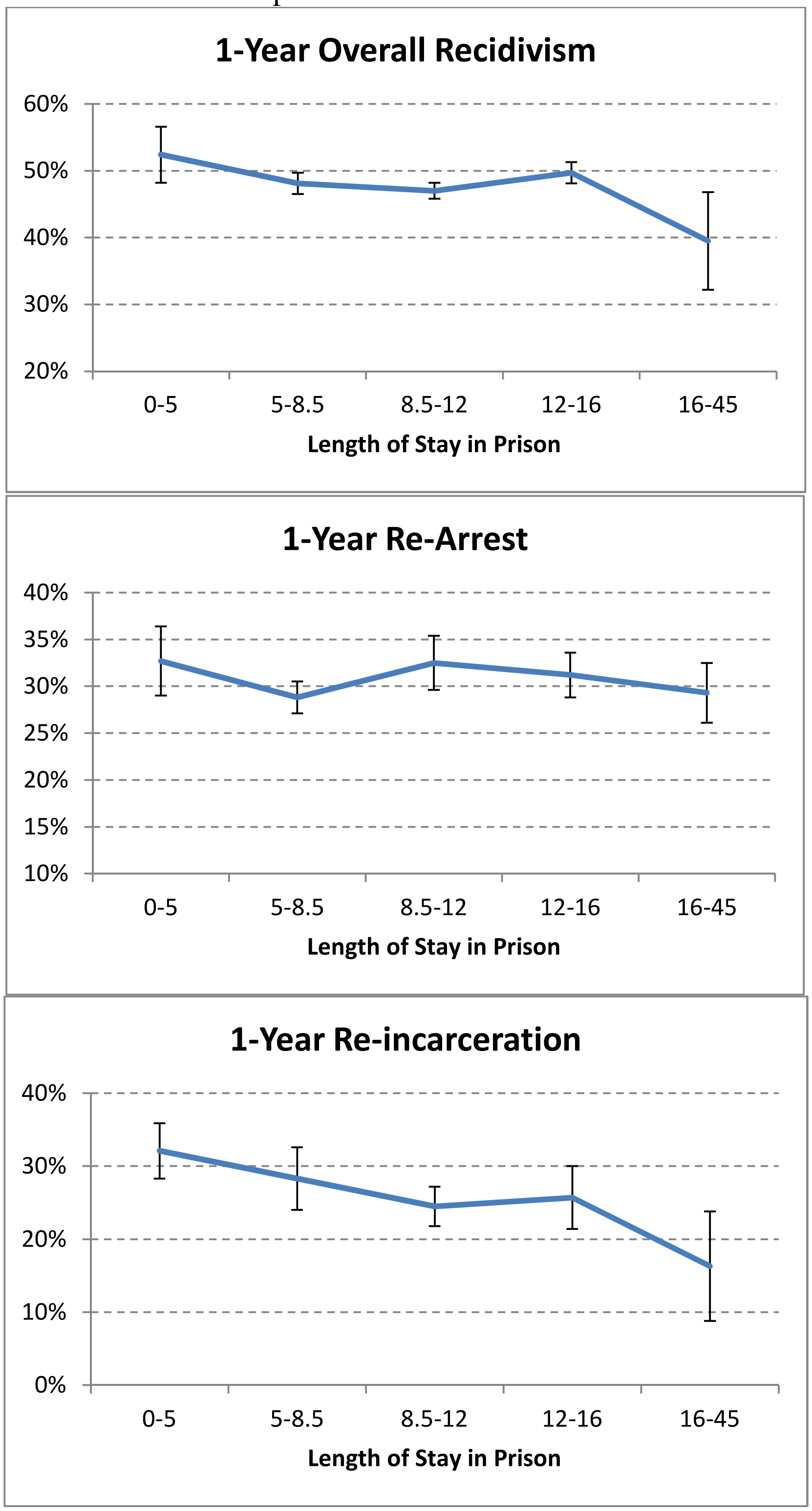
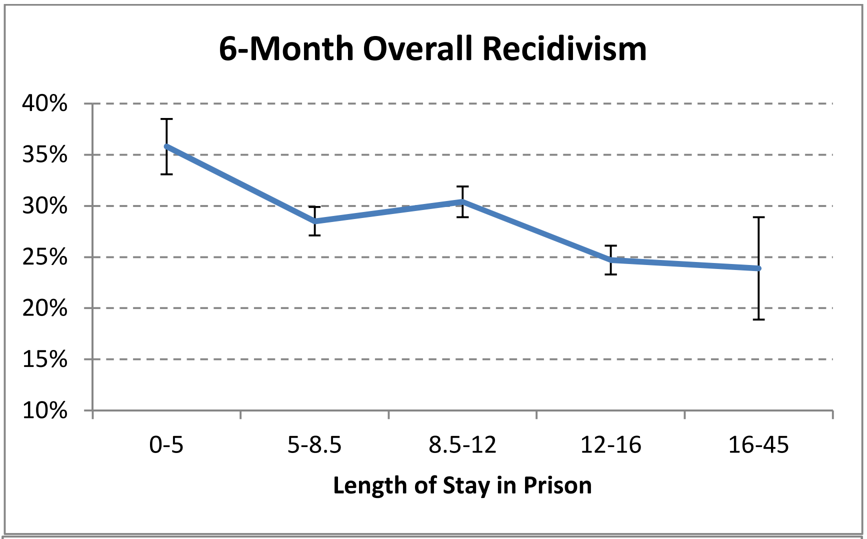
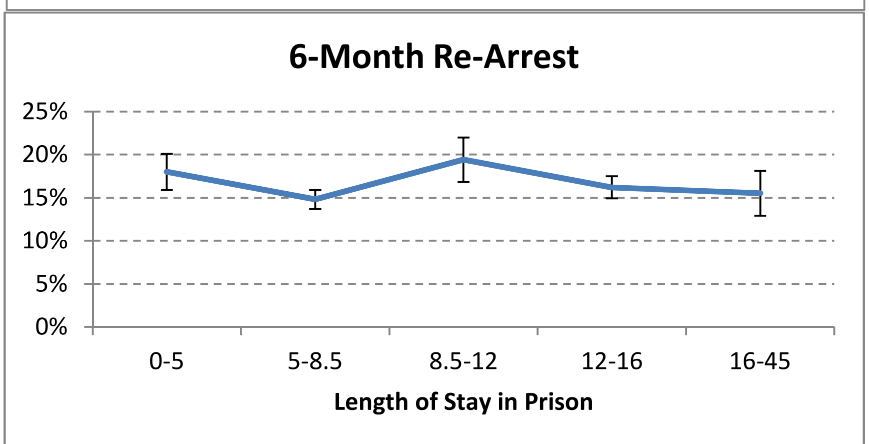


FIGURE 5.2 (continued)
Dose-Response Curves for Main Models





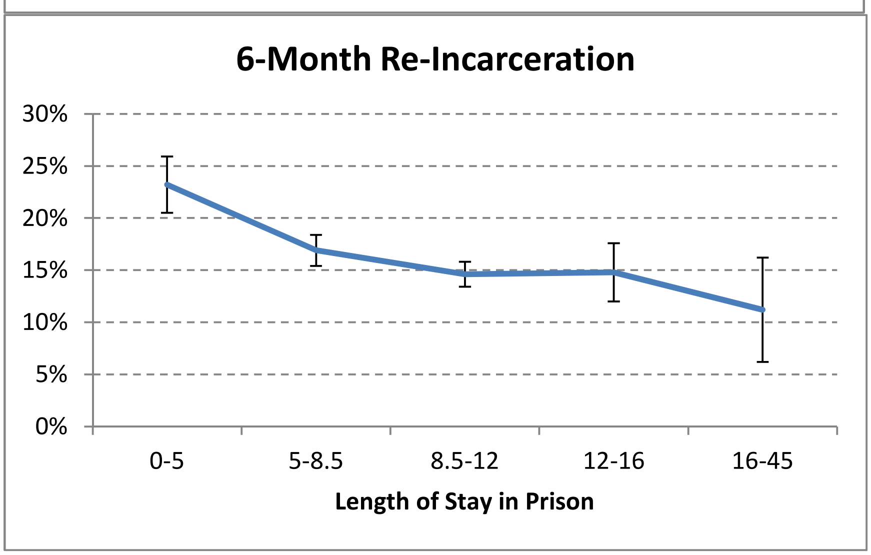


FIGURE 5.3
Dose-Response Curves Comparing Treatment Dosages to Control Group

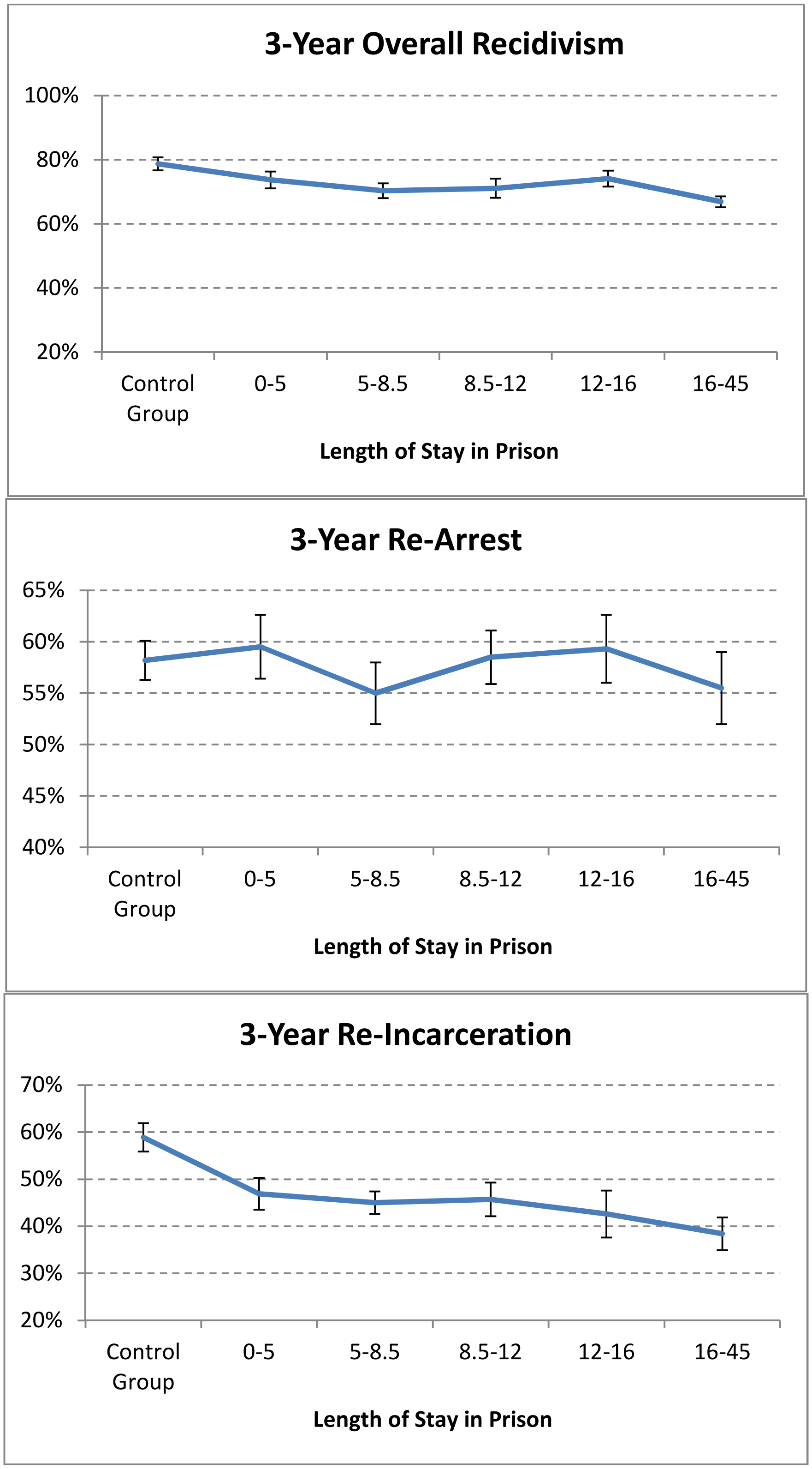
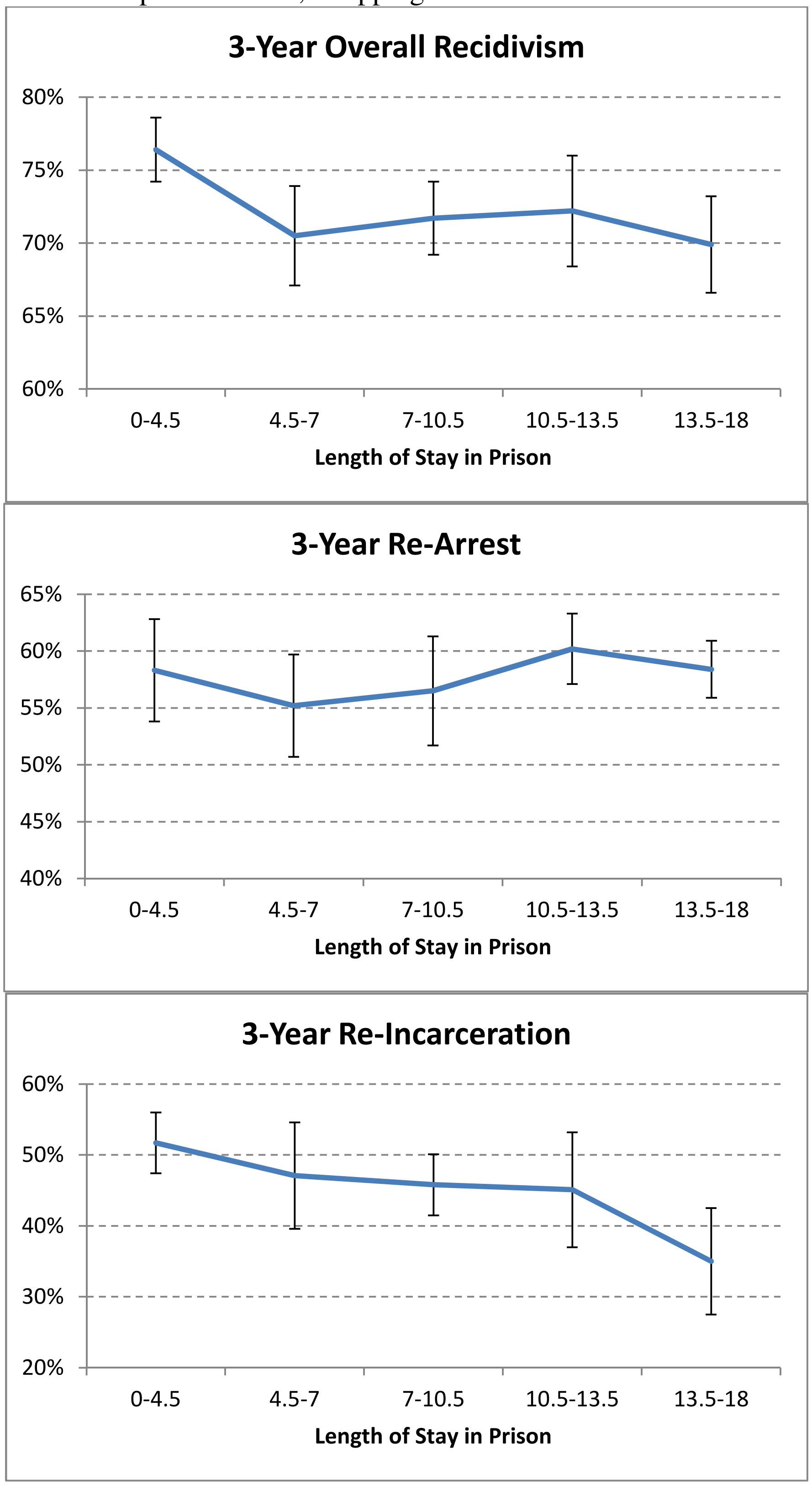


FIGURE 5.4
Dose-Response Curves, Dropping Doses Greater than 18 Months



REFERENCES

Alibrio, J. & Findley, W. (2005). *Technical Parole Violator Conditions Violated Study.* Pennsylvania Department of Probation and Parole.

Andrews, D. A., and Bonta, J. (1998). *The Psychology of Criminal Conduct*. New Providence, NJ: Matthew Bender & Company, Inc.

Anwar, S., & Loughran. T. A. (2011). Testing a Bayesian learning theory of deterrence among serious juvenile offenders. *Criminology, 49*(3)*,* 667-698.

Avi-Itzhak, B., & Shinnar, R. (1973). Quantitative models in crime control. *Journal of Criminal Justice 1*(3), 185-217.

Bachman, R., Paternoster, R., & Ward, S. (1992). The rationality of sexual offending: Testing a deterrence/rational choice conception of sexual assault. *Law and Society Review 26*(2), 343-372.

Barton, W. H., & Butts J. A. (1990). Viable options: Intensive supervision programs for juvenile delinquents. *Crime & Delinquency 36*(2), 238-256.

Beccaria, C. (1764). *On crimes and punishments.* Upper Sadler River, NJ: Prentice Hall.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76,* 169-217.

Becker, S. O., & Caliendo, M. (2007). *Mhbounds-sensitivity analysis for average treatment effects. No. 2542.* IZA Discussion Papers.

Bentham, J. (1781). *Principle of Morals and Legislation, vols. 1-4*. Amherst, NY: Prometheus.

Berecochea, J., & Jaman, D. (1981). *Time served in prison and parole outcome: An experimental study, Report 2.* Sacramento: California Department of Corrections.

Bergman, G. R. (1976). *The evaluation of an experimental program designed to reduce recidivism among second felony criminal offenders, vol. 2.* Detroit, MI: Wayne State University.

Berube, D. A., & Green, D. P. (2007). The effects of sentencing on recidivism: Results from a natural experiment. In *2nd Annual Conference on Empirical Legal Studies Paper.* New York, NY: NYU Law School.

Bhati, A. S., & Piquero, A. R. (2008). Estimating the impact of incarceration on subsequent offending trajectories: Deterrent, criminogenic, or null effect? *The Journal of Criminal Law and Criminology, 98,* 207-253.

Blumstein, A. (2011). Approaches to reducing both imprisonment and crime. *Criminology & Public Policy, 10*(1), 93-102.

Blumstein, A., & Beck, A. J. (2005). *Reentry as a transient state between liberty and recommitment*. New York, NY: Cambridge University Press.

Blumstein, A., Cohen, J., & Nagin, D. (1978). *Deterrence and incapacitation: Estimating the effects of criminal sanctions on crime rates*. Washington, DC: National Academy of Sciences.

Blumstein, A., Cohen, J., Roth, J. A., & Visher C. A. (1986). *Criminal careers and "Career Criminals."* Washington, DC: National Academy Press.

Bouffard, J. A. (2002). Methodological and theoretical implications of using subject generated consequences in tests of rational choice theory. *Justice Quarterly*, *19*(4), 747-771.

Braithwaite, J. (1989). *Crime, shame and reintegration.* New York, NY: Cambridge University Press.

Bucklen, K. B. (2005). Special focus on PA DOC's parole violator study (Phase 1). *Research in Review, 8*(1), 1-15.

Bureau of Justice Statistics. (2008). *Prisoners in 2007*, eds. H. C. West and W. J. Sabol. Washington, DC: Department of Justice.

Burke, P. J. (1997). An identity model for network exchange. *American Sociology Review, 62,* 134-150.

Civic Institute. (2005). *Sanction certainty: An evaluation of Erie County's adult probation sanctioning system,* eds. A. Amann, M. Beary, and E. Brown. Erie, PA: Mercyhurst College.

Deschenes, E. P., Turner, S., & Petersilia, J. (1995). A dual experiment in intensive community supervision: Minnesota's prison diversion and enhanced supervised release programs. *The Prison Journal, 75*(3), 330-356.

Devine, J. A., Sheley, J. F., & Smith, D. M. (1988). Macroeconomic and social-control policy influences on crime rate changes, 1948-1985. *American Sociological Review, 53*(3), 407-420.

Drago, F., Galbiati, R., & Vertova, P. (2009). The deterrent effects of prison: Evidence from a natural experiment. *Journal of Political Economy, 117*(2), 257-280.

Drake, E. K., & Aos, S. (2012). *Confinement for technical violations of community supervision: Is there an effect on felony recidivism?* Olympia, WA: Washington State Institute for Public Policy.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich College Publishers.

Exum, M. (2002). The application and robustness of the rational choice perspective in the study of intoxicated and angry intentions to aggress. *Criminology, 40*(4), 933-966.

Farabee, D. (2005). *Rethinking rehabilitation: Why can't we reform our criminals?* Washington, DC: Aei Press.

Franke, D., Bierie, D., & MacKenzie, D. L. (2010). Legitimacy in corrections. *Criminology & Public Policy, 9*(1), 89-117.

Geerken, M., & Gove, W. R. (1977). Deterrence, overload, and incapacitation: an empirical evaluation. *Social Forces, 56*(2), 424-447.

Gendreau, P., Cullen, F. T., & Goggin, C. (1999). *The effects of prison sentences on recidivism*. Ottawa, ON: Solicitor General Canada.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford, CA: Stanford University Press.

Grasmick, H. G., & Bursik Jr., R. J. (1990). Conscience, significant others, and rational choice: Extending the deterrence model. *Law and Society Review, 24*(3), 837-861.

Green, D. P., & Winik, D. (2010). Using random judge assignments to estimate the effects of incarceration and probation on recidivism among drug offenders. *Criminology*, *48*(2), 357-387.

Harrell, A., & Roman, J. (2001). Reducing drug use and crime among offenders: The impact of graduated sanctions. *Journal of Drug Issues, 31*(1), 207-231.

Hawken, A., & Kleiman, M. (2009). *Managing drug involved probationers with swift and certain sanctions: Evaluating Hawaii's HOPE: Executive summary*. Washington, DC: National Criminal Justice Reference Services.

Helland, E., & Tabarrok, A. (2007). Does three strikes deter? A nonparametric estimation. *Journal of Human Resources, 42*(2), 309-330.

Hipp, J. R., & Yates, D. K. (2009). Do returning parolees affect neighborhood crime? A case study of Sacramento. *Criminology, 47*(3), 619-656.

Hjalmarrson, H. (2009). System identification of complex and structured systems. *European Journal of Control, 15*(3), 275-310.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis, 15*(3), 199-236.

Horney, J., & Marshall, I. H. (1992). Risk perceptions among serious offenders: The role of crime and punishment. *Criminology, 30*(4), 575-594.

Howe, E. S., & Loftus, T. C. (1996). Integration of certainty, severity, and celerity information in judged deterrence value: Further evidence and methodological equivalence. *Journal of Applied Social Psychology, 26*(3), 226-242.

Jensen, G. F., Erickson, M. L., & Gibbs, J. P. (1978). Perceived risk of punishment and self-reported delinquency. *Social Forces, 57*(1), 57-78.

Katz, J. 1988. *Seductions of crime: A chilling exploration of the criminal mind- from juvenile delinquency to cold-blooded murder.* Perseus Books Group.

Kelling, G. L., & Coles, C. M. (1996). *Fixing broken windows: Restoring order and reducing crime in our communities.* New York, NY: Touchstone, Simon and Schuster.

Killias, M., Aebi, M., & Ribeaud, D. (2000). Does community service rehabilitate better than short-term imprisonment?: Results of a controlled experiment. *The Howard Journal of Criminal Justice, 39*(1), 40-57.

Kleck, G., Sever, B., Li, S., & Gertz, M. (2005). The missing link in general deterrence research. *Criminology, 43*(3), 623-660.

Kleiman, M. (2009). *When brute force fails: How to have less crime and less punishment*. Princeton, NJ: Princeton University Press.

Kramer, J., Silver, E., Van Eseltine, M., Ortega, C., & Rutkowski, A. (2008). *Evaluation of the Pennsylvania Board of Probation and Parole's violation sanction grid.* Harrisburg, PA: Pennsylvania Commission on Crime and Delinquency.

Legge, J. S., & Park, J. (1994). Policies to reduce alcohol-impaired driving: Evaluating elements of deterrence. *Social Science Quarterly, 75*(3), 594-606.

Lemert, E. M. (1951). *Social pathology: Systematic approaches to the study of sociopathic behavior.* New York: McGraw-Hill.

Levitt, S. D. (1996). The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. *Quarterly Journal of Economics, 111*(2), 319–351.

Liedka, R. V., Morrison Piehl, A., & Useem, B. (2006). The crime-control effect of incarceration: Does scale matter? *Criminology & Public Policy, 5*(2), 245-276.

Lochner, L. (2007). *Education and crime.* Ontario, Canada: University of Western Ontario.

Lu, B., Zanutto, E., Hornik, R., & Rosenbaum, P. R. (2001). Matching with doses in an observational study of a media campaign against drug abuse. *Journal of the American Statistical Association, 96*(456), 1245-1253.

MacKenzie, D. L. (2002). Reducing the criminal activities of known offenders and delinquents. In L. W. Sherman, D. P. Farrington, B. C. Welsh, and D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 334-421). New York, NY: Routledge.

Manski, C. F., & Nagin, D. S. (1998). Bounding disagreements about treatment effects: A case study of sentencing and recidivism. *Sociological methodology, 28*(1), 99-137.

Martin, B., & Van Dine, S. (2008). *Examining the impact of Ohio's progressive sanction grid.* Rockville, MD: National Institute of Justice.

Matsueda, R. L. (1992). Reflected appraisals, parental labeling, and delinquency: Specifying a symbolic interactionist theory. *American Journal of Sociology, 97*(6), 1577-1611.

Matsueda, R. L., Kreager, D. A., & Huizinga, D. (2006). Deterring delinquents: A rational choice model of theft and violence. *American Sociological Review*, *71*(1), 95-122.

Mead, G. H. (1934). *Mind, self, and society: From the standpoint of a social behaviorist*. Chicago, IL: University of Chicago Press.

Nagin, D. S. (1998). Criminal deterrence research at the outset of the twenty-first century. *Crime and Justice*, 1-42.

Nagin, D. S., Cullen, F. T., & Jonson, C. L. (2009). Imprisonment and reoffending. *Crime and Justice, 38*(1): 115-200.

Nagin, D. S., & Paternoster, R. (1993). Enduring individual differences and rational choice theories of crime. *Law and Society Review, 27*(3), 467-496.

Nagin, D. S., & Pogarsky, G. (2001). Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: Theory and evidence. *Criminology, 39*(4), 865-892.

Nagin, D. S., & Snodgrass, G. M. (2013). The effects of incarceration on re-offending: Evidence from a natural experiment in Pennsylvania. *Journal of Quantitative Criminology, 29*(4), 601-642.

O'Donoghue, T., & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making, 13*(2), 233-250.

Ostermann, M. (2011). Parole? Nope, not for me: Voluntarily maxing out of prison. *Crime & Delinquency, 57*(5), 686-708.

Paternoster, R. (1987). The deterrent effect of the perceived certainty and severity of punishment: A review of the evidence and issues. *Justice Quarterly, 4*(2), 173-217.

Paternoster, R., Brame, R., Bachman, R., & Sherman, L.W. (1997). Do fair procedures matter? The effect of procedural justice on spouse assault. *Law and Society Review, 31*(1), 163-204.

Paternoster, R., & Iovanni, L. (1989). The labeling perspective and delinquency: An elaboration of the theory and an assessment of the evidence. *Justice Quarterly, 6*(3), 359-394.

Paternoster, R., & Simpson, S. (1993). A rational choice theory of corporate crime. In R. V. Clarke and M. Felson (Eds.), *Routine Activity and Rational Choice, vol.* 5 (pp. 37-56). New Brunswick, NJ: Transaction Publishers.

Petersilia, J. (2003). *When prisoners come home: Parole and prisoner reentry*. New York, NY: Oxford University Press.

Petersilia, J., & Turner, S. (1993). Intensive probation and parole. *Crime and justice*, 281-335.

Pew Center on the States. (2007). *When offenders break the rules: Smart responses to parole and probation violations.* Washington, DC: The Pew Charitable Trusts.

Pew Center on the States. (2009). *One in 31: The long reach of american corrections*. Washington, DC: The Pew Charitable Trusts.

Piehl, A. M., & LoBuglio, S. F. (2005). Does supervision matter? In J. Travis and C. Visher (Eds.), *Prisoner reentry and crime in America* (pp. 105-138). New York, NY: Oxford University Press.

Piliavin, I., Gartner, R., Thornton, C., & Matsueda, R. L. (1986). Crime, deterrence, and rational choice. *American Sociological Review, 51*, 101-119.

Piquero, A. R., & Blumstein, A. (2007). Does incapacitation reduce crime? *Journal of Quantitative Criminology, 23*(4), 267-285.

Pogarsky, G. (2002). Identifying "deterrable" offenders: Implications for research on deterrence. *Justice Quarterly, 19*(3), 431-452.

Pogarsky, G., & Piquero, A. R. (2003). Can punishment encourage offending? Investigating the "resetting" effect. *Journal of Research in Crime and Delinquency, 40*(1), 95-120.

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L. E., & Madensen, T. D. (2006). The empirical status of deterrence theory: A meta-analysis. In F. T. Cullen, J. P. Vright, and K. R. Blevins (Eds.), *Taking stock: The status of criminological theory* (pp. 367-396). New Brunswick, NJ: Transaction Publishers.

Reinventing Probation Council. (1999). *"Broken windows" probation: The next step in fighting crime.* New York, NY: Manhattan Institute of Policy Research.

Rosenbaum, P. R. (2002). Sensitivity to hidden bias. In *Observational studies*. New York, NY: Springer Science and Business Media.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516-524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*(1), 33-38.

Rosenfeld, R., Wallman, J., & Fornango, J. (2005). The contribution of ex-prisoners to crime rates. In J. Travis and C. Visher (Eds.), *Prisoner reentry and crime in America* (pp. 80-104). New York, NY: Oxford University Press.

Sampson, R. J., & Laub, J. H. (1993). *Crime in the making: Pathways and turning points through life.* Cambridge: Harvard University Press.

Sampson, R. J., & Laub, J. H. (1997). A life-course theory of cumulative disadvantage and the stability of delinquency. *Developmental theories of crime and delinquency, 7*, 133-161.

Scheff, T. J., & Retzinger, S. M. (1997). Shame, anger and the social bond: A theory of sexual offenders and treatment. *Electronic Journal of Sociology, 3*(1).

Schneider, A. L. (1986). Restitution and recidivism rates of juvenile offenders: Results from four experimental studies. *Criminology, 24*(3), 533-552.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin Company.

Sherman, L. W. (1993). Defiance, deterrence, and irrelevance: A theory of the criminal sanction. *Journal of research in Crime and Delinquency, 30*(4), 445-473.

Shinnar, S., & Shinnar, R. (1975). The effects of the criminal justice system on the control of crime: A quantitative approach. *Law and Society Review, 9*(4), 581-611.

Simon, H. A. (1957). *Models of man; social and rational.* Oxford, England: Wiley.

Snodgrass, G., Blokland, A. A., Haviland, A., Nieuwbeerta, P., & Nagin, D. S. (2011). Does the time cause the crime? An examination of the relationship between time served and reoffending in the Netherlands. *Criminology, 49*(4), 1149-1194.

Solomon, A. L., Kachnowski, V., & Bhati, A. (2005). *Does parole work?: Analyzing the impact of post-prison supervision on re-arrest outcomes.* Washington, DC: Urban Institute.

Spelman, W. (2000). The limited importance of prison expansion. In A. Blumstein and J. Wallman (Eds.), *The crime drop in America* (pp. 123-125). New York, NY: Cambridge University Press.

Spelman, W. (2005). Jobs or jails? The crime drop in Texas. *Journal of Policy Analysis and Management, 24*(1), 133-165.

Stafford, M. C., & Warr, M. (1993). A reconceptualization of general and specific deterrence. *Journal of Research in Crime and Delinquency, 30*(2), 123-135.

Stemen, D. (2007). *Reconsidering incarceration: new directions for reducing crime.* New York, NY: Vera Institute of Justice.

Sykes, G. M. (1958). *The Society of Captives: A study of a Maximum Security Prison.* Princeton, NJ: Princeton University Press.

Taxman, F. S., Soule, D., & Gelb, A. (1999). Graduated sanctions: Stepping into accountable systems and offenders. *The Prison Journal, 79*(2), 182-204.

Tittle, C. R., & Rowe, A. R. (1974). Certainty of arrest and crime rates: A further test of the deterrence hypothesis. *Social Forces, 52*(4), 455-462.

Travis, J. & Lawrence, S. (2002). California's parole experiment. *California Journal*, *33*(8). Sacramento, CA: A State Net Publication.

Tyler, T. R. (1990). *Why people obey the law: Procedural justice, legitimacy, and compliance.* New Haven, CT: Yale University Press.

Urban Institute Justice Policy Center. (2008). *An evolving field: Findings from the 2008 parole practices survey,* eds. J. Janetta, B. Elderbroom, A. Solomon, M. Cahill, B. Parthasarathy, and W. D. Burrell. Washington, DC: US Department of Justice.

Van der Werff, C. (1979). Recidivism and special deterrence. *British Journal of Criminology, 21*(2), 136.

Villettaz, P., Killias, M., & Zoder, I. (2006). The effects of custodial vs non-custodial sentences on re-offending: A systematic review of the state of knowledge. *Campbell Systematic Reviews, 13*.

White, M. D., Mellow, J., Englander, K., & Ruffinengo, M. (2011). Halfway back: An alternative to revocation for technical parole violators. *Criminal Justice Policy Review, 22*(2), 140-166.

Wilson, J. Q., & Herrnstein, R. J. (1985). *Crime human nature: The definitive study of the causes of crime*. New York, NY: The Free Press.

Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic monthly, 249*(3), 29-38.

Wimer, C., Sampson, R. J., & Laub, J. H. (2008). Estimating time-varying causes and outcomes, with application to incarceration and crime. In P. Cohen (Ed.), *Applied data analytic techniques for turning points research* (pp. 37-60). New York, NY: Routledge.

Wolfgang, M. E., Figlio, R. M. & Sellin, T. (1972). *Delinquency in a birth cohort.* Chicago: University of Chicago Press.

Wright, B. R., Caspi, A., Moffitt, T. E. & Paternoster, R. (2004). Does the perceived risk of punishment deter criminally prone individuals? Rational choice, self-control, and crime. *Journal of Research in Crime and Delinquency, 41*(2), 180-213.

Yu, J. (1994). Punishment celerity and severity: Testing a specific deterrence model on drunk driving recidivism. *Journal of Criminal Justice, 22*(4), 355-366.

Zamble, E., & Porporino, F. J. (1988). *Coping, behavior, and adaptation in prison inmates*. New York, NY: Springer-Verlag Publishing.

Zanutto, E., Lu, B., & Hornik, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics, 30*(1), 59-73.