



---

Publicly Accessible Penn Dissertations

---


1-1-2013

# The Impact of Cognitive-Behavioral Therapy on the Recidivism of High Risk Probationers: Results from a Randomized Trial

Jordan Hyatt

*University of Pennsylvania*, [jhyatt@sas.upenn.edu](mailto:jhyatt@sas.upenn.edu)

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Criminology Commons](#), and the [Criminology and Criminal Justice Commons](#)

---

## Recommended Citation

Hyatt, Jordan, "The Impact of Cognitive-Behavioral Therapy on the Recidivism of High Risk Probationers: Results from a Randomized Trial" (2013). *Publicly Accessible Penn Dissertations*. 644.  
<http://repository.upenn.edu/edissertations/644>

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/edissertations/644>  
For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# The Impact of Cognitive-Behavioral Therapy on the Recidivism of High Risk Probationers: Results from a Randomized Trial

## **Abstract**

Community corrections are being used with increasing regularity for the supervision and management of serious and violent offenders. Attempts to increase the frequency and severity of conditions of supervision have not resulted in meaningful decreases in crime rates among this population. Some encouraging results, however, have been observed when a treatment component is integrated into supervision protocols. This dissertation first examines the theories and current research that inform this shift in strategies. Secondly, we evaluate for the first time, a cognitive-behavioral therapy intervention developed to reduce recidivism within a high-risk, male probation population.

This dissertation begins with a review and synthesis of the literature, both in criminology and psychology, regarding the development of cognitive-behavioral techniques designed to reduce recidivism. Next, the unique characteristic of the intervention being evaluated are set out in Chapter 3. The logistics and characteristics of the randomized trial itself are discussed in Chapter 4. This section includes an overview of the risk forecasting procedures used to identify the experimental sample and the randomization scheme employed. In the following section, the impact of the cognitive-behavioral intervention delivered in Philadelphia is evaluated. Using techniques standard within experimental research, a significant reduction in the prevalence of non-violent offending and some forms of drug use are identified. An instrumental variable analysis is then used to better specify effect sizes in light of relatively high treatment dilution. Finally, implications for future research and public policy are discussed in Chapter 6.

After 12 months, there were some significant and meaningful differences within the measures of prevalence of offending. Fewer offenders assigned to the treatment group (33.9%) than control (40.5%) were charged with an offense of any kind ( $p=.041$ ). Therefore, assignment to the Life Skills program caused a 7.5% decrease in the number of offenders committing non-violent crimes. Significant reductions were also noted in the proportion of urinalysis screenings that were positive for PCP and time-to-failure for non-violent offending. Using randomization as an instrumental variable to compensate for treatment dilution, the reduction in the prevalence of non-violent offending was estimated at 18.8%.

This research contributes to the broader literature by reinforcing the hypothesis that an integrated treatment-control supervision strategy is a viable approach for probation agencies seeking to both increase levels of control and reduce recidivism. Specifically, the results reported here represent the first, randomized outcome evaluations of an innovative form of cognitive-behavioral therapy with that specific goal. Secondly, the integration of these findings into the literature using meta-analytic techniques may better inform our understanding of the actual effects and promises of community-based recidivism-reduction programming. Finally, the innovations in experimental design and implementation developed during this project may serve as both an inspiration and a caution for other experimental criminologists.

## **Degree Type**

Dissertation

## **Degree Name**

Doctor of Philosophy (PhD)

---

**Graduate Group**

Criminology

**First Advisor**

Adrian Raine

**Keywords**

CBT, Cognitive-behavioral therapy, Probation, Randomized Control Trial, RCT, Recidivism

**Subject Categories**

Criminology | Criminology and Criminal Justice

THE IMPACT OF COGNITIVE-BEHAVIORAL THERAPY ON THE RECIDIVISM  
OF HIGH RISK PROBATIONERS:

RESULTS FROM A RANDOMIZED TRIAL

Jordan Michael Hyatt

A DISSERTATION

In Criminology

Presented to the Faculties of the University of Pennsylvania


In

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation



Dr. Adrian Raine  
Professor of Criminology,  
Psychology & Psychiatry

Graduate Group Chairperson



Dr. Adrian Raine  
Professor of Criminology,  
Psychology & Psychiatry

**Dissertation Committee:**

Geoffrey Barnes

Research Assistant Professor of Criminology

Robert Boruch

University Trustee Chair Professor of Education &  
Statistics

William Laufer

Julian Aresty Professor of Legal Studies, Business Ethics,  
Sociology & Criminology

Adrian Raine

Richard Perry University Professor of Criminology,  
Psychology & Psychiatry

In memory of

Bernice, Alvin and Ralph

Each, in their own way, an inspiration.

## ACKNOWLEDGMENTS

This dissertation could never have been completed without the support of many, many people. I am both appreciative of and indebted to:

Geoffrey Barnes. You have been my supervisor and guide for the duration of this project and much of what I know about experimental research has been gleaned under your watch. Perhaps more importantly, the passive voice was (mostly) beaten out of my writings by you.

Adrian Raine, for recognizing my potential and taking on the duties as my chair. Without your support and foresight, this dissertation and my career at Penn would never have been possible.

The research staff and leadership of the Adult Probation and Parole Department, including Ellen Kurtz, Robert Malvestuto, Charlie Hoyt, Kathy Intenzo, Steve Austin, Vince Fiorintino, Kathres Davis, Amanda Tudor and Lindsay Ahlman. Without your support, this project would never have gotten off the ground and could not have run as smoothly as it did.

“Team CBT,” Ron Noble, Jill Eidson, Carrinnia Woodson, Chiquita Brice and Keith Nigro. Your hard work and dedication are truly the backbone of this research.

John MacDonald. Much of this writing has been fueled by coffee. Even more important than the caffeine, your advice and guidance during those breaks was invaluable. You have been a mentor, even when not your obligation, and I remain appreciative.

Lawrence Sherman. Your expertise and relationship building skills are the foundation of this project and for our continuing association with APPD. Your decision

to bring me on to the Jerry Lee Center was the spark that set this dissertation, and my career, in motion.

Richard Berk. Without your forecasting models and statistical expertise we would all still be stumbling around in the dark, looking for our sample.

Freda Adler. Ever since I was your TA during both of our first terms at Penn, you have been a tireless supporter of my potential to do meaningful work. I'll always be grateful.

Professors Rosenbaum and DeRubeis, for pointing me back in the right direction when I was feeling especially lost.

Janel Baslice and Knaykia Hagans. We all know that, without the two of you, this dissertation would be forever a pile of unfiled pages and a whole lot of wishful thinking.

My colleagues in the doctoral program, Seunghoon Han and Ben Grunwald. You've both kept me grounded, listened to my ramblings and been immensely supportive. Your advice on statistics (and where to find Stata and tea on short notice) was irreplaceable.

My family, whose patience never completely waned, even when I decided to go back for "one more" degree. This really is the last one, I promise.

And, finally, Lindsay Wilde, my spouse. You've probably suffered most of all through this process. Between the extra dog walking, putting up with an epic finals beard and the endless proofreading, you've earned this degree as much as I have. Though we both know the greatest adventure is coming up, your unwavering support has made this one possible.

## **ABSTRACT**

### **THE IMPACT OF COGNITIVE-BEHAVIORAL THERAPY ON THE RECIDIVISM OF HIGH RISK PROBATIONERS: RESULTS FROM A RANDOMIZED TRIAL**

Jordan Michael Hyatt

Adrian Raine

Community corrections are being used with increasing regularity for the supervision and management of serious and violent offenders. Attempts to increase the frequency and severity of conditions of supervision have not resulted in meaningful decreases in crime rates among this population. Some encouraging results, however, have been observed when a treatment component is integrated into supervision protocols. This dissertation first examines the theories and current research that inform this shift in strategies. Secondly, we evaluate for the first time, a cognitive-behavioral therapy intervention developed to reduce recidivism within a high-risk, male probation population.

This dissertation begins with a review and synthesis of the literature, both in criminology and psychology, regarding the development of cognitive-behavioral techniques designed to reduce recidivism. Next, the unique characteristics of the intervention being evaluated are set out in Chapter 3. The logistics and characteristics of the randomized trial itself are discussed in Chapter 4. This section includes an overview of the risk forecasting procedures used to identify the experimental sample and the



randomization scheme employed. In the following section, the impact of the cognitive-behavioral intervention delivered in Philadelphia is evaluated. Using techniques standard within experimental research, a significant reduction in the prevalence of non-violent offending and some forms of drug use are identified. An instrumental variable analysis is then used to better specify effect sizes in light of relatively high treatment dilution. Finally, implications for future research and public policy are discussed in Chapter 6.

After 12 months, there were some significant and meaningful differences within the measures of prevalence of offending. Fewer offenders assigned to the treatment group (33.9%) than control (40.5%) were charged with an offense of any kind ( $p=.041$ ). Therefore, assignment to the Life Skills program caused a 7.5% decrease in the number of offenders committing non-violent crimes. Significant reductions were also noted in the proportion of urinalysis screenings that were positive for PCP and time-to-failure for non-violent offending. Using randomization as an instrumental variable to compensate for treatment dilution, the reduction in the prevalence of non-violent offending was estimated at 18.8%.

This research contributes to the broader literature by reinforcing the hypothesis that an integrated treatment-control supervision strategy is a viable approach for probation agencies seeking to both increase levels of control and reduce recidivism. Specifically, the results reported here represent the first, randomized outcome evaluations of an innovative form of cognitive-behavioral therapy with that specific goal. Secondly, the integration of these findings into the literature using meta-analytic techniques may better inform our understanding of the actual effects and promises of community-based recidivism-reduction programming. Finally, the innovations in experimental design and

implementation developed during this project may serve as both an inspiration and a caution for other experimental criminologists.

## TABLE OF CONTENTS

Acknowledgments.....	iii
Abstract .....	v
Table of Contents .....	viii
List of Tables .....	xi
List of Figures.....	xiii
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
<b>CHAPTER 2: LITERATURE REVIEW .....</b>	<b>5</b>
I. Trends in Community-Based Supervision .....	5
II. Risk Prediction & Random Forest Forecasting Procedures .....	9
III. Development of Cognitive-Behavioral Therapy .....	13
a. Theoretical Foundations .....	14
b. Mechanisms .....	22
c. Cognitive-Behavioral Therapy and Criminological Theory .....	24
d. Models Of Cognitive-Behavioral Therapy .....	33
IV. Cognitive-Behavioral Therapy and Crime Prevention .....	34
a. The Implications Of Meta-Analysis .....	34
b. Meta-Analyses Of Cognitive-Behavioral Therapy & Crime .....	39
c. Prior Evaluations Of Similar Interventions .....	48
V. Conclusion .....	53
<b>CHAPTER 3: THE COGNITIVE-BEHAVIORAL SKILLS</b> <b>INTERVENTION .....</b>	<b>55</b>
I. Background.....	55
II. Thematic Structure of the Program .....	57
III. Session Scheduling .....	60
IV. Curriculum Design .....	62
V. Incentive Structures .....	64
VI. Course Materials .....	67
VII. Facilitators.....	68
VIII. Evaluating <i>Choosing to Think and Thinking to Choose</i> .....	70
IX. Conclusion .....	75

<b>CHAPTER 4: METHOD</b> .....	77
I.    Background .....	77
II.   Risk Assessment Protocol .....	79
III.  Setting .....	82
IV.  Treatment Groups .....	87
V.    Random Assignment Procedures .....	88
VI.  Subject Eligibility .....	93
VII. Sample Size and Power Calculations .....	97
VIII. Participant Recruitment Procedures .....	101
IX.  Treatment Rates .....	109
X.    Outcome Measures .....	111
 <b>CHAPTER 5: RESULTS</b> .....	 113
<b>I.    Intention to Treat (ITT) Analysis</b> .....	114
A. Post-Randomization Group Comparisons.....	115
i. Age .....	116
ii. Race .....	116
iii. Risk Scores .....	117
iv. Juvenile Criminal History .....	118
v. Adult Criminal History .....	119
vi. Prior Incarceration Experiences .....	120
vii. Instant Offense .....	121
viii. Instant Sentence .....	122
ix. Experience while on Supervision .....	123
B. Results: Intention to Treat.....	127
i. Absconding .....	128
ii. Drug Use .....	129
iii. Offending .....	132
iv. Incarceration .....	135
v. Time to Failure .....	136
C. Summary .....	140
 <b>II.   Treatment On The Treated (TOT) Analysis</b> .....	 142
A. Treatment on the Treated Sample Construction .....	144
B. Post-Randomization Group Comparisons.....	144
i. Risk & Age at Assignment .....	144
ii. Offending History and Onset .....	145
iii. Prior Sanctioning History .....	148
iv. Instant Offense and Sentence Characteristics .....	149
v. Experience while on Supervision .....	150
C. Results: Treatment on the Treated .....	151
i. Absconding .....	152

ii. Drug Use .....	152
iii. Offending .....	155
iv. Incarceration .....	156
v. Time to Failure .....	157
D. Summary .....	161
<b>III. Instrumental Variable (IV) Analysis .....</b>	<b>163</b>
A. Absconding (and an example) .....	168
B. Drug Testing and Use .....	169
C. Offending .....	172
D. Incarceration .....	174
E. Summary .....	175
<b>CHAPTER 6: SUMMARY AND IMPLICATIONS .....</b>	<b>176</b>
I. Summary .....	176
II. Discussion of Results .....	179
III. Integration of Findings .....	186
IV. Limitations .....	189
V. Implications for Future Interventions and Public Policy .....	193
VI. Directions for Forthcoming Research .....	194
VII. Conclusions .....	196
<b>Appendices .....</b>	<b>198</b>
Appendix A: Complete Life Skills Course Schedule .....	199
Appendix B: Judicial Letter and Revised Conditions of Supervision .....	200
Appendix C: APPD Departmental Organization and Hierarchy .....	202
Appendix D: Risk Prediction Variables .....	203
Bibliography .....	212



Table 5.21: TOT Ages of Onset for Any Charge and for Violent Charges .....	148
Table 5.22: TOT Prior Sanctioning History .....	149
Table 5.23: TOT Instant Offense .....	149
Table 5.24: TOT Instant Sentences and Sentence Characteristics .....	150
Table 5.25: TOT Supervision Contacts within 12 months .....	151
Table 5.26: TOT Results, Absconding .....	152
Table 5.27: TOT Results, Urinalysis Screening Rates .....	153
Table 5.28: TOT Results, Count of Positive Urinalysis Screenings .....	153
Table 5.29: TOT Results, Prevalence of Positive Urinalysis Screenings .....	154
Table 5.30: TOT Results, Proportion of Screenings with a Positive Result .....	154
Table 5.31: TOT Results, Average Number of Charges .....	155
Table 5.32: TOT Results, Proportion of Each Group Charged, by Offense .....	156
Table 5.33: TOT Results, Incarceration Length and Characteristics .....	157
Table 5.34: TOT Results, Cumulative Probability of Failure (Log Rank Test) .....	158
Table 5.35: TOT Results, Average Time to First Incident Among Failures .....	161
Table 5.36: IV Results, Absconding .....	169
Table 5.37: IV Results, Count of Positive Urinalysis Screenings .....	170
Table 5.38: IV Results, Prevalence of Positive Urinalysis Screenings .....	171
Table 5.39: IV Results, Proportion of Screenings with a Positive Result .....	172
Table 5.40: IV Results, Average Number of Charges .....	173
Table 5.41: IV Results, Proportion of Each Group Charged, by Offense .....	174
Table 5.412 IV Results, Incarceration Length and Characteristics .....	175

## LIST OF FIGURES

Figure 3.1: Life Skills Scheduling Schema .....	60
Figure 4.1: Changes in APPD Active Population, August 2010 to August 2012 ....	85
Figure 4.2: Final RCT Case Allocation .....	92
Figure 4.3: Forecasts, Screenings and Enrollment Rates .....	96
Figure 4.4: Sample Size and Enrollment .....	98
Figure 4.5: Relative Power Calculations Over Time .....	100
Figure 4.6: Enrollment and Attrition of Life Skills Participants .....	109
Figure 5.1: ITT Results, Average Number of Charges .....	134
Figure 5.2: ITT Results, Survival Functions for Non-Violent Offending .....	138
Figure 5.3: ITT Results, Survival Functions for Drug Offending .....	139
Figure 5.4: TOT Results, Survival Functions for Non-Violent Offending .....	159
Figure 5.5: TOT Results, Survival Functions for Violent Offending .....	159
Figure 5.6: TOT Results, Survival Functions for Absconding .....	160



## **CHAPTER 1: INTRODUCTION**

Probation, as well as other community correctional sanctions, is being relied upon with growing frequency in the modern criminal justice system. For example, in Pennsylvania alone, 258,905 individuals were on probation or parole in 2007, a figure more than 5.6 times the inmate population in the state's correctional institutions (Emery, Gasswint, Hartman, & Lategan, 2008). In light of increasingly pressing concerns regarding budgetary constraints and prison overcrowding, the shift towards community corrections is unlikely to abate. The increased reliance on non-penal sanctions has placed the use of probation and parole into the vanguard of the modern crime policy, but has done little to diminish perceptions that community corrections are an ineffective set of sanctions, especially with regard to the prevention of serious crime.

In Philadelphia, the Adult Probation and Parole Department (APPD) is the primary agency in the First Judicial District (FJD) responsible for the supervision of offenders on probation and county-level parole within the city. Working with researchers from the Jerry Lee Center of Criminology at the University of Pennsylvania (JLC), APPD has worked to evaluate and implement evidence-based supervision policies. In recent years, this has included the development of a risk assessment tool that is used to forecast offender conduct while on supervision (Berk R. A., Sherman, Barnes, Kurtz, & Ahlman, 2009). This approach, now used to classify all incoming cases, is a key component of APPD's supervision strategy (Barnes & Hyatt, 2012). Additionally, risk stratified supervision allows for the maintenance of supervision levels, even in light of decreased resource availability and fiscal constraints (Elliott-Engel, 2011).

In response to increasing numbers of serious offenders being placed on community supervision, and ostensibly to ensure public safety, some agencies have increased the strictness and frequency of supervision characteristics. This practice, known as intensive supervisory probation (ISP), is often used for offenders considered to be at an increased risk for criminal conduct (Petersilia & Turner, 1993). Since the literature suggests that increase in supervision intensity may, in fact, have little effect on crime rates, some researchers have begun to seek out interventions that, when delivered in tandem with ISP, could result in crime reduction.

Advances in risk forecasting procedures have allowed for the use of standardized, actuarial forecasting procedures in order to make policy decisions. Newer methods, including the random forest model employed here, can, as Berk notes, “[address] important concerns that result from model selection methods, symmetric loss functions, and overreliance on linear models” (Berk R. A., 2008, p. 236). These approaches to forecasting can then be used to assist probation agencies as they consider the amount and nature of supervision that they can deliver (Austin, 2010). Risk prediction tactics can be used to allocate supervision resources and to identify individuals and subgroups for whom the inclusion of a treatment component could return some benefits (Sherman L. W., 2007).

One of the most promising of these complimentary programs that has been both identified and operationalized is Cognitive-behavioral Therapy (CBT). Most generally, CBT is a family of psychological interventions derived from the principles of traditional psychotherapy and behavior learning theories (Beck A. T., 1995). With decades of positive results in treating depression and anxiety, a critical mass of research regarding

the impact of CBT on crime-related outcomes has developed (Wilson & Lipsey, 1993). Since a wide variety of programs fall under the umbrella of CBT-derived interventions, direct comparisons are not always illustrative of the relative promise of the approach. Meta-analytic techniques, however, indicate that CBT has an overall positive impact on recidivism, especially when compared to an absence of treatment or other types of non-cognitive, pro-social programming (Lipsey & Landenberger, 2005). In an area of the criminal justice system where few programs are shown to be effective, especially for a population at a high risk of violence, CBT is a promising intervention worth exploring.

A unique intervention, based upon the principles of CBT, was developed for use in this project. Specifically, the program, called *Choosing to Think, Thinking to Choose*, addresses the needs of the target population: high-risk, urban males on probation. This is accomplished through the use of media clips, examples and conversation prompts that are relatable for the target population. The full program lasts for 14 weeks and is conducted in a classroom environment managed by trained probation officers. Within the Department, the course was known as the ‘Life Skills’ program.

A randomized trial was conducted to evaluate the impact that the CBT program had on recidivism. 904 probationers were assigned to receive both CBT and ISP or to only receive the ISP component. Outcomes, including criminal recidivism, drug use, absconding and time-to-failure, are reported for conduct within 12 months of random assignment. Between-group differences are reported, as well as those from an instrumental variable analysis conducted to adjust for incomplete treatment delivery.

This project contributes to the literature in three meaningful ways. First, as a randomized control trial (RCT) conducted in a field setting, this evaluation provides a

new and rigorous evaluation of a CBT-derived program. As evidenced within a review of the literature, there are few field-based, experimental evaluations of CBT within the community corrections context. Probations agencies, by virtue of the duration and nature of supervision, are well suited to deliver CBT; these findings may be of value to practitioners. They will also increase the extent to which community corrections-based programs can be included in and weighted in future meta-analyses.

Secondly, this dissertation considers the potential theoretical mechanisms, grounded in criminology, which may help to explain the impact of CBT-based programs on recidivism. This is relevant to the development of newer interventions and to allow for the refinement of current interventions, including the one evaluated here.

Lastly, strategies developed during the implementation phase of this evaluation highlight the challenges in field trials and the need, stressed throughout the criminology, for a greater emphasis on experimentally-derived evidence of program effectiveness.

## **CHAPTER 2: LITERATURE REVIEW**

This chapter describes the literature relevant to this evaluation. Beginning with an overview of developments in community corrections, subsequent sections describe the risk forecasting methodologies utilized during this project, as well the development of CBT within both the psychological and criminological literature. The concluding sections review studies relevant to this dissertation, including meta-analyses and evidence derived from evaluations of similar interventions.

### **I. Trends in Community-Based Supervision**

Community corrections, most notably probation and parole, are some of the most frequently relied upon criminal sanctions in the American justice system. At any given time, approximately 1 in every 45 adults in the United States is under some form of community correctional supervision (Pew Center on the States, 2009), far exceeding the 1% of adults representing the penal population (Pew Center on the States, 2008). Further, the use of probation has shown regular and sustained growth. The total community corrections population swelled by nearly one million individuals between 1995 and 2006 (Glaze & Bonczar, 2009). This growth places enormous pressure on community corrections agencies, especially as historical data indicates that staff and budgetary increases have not kept pace with an exploding population (Gifford, 2002).

Despite this increasing reliance, community corrections agencies are often faced with criticism that their approach is “soft on crime” and cannot effectively prevent criminal conduct (Petersilia 1997). This perception may also contribute to chronic under-funding of probation agencies, resulting in an increasing inability to deliver effective supervision and to protect public safety (Beto, Corbett, & Hinzman, 1999). At

the same time, recidivism rates are high among probationers and parolees. Some studies have placed the recidivism rate as high as 65% (Petersilia, 1985). The large majority (77%) of community supervision violators sentenced to incarceration were returned to prison for the commission of a new felony while under supervision (Cohen, 1995). These high rates have caused some to “question the ability of community supervision to effect meaningful behavioral change in a direction favorable to public safety” (Lowencamp, Latessa, & Smith, 2006, p. 576). One way that probation agencies have responded to these critiques is to intensify probation for certain groups of offenders.

Intensive supervision probation (ISP), as in the case of Philadelphia’s Anti-Violence units, most often consists of increased office visits, more frequent drug testing and a zero-tolerance policy towards minor infractions (Gill, 2010). Beginning in the 1950s, this model of supervision was presumed to result in lower levels of recidivism and increased employment. It was suggested that smaller caseloads and more frequent contacts would lead to lower recidivism rates. Early evaluations found little evidence that the increase in intensity reduced recidivism (Carter & Wilkins, 1976).

ISP reemerged in the 1980s, as prison overcrowding necessitated the supervision of increasingly serious offenders within the community. In addition to promising reduced crime rates, the ISP approaches popularized at this time were also supposed to conserve resources. However, a large, multi-site randomized trial found little difference between this approach and traditional protocols or incarceration. In fact, the evaluation found an increase in technical violations, possibly increasing the number of offenders being returned to prison (Petersilia and Turner 1993). Subsequently, intensive probation was classified as an approach to supervision that “doesn’t work” in preventing crime

(Sherman L. W., 1997), effectively ending much of the academic debate about the utility of the control focused ISP.

Despite this lack of convincing empirical support, ISP continued to be developed as a community-based supervision program for offenders thought to pose a danger to the community. In some cases, this approach was pursued as a sufficiently harsh alternative such that it could approximate incarceration (Petersilia & Turner, 1990). Though research has found that a quantitative increase in the number of probation contacts had little effect on offending, there are other aspects of ISP programs that may be promising. As Doris Mackenzie (1997) notes,

"Although research has not revealed a significant relationship between levels of surveillance and recidivism, there was some evidence that increased treatment of offenders in ISP programs may be related to significant reductions in rearrests. Follow-up analyses by the RAND researchers (Petersilia & Turner 1993a,b) and also researchers evaluating ISP programs in Massachusetts (Byrne & Kelly 1989), Oregon (Jolin & Stipack 1991) and Ohio (Latessa, 1993a,b) had found evidence that rearrests are reduced when offenders receive treatment services in addition to the increased surveillance and control of the ISP programs."

Much of the early research on ISP largely focuses on the quantity, and not the quality of the supervision. Taxman (2002), however, suggests that the use of therapeutic techniques, in conjunction with more traditional means of supervision, may be essential to recidivism reduction. This approach is also supported by Petersilia and Turner's (1993) evaluation of ISP, as those offenders who received some form of auxiliary counseling services tended to perform better than those receiving similar protocols lacking in these elements. Despite this, there remains a "tension in probation between the goals of protecting community safety ('control') and promoting offender rehabilitation ('care')" (Skeem & Manchak, 2008). Pragmatically, a supervision strategy that combines the two approaches can be logistically challenging for a community

corrections agency. The duality can also be philosophically challenging for some, as the allocation of resources and organizational priorities are distributed differently when seeking to both supervise and treat than when the focus is simply control.

Despite the inherent difficulties, the consideration of the relationship between offender needs and offender risk should be a viable component of supervision. At the heart of this approach lies the idea of responsivity, that is providing targeted services to those individuals known to have the potential to best benefit from them (Thanner & Taxman, 2003). Similarly, the principle that underlies an approach to supervision known as the “principles of effective intervention” (Andrews, Bonta, & Hogue, Classification for effective rehabilitation: rediscovering psychology, 1990) reinforces the notion that programming, regardless of context, should be targeted to an offenders’ specific risk and need levels. Empirical evaluations have been supportive of this principle (Lowenkamp, Latessa, & Holsinger, 2006). In practice, this suggests that both intensive supervision and treatment opportunities should be allocated in a manner through which “the risk and needs of the offender should drive the selection of an appropriate program that can address the criminogenic factors (Taxman, Thanner, & Weisburd, 2006).

ISP, despite posing challenges, also creates the opportunity for the delivery of time-intensive interventions. Some research has suggested that ISP programs consisting of both treatment and control components can be more effective than approaches to supervision that focus on a single dynamic (Fulton, Stone, & Gendreau, 1994). These programs may encourage short-term compliance with supervision while, at the same time, allowing for longer-term behavioral changes. Other approaches have encouraged



the use of deterrent measures to achieve additional compliance with the treatment components of the ISP program (Petersilia & Deschenes, 1992).

The increase in the number and regularity of supervision contacts ensures that offenders will physically report to the agency with the frequency necessary to permit enrollment in, for example, a classroom-based program. Less intense forms of supervision would require a program framework with larger delays between meetings, potentially decreasing the total number of treatment hours or increasing the amount of time necessary to complete the full program.

## **II. Risk Prediction & Random Forest Forecasting Procedures**

The allocation of any scarce resource first requires the accurate identification of the targeted population. In community corrections, and much of criminal justice, this responsibility traditionally fell within the broad discretion accorded to individual officers. Evidence has shown, for quite some time, that actuarially-developed forecasts can outperform subjective human judgments in most situations (Gottfredson & Moriarty, 2006). Each approach to assessment represents an attempt to situate offenders “on a continuum of risk using risk-related attributes, such as drug abuse, criminal offense history, employment status, and childhood exposure to physical or sexual abuse” (Silver & Miller, 2002) in a consistent and inclusive manner. Recent advancements in the statistical procedures that underlie such tools have allowed for more precise identification of subgroups within the probation population.

Risk assessment has long been a part of criminology, though the extent to which research on the subject has been accepted, both by other scholars and practitioners, has

varied. For example, Hart (1923) examined demographic and criminal history variables for 680 parolees, finding 15 factors significantly correlated with rearrest ( $p < .01$ ). Later, Burgess (1928) developed a 21-variable risk instrument and used it to evaluate over 3,000 parolees. In his work, Burgess found that, of those men who scored poorly on his scale, 76% violated parole, while only 1.5% of the lowest risk group violated. Shortly thereafter, Glueck and Glueck (1930) identified seven variables from their data set that were identified as being highly correlated with subsequent criminal behaviors.

Risk assessment techniques have continued to develop, becoming increasingly more specialized and accurate over time. During this time, risk assessment has become more trait-focused and relies more heavily upon data gathered throughout the criminal justice process. Clement, in his review of the development of assessment approaches, concludes that, “these assessments typically involve both clinical and actuarial approaches sent against ‘political second guessing’” (Clements, 1996, p. 123). Often, instruments developed during this time assigned points for criminal history, offense characteristic and, prior behaviors, as well as for psychological factors, social history and observed personality (Clements, 1996). Over recent decades, the inclusion of increasingly dynamic data, from a wider range of sources, led to an increase in relative accuracy of the forecasting methods (Andrews, Bonta, & Wormoth, 2006). Research has consistently shown that certain variables, including offender age, number of convictions, pro-criminal attitudes and associations, and measures of antisocial personality predict reoffending. Recent meta-analytic reviews have found that these variables reliably predict general recidivism among juvenile delinquents, adult sex offenders, general adult offenders and mentally disordered offenders (Barbaree, Seto, Langton, & Peacock, 2001).

This is especially relevant for practitioners since, when comparing statistical assessments to clinical decisions, the evidence-based, actuarial approaches are consistently more accurate across multiple contexts (Gottfredson & Moriarty, 2006).

Risk assessment is more than a tool for the screening of offenders. Forecasts are key in making certain that evaluation results reflect the potential impact of a program, as accurate assessment is necessary to ensure that “the most costly and intensive services should be reserved for those individuals who present the most serious challenges to public safety and are apt to be in need of the interventions” (Taxman & Marlowe, 2006, p. 3). Meta-analytic evidence supports the relationship between risk-targeted interventions and larger effect sizes (Andrews, Bonta, & Hogue, 1990).

Sitting close to the leading edge of current prediction methods, the risk forecasting procedures used in the this project allow for the more accurate identification of the probationers best suited for the CBT focused, violence reduction intervention. Random forest modeling techniques, a machine learning-based approach for prediction, has a number of advantages over earlier, more traditional methods. Notably, these statistical techniques allow for the inclusion of asymmetrical costs for errors and, capitalizing on data-mining techniques, predictions can be made based on the untapped power in large, machine-readable datasets (Berk R. , 2012). These “ensemble” approaches also permit forecasting in absence of a causally specified model (Berk R. , 2005).

A machine learning approach to prediction in a criminal justice context has already proven successful. Advanced forecasting models using this approach have been successfully designed to predict homicide (Berk R. A., Sherman, Barnes, Kurtz, &

Ahlman, *Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning*, 2009), violence in a correctional setting (Berk & de Leeuw, 1999) and the role of race in capital punishment (Berk, Azusa, & Hickman, 2005). Similar models have also been used, within the same probation population as in this project, to identify those offenders who did not, at the time they began their sentence, pose a threat of serious recidivism (Barnes, et al., 2009). Perhaps most significantly (at least for the current endeavor), a random-forest prediction model developed by Dr. Richard Berk was used to both identify and pre-screen probationers for this project, ensuring that the intervention would be delivered, as designed, to high risk probationers (Barnes & Hyatt, 2012).

Recently, the relationship between risk assessment and effective program evaluation has been reinforced with meta-analytic evidence. Andrews, Bona and Hogue (1990), completed a study examining the relationship between risk, needs, responsivity and professionalism of treatment programs. They note that there is a significant interaction between risk classification and outcomes. That is, programs that focus on higher risk individuals are more likely to have larger, and more statistically significant, results. For example, they note that the studies of adult probationer recidivism that used LSI scores had RIOC (Relative Increase Over Chance) index rates of 43% to 56%<sup>1</sup> Risk of recidivism, in another recent meta-analysis, was shown to correlate with the magnitude of effect sizes. At the same time, the targeting of a higher risk population resulted in significantly smaller effects (Lipsey, Landenberger, & Wilson, 2007).

---

<sup>1</sup> RIOC is an index of association that corrects for chance and limitations in predictive ability due to unequal distributions in the margins of a 2 x 2 binary prediction table (Copas & Loeber, 1990).

Both Risk Needs Responsivity and the Principles of Effective Intervention require the selection of an intervention appropriate for the designated population. The use of actuarial risk predictions allows for the identification of the desired high-risk population. This match allows for the generalization of findings to populations identified with similar instruments in variable contexts. Additionally, accurate risk forecasting ensures that resources are expended in efficient ways, including the enrollment of only those offenders likely to commit a violent act in a violence reduction program or likely drug users in an early stage addiction intervention.

Population identification is only the first step; an appropriate and relevant intervention must still be developed. There are innumerable programs designed to encourage desistance from crime, it has been suggested that better understanding the cognitive processes that promote both continued offending and desistance from crime may have significant utility in the development of effective interventions (Ward, Hudson, Johnston, & Marshall, 1997). Though there are many programmatic approaches that view the issue from that perspective, cognitive-behavioral therapy has received a significant amount of attention.

### **III. Development of Cognitive-Behavioral Therapy**

Over the past several decades, Cognitive-behavioral Therapy (CBT) has become the most dominant, and perhaps most studied, approach to psychotherapy and behavior modification (Baker, McFall, & Shoham, 2009). The foundation for this approach is the principle that emotions arise from the operation of cognitive processes, both automatic and conscious, and that these thought patterns can be observed, managed and reformed.

Psychological distress and nonconforming behaviors are, therefore, suggested to result when “non-conscious [situational] evaluations have become sufficiently strong to overcome . . . conscious attempts to control them” (Matthews, 1997, p. 48). The development of “improved social and cognitive skills may result in the establishment of stronger social bonds and increased social integration” (Kazemian, 2007). Ultimately, these social skills may play a key role in the reduction of recidivism. A number of promising interventions have been developed to apply these principles to criminal and delinquent behaviors, potentially representing a shift in the intervention paradigm in community corrections.

#### **A. Theoretical Foundations**

The role of thought processes in criminal behavior is not new to criminology. A number of recognized, prominent theoretical approaches have key components that include cognition, including Sutherland’s “definitions favorable to crime” (1939), Sykes and Matza’s techniques of neutralization (1957) and the formulation of self-control under Gottfredson and Hirschi’s theories (1990). Psychological research has linked deficiencies in problem solving, self-control, anger management and decision-making to criminal conduct (Gendreau, Little, & Goggin, 1996). Numerous treatment programs have been developed to address these needs; cognitive-behavioral therapy has been one of the most successful and widely studied.

CBT effectuates change through modifying the way that participants respond to external stimuli. Cognitive structures, referred to as schemata, are used to process and organize information and are thought to form during psychological development. Logical errors formed at that time, especially those reinforced with negative stimuli, become the

foundation of later emotional problems. Generally, these stimuli arise during stressful social situations and overt confrontations. This cognitive architecture, though it can be managed or unlearned, may cause an individual to distort reality and predisposes them to depression, anger or other psychological dysfunction, depending on the nature of the distorted schema (Dobson & Block, 1988).

Cognitive-behavioral therapy is less a defined approach to behavior modification and more a collection of beliefs about the relationship between thoughts and actions, with a commonly used toolkit designed to impact both facets. Most forms of CBT feature “an emphasis on broad human change, but with a clear emphasis on demonstrable, behavioral outcomes achieved primarily through changes in the way an individual perceives, reflects upon, and, in general, thinks about their life circumstances” (Dobson & Khatri, 2000, p. 908). The extent to which self-reflection characterizes the interventions, as well as the predominant theoretical mechanisms thought to link cognition to action, has developed over the past several decades.

In the large majority of instances, CBT focuses on identifying dysfunctional cognitive-behavioral processes and replacing these routines with more acceptable and adaptive practices (Dobson & Dozois, 2001), though the methods and contexts vary significantly.

Cognitive-Behavioral Therapy is not uniformly defined. There are, however, core tenets that connect each form of the therapy and create the “fundamental propositions” of the approach:

1. Cognitive activity affects behavior;
2. Cognitive activity can be monitored and altered;
3. Desired behavior change can be affected through cognitive change (Dobson & Block, 1988, p. 4)

Taken together, these assumptions allow for an approach to behavior modification based on the thought that irrational or maladaptive cognitive schemata (attitudes and beliefs), cognitive products (thoughts and images), and operations (processing) influence problematic behaviors (Grave & Blissett, 2004).

The first proposition is a simplified restatement of the psychological mediation model. In this general model, an antecedent variable affects a mediator variable and the mediator variable subsequently affects a dependent variable, thus forming a chain of relations among the three variables (MacKinnon & Fairchild, 2009). In this case, cognitive activity acts as the antecedent variable and the restructuring as a mediator on the dependent variable, here being criminal conduct. The exact relationships and pathways of influence between antecedents, mediators and outcomes can be difficult to test empirically (Hoffman, et al., 2007). The presence of that relationship, in some form, is no longer a strongly contested issue within the psychological literature, although the form of the relationship is constantly being reevaluated (Dobson & Block, 1988).

The second proposition assumes that participants have the meta-cognitive skills to observe their own thought processes in situations. In practice, this skill is often generated through analogy or vicarious examples. As Grave and Blissett (2004) note, “[i]t can be argued that the cognitive capabilities required to understand and participate in this therapeutic approach are likely to be self-reflection, perspective taking, understanding causality, reasoning, and processing new information, as well as linguistic ability and memory.” Since much of the CBT model requires self-reflection and self-reporting, this aspect is difficult to evaluate empirically. Working with a criminogenic population with little prior exposure to these concepts only complicates matters further. Without the



ability to both monitor and reform thought processes, however, a CBT-based treatment modality would not be effective.

The final assumption of cognitive-behavioral therapy restates that CBT, as an interventional approach, has the potential to affect outcomes through the manipulation of cognitive thought and attendant mental processes (Beck A. T., 1995). This emphasizes the role of reformation of the negative cognitive mechanism through learning, as opposed to reinforcement strategies focusing on supporting positive aspects, in effecting behavior (Dobson & Block, 1988). As suggested by the mediation model approach, this is a requisite assumption in order for a cognitively-focused intervention to, through the suggested mechanisms, have an influence on behavioral outcomes.

These assumptions are neither uncommon nor unique to crime-prevention interventions. Programs that fall under the umbrella of cognitive-behavioral therapy all assume that internal, cognitive processes are employed in a manner that allows them to be both responsive to and control external stimuli. Therefore, not only can the restructuring of cognitive activity modify behavior, but, as some researchers have argued, it may be an automatic and reflexive change. If that is the case, “behavioral change [can] then be used as an indirect index of cognitive change” (Dobson & Block, 1988, p. 6).

It is important to distinguish cognitive-behavioral therapy from interventions that are simply behavioral or cognitive in nature.<sup>2</sup> As highlighted in the historical development below, modern CBT is derived from tenets of both traditions, but it has developed into a distinct, and stand-alone approach to treating psychological and behavioral distress. Behavior therapy is focused on environmental determinates of

---

<sup>2</sup> From a clinical perspective, the term “cognitive behavioral therapy” is often used to refer to both cognitive and behavioral programs, as well as interventions based upon a combination of the approaches.

behavior while cognitive approaches view only disposition and cognition as causal (Arkowitz & Hannah, 1989). Behavioral therapy, which draws on classical and operant conditioning models (Skinner, 1953), holds that behavior is structured through social learning and reinforcement (Bandura, 1977).

On the other hand, purely cognitive approaches see behavior as exclusively and directly derivative of cognitive dissonance (Arkowitz & Hannah, 1989). Similarly, social learning approaches perceive cognition as a mediator, while in cognitive therapy it is both mediational and causal (Schwartz, 1982). They also vary in their conceptualization of the ideal form of the therapeutic relationship and the role of technique in effecting change (Arkowitz & Hannah, 1989). Importantly, under both approaches, the role of cognition, and its suitability as a target for behavior modification, is recognized and central (Meichenbaum D. , 1977, p. 107). The hypothesized form of this relationship has developed along with the advancement of the CBT approach.

Cognitive-behavioral therapy is under near-constant adaptation and refinement. This “progressive, natural evolution,” is the result of developing technologies and increasingly sophisticated methods of assessment, both of programs and of participants (Herbert & Foreman, 2012, p. 4). CBT, like many other practice-driven approaches (as opposed to those derived from a single, unifying theory), has developed in multiple waves, each characterized by shifting emphases within the cognitive-behavioral process.

Hayes (2004) provides a useful framework to categorize and distinguish these stages of development. With roots in classical behaviorism (Skinner, 1953, p. 5), the first “wave” of CBT was developed in reaction to the weak support, within the psychological literature, for the psychoanalytic tradition. Instead, an approach was developed that was

based on the empirically-supported principles of learning theory. The focus of the approach was behavior modification through methods similar to those used in experiments of classical and operant conditioning. As Hayes (2004, p. 641) notes, the approach “focused directly on problematic behavior and emotion, based on conditioning and neo-behavioral principles. The goal [was] not to resolve the hypothesized unconscious fears and desires ...,” but to isolate associations between thoughts and behaviors. By doing so, early therapies attempted only to encourage and reinforce situationally appropriate conduct, not fundamentally change the way that thoughts were managed. This suggests that, though the relationship between cognition, emotion and action was understood, the direction of association was thought to be direct and unidirectional and not cyclical and reflexive (Dobson & Block, 1988).

The second wave of CBT took shape in the 1960s when the focus shifted to the interpretation of emotionally-charged and contextually-relevant patterns of cognition. In this period, “neo-behaviorists began to abandon simple associative concepts of learning in favor of more flexible mediational principles and mechanistic computer metaphors” (Hayes, 2004, p. 642). The clinical practices and theoretical conceptualizations of this era form the backbone of modern, commonly-accepted formulations of CBT.

Building on Bandura’s social learning framework, early cognitive-behavioral therapy rejected the stimulus-response construct as overly simplistic. Basic behavioral theory, developed by Pavlov, and later advanced by Skinner, was problematically limited to observable behaviors. Additionally, the newer approach precluded the “psychoanalytic emphasis on unconscious processes, historical material... and the need for long-term therapy [focused] on the transference-countertransference model” (Dobson & Block,

1988, p. 9). Basic psychoanalytical models of behavior, though still thought to play a role in the mediation of anxiety on cognition and cognitive processes, were overtly rejected and so were not applied to new constructs of cognitive restructuring by design (Doizos & Beck, 2012).

The primary treatment modality developed in this period is, in fact, the synthesis of two unique but theoretically compatible traditions. Aaron Beck, in his research on depression, developed an approach that was both a refutation of psychoanalytic treatment and based upon the (very) early Stoic traditions (Beck, Rush, Shaw, & Emery, 1987, p. 8). In the 1960's, Beck and others worked to integrate this methodology with Rational-Emotive Therapy (RET), a primarily behaviorist approach advanced by Albert Ellis (1969). RET similarly holds that evaluative beliefs play a translational role between adverse events and resulting consequences (Ellis, 1957). This conceptualization of a causal mechanism, focusing on the use of behavioral techniques for the management of emotions, was used to expand the range of treatment options, primarily for major depressive disorders (Beck A. T., *Thinking and depression II: Theory and therapy*, 1964). CBT and RET share many of the same outward characteristics, but are derived from different psychological and theoretical traditions and place varying emphases on treatment modalities. For example, RET favors a more aggressive role for therapists in challenging errors while CBT focuses on problem-solving frameworks (Ellis, 1980).

Under the CBT and RET approaches, impulses to commit crime are considered to be responsive to cognizable thoughts, not only environmental or emotional cues. These distortions in thinking include self-centeredness, misinterpretation of social cues, skewed moral reasoning, and perceptions of entitlement (Clark, Beck, & Alford, 1999). Beck

suggested that the automatic and antecedent thought processes and belief systems, not the uncontrolled emotional reaction, drive anti-social and depressive behaviors.

Although more recent, academic and theoretical developments in some areas of cognitive-behavioral therapy have begun to move away from Beck's conceptual model of cognition, this remains the dominant paradigm for CBT-based interventions, especially in criminology and criminal justice. This third wave of CBT programs, according to Hayes (2004), seeks to incorporate constructs of mindfulness and psychosocial acceptance into the general CBT model. These approaches, therefore, are less attuned to situational cognitive management, and instead intend to provide participants with a broad range of tools to enhance overall wellbeing (Herbert & Foreman, 2012, p. 5). As Hayes notes, therapies developed as a part of this third wave

tend to seek the construction of broad, flexible, and effective repertoires over an eliminative approach to narrowly defined problems, and to emphasize the relevance of the issues they examine for clinicians as well as clients (2004, p. 648).

For criminologists interested in applied and targeted behavior modification, as well as for those hoping to evaluate a program with discrete (and measureable) outcomes, these more recent iterations of CBT are not of significant interest. From a practical perspective, these newer approaches have been less utilized to manage specific and situational conduct, like criminal conduct or to control impulsive anger. They focus instead on building broad, flexible repertoires (Goldiamond, 2002, p. 121) that facilitate change across contexts. However, these approaches still mirror Beck's understanding of anger (Beck A. , 1999), leaving open the possibility for additional avenues for crime-related inquiry in this area.

Overall, Beck recognizes the relationship between his model of cognitive change and later iterations. He notes that "[a]lthough there have been many definitions of cognitive therapy, I have been most satisfied with the notion that cognitive therapy is best viewed as the application of the cognitive model of a particular disorder with the use of a variety of techniques designed to modify the dysfunctional beliefs and faulty information processing characteristic of each disorder" (Beck A. T., 1993, p. 194). Even though there are multiple ways in which the CBT model has been operationalized, the mechanism, as understood by its architects, remains sound.

## **B. Mechanisms**

The models of CBT that form the foundation for this research are derived from Beck's foundational research but incorporate aspects from the more recent models as well. Although the techniques have been adapted for use in non-traditional populations and for new outcomes, the underlying and basic mechanisms remain the same. Chronic offenders, much like depressed individuals, can be characterized by their faulty cognitive habits. These patterns, including self-justification, perceptions of dominance and victimization, misinterpretation of social cues and failures in moral reasoning, cause offenders to respond inappropriately to stimuli (Beck A. , 1999). Individuals operating under this schema tend to misinterpret actions, leading to apparently unprovoked, socially unacceptable and potentially violent reactions.

This abnormality of cognitive habits directly correlates with behavioral outcomes. Beck's model of cognitive discord, especially of the type that often leads to anger or depression, emphasizes "the cognitive content of [an individual's] reaction to the

upsetting event or stream of thought” (DeRubeis & Beck, 1988, p. 273). These thought patterns are assumed to be both causally related and antecedent to observable behavioral reactions; that is, the presence of a certain cognitive pattern has the ability to influence conduct independent of external stimuli (Teasdale, 1997). There is a degree of reciprocity between thought processes, control and action. Cognition plays a causal role in the development of maladaptive and antisocial behavioral patterns; mental processes both give rise to socially destructive behaviors and reinforce them over time (Beck A. T., 1964).

Since learning processes are regarded as the result of internalized thoughts and not external stimuli, an awareness of an individual’s cognitive structure allows for the isolation and remediation or replacement of problematic patterns within the therapeutic environment (Meichenbaum D. , 1977, p. 184). In many CBT interventions, this change is accomplished by attempting to introduce or simulate stressful situations and then training students to identify the antecedent cognitive processes and, subsequently, providing them with the coping skills necessary to avoid antisocial or counterproductive reactions (Mahoney & Arnkoff, 1978). Participants are also taught to identify cognitive errors, repeated patterns of near-automatic thought, that lead individuals to attach incorrect meaning to events or to draw- and act out on- inaccurate conclusions (DeRubeis & Beck, 1988, p. 276). CBT focuses on preparing participants to exert control over these cognitive processes. This is an especially important ability for offenders, as cognitive errors reinforce feelings of victimization and isolation, both of which are frequent predecessors to violence (Beck A. , 1999).

Cognitive-behavioral therapies attempt to change subsequent behaviors by altering thoughts, interpretations, and assumptions regarding external, behavioral stimuli. There are two approaches to behavioral modification from a cognitive perspective: cognitive restructuring and cognitive development. The program being evaluated incorporated elements of both approaches. Restructuring programs focus on distortions in thought, especially those noted by Beck, while developmentally focused programs work to address deficits in problem solving and social skills (Baro, 1999). Most effective CBT programs include elements from both approaches. As Dobson and Block (1988, p. 4) note, the assumption that individuals can recognize, as well as regulate, these cognitive processes is essential to the approach. This conforms to earlier research on self-control, especially with regard to the internal generation of the perceived goal structures and their role in mediating, or failing to mediate, impulsive conduct (Ainslie, 1975).

Cognitive regulation, therefore, can be perceived as a measure of an individual's ability to orchestrate environmental and personal variables to achieve desirable levels of behavioral regulation (Dobson & Block, 1988, p. 10). As Beck notes, "[e]ven when a person is highly aroused to engage in antisocial behaviors, he usually must contend with an inner deterrent to such behavior" (1999, p. 267). Accordingly, self-control can act as a check on violent, and potentially criminal, conduct and serves as a key theoretical mechanism.

### **C. Cognitive-Behavioral Therapy and Criminological Theory**

In many ways, criminogenic thought patterns and conduct, from a CBT perspective, are not unique from those that underlie depression and other psychological



disorders (Beck A. , 1999). When dealing with an offender population, the focus of the intervention is on interpersonal and social skills, two distinct skill-sets thought to influence the propensity to commit crime. This includes reinforcing attitudes necessary to encourage responsible conduct, to develop empathy and to gauge consequences (Little, 2000), as well as reducing the influence of automatic thought patterns making individuals prone to repeated, criminal and negativistic outcomes (Wanberg & Milkman, 2006).

Cognitive therapies, in some form, have been shown to be successful in ameliorating a number of non-depressive symptoms across different contexts and for a wide range of behavioral outcomes. For example, CBT has been shown to reduce suicidal ideation (Stewart, Quinn, Pleyer, & Emmerson, 2009), to aid in smoking cessation (Killen & al., 2008) and has shown promise in mediating adolescent drug use (Liddle, Dakof, Turner, Craig, & Paul, 2008). Of most interest to criminologists are those programs designed to address violence, delinquency and, in some cases, anger.

From a psychological perspective, self-control manages the cognitive regulatory process, allowing an individual to override automatic or instinctive impulses. This is especially relevant when deferring gratification or when perceptions of consequences have changed (Ainslie, 1975). More recent psychological research has shown that self-control, like a muscle, can be made stronger through repeated practice (Muraven, Baumeister, & Dianne, 1999). There is also evidence suggesting that the associations between self-control and behavior are significantly stronger for automatic behavior and for imagined, not actual, behaviors (de Ridder, Lensvelt-Mulders, Finkenauer, Stolk, & Baumeister, 2011). Therefore, the ability to resist automatic thoughts and resulting behaviors could be reinforced using hypothetical situations and relevant examples.

The role of impulse control in mitigating potentially criminal behaviors is well established in the criminological literature. Self-control can be expressed as “an outcome of the interaction between individual executive capabilities and the environmental settings” (Wikström & Treiber, 2007, p. 238). This cognitive-oriented skill is relied upon, therefore, when an individual deliberates, however quickly, about the commission of a deviant act.

Gottfredson & Hirschi’s General Theory of Crime (GTC) (Gottfredson & Hirschi, 1990) provides a framework to link internal self-control to criminal conduct. Gottfredson and Hirschi characterize an average offender as “impulsive, insensitive, physical (as opposed to mental), risk-taking, short-sighted, and nonverbal” (1990, p. 90). The propensity to commit crime and a lack of self-control are rough analogs under the General Theory of Crime. Since “people with high self-control are less likely under all circumstances throughout life to commit crime,” a deficit in that ability correlates with increased and sustained criminality (Gottfredson & Hirschi, 1990, p. 118). However, low self-control only creates a propensity for crime; the predicate circumstances for a particular crime must also be present (Pratt & Cullen, 2000). The same could be said for attendant cognitive processes; even in the presence of low levels of self-control, certain beliefs must be activated, through environmental cues, before crime can result.

As Grasmick and others have noted (Grasmick, Tittle, Bursik, & Arneklev, 1993), self-control can be broken down into several key dimensions, including a need for immediate gratification and indifference to the needs of others. There is some variability in the influence self-control has on individual-level actions, as a “lack of self-control does not require crime and can be counteracted by situational characteristics or properties of

the individual' (Gottfredson & Hirschi, 1990, p. 90).” Tittle et al. (2003) examined this variation in the effects of self-control across typically-used classifications, including age and offense. Although their findings “challenge the theory with respect to self-control being the primary, or perhaps only, cause of misbehavior and the implication that its effects are universal and similar in magnitude in all conditions, that is, that self-control operates without contingencies (2003, p. 448),” they recognized that the relationship between self-control was meaningful.

The effect of low self-control extends beyond crime and includes other non-criminal, but still deviant activities and the establishment of social bonds (Gottfredson & Hirschi, 1990, p. 191). Evans, et al. considered additional correlates, including “educational and occupational attainment, quality of interpersonal relationships with family and friends, marital status, association with criminal friends, including criminal values, and having a preference for time spent outside the home” (1997, p. 478). They found, when using measures of anti-social conduct, including self-reports, not captured in outcomes limited to arrests or convictions, a negative correlation between deviance and self-control. This relationship persisted, even when controlling for the influence of social factors” (1997, p. 493).

Prior research has been generally supportive of the GTC (Nagin & Paternoster, 1994) (Grasmick, Tittle, Bursik, & Arneklev, 1993), establishing an empirical link between crime and self-control. A meta-analysis, conducted by Pratt and Cullen (2000), evaluated the GTC across 21 studies containing 49,727 individuals. They found a mean effect size of .27, attributable to the link between measures of self-control and crime, even across studies that varied in the location of the study (e.g. community or custody),

by racial composition, age and dependent variables. As Geis notes, though there is no shortage of methodological or theoretical critiques of the GTC, “researchers typically find that there is a better-than-average chance that persons who commit traditional kinds of criminal acts lack self-control, however defined (Geis, 2000, p. 46).”

These conceptualizations of the role of self-control are related to the traits of impulsivity and insensitivity that cognitive-behavioral interventions were designed to directly address. In practice, Moral Reconciliation Therapy (MRT), another CBT-based intervention, attempts to develop the ability to delay gratification, encourages participants to end relationships with delinquent peers and overtly encourages the consideration of consequences before action (Armstrong T. A., 2003).

The fit between the GTC and the basic tenets of CBT is not perfect. According to Gottfredson and Hirschi, self-control abilities develop, and are stable, prior to adolescence and are the direct result of a lack of parental monitoring of, and punishment for, deviance (Gottfredson & Hirschi, 1990, p. 97). If that were the case, self-control would remain stable over time, inhibiting the potential efficacy of CBT. Research has shown that a supportive school atmosphere, association with pro-social peers, and parental skill improvement have been correlated with positive gains in self-control measures (Burt, Simons, & Simons, 2006). Pratt and Cullen also noted that social learning factors, thought to be incompatible under the assumptions of the GTC, were also significant predictors of crime and that controlling for the relationship of one theory did not reduce the impact of the other (2000, p. 148).

The impact of self-control deficits is not limited to criminal conduct; neither is CBT. As Evans, et al., found (1997), the deficits characterized under the GTC are also

associated with limitations in life chances, life quality, and other measures of social disadvantage. Cognitive-behavioral therapy was developed to address the same, basic underlying processes and outcomes and has been successful. CBT has been shown effective in the treatment of social disorders often attendant to and associated with crime, including post-traumatic stress (Foa, Hearst-Ikeda, & Perry, 1995) and anxiety (Bryant, Moulds, & Nixon, 2003). Additionally, by conceptualizing the role of self-control as one focused on the delay of only cognitive processes, and not physical impulsivity, GTC can be seen to include distinct aspects of control as separate from the propensity towards the commission of a crime. This may help in circumventing a long-standing critique of the GTC (Akers, 1991).

Given that there is a solid theoretical foundation for a relationship between crime and self-control, it is unsurprising that this relationship is evident in the subset of interventions targeting anger. Many of these interventions directly reflect conceptualizations of the interactions between internal cognition and self-control. Platt and Prout (1987), in discussing the origins of the psychological self-control literature, observe that early iterations of the perspective defined the process of self-regulation of behavior as occurring in three stages: first, through the commands of others, second, through self-directed, verbal commands and, finally, through internalized self-instruction. Over time, overt verbalizations and self-instruction becomes less essential, but modified behavioral patterns have the potential to remain (Meichenbaum & Cameron, 1971). Through this process, CBT, as well as other behavioral interventions, reinforces self-control measures as a mechanism to reduce the influence of faulty belief systems and, accordingly, externalized, deviant conduct.

Self-control is not the sole theoretical mechanism within the criminological literature that would support the impact of a well-designed CBT program on recidivism. Sutherland's (1939) theory of differential association, a social learning approach to crime, is another potential theoretical explanation for the impact of the intervention. Under basic Differential Association, learned definitions, normative evaluations of suitable behavior, caused crime when an individual had learned more definitions favorable to crime. Individuals learn these definitions by observing other individuals and other social learning processes (Akers, Krohn, Lanza-Kaduce, & Radosevich, 1979). Hostile behaviors continue unabated when acceptable means of handling the demands of relationships, and other interpersonal demands, within broader society, have not been learned (Fehrenback & Thelen, 1982). Cognitive-behavioral therapy offers an environment and an intervention that can address the specific learning processes regarding these criminogenic definitions.

Cognitive-Behavioral Therapy is, at its heart, designed to help individuals make better decisions through the management of their automatic thoughts and belief systems. The impact of cognitive restructuring mirrors the processes that underlie rational choice theory. Rational choice, refined by Clarke and Cornish, suggests that individuals make calculated decisions that are perceived as rational under the immediate circumstances, when deciding to engage in criminal acts (Clarke, 1997). Offenders must (1) understand the risk and rewards of a crime, (2) consider the expected costs, and (3) must evaluate these factors subjectively (not as objective society would) (Pilliavin, Thornton, Gartner, & Matsueda, 1986). These calculations are similar, in structure and nature, to Beck's (1995) understanding of the mental processing that precedes any action.

Rational choice focuses less on stable, individual level characteristics and considers the relationship between situational factors and the (potential) offender's perception of the circumstances and potential outcomes. Central to this notion is that the probability of a given choice can be manipulated by influencing the cost-benefit analysis regarding of the likelihood or severity of potential consequences and degree of hedonistic pleasure (Hirschi, 1986). By teaching participants to better think through potential outcomes, CBT can discourage crime by focusing on non-immediate gratification, alternative rewards (including intangible benefits like self-esteem) and expand perceptions of potential consequences (such as loss of family connections during incarceration) beyond those most often considered by probationers.

Lastly, cognitive-behavioral therapy, applied to criminal behaviors, may provide support for basic deterrence theory. As Wright, et al., note, many attempts to “deter crime with punishment may be ineffective because those individuals most prone to commit crime often act impulsively, with little thought for the future, and so they may be unmoved by the threat of later punishment” (2004, p. 180). The basic tenets of CBT, whether the active theoretical mechanism is self-control, differential association or rational choice, encourage deliberation and the suppression of automatic and uncontrolled actions. Under any of these approaches, CBT participants would be more likely to both consider longer-term consequences and evaluate alternatives prior to acting. This influence is supported though economic modeling approaches to deterrence theory, as the cognitive restructuring processes allows a participant to better assess the probability of detection and so influence the calculation of optimal behaviors within the “market” for crime (Becker, 1968).

Research has shown that individuals with low reported levels self-control are significantly more prone to offending when presented with instances of situational crime akin to those discussed under the rational choice approach (Nagin & Paternoster, 1993). This suggests that cognitive-behavioral therapy may impact criminal behavior through multiple, non-exclusive theoretical mechanisms. Though this evaluation is largely atheoretical, the crime suppression effects hypothesized are well supported within the criminological literature. Further exploration, moving beyond causality, will be necessary to specify the mechanisms through which CBT, in this instance, may be affecting criminal conduct.

#### **D. Models Of Cognitive-Behavioral Therapy**

Cognitive-behavioral therapy does not represent a single therapeutic approach, but rather can be considered to encompass a broad family of approaches, each united by a common hypothesis about the relationship between cognition and action. This inherent flexibility in the structure and theory of CBT-based programs allows them to be delivered across contexts that vary in the frequency or duration of treatment delivery or the skill level of available staff. As Herbert and Foreman (2012, p. 6) note, “comparing ‘CBT’ to a particular therapeutic model ... represents a category error, analogous to comparing ‘trees’ with ‘oaks.’” While this flexibility has allowed for multiple applications, and evaluations, of the basic theoretical approach, it also makes the generalization of results, especially within between-program comparisons, difficult.

With regard to offender treatment, the CBT approach has been operationalized through a number of different programs. Some of the more well known of these



interventions include the Reasoning and Rehabilitation (RnR) program (Ross, Fabiano, & Ewles, 1988), Moral Reconciliation Therapy (MRT) (Little & Robinson, 1993) and Aggression Replacement Training (ART) (Goldstein & Glick, 1994). Thinking for a Change (TFAC), a more recent program, was developed specifically for use in a correctional environment, but has not yet been as extensively evaluated (Bush, Glick, Taymans, & Guevara, 2011).

Though the exact components of the programs vary, these approaches share similar goals and, in many cases, tools and activities. As an example, the Thinking for a Change program,

is a cognitive-behavioral therapy developed to integrate cognitive skills and cognitive restructuring modalities of offender treatment. At its core, TFAC uses problem solving to teach offenders prosocial skills and attitudes. Consisting of 22 lessons, each lesson teaches participants important social skills, such as active listening and asking appropriate questions to more complex restructuring techniques, such as recognizing the types of thinking that leads them into trouble and understanding the feelings of others. (Lowencamp, Hubbard, Makarios, & Latessa, 2009, pp. 139-140)

Reasoning and Rehabilitation, on the other hand, as Tong and Farrington (2006, p. 5) note, is designed to,

teach offenders “how to think, not what to think”. The idea is that, when they are equipped with thinking skills, offenders will make prosocial behavioral choices that will allow them to move out of an offending lifestyle, which had previously been reinforced by poor thinking skills or criminal thinking. The R&R programme consists of 36 two-hour sessions... and has nine components: problem solving, social skills, negotiation skills, management of emotions, creative thinking, values enhancement, critical reasoning, skills in review and cognitive exercises. These components are interlinked, allowing for repetition so that the skills can be practised in different. (*internal citations omitted*)

Both programs, as are most cognitive-behavioral interventions, are characterized by a focus on identifying and resolving deficits in social skills and information

processing, as well as providing students with cognitive tools to prevent situationally, as well as legally and socially, inappropriate conduct.

The intervention delivered in Philadelphia, though developed specifically for the APPD population, was based upon approaches developed and validated as part of the aforementioned programs (Noble and Hyatt 2010). Characteristics of both programs were incorporated into the Philadelphia program. Notably, the didactic dialogue structure, embedded in a group context, was derived from Reasoning and Rehabilitation, while the use of video and multimedia examples to facilitate discussion was modeled on the approaches used in Thinking for a Change. The final program in Philadelphia, as discussed below, is unique, both in its delivery methods and in content.

#### **IV. Cognitive-Behavioral Therapy and Crime Prevention**

Given the promise of the CBT approach, as well as the relatively low level so invasiveness, it is unsurprising that a number of researchers, in criminology and elsewhere, have attempted behavior modification using interventions derived from the cognitive model. These evaluations have taken place across multiple contexts and varied in both their methodologies and outcomes of interest. Meta-analytic approaches, however, allow for the aggregation of the relevant studies and for the exposure of trends and broader findings in this literature.

##### **A. The Implications Of Meta-Analysis**

Cognitive-behavioral therapy has been evaluated extensively, especially as the approach has gained in popularity (Beck & Beck, 2011). This has become increasingly true in criminal justice contexts, as CBT is viewed as a cost-effective intervention with a

high potential for success. As noted above, there are multiple programs and interventions, which are derived, either wholly or in part, from the principles that underlie CBT. Though the individual program evaluations, a select and relevant few of which are discussed below, can allow for only limited conclusions about the potential efficacy of the Philadelphia program, the meta-analytic results better capture the effect expected from CBT-based programs.

Traditional cognitive-behavioral therapy and CBT-based rehabilitative programs have been in use for a number of years. There is an enormous amount of variability among these programs, including in the nature of the psychological skills being taught, the setting of the intervention, the structure of the program and the duration, among others. The outcomes of interest within the literature also vary significantly, as some interventions are designed to prevent recidivism, while others focus primarily on non-conduct behaviors, including anger and psychosocial measures of deviance (Lösel, 1995). This makes directly comparing program effectiveness difficult and drawing meaningful, externally valid conclusions about relative effectiveness complicated.

Meta-analytical approaches may also threaten construct validity, as the relationship between outcomes measured and hypothesized processes within each intervention may also vary. This can be problematic when dealing with a relatively heterogeneous sample of studies with variation in the measures, population, and intervention characteristics, as is the case with many meta-analytical evaluations of CBT (Nurius & Yeaton, 1987).

Initially controversial for methodological reasons, meta-analyses have largely become an accepted, and relatively mainstream, component of the literature, both in

criminology and the general social sciences (Durlak & Lipsey, 1991). Developed in the 1970s and first dubbed “metaanalysis” by Glass (1977), this approach has been relied on extensively in the psychology and treatment literature, for many of the same reasons it is useful in the present evaluation. For example, a 1993 survey of the prevalence of reported meta-analyses in the psychological and behavioral sciences literature identified over 300 individual analyses (Wilson & Lipsey, 1993). There, the authors note that this number reflects an accelerating methodological trend, with the primary limitations being those related to the assimilation of dissimilar methodologies, including in construct and experimental design (Hunter & Schmidt, 2004, p. 4), as well as an overreliance on the arbitrariness of statistical significance testing (Schmidt & Hunter, 1997). This limitation can be overcome, in part, through the use of other measures, including effect sizes and confidence intervals, and allows for the synthesis of research findings and the communication of those findings (Armstrong J. S., 2007).

The use of meta-analyses permits conclusions to be drawn about the overall effectiveness of a program, while overcoming the general limitations in external validity that characterize many field experiments and individual program evaluations. For example, not all randomized trials have a sample size sufficient to provide the power necessary to detect an effect. Meta-analytic techniques attempt to overcome this limitation, especially in program evaluation, as they allow for the statistical aggregation of results from independent studies (Garrett, 1985). The limited numbers of studies included in an analysis, possibly influenced by publication biases, and a lack of commonality across predictor variables are statistical hurdles that many meta-analyses must also address (Cottle, Lee, & Heilbrun, 2001). In meta-analysis, the suitability of

comparisons may be subjective, a bias difficult to statistically manage. There are approaches that may diminish the extent to which these factors influence results, though no amount of statistical manipulation can overcome the design limitation of the primary source studies (Hunter & Schmidt, 2004). If these limitations can be codified and coded, it is still possible to obtain an unbiased estimate of effect sizes across research designs of varying quality (Sanchez-Meca, 1997).

Lipsey and Wilson caution about the over-generalization of meta-analytic results. Methodological quality, availability and publication biases, as well as small sample bias, all pose challenges in the interpretation of results (Lipsey & Wilson, 1993, pp. 1192-1196). Garrett makes clear the relationship between methodological rigor and the relative effect sizes in an early correctional meta-analysis, noting that “the magnitude of the effect size is inversely related to the design of the study” (Garrett, 1985, p. 294). In that assessment, the more rigorous, and largely randomized, studies had a mean effect size of .24, while the remainder of the “weaker” studies had  $r = .65$ ; less stringent studies tended to overestimate effects. Overall, Garrett concludes that, though relatively small sample sizes limited the ability to evaluate all combinations of program characteristics and rigor, there was sufficient evidence to suggest that correctional programming could be effective, in the aggregate. This conclusion, strengthened through the meta-analytical framework, also suggests that cognitive-behavioral options were among the most promising.

Generally speaking, meta-evaluations focus on determining the mean effect size for multiple homogeneous, or at least similar, programs. Standard comparison statistics, including  $t$  and  $F$ , are not ideal for cross-study comparisons, as the size of the test statistic

depends directly on the amount of sampling error in each individual study and the sample size(s). Some researchers prefer the point biserial correlation ( $r$ ), as the measure can be used in path analysis and for analysis of covariance (Hunter & Schmidt, 2004, p. 275). Glass (1977) defines the effect size as the difference in the group means divided by the standard deviation of the control group. While discussing a meta-analysis of psychotherapy treatments, Smith and Glass note that since “effect sizes are identified by type of outcome, the magnitude of effect can be compared across type of outcome to determine whether therapy has greater effect on anxiety, for example, than it does on self-esteem,” as well as to compare similar outcomes across interventions (1977, p. 753). The reporting of effect sizes is representative of a general shift away from significance testing and towards an “emphasis on reporting the magnitude of experimental effects obtained” (Rosenthal & Rubin, 1982), though p-values remain pervasive in the research. Glass’  $d$  is most often used to express the difference between treatment and control groups in experimental comparisons (Hunter & Schmidt, 2004, p. 281), and so is a primary test statistic reported in the majority of research studies.

Standardization is key when comparing results derived from sources using different evaluations, analytical techniques and/or populations. An effect size provides an estimation of the difference in recidivism rates between the treatment and control groups that can be standardized across evaluations with differing outcomes on variable scales. For example, when looking at the point biserial correlation ( $r$ ) an effect size of .25 would equate to a difference of 25 percentage points between the arms of the trial. Therefore, if the recidivism rate for the control group was at a baseline of 80%, an effect size of .25 would translate to a reduction in recidivism of 55% for those offenders

receiving CBT. A negative effect size would indicate that, contrary to expectations, the CBT treatment actually increased the recidivism rate for those receiving the intervention.

Meta-analytical evaluations of CBT and other cognitive skills programs have a relatively long history within the treatment literature, only a fraction of which focuses on delinquency and crime-related outcomes. There are a number that sample from correctional and treatment programming (Lipsey & Wilson, 1993). This review focuses on those meta-evaluations that include identifiable cognitive-behavioral interventions in the study sample and an outcome that captures some measure of recidivism. Though the focus of this research is on an adult population, some juvenile studies are included where other characteristics suggest their inclusion has utility.

### **B. Meta-Analyses Of Cognitive-Behavioral Therapy & Crime**

Since CBT encompasses a broad range of interventions, meta-analyses are frequent used in order to synthesize the results from diverse programs. An early meta-analysis, both generally and within criminology, conducted by Garrett (1985), included 111 studies, published between 1960 and 1983, and examined the effect of residential treatment on adjudicated delinquents. Garrett found a positive, overall effect size of  $r = .37$ . The cognitive-behavioral interventions, making up only 14% of the total studies sampled, indicated effects that were, across outcomes, .58 standard deviations larger than the comparison groups. Though screening studies based on methodological rigor often decrease effect sizes (Lipsey M. W., 1992), these results largely remain consistent when the analysis is limited to more rigorous studies, though the differential is reduced to .44 standard deviations.

Mark Lipsey conducted a large meta-analysis of studies of juvenile treatment and delinquency diversion programs, both published and not, reported over 40 years (Lipsey M. W., 1992). Although only 24 of the total effects reported were attributable to cognitive-behavioral treatments, Lipsey notes that, when studies were clustered by design elements, these treatment-oriented approaches were “associated with larger effect sizes than other treatment approaches (Lipsey M. W., 1992, p. 120). In this case, the  $r^2$  increased by .11 over a regression model fitting all types of programs, supporting the idea that CBT, by itself, could be an empirically promising program framework.

Over the next several years, cognitive and behavioral interventions continued to be recognized as some of the most successful approaches to reducing recidivism. Lösel (1995) reviewed twelve meta-analyses on multiple modes of correctional treatment. The included analyses each had at least one outcome measure capturing crime or recidivism, though the target populations varied. Across each of the included meta-analyses, reported effect sizes ranged between  $r=.05$  and  $r=.36$ , with an estimated mean effect for all assessed studies of  $r=.10$ . Among these studies, the cognitive-behavioral interventions were among those with the highest success rates; 44% of the included studies had an effect size greater than .20. Lösel notes, however, that there are a number of other similar programs with negative effect sizes. Although an aggregated analysis of meta-analyses, where the unit of analysis is comprised of an already aggregated measure, provides little information on individual program effects, Lösel’s review suggests that CBT programs are effective across contexts and in a field where “nothing works” was the dominant paradigm for decades (Martinson, 1974).



Mark Lipsey, in the same volume, published a review of almost 400 treatment programs, each targeting an exclusively juvenile population and which focused directly on recidivism (Lipsey M. W., 1995). Lipsey found that that the entire sample of programs (n=397), reported between 1950 and 1995, on average, reduced overall recidivism by 10%. Although classified with a recognized “inherent fuzziness,” programs considered “skill-oriented” (approx. 33% reduction) and behavioral (20% reduction) were among the most effective (Lipsey M. W., 1995, p. 74).

A 1995 review of the impact of CBT on sexual offender recidivism also found promising effects for a narrow range of outcomes. The 5 CBT-based interventions included were designed to treat both juveniles and adults and outcomes included the commission of rape, attempted rape or child sexual offenses. The average mean effect of the CBT programs, when compared to controls, was 0.35, roughly equivalent to the effect of hormonal therapies (Hall, 1995). As Butler et al. (2006) note, though the CBT effect sizes were small, “given the impact on victims, it is arguable that any reduction in sexual offender recidivism is clinically meaningful.” The patterns of behavior that lead to recidivism, and so increase the danger to the public, are also notoriously difficult to reform in this population (Hanson & Bussière, 1998). Recent research has identified dynamic, psychosocial risk factors that predict sexual recidivism (Hanson & Morton-Bourgon, 2005), suggesting a possible mechanism for the relative effectiveness of CBT.

Meta-analytic evidence also supports the effectiveness of CBT in a criminal population, even when not targeting traditional measures of recidivism. Anger is one of the key cognitive components that feeds into criminogenic thought patterns, leading, in turn, to actual criminal behaviors. Many CBT-based deviance reduction programs

include explicit anger management strategies (Beck A. , 1999). The majority of interventions designed to treat anger, therefore, have implications for recidivism, though the two outcomes are not identical. Under one approach, known as stress inoculation training (SIT), cognitive restructuring is accomplished through exposing individuals to triggers known to provoke anger and aggression. Individuals are then taught to couple “cognitive self-statements” with an attempt to “mentally and physically soothe themselves” and avoid hostile conduct (Beck & Fernandez, 1998, p. 64). Beck and Fernandez conducted a meta-analysis of 50 studies ( $n=1,640$ ) and found a weighted mean effect size of 0.70. Though the analysis included diverse samples and interventions, at a minimum, it further supports the proposition that cognitive-behavioral techniques can impact key constructs related to anger and crime.

Meta-analytical evaluations of CBT have also been conducted at the international level. Redondo, Sanchez-Meca, & Garrido’s meta-analysis (1999) aggregated 32 European interventions directly addressing recidivism, 9.4% of which were explicitly cognitive-behavioral in nature. Aside from variations in theoretical models, there was significant heterogeneity in the programs. For example, the analysis included only 3 randomized studies, and multiple settings, sample ages and intervention characteristics; each subgroup represented only a fraction of the total sample. Despite these limitations, cognitive-behavioral interventions were the most successful type of programs overall. In this case, the effect size of the CBT programs (Pearson correlation coefficient  $r = .226$ ) was nearly double the impact of the mean result across all types ( $r = .120$ ). In a subsequent multivariate regression analysis, treatment type accounted for 48% of the explained variance in the model, with the partialised, unstandardized regression

coefficients for the CBT programs ( $\beta = .785$ ) remaining the most effective with regard to overall recidivism (1999, p. 271).

Lipsey et al. (2001), in a relatively early meta-analysis, were able to identify 14 studies that evaluated the impact of cognitive-behavioral programs specifically designed to reduce recidivism. Eight of the included programs were experimental in nature. Lipsey et al. reported a weighted mean odds ratio of .66 ( $\alpha = .05$ ) (Lipsey, Chapman, & Landenberger, 2001, p. 152). Taken individually, each of the studies reported a positive effect of the treatment, though few were statistically significant ( $\alpha = .05$ ). The overall significance, the authors note, may have been due to a single study, discussed below (Robinson D. , 1995), with a very large sample size. Overall, Lipsey, et al. consider CBT-based programs promising, though they question the generalizability of small sample studies with significant heterogeneity (2001, p. 155). Lipsey and Landenberger updated this study several years later, identifying a broader sample of 14 randomized evaluations, only some of which were included in the prior version (Lipsey & Landenberger, 2005). Though the focus was on the contribution of program characteristics, the study found a mean reduction in recidivism of approximately 27%. However, the make-up of the components of the evaluations was strongly predictive of effect size. For example, Lipsey and Landenberger note that, while the demonstration projects, led by researchers, returned a 49% reduction in recidivism, the practitioner-led programs did not reach that level, with a mean reduction in offending of approximately 11%.

Pearson, Lipton, Cleland, and Yee (2002) identified a group of 69 studies and program evaluations that were predominantly cognitive-behavioral in nature and focused

on recidivism reduction. Across all of these studies, the mean weighted  $r$  was .144, an approximately 14% reduction in offending. However, among those programs rated excellent from a methodological perspective ( $n=3$ ), the average effect size rose to  $r=.254$ . Overall, the authors, despite relatively strong findings, conclude that cognitive-behavioral interventions are an approach to recidivism reduction that lie in a “borderline area of verification” (2002, p. 492) due to concerns about statistical homogeneity.

Wilson, Bouffard, and MacKenzie (2005) compiled data on evaluations of structured, group-oriented CBT programs for offenders. Almost 2/3 of the evaluations assembled were for programs derived from either the Moral Reconciliation Therapy (MRT) or Reasoning and Rehabilitation (R&R) models. From these studies, they were able to calculate a total of 74 different effect sizes across the 20 studies, as some studies included multiple treatment-control comparisons. The mean effect size for R&R programs, including experimental and high quality quasi-experimental studies was positive (8% reduction) and statistically significant ( $r = 0.16, p < .05$ ). MRT, with only 6 high quality evaluations, had mean effect size  $r= 0.369, p < .05$  overall, an effect that drops to  $r=.33$  ( $p < .001$ ) when limiting the analysis to the 4 true experiments. Overall, all of the higher quality studies had positive effects favoring the cognitive-behavioral treatment programs, with a mean effect of  $r=.32$  ( $p < .001$ ). Unlike many other meta-evaluations, Wilson et al. were able to compare the relative effectiveness of the two most popular CBT programs, as “comparing the mean effect sizes across higher quality MRT, R&R, and other cognitive-behavioral programs suggests that R & R might be less effective than the other two (mean effect sizes of 0.33, 0.16, and 0.49, respectively; all are statistically significant

at  $p < .05$ ),” though this difference may be attributable to the larger variation in R&R-related outcomes (Wilson, Bouffard, & Mackenzie, 2005, p. 198).

Landenberger and Lipsey (2005) conducted a meta-analysis of 58 experimental and quasi-experimental CBT studies in order to examine which factors and program covariates were the most strongly associated with successful outcomes. This meta-analysis confirmed the positive findings of CBT effects on recidivism; some aspect of the cognitive-behavioral approach has an impact on subsequent recidivism. The mean odds ratio indicated that the odds of not recidivating in the 12 months after intervention for individuals in the treatment group were 1.53 times as great as those for individuals in the control group. This represents a reduction from the .40 mean recidivism rate of the control groups to a mean rate of .30 for the treatment groups, a 10% decrease. The most effective configurations of CBT produced odds ratios nearly twice as large as the mean, corresponding to recidivism rates of around .19 in the treatment groups, though these were not necessarily observed in the “brand name” programs.

Although designed to evaluate the broader impact of risk-guided correctional programs, Lowenkamp, Latessa and Holsinger examine the relationship between treatment characteristics and outcomes (Lowenkamp, Latessa, & Holsinger, 2006). Of the 97 programs identified in this analysis, all were delivered in some form of community-based program. The relationship between treatment type and program effectiveness was significant ( $p < .05$ ) and the direction indicated that those programs coded as either cognitive-behavioral or behavioral were more effective than those focused strictly on supervision intensity. Notably, those programs that included a

cognitive component and focused on risk principles were even more effective in reducing recidivism (Lowenkamp, Latessa, & Holsinger, 2006, p. 86).

In a later Campbell Systematic Review, Lipsey and colleagues were able to identify 58 randomized or quasi-experimental studies on juvenile or adult offenders where cognitive-behavioral interventions with recidivism as an outcome were tested (Lipsey, Landenberger, & Wilson, 2007). Overall, the study identified a 25% (.40 offenses to .30) decrease in mean offending rates within the 12 months post-treatment between the CBT and control groups. Lipsey et al. conducted a number of moderator analyses as well. From a methodological perspective, they found a significant negative correlation only between the use of an ITT framework and effect sizes ( $-.24, p < .05$ ); randomization was not significantly related to the magnitude of effect sizes. Factors relating to the class structure (number of sessions, number of hours per week and total hours) were also significant ( $p < .10$ ), as was the use of a risk screening process ( $\beta = .27, p < .10$ ), and the proportion of program dropouts ( $\beta = -.28, p < .05$ ). Interestingly, the “brand name” and generic programs fared equally well. When looking at the components of each of these programs, cognitive restructuring techniques ( $\beta = -.27, p > .05$ ) and anger control ( $\beta = -.32, p > .05$ ) were found to have the only statistically significant, negative relationship with post-treatment recidivism rates.

A review of these meta-evaluations hints at the effect sizes that could be observed in Philadelphia. However, the impact of CBT programs varies considerably, depending on the many characteristics, including program size. This difference has been attributed to the difficulties in delivering treatments to larger groups of offenders or over extended periods of time (Spruance, van Voorhis, Listwin, Pealer, & Seabrook, n.d.) and to the fact

that smaller programs are more likely to be demonstration projects run by academics (Landenberger & Lipsey, 2005). Multiple evaluations have also noted the relationship between the length and intensity of the program (Redondo, Sanchez-Meca, & Garrido, 1999). Taken broadly, the overall direction of more post-treatment effects sizes, while often small or statistically indistinguishable from chance, favors cognitive-behavioral interventions.

This heterogeneity in the literature also poses a challenge for researchers seeking to draw a conclusion regarding the overall efficaciousness of CBT and CBT-based programs. In some cases this variation is due to methodological limitations, while, in many other cases, there are simply not enough completed studies that combine specific programmatic elements, analytical design and analogous settings to provide a close parallel.

Relying on meta-analytic research poses challenges for those seeking to draw conclusions about the potential efficacy on individual approaches. As Wilson et. al, (Wilson, Bouffard, & Mackenzie, 2005, p. 200) note,

What cannot be determined from the preceding literature are the specific elements or combinations of elements that are critical in producing positive effects on offenders' behaviors. The evidence suggests that both deficit and distortion approaches can be effective as well as programs that emphasize moral teachings and reasoning. Further research is needed to gain insight into the "active ingredients" of these programs.

What can be gleaned is that cognitive-behavioral approaches are a promising approach to reducing recidivism. Closer scrutiny of individual programs similar to Philadelphia provides additional support and suggests about the promise of the intervention.

### **C. Prior Evaluations Of Similar Interventions**

As highlighted in the meta-evaluations, there are many programs that incorporate ideas of cognitive-behavioral therapy into a treatment protocol of some kind. There have been few evaluations that employ a relevant population and experimental design similar to this project. Notably, this project employs a (1) large-scale (2) randomized design to evaluate the impact of a (3) practitioner-led (4) CBT-based intervention designed to be delivered in a community correctional setting for (5) high risk offenders identified with a powerful, actuarial forecasting tool.

Two programs, Moral Reconciliation Therapy (MRT) and Reasoning and Rehabilitation (R&R), account for approximately two thirds of all available evaluations of cognitive-behavioral programs (Wilson, Bouffard, & Mackenzie, 2005, p. 177). The identification of comparable programs is difficult, as there are relatively few that have been evaluated using rigorous methodological approaches. For example, Landenberger and Lipsey (2005), though they identified 58 CBT programs that focus on reducing crime, found that only 33% were evaluated using a randomized design, only 13% of which “maintained sufficiently low attrition from outcome measurement to yield results with high internal validity” (Landenberger & Lipsey, 2005, p. 471). At the same time, approximately only half (53%) of the programs analyzed were implemented in a community corrections (as opposed to a detention) setting. The overlap between these two categories, representing the studies methodologically comparable to this evaluation, was understandably small.

Another review of 30 CBT programs for offenders, using different selection criteria and published the same year, found that only 4 (20%) of the studies employed



random assignment and only 7 (35%) involved probationers (Wilson, Bouffard, & Mackenzie, 2005). This prevalence suggests that there are relatively few CBT programs that have been evaluated through a randomized trial set in a probation agency. Landenberger and Lipsey (2005) do not identify the manner in which each program was classified, however, Wilson, Bouffard, & Mackenzie's review (2005) identified a single study, published 24 year ago, that evaluated a probation-based CBT program using a true experimental design.

The identified study, completed by Ross, Fabiano, and Ewles, was an evaluation of a Canadian implementation of the Reasoning and Rehabilitation program (1988). Aside from being noteworthy for design and context, the Reasoning and Rehabilitation program shares a number of characteristics with the intervention designed in Philadelphia. Significantly, the intervention was delivered by trained probation officers, audio-visual presentations were used to stimulate discussion, and the program targeted high-risk offenders. Additionally, like the current program, the Reasoning and Rehabilitation program is,

focused on modifying the impulsive, egocentric, illogical and rigid thinking of the offenders and teaching them to stop and think before acting, to consider the consequences of their behavior, to conceptualize alternative ways of responding to interpersonal trials and to consider the impact of their behavior on others, including victims. (Ross, Fabiano, & Ewles, 1988, p. 31)

After nine months, a smaller percentage of the probationers participating in the life skills program were convicted of any new offense (47.5% v. 69.5%) and, of those that were convicted, a lower ratio were sentenced to incarceration (30% to 0%) (Ross, Fabiano, & Ewles, 1988, p. 34). The lower rates of incarceration suggest that the

offenses committed by the treated probationers were less serious, but no offense data or significance levels were reported.

Despite these results, the Reasoning and Rehabilitation evaluation is not without problems. First, the sample was relatively small. Although the researchers intended to assign 25 offenders to each of the relevant treatment arms, only 22 offenders were enrolled in the cognitive-behavioral program and 23 into the standard control condition (Ross, Fabiano, & Ewles, 1988, p. 32). Secondly, the study reports 9-month outcomes but relies only on official conviction records. It is possible that many offenders, especially those committing serious offenses, would not have reached the conviction stage at the end of the follow-up period; the measures used may undercount the prevalence of offending. Lastly, although the direction of the results suggests a positive effect of the program, significance tests are not reported, for either group equivalence or outcomes.

The Georgia Cognitive Skills Experiment, which was not included in the above meta-analyses, was a later randomized, community corrections-based evaluation. The Georgia program, another adaptation of the Reasoning and Rehabilitation curriculum, was a much larger evaluation, including probationers being supervised in 16 parole districts in 1997-1998 (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004). The overall sample ( $n=468$ ), was divided between parolees assigned to receive the course (232), taught by trained officers, or a control condition. Randomization was successful; there were no significant differences in the usual array of variables (van Voorhis, 1999). Approximately 60% of those assigned to the course completed the entire program (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004, p. 292). Although treated

parolees had lower rates of arrest, the study found no significant group differences in returns to prison, arrests, revocations or employment. Using a quasi-experimental design to examine the impact of completing the entire program, van Voorhis did find significant effects in incarceration, arrests and technical violations when controlling for group differences (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004, p. 297).

A follow-up randomized study to the Georgia Cognitive Skills Experiment was conducted from July 1998 to April 2000 and included 1,193 randomly assigned parolees. Using an event history approach, researchers found a 3.3% difference between experimental and comparison group returns to prison after 30 months. After 12 months, the control group arrest rate was only 2.5% greater ( $p < .05$ ). These measures remained insignificant, even when combined with the sample gathered during the first study (Spruance, van Voorhis, Listwin, Pealer, & Seabrook, n.d.)

Both Phases I and II of the Georgia Cognitive Skills Experiments, though they focused on high-risk probationers, assessed risk using subjective determinations of which offenders were “problematic” (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004, p. 288), and not with an actuarial or validated instrument. Risk “screening” was done pre-random assignment, potentially influencing the external validity of results. Only post-hoc analysis considered the relationship between risk (using another instrument) and outcomes (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004, p. 290).

Perhaps the largest evaluation of a cognitive-behavioral skills program was carried out in Canada and is also absent from the meta-evaluations above. The evaluation, following several years of pilot testing, included 4,072 offenders who

completed the Cognitive Skills Training or were eligible for the program between 1990 and 1994 (Robinson D. , 1995). The intervention consisted of 36, two-hour group sessions offered in institutional and community settings that were led by trained correctional staff. Participants were randomly assigned to receive the course, or to a waitlist for potentially delayed treatment or to a no-treatment control group. Only 5.5% of the sample (225 offenders) completed the program in a community setting, and, since only 13 community wait-list offenders remained enrolled, the full wait-list sample was used for comparisons. The effect of the community-based programs was larger than that for the correctional program. Although there was only a 16.2% reduction in reconvictions among program completers from institutional programs, there was a 66.3% reduction in reconvictions among graduates from community programs. There was also a 39.1% reduction in any readmissions for offenders who had completed the program in the community (Robinson D. , 1995, p. 50). This effect remained significant, even when including those offenders who had enrolled in, but failed to complete, the program.

The Canadian program, however, screened only for preexisting cognitive deficits, not actuarial risk levels, though risk is considered in a sub-group and moderator analyses. Although location in the sanction process, here for pre-release prisoners, is different from the Philadelphia program, the relative effect sizes are what would be expected from large-scale cognitive programs. The use of an artificial control group also raises some methodological concerns. The Canadian program, like Georgia and Philadelphia, faced implementation challenges, with approximately one-third of enrolled offenders failing to complete the program (Robinson D. , 1995, p. 50). This is a likely contributor to the relatively small, but replicable, effect sizes (approximately 5%), observed in

interventions of this nature and scale (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004)

Overall, the results for CBT interventions have varied in magnitude and significance, but have consistently favored the experimental groups; offenders receiving CBT tended to offend less than their untreated peers. Treatment effects, the difference in the percentage of each group that reoffends, for some programs have been reported as high as 52%, but more typically have been in the range of an 18% to 25% reduction (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004). Within the framework of recent meta-analyses, programmatic effects were reported from 8% to 16% and from 4% to 5% in large programs (Wilson, Bouffard, & Mackenzie, 2005). This suggests that some mechanism common to many CBT and CBT-like programs has the potential to meaningfully impact recidivism rates. Few studies prospectively explore potential moderators or dose-response relationships (van Voorhis, et al., n.d.), but the need for more rigorous evaluations and replications is clear.

## **V. Conclusion**

The relevant literature, highlighted above, strongly outlines the contours of the current research and policy landscape. Probation agencies are being faced with increasingly severe challenges and budget restrictions. At the same time, community-corrections agencies, being used as both a proxy and overflow valve for overcrowded prisons, are looking to increasingly restrictive supervision protocols. Contrary to policy goals, these programs have been shown to increase rule and legal violations, or, at best, to fail to have a meaningful impact despite higher costs. CBT offers one potentially

efficacious intervention that could be used with this population, as theory suggests that interventions should target high-risk populations (Andrews, Bonta, & Hogue, Classification for effective rehabilitation: rediscovering psychology, 1990).

Relatively little is known about the impact of CBT programs delivered in a community-corrections environment. Despite having been evaluated many times, heterogeneity in prior research design has limited the generalizability of the findings. Cognitive-behavioral therapy, although shown to be one of the more successful approaches to recidivism reduction, is less frequently evaluated in a community corrections setting. This research, in addition to providing a randomized field trial to include in future meta-evaluations, suggests theoretical mechanisms and implementation strategies to begin to fill these gaps in the literature.

## CHAPTER 3: THE COGNITIVE-BEHAVIORAL SKILLS INTERVENTION

The chapter describes the unique characteristics of the CBT-based intervention developed expressly for this project and evaluated for this dissertation.

APPD, at the outset of the project, recognized that a control-only supervision approach was not in keeping with the full scope of their mission. This presented an opportunity, working with JLC researchers, to develop an intervention that, while based on the tenets shared by most CBT programs, addressed the unique situations and pressures faced by offenders on probation in Philadelphia. This program is further distinguished from other approaches through the use of specific language and phrasing, as well as hypothetical scenarios, designed to be relevant to the target population. This evaluation represents the first, and a preliminary, assessment of the program's impact on recidivism and supervision compliance amongst high-risk probationers.

### I. Background

Cognitive-behavioral Therapy has been used as a theoretical foundation for interventions addressing multiple types of behaviors, many of which were designed to be used under specific circumstances or with a distinct population. There are several CBT-based programs that focus on conduct relating to criminality and deviance. Of the better known, "brand name" programs, *Moral Reconciliation Therapy* (Little & Robinson, 1986), *Reasoning and Rehabilitation* (Ross & Fabiano, 1985), and *Thinking for a Change* (Bush, Glick & Taymans, 1997) are among the most frequently evaluated. Other programs, such as *Aggression Replacement Training* (Goldstein & Glick, 1987), share similar theoretical mechanisms, but are more focused on the reduction of anger-related behaviors, and less

on criminal activity. Research has shown that these programs are among the most effective applications of the cognitive-behavioral model, especially with regard to offender populations (Lipsey, Chapman, & Landenberger, 2001). These programs serve as the foundation for the intervention that was developed as part of this project. The newly minted, distinct program *Choosing to Think and Thinking to Choose* (CtT) has been an integral component of APPD's high-risk supervision protocol. Within the Department, the intervention is referred to as the "Life Skills" program.

*Choosing to Think* was not the first version of a CBT-based intervention to be developed for high-risk probationers in Philadelphia. Initially, APPD leadership, recognizing the need to address high rates of serious recidivism within certain subgroups of their population, created a special unit intended to provide both increased supervision and CBT-based skills training. Established in 2006, with the assistance of Penn researchers, the "Strategic Anti-Violence Unit," or "SAV-U," was established. In addition to increased levels of basic supervision, including regular home visits, probationers enrolled in SAV-U participated in weekly, one-on-one CBT sessions. The sessions, initially run by a Penn-trained clinical psychologist and later with the assistance of trained officers, were designed to meet specific needs, including anger management and prosocial communication skills. Unlike the rest of the Department, where all offenders received approximately the same level of supervision under caseloads



approaching 180 offenders per officer,<sup>3</sup> SAV-U caseloads were initially designed to be capped at 15.

The SAV-U model could not be maintained. In light of limited resources and caseload requirements, the supervision component of the unit was simply unsustainable. The CBT component also failed, but for different reasons. As Sherman (2006) notes, the therapeutic model broke down because,

[t]hey found one-on-one communication to be difficult, largely because the offenders did not define themselves in need of any help with psychological problems. The help they wanted was to deal with concrete issues in their daily lives, such as work, education and interpersonal communication skills.

The CBT component of the SAV-U protocol was terminated in December 2007. At that time, JLC researchers and APPD staff began to redesign the intervention, focusing on group-based instruction and facilitated dialogues. This reshaped curriculum would become the heart of the *Choosing to Think and Thinking to Choose* program.

## **II. Thematic Structure of the Program**

There are several overarching themes that are addressed from multiple perspectives, during the program. These include anger management, dealing with stressful situations, successful management of criminal justice and community correctional interactions and management of interpersonal and professional relationships (Noble & Hyatt, 2010). The themes, illustrated through topically relevant scenarios and

---

<sup>3</sup> In actuality, the entire caseload was not supervised under uniform protocols. From 2007 through 2008, 799 low-risk probationers were randomly assigned to be supervised, in another special unit and under a less restrictive protocol, under caseloads approaching 400 probationers per officer (Barnes, et al., Low-intensity community Supervision for low-risk offenders: a randomized, controlled trial, 2010). These probationers, and those enrolled in the SAV-U program, were the exceptions (along with judicially-mandated conditions) to the Department's one-size-fits-all supervision strategy.

examples, are used to both demonstrate and practice the necessary skills. The behavioral patterns encompassed the identification of automatic thoughts, the management of these beliefs, and, ultimately, a reliance on non-criminogenic belief structures.

There are 14 distinct lessons included in the program. Each of the topical sessions in CtT focuses on a particular aspect of behavior or cognition that is considered to be theoretically related to criminal behaviors. The application of these skills in contexts known to be problematic, criminogenic or challenging was also included. The thematic titles of these lessons are:

1. Introduction
2. Hopefulness
3. New Thinking I
4. New Thinking II
5. Choices and Consequences
6. Goal Setting/ Education
7. Employment/ Time Management
8. Anger (I)
9. Assertiveness (Anger II)
10. Dealing with Triggers (Anger III)
11. Interacting with the Community/ Social Skills
12. Stressful Conversations
13. Dealing with Setbacks
14. Wrap-up/ Graduation

The first class session consists of an overview of the program, including class rules, the benefits of completion and the consequences of misbehaviors. The second week is centered around the need for positive life changes, in light of contact with the criminal justice system, and is intended to serve as a motivation for engagement in the course. The basis of CBT is presented in weeks 3 and 4, as students are introduced to the relationship between thinking, feeling, and, ultimately, behavior. During week 5, facilitators focus on developing the notions of choice and consequences. Importantly, the

lesson stresses the importance of thinking about options before taking any action. Classes 6 and 7 are both designed to help students to identify prosocial, achievable goals and to identify a realistic plan for meeting them. Special attention is paid to the commonly held goals of furthering educational aims and securing legitimate employment. The 3 session anger management block begins in the eighth week. These classes will fulfill any Court stipulated anger management condition. Week 9 breaks down the differences between anger, aggression and assertiveness, while the tenth class works to provide students with a plan for dealing with common, difficult situations. Week 11 highlights the social skills necessary for successfully interacting with members of the community. In week 12, the discussion is focused on cognitive strategies to make stressful conversation, including those with probation officers and law enforcement figures, less difficult. The thirteenth session is used to work with students to both plan ahead and develop patterns of behavior that will assist them in continuing to make positive progress. The final class meeting is comprised of a course review and graduation ceremony.

Topics presented in the course are not mutually exclusive; each session begins with a review of the main points of the prior session and of cognitive and behavioral elements that unify the overall program. Notably, 6 sessions (42%) directly address the relationship between beliefs, cognition and action (sessions 2, 3, 4, 5, 6 and 13). Anger and anger management are the primary focus of 5 (35%) other sessions (5, 8, 9, 10, and 13) and pro-social communication skills, and the application of those tools to significant relationships, are addressed in 42% (6 sessions; 5, 6, 7, 11, 12, 13).

### III. Session Scheduling

Courses are arranged to provide the greatest flexibility and to encourage participation. During the evaluation period there were, at any given time, two sessions running simultaneously, one in the morning and one in the afternoon. Morning sessions ran from 10:00 AM until noon; afternoon sessions ran from 2:00 PM until 4:00 PM. These sessions were staggered by seven weeks. As such, when the morning class was participating in their first lesson, the afternoon class was learning the material in the seventh week of the course. Figure 3.1 illustrates a sample of Life Skills session scheduling; a full listing can be found in Appendix A. The weeks labeled A, B and C represent the time set aside to complete the subject recruitment process for each session.

**Figure 3.1: Life Skills Scheduling Schema**

	5/17	5/24	5/31	6/7	6/14	6/21	6/28	7/5	7/12	7/19	7/26	8/2	8/9	8/16	8/23	8/30	9/6	9/13	9/20	9/27	10/4	10/11	10/18	10/25	11/1
Session 1 (AM)	A	B	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14								
Session 2 (PM)									A	B	C	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Session 3 (AM)																A	B	C	1	2	3	4	5	6	7

By staggering the start dates between the morning and afternoon sessions, it was possible to deliver the course continuously and without a complete interruption when participants were being enrolled. As the facilitators concluded teaching a cohort of students, the recruiter was already in the process of identifying and enrolling the next group of participants. A single JLC staff member was responsible for this recruiting process throughout the evaluation period.<sup>4</sup> This flexibility also provided a higher

<sup>4</sup> A system of back-ups for this recruiter was put in place and used when the primary individual was out of the office, ill or otherwise occupied. A set of standardized procedures and a reporting protocol, including a script for communication with potential participants, was utilized to ensure homogeneity in the recruiting process.

probability of successfully scheduling offenders who had regularly scheduled concurrent drug treatment or other, court-ordered obligations, since the course was offered across a wider variety of times.

**Table 3.1. Class Start and End Dates**

Class Number	Class Start	Class End	Start Time
1	6/7/2010	9/6/2010	Morning
2	8/2/2010	11/1/2010	Afternoon
3	9/20/2010	12/20/2010	Morning
4	11/15/2010	2/14/2011	Afternoon
5	1/3/2011	4/4/2011	Morning
6	2/28/2011	5/30/2011	Afternoon
7	4/18/2011	7/19/2011	Morning
8	6/13/2011	9/12/2011	Afternoon
9	8/1/2011	10/31/2011	Morning
10	9/26/2011	12/26/2011	Afternoon
11	11/14/2011	2/13/2012	Morning
12	1/9/2012	4/9/2012	Afternoon
13	2/27/2012	5/28/2012	Morning
14	4/23/2012	7/23/2012	Afternoon

Each session included three classes offered each week (Tuesday, Wednesday and Thursday), with a fourth (often Friday), used as a make-up session. Participants from any class who missed their assigned class due to a pre-approved reason were allowed to attend on Friday. The same lessons were taught during each class. Mondays were

reserved for administrative work, including the updating of electronic case files, meeting with the JLC project manager and participation in APPD's staff training programs.

Overall, this schedule was designed to allow for the enrollment of the largest number of offenders. The staggered starts and make-up classes, for example, provided an opportunity for those students that missed a class to make up the lesson without penalty, and allowed for sufficient time to pass between session start dates for enrollment of randomly assigned offenders.

#### **IV. Curriculum Design**

There are many different interventions based on the principles of cognitive-behavioral therapy. Despite significant variation in application and implementation, many programs share similar design features; overall, the Philadelphia-based program was not remarkably different. In most cases, CBT interventions are designed to be delivered in a structured, classroom-based environment, to groups of 8-12 individuals (Dobson & Khatri, 2000). In this case, 15 participants were enrolled in each class. Since three classes were run simultaneously (Tuesday, Wednesday and Thursday), each cohort of participants included a maximum of 45 probationers.<sup>5</sup> To encourage dialogue and streamline record-keeping, each participant was assigned to a single session and was expected to report only to their assigned section each week. Modifications to

---

<sup>5</sup> During the initial weeks of the random assignment period, the classes were comprised of both RCT participants and other high-risk probationers meeting the eligibility requirements of the Life Skills program. This was necessary, as it took several months for enough RCT participants to be assigned into the treatment condition to completely fill the available slots in a class. Once this tipping point was reached, the majority of each class was comprised of experimental offenders. This same pattern was repeated, in reverse, at the end of the enrollment period.

this timetable due to changes in the participant's schedule were made at the facilitator's discretion.

The Life Skills program was designed to be delivered on a weekly basis for fourteen weeks. Each session was intended to last for approximately 2 hours, with a mid-session break lasting approximately 15 minutes. Depending on the lesson, class time was spent engaged in classroom dialogue, listening to instructor monologues, watching and discussing movie and video clips relevant to class topics and completing classroom-based writing exercises. Specific sessions also included mock job interviews and simulated encounters with probation and law enforcement officials. Homework was assigned on a limited basis, though was not considered a key component of the program. Compliance with out of class exercises was consistently low, though participants were expected to check an email address, set up by the course facilitators to encourage communication, on a regular basis.

While enrolled in the class, probationers were not expected to meet with their regular officer. Instead, they reported to a designated and dedicated classroom space. Although the non-participating offenders could not enter the classroom, this floor also contained the Department's single urinalysis laboratory; the communal waiting area was often quite full. Despite this, every attempt was made to keep the class participants separated from the general population. The facilitators also had their office in that area, further limiting the need for participants to interact with nonparticipating probationers and other APPD staff members. Complete separation, however, was practically impossible. All offenders under APPD supervision, regardless of their RCT status, waited in the same lines, most notably to enter the building's security checkpoint and

while waiting for a drug screening. Non-experimental, but repeated, qualitative observations of these common areas produced no evidence that offenders were discussing the relative differences in experimental and standard conditions of supervision, though the possibility of some, non-incidental discussions could not be fully controlled.

## **V. Incentive Structures**

Encouraging probationer participation in the program was one of the most challenging aspects of the implementation process. Programs that make participation mandatory, including as a condition of incarceration or parole, have, quite obviously, reported treatment compliance rates that approach 100%. Programs that cannot, or choose not to, incentivize participation often have treatment uptake rates in the 30- 40% range (Ebener & Kilmer, 2003). The early failures of the voluntary, one-on-one CBT sessions served to underscore the practical problems with non-mandatory participation. Optional enrollment also presents concerns for the analysis of experimental data. Since only the treatment arm of the trial could decline to participate, this unequal loss of participants could result in attrition bias and influence the validity of conclusions. (Jüni, Altman, & Egger, 2001).

Working with APPD leadership, the research team aimed to develop a system of incentives and sanctions designed to encourage voluntary participation. After a series of lengthy negotiations, including a consideration of the extent to which any incentive structure was permissible under departmental guidelines, an initial attempt was made to reduce an offender's total sentence length upon completion of the program. Delivering this reward unilaterally was, however, outside of APPD's authority. Instead, the



Department could only offer to write a letter to the sentencing judge explaining the nature of the offender's accomplishment and recommending a reduction in total time to be served on probation. During the "dry run" of the experiment, this incentive was found to be lacking in two significant ways. First, some judges, did not feel bound by the recommendation, ignored the letter and effectively rendered the reward moot. Secondly, participants seemed to have difficulty in comprehending the abstracted and uncertain value of the reward. These attempts were ultimately deemed unsuccessful.

In order to encourage participation for enrolled offenders during the evaluation, APPD modified the Terms and Conditions document that every probationer signs at the outset of their supervision. This was a significant accomplishment, in itself, as it required negotiating with the leadership of the Philadelphia judiciary in order to secure written approval of the change. As a result of this change, in addition to rules regarding the carrying of firearms and the use of drugs, every probationer "agreed" that, if they were asked to participate in the class, they would do so. Though the large majority of the individuals who signed the form had almost no chance to participate, the modification had significant consequences for those study participants in the treatment arm. For them, a refusal to physically attend the class<sup>6</sup> was the practical equivalent of violating any of the conditions of their supervision. For those offenders, failing to attend the class- or being disruptive during it- would result in the listing of their case for a technical

---

<sup>6</sup> The distinction between physical attendance and meaningful participation is important here. Although APPD felt that it could require offenders to go to the class, requiring them to engage in the class material was beyond their authority. Qualitative experiences of the research team suggests that, while many individuals, when faced with the enrollment requirements, stated that they would attend but would not talk at all, many, over the course of the program, did engage in dialogue with the facilitators and other students.

violation.<sup>7</sup> In addition to clearly reinforcing the Department's valuation of the program, this created a sanction that was clear, unambiguous and understandable to the target population.

In order to further emphasize this message, each enrolled participant received a letter from the Judge assigned to oversee any subsequent violation hearings. This Judge was familiar, by name or prior experience, to many of the program participants. The letter stressed the importance of program compliance, provided encouragement and clearly set out the result of failing to comply with APPD regulations, including attendance in the Life Skills program. Both the full text of the letter and the revised conditions of supervision are included in Appendix B. The letter was automatically populated with the offender's name and was printed out on judicial letterhead. Although there is no counterfactual with which to compare, APPD leadership felt that this was the strongest message that they, as an agency, could send to potential participants about the significance of participation.

Enrollment in the program was not without reward. Despite initial setbacks, APPD determined that it was able to incentivize participation by reducing the reporting requirements for probationers who completed all 14 weeks of the course and successfully met all of their conditions of supervision prior to graduation. This benefit was extended to all class participants, not just those in the experiment. Practically, and for the large majority of participants, weekly reporting requirements (4x per month) were reduced to

---

<sup>7</sup> In many cases, a failure to comply with the Life Skills requirements usually occurred in conjunction with multiple other technical violations, including failed drug tests or absconding from supervision. There were few instances where the sole justification for a violation hearing was related to the program; results of these hearings were mixed.

bi-weekly (2x per month). Exit interviews, conducted on the last day of the course, provided qualitative evidence that the reward served to encourage participation.

## **VI. Course Materials**

Most CBT programs, as Cullen and Gendreau (2000) note, attempt to help offenders to identify the problems or situations that led them into conflict with authorities, choose goals and to create and implement prosocial solutions to their problems. However, while many CBT programs rely on a dialogue between therapist and patient to uncover cognitive errors (Free, 1999), the Philadelphia program relies heavily on vignettes, video clips and abstracted, facilitated discussions. The influence of the balancing of media and dialogue in the Philadelphia program is of interest, as the literature suggests competing results: the videos may increase engagement by offering students the ability to discuss difficult, cognitive issues in a group setting while protecting dignity and confidentiality (Sheldon, 1987) or, as DeRubeis and Beck suggest, it could interfere with the “collaborative” nature of the therapeutic relationship (DeRubeis & Beck, 1988, p. 277).

Notably, the mass media clips and excerpts used were selected to be relevant for this population: fatherhood, respect, probation supervision and crime are used as frequent examples.

The application of examples relevant to the target population, as well as the use of a common vernacular, makes this approach to cognitive-behavioral therapy unique. For example, Ronald Noble (Noble R. , 2012), the psychologist involved in the development of the program, notes,

Two film clips shown during different sessions are of characters discussing the issue of achieving or regaining respect by resorting to violence. Both of these clips have been chosen because of their relevance to the street culture to which many of the probationer students belong. One clip is from the movie Juice. In the clip the character Bishop, played by Tupac Shakur, advocates the idea that one should be willing to kill and to die for respect. Another character, Q, played by Omar Epps, argues for a contrary point of view. The students are for the most part familiar with the film, and are aware that the character Bishop dies as a result of his obsession with achieving respect through violence.

Given that earlier attempts at one-on-one dialogue had broken down, in large part, due to probationers' lack of desire to talk about themselves (Noble R. , 2012), the use of the mass media clips was an attempt to demonstrate an application of CBT-based cognitive skills, without requiring the disclosure of personal information. Class participants were taught to identify cognitive errors in the abstract and, in later portions of the program, were given the tools to bring these skills together and practice applying these behaviors to simulated interactions.

## **VII. Facilitators**

The unique aspects of the Philadelphia experiment go beyond the methodological rigor of the trial. The intervention developed was designed to be run by practitioners, not academic researchers. This aspect of implementation is important, because, “program effects do not generalize from carefully controlled trials in which the researcher is heavily involved to ‘real-world’ settings, then the policy implications of carefully controlled trials are substantially diminished” (Armstrong T. , 2003). The balancing act between academic research (i.e. demonstration projects) and real-world trials (i.e. program evaluation) is a difficult one. Though more academic involvement may increase fidelity to treatment, external validity may be lost, and it can be difficult to generalize

findings. This is a key component of assessing the quality of implementation, and can have meaningful implications for effect sizes (Lipsey, Landenberger, & Wilson, 2007).

During the course of the experiment, there were three officers assigned to the program at all times.<sup>8</sup> Each facilitator participated in a rigorous training process, including written examinations in the principles of CBT and teaching observations, to ensure an understanding of the curriculum and the cognitive skills being taught. Program fidelity was also increased through the creation of rough scripts for each class that provided bulleted talking points and discussion questions to accompany the PowerPoint slides used during instruction. Regular observation, during the early stages, by a psychologist, as well as regular, qualitative observations did not indicate significant deviations from the intended lesson plans.

The degree of fidelity that can be expected during a field evaluation is directly related to the characteristics of the program. Measures of the extent to which a treatment, as delivered, adheres to the idea can be considered as a “continuous variable whose strength is relative to antecedent conditions” (Fagan & Forst, 1996). The use of paraprofessional instructors, in this case probation officers, allows for the externalization of findings, though it may presents challenges to the systematic and consistent delivery of a set protocol (Rezmovic, 1984). In this experiment, this represented a compromise necessary to ensure that the trial could be conducted within the parameters of the agency

---

<sup>8</sup> Initially, only two officers were tasked with supervising the course, with the third spot filled by fully-trained psychologists. At the conclusion of the “dry run” phase of the evaluation, since the curriculum was no longer being revised, a third officer was assigned to the Life Skills program. Three individuals were needed at all times in order to ensure there were sufficient resources to manage the classroom environment, provide back-up coverage for vacation and sick days and complete administrative tasks. These tasks included course scheduling, drug testing oversight and communication with unit supervisors.

(and labor union) rules and could continue to be used after the conclusion of the evaluation.

### **VIII. Evaluating *Choosing to Think and Thinking to Choose***

The number of CBT interventions available, as well as the variation within these programs, makes it difficult to compare programs. However, relatively recent developments in the juvenile justice and treatment arenas offers a yardstick with which to gauge the relative merits of- and evidence-based support for- the Philadelphia program. Although imperfect, this approach offers the best opportunity to assess the intervention in a standardized, quantitative manner.

Evidence-based policy, though infrequently the norm in criminal justice circles, has long been the standard in medicine and “other fields dedicated to the betterment of society” (Sherman, Farrington, Welsh, & MacKenzie, 2002 , pp. 1-2). Sherman and his co-authors note that, although being evidence-based does not, in itself, guarantee that an intervention is effective, it does ensure that the program is in keeping with the current understanding of what is more likely to be successful (Sherman, Farrington, Welsh, & MacKenzie, 2002 ). A number of scales have been developed to “grade” programs on the extent to which they reflect the leading edge of social science research. Some focus on the internal validity of the research, rating interventions by methodologies deployed (Sherman, et al., 1997). Others advocate for a scale that uses the relative cost-benefit ratios to rank the value of individual interventions (Welsh & Farrington, 2000) (Dodge & Mandel, 2012). Lipsey (2008) has developed an approach to ranking interventions based

on the extent to which they comport with the collective aspects of other recidivism reduction programs already known to be effective.

Lipsey's Standardized Program Evaluation Protocol (SPEP) was developed to evaluate the effectiveness of interventions targeted for juvenile offenders with outcomes relating to recidivism. The SPEP scale was constructed using meta-analytic techniques on a sample of over 600 controlled studies. After completing the initial analysis,

the characteristics of the programs with the largest effects on recidivism [were] identified from that research and translated into guidelines for effective interventions. Based on those guidelines, the SPEP is designed to rate programs according to how closely their characteristics resemble the characteristics shown by research to be most strongly associated with recidivism reductions. (Lipsey M. W., 2008, p. 4)

A 2008 analysis of programs in Arizona found that SPEP scores had a significant and relatively strong relationship with the risk-adjusted recidivism outcomes. That is to say, those programs with a higher SPEP score had lower recidivism rates than those with low scores. Another analysis, conducted using 163 programs in North Carolina, found a more modest, but still significant, correlation between "high" SPEP scores and larger decreases in expected recidivism (Lipsey, Howell, & Tidd, 2007). The SPEP scale offers a relatively efficient framework to judge the evidence-based merits of an intervention designed to reduce recidivism in an exclusively juvenile population.

The basic tenants of life course criminality highlight that juvenile and adult offenders behave differently, from a criminological standpoint, as they age (Sampson & Laub, *Crime and Deviance over the Life Course: The Salience of Adult Social Bonds*, 1990). The fact that adult offenders have had different experiences than juveniles influences their propensities towards crime (Warr, 1998) and likely impacts their

susceptibility to crime prevention interventions. However, in their meta-analysis of 58 studies of CBT, Landenberger and Lipsey (2005) found no relationship between effect size and whether the treated offenders were juveniles or adults. This suggests two things: first, that the SPEP scale, though imperfect, can be used to gauge adult programs and second, that a similar scale could be constructed for adult-limited interventions.

The SPEP scale ranks each program on 5 factors, each receiving a proportion of the 100 possible points, weighted to represent the proportion of the outcome attributable to that factor. The factors relevant for a probation-based program are, as summarized in *Improving the Effectiveness of Juvenile Justice Programs: A New Perspective on Evidence-Based Practice* (Lipsey, Howell, Kelly, Chapman, & Carver, 2010):

Primary Service (the type and goals of the program evaluated)	35 points
Supplemental Services (the presence of additional treatment types)	5 points
Treatment Amount	
% of population participating in full treatment, in weeks	10 points
% population receiving target number of hours (dosage)	15 points
Treatment Quality (delivery of services)	15 points
Risk Level (proportion high risk)	20 points

The scoring of any individual program, even given the standardized format of the SPEP, is inherently subjective. Judgments must be made, especially with regard to adherence to programmatic goals that require a measure of qualitative evaluation.

The Primary Service category is designed to evaluate the “philosophy” of the program. Lipsey (2009) notes that programs with a treatment modality, as opposed to a surveillance or deterrence-based approach, have a larger mean reduction in recidivism. The CBT intervention, considered alone, focuses primarily treatment and training. Since the Life skills classes do not require an increase in supervision intensity, relative to other high-risk probationers, the intervention should not be characterized as a control-based



approach. In this category, the Life Skills component of the Philadelphia program would receive 35 points in the primary category.

The foundational meta-analysis noted that increased supervision had a slight, positive effect on outcomes (Lipsey M. W., 2009). Supplemental Services, the small category designed to capture the additive impact of mixed modality treatment programs, could include the aspects of ISP included in the high-risk protocol. However, since the impact of recidivism is slight, at best, and increased supervision is not a component of the program itself, the Philadelphia program would receive 0 points in this category.

The Treatment Amount classifies programs by the extent to which they are able to deliver the intervention to the target population. Within one year of random assignment, 60.3% of those offenders assigned to participate in the program had enrolled in one class and 54.0% of those assigned had completed the entire program. Since SPEP considers only those offenders reaching “target levels” of treatment, here graduating from the class, the intervention would receive 5 points for duration and 7.5<sup>9</sup> for contact hours.

As discussed in Chapter 4, the intervention itself, though designed by a trained clinical psychologist, was delivered by probation officers. This complicates the assessment of the Treatment Quality category, already considered to be “the most difficult SPEP factor to rate based on actual program data” (Lipsey, Howell, Kelly, Chapman, & Carver, 2010, p. 30). Lipsey and colleagues (2010) hold that this element consists of 4 sub-factors, each designed to measure implementation quality:

- (1) a written protocol describing the intended service,
- (2) provision of training on the intended service for those delivering it,

---

<sup>9</sup> The SPEP scale includes thresholds at the 40% and 60% levels. Since actual treatment rates were slightly above 50%, the mean of the two categories was used.

- (3) a regular procedure for monitoring service to assess whether it is being delivered as intended, and
- (4) a procedure for taking corrective action when service delivery strays from what is intended.

In Philadelphia, and during this project, there was an extensive written protocol, including scripts for lectures, discussion questions and standardized classroom management procedures. Each of the four officers responsible for teaching the classes was trained by the same psychologist and passed a series of tests designed to “certify” their preparation to teach the class and confirm their understanding of the tenets of CBT. Researchers from the Jerry Lee Center and staff from the Adult Probation and Parole Department observed classes to ensure fidelity to treatment. Though not all sessions were continuously monitored, efforts were made to observe a randomly selected set of classes in each session. There were, however, no formalized procedures for taking corrective action when the delivery method deviated from the script. Informal discussions did take place, but it is likely that the program, as implemented, would fail to meet the criteria of the fourth sub-factor. Therefore, the program could be awarded either 10 points (“medium”) or 15 points (“high”); it is here that subjectivity enters the evaluation. For these purposes, the program can be awarded 15 points, as 75% of the requirements were met and the coarse granularity of the scale fails to capture the extent to which implementation was both structured and supervised.

The Philadelphia program would receive full credit under the final category, Risk Level. As discussed above, each of the participating offenders was assessed as high-risk using a random forest prediction model developed by Dr. Richard Berk (Barnes & Hyatt, 2012). There were no exceptions or complications due to treatment crossover. Notably,

this iteration of the program was designed to be delivered only to a high-risk urban, male population. Therefore, the program would receive 20 points under this factor.

The Life Skills intervention would receive 82.5 points on the SPEP scale, broken down as follows:

Primary Service (the type and goals of the program evaluated)	35 points
Supplemental Services (the presence of additional treatment types)	0 points
Treatment Amount	
% of population participating in full treatment, in weeks	5 points
% population receiving target number of hours (dosage)	7.5 points
Treatment Quality (delivery of services)	15 points
Risk Level (proportion high risk)	20 points
TOTAL:	82.5

Relatively few programs score as highly on SPEP as this program. For example, of the all programs included in one analysis, of a maximum possible total score of 85 (the Treatment Quality factor was unscored), 73% of the providers scored under 50. Only 6% of the providers scored 70 or higher (Lipsey M. W., 2008). This highlights one of the successes of the project- the development and delivery of an evidence-based cognitive-behavioral therapy intervention.

## **IX. Conclusion**

In Philadelphia there is strong evidence that suggests that a program, with these characteristics would have a positive effect on criminal offending. This is a rarity within the current literature on CBT and crime; this analysis presents an opportunity to advance our understanding of the impact that cognitive-behavioral therapies could have in the “real world.” The Philadelphia program, and this evaluation, is unique in several ways, including the probation-based classroom environment. Overall, this analysis, though not

designed to parse out the impact of these factors, will provide new insight into the effectiveness of this approach to delivering CBT to an urban, male population and the impact of new approaches to treatment. This perspective has been identified as lacking within the current literature (Lipsey, Landenberger, & Wilson, 2007, p. 27).

## **CHAPTER 4: METHOD**

Working together since 2005, the Philadelphia Adult Probation and Parole Department (APPD) and the Jerry Lee Center of Criminology (JLC) have collaborated on a number of projects designed to explore the relationship between community supervision and recidivism. The Philadelphia Anti-Violence Experiment is the most recent in a series of randomized trials and field evaluations conducted by the partnership. Through these efforts, researchers and practitioners seek to develop experimental, field-based evidence regarding the efficacy of community corrections-based harm prevention strategies.

This chapter describes the methodology of the Philadelphia Anti-Violence Experiment, a evaluation focusing on the effects of two aspects of APPD's high-risk supervision protocol on the conduct of high-risk offenders: cognitive-behavioral therapy and intensive probation.

### **I. Background**

Since 2005, APPD and researchers at the University of Pennsylvania have worked together to conduct a number of experimental evaluations of local policy shifts. With the development of a forecasting model, the partnership has focused on the identification of evidence-based supervision strategies that have shown the ability to protect public safety and increase efficiency within the department.

This high-risk project was not the first evaluation completed by the partnership. One of the most notable projects was a randomized trial assessing the impact of highly reduced supervision on low-risk offenders. Specifically, identified low risk offenders

were moved into to larger units with less restrictive supervision protocols. In that project, 1,558 offenders identified as low-risk were enrolled in the evaluation. Of these offenders, 799 were assigned to experimental caseloads of up to 400 offenders per officer and reported only once every six months. After one year, their offending patterns were compared to similar offenders (759) randomly assigned to receive the standard supervision protocol, which required only monthly reporting and consisted of much smaller caseloads. No significant differences in arrests rates were found after 12 months; 16% of the control group and 15% of the treatment group had been arrested for any new offense ( $p = .593$ ) (Barnes, et al., 2010). The lack of meaningful differences was found to persist over time, with no significant differences in the prevalence of offending emerging after three years ( $p = .874$ ) (Barnes G. C., Hyatt, Ahlman, & Kent, 2012).

These findings demonstrated the safety and feasibility of the low-risk protocol and prompted APPD to update the structure and organization of the entire department based on actuarial risk scores. Currently, all offenders are supervised in homogeneous, risk-based groups. These units have supervision programs tailored to the relative dangerousness of the offenders, instead of a traditional “one-size-fits-all” approach. Appendix C sets out the full hierarchy and organization of the Department. The decision to restructure was made in consultation with JLC researchers, but represents an internal, APPD-driven attempt to consider the policy and public safety implications of supervision strategies. This structural reformatting presented an opportunity to further evaluate the nexus between characteristics of community-based supervision and crime.

The supervision of potentially serious offenders presents an opportunity for APPD to implement a more intensive treatment component than was previously possible. The prevention of recidivism within this smaller subpopulation may return the greatest impact on public safety. This project represents an effort to develop new evidence, within an experimental framework, regarding two aspects of Philadelphia's high-risk supervision program: the impact of the intensive high-risk protocol and an exploration of a new and promising CBT-based crime-reduction program.

## **II. Risk Assessment Protocol**

Accurate and readily available risk determinations play a key role in shaping the structure of the Department and the manner in which individual offenders are supervised. A risk assessment protocol, therefore, is an essential component of the system. To that end, a risk forecasting model, designed by Dr. Richard Berk, was first implemented in 2009 in order to allow for the prediction of offender behavior while under APPD supervision. This model is based upon random forest prediction methods, a specialized classification and regression tree (CART) approach (Berk R. A., 2008). A series of models has been developed for use at APPD in order to reflect a developing capacity for risk stratified supervision and the availability of new data sources (Barnes & Hyatt, 2012). The methodology represents an approach to risk assessment that captures, in addition to traditional measurements of prior conduct, both the measurable and unknown non-linear interactions between predictor variables.

The identification and development of the most accurate and appropriate model for APPD, given limitations on their supervision capabilities, was an iterative process.

Over a period of several years, a series of prediction models was developed and refined (Barnes & Hyatt, 2012). The prediction model used during this experiment was constructed in late 2009 and was used to assess all cases, at their outset, from April 2010 through November 2011. Each forecast was designed to categorize an offender's statistically likely conduct for the two years following the start date of the assessment. Although this evaluation focuses on the most serious offenders, the model was designed to classify each case into one of three, mutually exclusive categories necessary for case management:

**High Risk:** the offender was predicted to commit at least one serious offense (murder, attempted murder, aggravated assault, robbery, or sexual crime) during the first two years of supervision; or

**Moderate Risk:** the offender was predicted to commit only non-serious offenses during the first two years of supervision; or

**Low Risk:** the offender was not predicted to commit offenses of any kind, during the first two years of supervision.

The classification of serious and non-serious offending encompasses a majority of the criminal conduct committed in Philadelphia County. The full catalog of offenses was derived from the Pennsylvania State Criminal Code (Title 18), as well as from state administrative law. This list was developed by APPD and JLC researchers to reflect a consensus, both within public policy literature and at the local, political level, regarding the severity of particular offenses. The research team reviewed, on multiple occasions, the developing criminal code and classified over 22,000 individual offenses. It is worth noting that the same classification schema used to determine prediction outcomes was used when classifying participants' post-random assignment criminal activity for the construction of categorical outcome variables.



The predictors used in the model reflect data routinely and electronically available at intake, and include criminal history, prior sentences, and demographic information. Berk et al. (2009) provides a comprehensive explanation of the statistical techniques used to forecast risk during the RCT, while an inclusive description of the model, including the predictor variables used, accuracy and cost ratios, can be found in Barnes and Hyatt (2012). A summary of the variables included in the prediction process during the course of this experiment is included in Appendix D. The data used to make forecasts, including measures of demographic characteristics, criminal history and prior conduct on supervision, are all available in machine-readable format and were collected as part of standard, administrative processes within the local and state court systems.

The risk forecasting model and the computer programs needed to make live predictions were integrated into the APPD intake department as part of the procedures used to manage all incoming cases of probation. This allowed the intake department to complete multiple actions simultaneously, all of which were necessary for both risk-based supervision and this randomized trial. Notably, this system allowed for the automation of the intake process, minimizing the opportunities for error and allowing for the blinding necessary during the experiment.

In practice, when an offender was sentenced directly to probation, they were in most cases given a paper copy of the judicial order and told to report to directly to the Intake Department. Located on the lowest levels of the Courthouse, the APPD staff in the office was responsible for entering the criminal case and sentence information into the Department's internal case tracking system ("Monitor"). Using the JLC program, these staff needed only enter the docket number of the case and the offender's Police

Photo Number (PPN).<sup>10</sup> The computer program then gathered all of the data necessary for the generation of a risk forecast, ran this information through the random forest prediction model, retrieved the prediction and assigned the offender to the appropriate officer (Barnes & Hyatt, 2012). During the experiment, this same set of programs, after retrieving the risk score, conducted eligibility checks, randomized participants and, where necessary, blinded the assessor and officer as to the “true” risk score.

This system, with the exception of the random assignment process, remains in effect at APPD and is used to forecast all incoming cases.

### **III. Setting**

The Adult Probation and Parole Department is the largest department within the Criminal Trial Division of the First Judicial District of Pennsylvania (Adult Probation and Parole Department, 2012). APPD remains the primary agency responsible for supervising criminal offenders in the community. To that end, their mission is “to protect the community by intervening in the lives of offenders.” They also seek to “hold [offenders] accountable by enforcing the orders of the Court... [and providing] a balance of enforcement and treatment strategies” (Adult Probation and Parole Department 2012).

The management of post-trial, adult community supervision has become a pressing need in Philadelphia. In Pennsylvania, sentences including less than two years of incarceration are served in County facilities. When offenders with such sentences are released within Philadelphia County, they fall under APPD’s parole authority. The

---

<sup>10</sup> In Philadelphia, Police Photo Numbers (PPN) are used as the primary person-oriented identification number. The first time an individual is arrested they have their photograph taken during the booking process. This number is used to link multiple types of records throughout the local criminal justice system and remains with the individual for the rest of their life. Absent subversion or an error in data processing, an individual should have one, and only one, PPN.

offender population under community supervision contains multiple types of sentences, including individuals sentenced to probation, to county parole or to both. Often, these individuals remain active offenders and contribute to the near constant growth in the local prison system (Shusik-Richards, 2010). These same pressures have encouraged the increased use of community corrections, leaving APPD responsible for the supervision of a larger and more serious group of offenders.

Since 2009, probationers have been supervised in three risk-based supervision divisions. After the “Low-Risk Experiment” was completed, every probationer was transferred within the restructured department to either high- (Anti-Violence), moderate- (General Supervision) or low- risk (Administrative Supervision) units.<sup>11</sup> Each unit includes only offenders of a particular risk score and has a distinct supervision protocol.

Units are supervised by a single Director and formed into larger groups known as Divisions. Like the component units, these Divisions are designed to supervise offenders of only one risk-level. APPD regularly audits their caseloads to ensure that, as active probationers receive a new sentence and, possibly, risk score, they are re-classified appropriately. The overall size and proportion of the risk scores, reported from the population census on one day, is set out in Table 4.1.

---

<sup>11</sup> Probationers and parolees with a Court-mandated condition of supervision, such as house arrest or drug counseling, were not included in this transfer process. These offenders remain supervised in mixed-risk units structured around the specific requirements of the condition.

**Table 4.1: Snapshot of Unit Caseloads, July 2012**

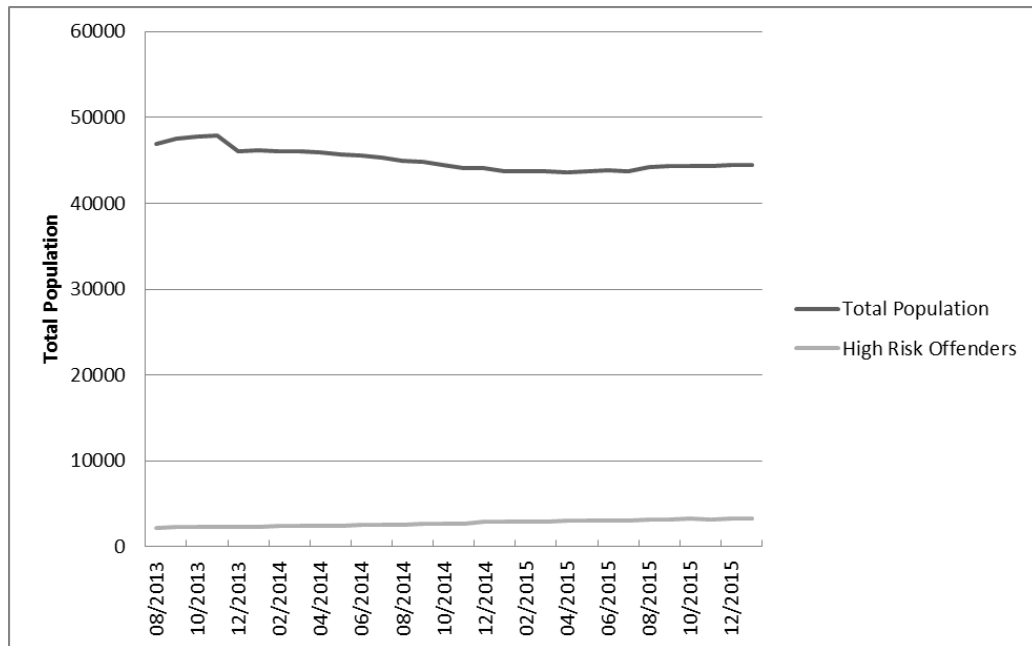
Division	No. Offenders	No. Officers	Mean Caseload
Low Risk	14,683	42	350
Moderate	8,911	57	156
High Risk	4,007	74	54 (Anti-Violence: 62; YVRP: 38)
Specialized	9,433	75	126 (range: 40 to 341)

The shift to risk-stratified supervision allowed APPD to move away from a one-size-fits-all approach to supervision to one that better adheres to the agency’s policy goals. Risk stratified supervision allows, to the extent possible, stricter supervision for probationers who may pose a danger to public safety, while reducing the intensity and cost of supervision for those that do not. Those offenders with a Court-ordered supervision condition, including drug treatment, sex offender monitoring or domestic violence education, remain supervised in separate units.

Philadelphia’s county correctional system relies heavily on probation and parole and current usage patterns are not an anomaly. As illustrated in Figure 4.1, the reliance on ISP-like increased levels of supervision for high-risk offenders has continued steadily, even when the overall population levels have fluctuated. Quite simply, there are more serious offenders being placed under community-based supervision. In August of 2010, APPD was supervising 46,965 distinct offenders, 2,254 of whom were classified as high-risk. Two years later, although the total number of offenders had dropped slightly, to 44,159, the high-risk population had grown to 3,160 offenders, a 140% increase. High-

risk offenders went from comprising 4.7% of the total caseload to 7.1% just two years (Tudor, 2013).

**Figure 4.1: Changes in APPD Active Population, August 2010 to August 2012**



These rates had significant implications for the department's allocation of resources. Due to a hiring freeze, few new officers were added during this period and so many divisions, including high-risk, were operating at less than full staffing capacity forecasting (Elliott-Engel, 2011).

Higher risk offenders require a larger amount of office time in order to complete basic case management functions. In many ways, this is the result of the volume of cases and the fact that these are the most criminally active offenders (Guynes, 1988). Over the same time period, high-risk officers were responsible for the administration of an increasingly large number of criminal cases. Since an offender could have more than one open case, this required the coordination of multiple sentences, types of judicial

oversight, mandatory conditions and, in the case of a violation, hearings in different courtrooms. The use of risk-stratified supervision allows for a degree of supervision and control to be used on potentially dangerous offenders, but comes at a cost.

The use of random-forest forecasting and increasingly restrictive supervision programs for higher risk offenders has had the desired effects. APPD has identified a population likely to commit a serious offense and has assisted in removing these offenders from the community. Though not explicitly designed to result in reincarceration, the zero-tolerance component ensures that dangerous offenders who commit new crimes, of any seriousness, or fail to follow the rules of probation are reincarcerated. During January 2013, for example, 41% (1777) of the high-risk offenders were in the local jail, while only 27% (2566) and 4% (537) of the moderate and low-risk offenders, respectively, were in police or local correctional custody (Tudor, 2013).

The high-risk division is comprised of two types of units: regional units, each of which is assigned to a specific geographic area of the city, or a single unit covering the entire County. All of the probationers assigned to receive CBT were placed in the single, non-geographically limited unit (“CityWide”). High-risk control offenders were evenly distributed across all of the regional units. Moderate risk units are all city-wide and participants were randomly assigned to each of the five units.

Regardless of their risk scores or unit assignments, all offenders report to one building, located in central Philadelphia. There is a single entrance, and all probationers and parolees wait in the same lines to enter the facility and proceed through security. Within the building, each floor generally includes offices for units of a single risk score,

reducing the level of interaction between probationers under different levels of supervision. Each floor has a waiting area only for use by those units and includes with approximately ten private interview rooms used for the majority of officer-offender interactions.

#### **IV. Treatment Groups**

The Philadelphia Anti-Violence Experiment consisted of three distinct comparison groups. As noted above, every offender beginning a case of probation or parole has some probability of being assigned to any of the experimental conditioners.

These conditions are:

1. **Intensive probation:** High-risk offenders receive an intensive form of probation, featuring standard weekly office visits and drug tests, as well as periodic home visits. Offenders under this protocol operate under a 'zero-tolerance' policy for rule violations and all technical violations should be prosecuted fully. These offenders are supervised in one of the three, geographically organized Anti-Violence Units, each with a different supervising probation officer.
2. **Intensive probation with CBT:** High risk offenders are supervised under the same protocol as the above, but are also expected to attend CBT classes when scheduled to do so. These probationers are all supervised in a single unit, AV Citywide, and are managed by one supervisor. Both the Citywide and regional units fall under the supervision of a single Director.
3. **Control:** High-risk offenders (having been forecasted as high) are labeled as moderate and receive the standard level of supervision. Notably, office visits are required once a month, there are no home visits and drug tests are ordered less frequently and not as a matter of policy. During this experiment, these offenders are supervised within

multiple units in the moderate-level General Supervision units.<sup>12</sup> This protocol is similar to the one-size-fits all strategy used for all offenders prior to the risk-based reorganization.

The focus of this dissertation, and for each of the comparisons reported below, is on the differences in post-randomization conduct between the intensive probation with CBT (treatment) and intensive probation (control) groups. The comparisons between the groups of high risk probationers receiving ISP (secondary treatment) and those receiving standard, moderate levels of supervision (secondary control) will be reported elsewhere. Discussions of experimental design, including randomization and power, will include all three of the comparison conditions to better reflect the overall structure and scope of the project.

## **V. Random Assignment Procedures**

The identification of a valid and consistent procedure for assigning participants into conditions was crucial to ensuring that sampling assumptions were supported and that analyses dependent on the  $F$  or  $t$  distributions could be used (Dean & Voss, 1999, pp. 3-6). In this case, random assignment procedures were automated and integrated into the computer program used during the intake process.

The integration of preliminary screening and random assignment into the intake process offered several advantages. First, the completely automated process ensured that

---

<sup>12</sup> During the RCT, the officers and directors responsible for supervising the control case offenders were not aware of which of the individuals on their caseloads were included in the experiment or, in many cases, that an experiment was being conducted at all. The officers in the Intensive Supervision group were similarly situated. The officers in the CBT unit were aware, given the nature of the intervention, that they were departing from standard procedures, but only the Supervisor was aware of the nature of the project. During the course of the experiment, the CBT-unit officers also had non-experimental cases (e.g. those assigned before RA began) and so were not aware exactly which of their offenders were enrolled in the RCT.



each offender was initially assigned the officer representing their appropriate experimental condition. Although offenders could, and often were, transferred for a variety of reasons post-assignment, the integrity of the assignment was preserved during these initial stages. The automation of the process, however, also prevented potential participants from being evaluated for their suitability for the intervention on any metrics not reflected in the databases. The assessment of language proficiency or for medical issues was not possible. Ultimately, the sample enrolled reflected the proudest possible selection, allowing for the evaluation of the impact of a *policy* of delivering CBT to as many offenders as practicable.

The automation of this process was especially significant during the evaluation, as it allowed for double-blinding and ensured that the probationer could be supervised in accordance with the experimental protocol and would not receive any extra attention from their assigned officer. Neither APPD line staff, nor the probationers themselves, were aware of which offenders were to be part of the RCT. In fact, in many cases, participants in the control conditions were not aware of the experiment at all.<sup>13</sup> As both experimental and non-participating probationers were supervised within each of the high-risk units, the supervising officers were not advised which of their cases were part of the trial and, in many cases, were also not aware that an evaluation was in progress. This limited the introduction of bias into individual-level treatment while on supervision and into subsequent analytical results (Jadad, et al., 1996).

---

<sup>13</sup> The need for informed consent was waived during this project as a result of the allocation of work between the JLC and APPD. JLC was responsible for the development and implementation of the risk forecasting model and assessment protocol. JLC also assisted in the development of the CBT intervention and implementation, as well as for the secondary analysis of outcome data. The allocation of supervision resources, including CBT and ISP, was at the sole discretion of APPD staff. Since this conduct falls under the general APPD operations, and not as research on behalf of JLC, the informed consent requirement was waived. At all times during active research, each of these projects operated under separate approvals from the University of Pennsylvania's Institutional Review Board (IRB).

Every offender who passed through intake during the enrollment was screened for enrollment in the trial. There were no systematic exclusions of qualified participants that could threaten the external validity of results. Periodic checks by researchers ensured that offenders were reporting to the appropriate units and all procedures for assignment were being followed.

The proportion of forecasted high risk offenders assigned to each of the conditions varied throughout the course of the experiment, though the criteria for enrollment remained constant. This was necessary to ensure a manageable caseload size within each of the treatment conditions and represented a necessary compromise with APPD. High-risk units, even those supervising CBT-assigned offenders, included probationers not eligible for the experiment. These offenders may have been under supervision prior to the evaluation period or failed the eligibility screening. The presence of these offenders complicated random assignment. Assignment rates had to take into account the maximum capacity of each unit, some who which could not be controlled through manipulation of the random assignment procedures.

Probationers who were not a part of the evaluation were assigned to officers on a rotating basis. When it was determined that an offender needed to be supervised by a specific unit, the assignment program selected the officer meeting that criteria who had not received a new case for the longest amount of time. Both RCT-eligible and non-eligible offenders, whose allocation was not directly influenced by the evaluation, were supervised within the same units. Therefore, the random assignment procedure needed to take into account influx of all cases to avoid exceeding the operational capacity of the participating units. Additionally, an attempt was made to “front-load” the CBT

treatment group so that the early sessions could include as many RCT participants as possible.

For those probationers deemed eligible for evaluation, each offender assessed had a chance to be placed in both the experimental and control units. For example, the probability of assignment to the CityWide Unit (CBT treatment) fluctuated between .5 in May 2010 as the unit was “filled” with offenders eligible for treatment, to .2 in the later months of the project. Table 4.2 shows the probability of offenders being assigned into each of the conditions over the course of the experiment. It is important to note that, although there was some variation in the relative probability of being placed in each group, every offender who was eligible to participate in the trial had at least a 20% chance of being assigned to any particular group.

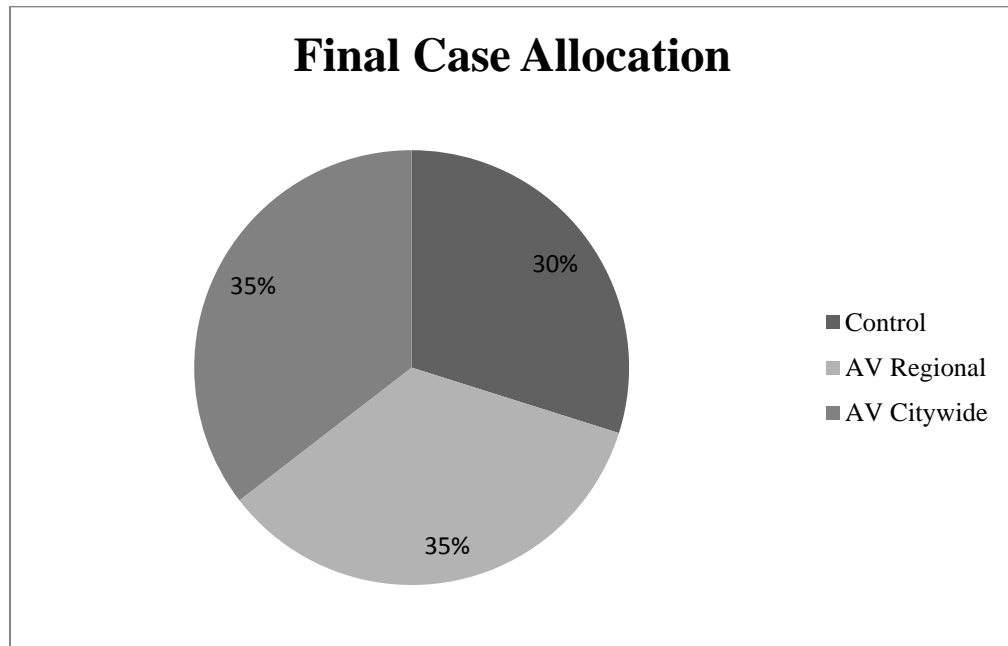
**Table 4.2: Random Assignment Allocations over Time**

Date	Percent to Moderate Control	Percent to High Control	Percent to CBT Treatment
5/1/2010	0.163	0.628	0.209
5/13/2010	0.2	0.3	0.5
7/19/2010	0.3	0.3	0.4
8/9/2010	0.3	0.2	0.5
11/1/2010	0.3	0.3	0.4
12/9/2010	0.3	0.4	0.3
2/2/2011	0.3	0.5	0.2

At the conclusion of random assignment, 34.6% of all eligible high risk offenders (447) had been assigned to the high-risk control condition and 35.4% (457) were assigned to receive the CBT intervention. Though not included in this analysis, 29.9% (385) of the

eligible, high-risk offenders were also assigned to receive ‘moderate’ levels of supervision.

**Figure 4.2: Final RCT Case Allocation**



Throughout the course of the evaluation, JLC conducted regular audits of the random assignment procedure and assignment mechanisms. Weekly reports were prepared, detailing the intended supervision unit of all enrolled offenders, where those offenders were actually being supervised and, in the case of a disjunction, the reasons why offenders had been transferred out of their assigned experimental condition. Throughout the course of the enrollment and evaluation period, there were no noted issues of treatment crossover. That is, no control condition offenders were enrolled in the Life Skills treatment program and none of the treatment group offenders were supervised

in the regionally-oriented control high-risk units.<sup>14</sup> High levels of treatment and supervision integrity ensured that the conduct of participants in one treatment arm did not influence the behavior of participants assigned to alternate conditions. Rubin (1978) refers to the lack of interference between units as the Stable Unit Treatment Value Assumption (SUTVA). SUTVA cannot be controlled through randomization alone, and “implies that the potential outcomes for a given unit do not vary with the treatments assigned to any other unit,” as well as that “there are no different versions of treatment” (Sekhon, 2007, p. 276). The Philadelphia Anti-Violence Experiment satisfies these assumptions with a near-perfect level of randomization compliance and through a regular auditing process that limited treatment crossover.

## **VI. Subject Eligibility**

The Anti-Violence Experiment was specifically designed to focus on high-risk probationers, as identified using the random forest model discussed above. Every probationer who was screened as high risk, however, was not ultimately included in the evaluation. A number of factors were considered when determining if a probationer was eligible for enrollment into the experiment. As with the risk forecasting and random assignment processes, these screenings were conducted “behind the scenes” and could not be influenced by the actions of APPD intake staff, supervising officers or the JLC researchers.

---

<sup>14</sup> As APPD, at the organization level, transferred cases between units in order to insure uniform caseload sizes, experimental offenders were occasionally moved into other high-risk units. However, as the JLC audits were conducted on a continuously and weekly basis, no offender spent more than 5 consecutive days (representing, at a maximum, a single appointment) in the incorrect unit.

First, the probationer had to have a new case of probation that began during the study enrollment period (5/1/2010 – 4/30/2011). Risk screening in Philadelphia is case based, so the opening of a new case file was treated as the initiating event date for both supervision and enrollment in the trial. Secondly, enrolled probationers could not have had a standing judicial order assigning them to a specialized unit (e.g. Domestic Violence, Sex Offender or drug treatment).<sup>15</sup> Third, they must also not have been eligible for the Youth Violence Reduction Program (YVRP), a multi-agency, grant-funded program that included intensive supervision and mentoring to young adult (< 25 years old) offenders living in certain areas of the city.

Fourth, the risk assessment and score that triggered consideration for enrollment had to have been the only such score within the 12 months preceding the case for which eligibility was being determined. The fifth criteria required that the probationer reside in Philadelphia County. Sixth, the subjects could not have been supervised under the high-risk protocol, or in any of the Anti-Violence units, at the time their new case began. Prior cases supervised by these units, if terminated or closed prior to the random assignment date, did not prevent inclusion. This was necessary to ensure that the effect of the ISP protocol and CBT intervention could be distinguished from prior experiences with the same increased levels of control and supervision. This criterion was also necessary to ensure the integrity of the blinding process, as the Department rules prohibited the transfer of active cases from high to moderate, a transfer that was necessary for offenders assigned to the control condition.

---

<sup>15</sup> In some cases, a probationer had been assigned to a specialized condition of supervision, but this qualification was not reflected on the Order in the offender's possession. In this case, the probationer was enrolled in the trial and, once the officer obtained an updated order, the participant was transferred from their randomly assigned condition and into another unit. There was no practicable way to screen for this situation, as the requisite information was simply not available at the time of screening and assignment.

Next, the seventh criteria required that eligible probationers also needed to be appropriately represented in the databases used to make predictions. A Police Photo Number (PPN) was required, as this unique identifier is required to link together records, each of which refer to the same offender, across the multiple data sources used in Philadelphia. Without a PPN, the risk software could not make predictions with the same level of confidence, as key data may be missing. A PPN was also necessary for the retrieval of outcome data, some of which were stored in Court and local jail databases using the same identifier.

The logistics of the Life Skills intervention also required the imposition of a number of additional screening criteria. For both the experimental and treatment groups, the eighth criteria required eligible offenders had to have been sentenced to a term of probation of at least 9 months. This was determined to be the minimum amount of time necessary to allow the offender sufficient opportunities to enroll in the program. Probationers who had sentences that, when aggregated with additional terms of probation beginning on the same day or before, exceeded the 9 month threshold were considered eligible for the experiment.

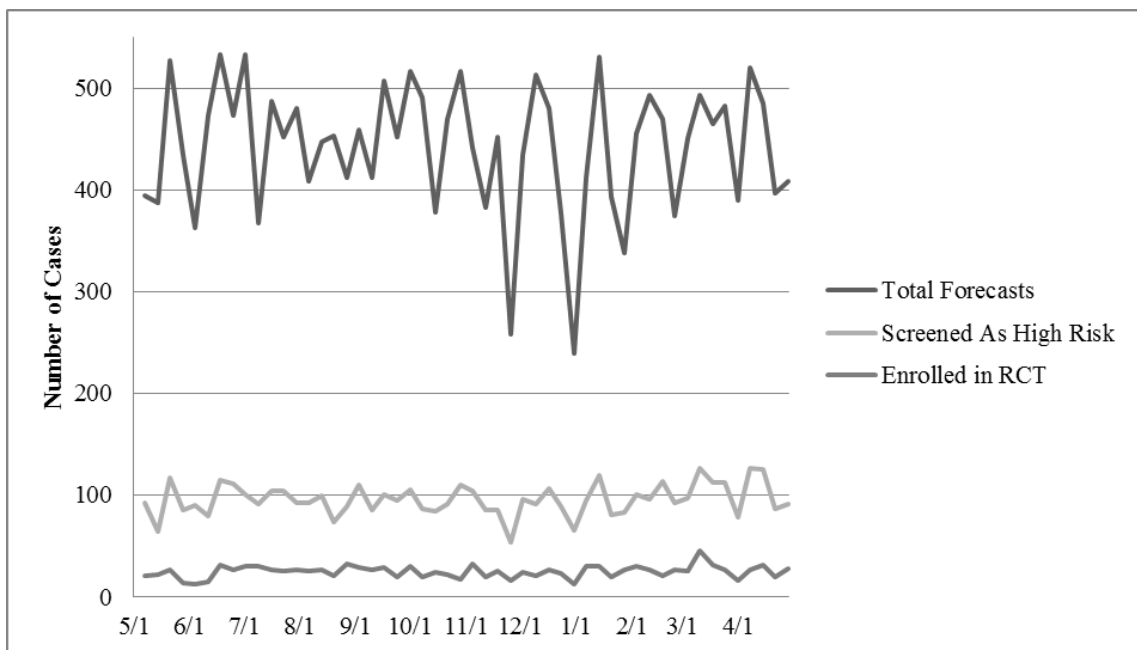
Finally, women were excluded from the program, as the intervention's developers, as well as APPD leadership, felt that the course was best suited for men. The program itself, from a learning standpoint, was better tailored to the types of experiences, including fatherhood, most often experienced by men. From a management perspective, APPD leadership, as well as the classroom-based facilitators, felt that a male-only environment would be safer, more productive and easier to manage.

**Table 4.3: Enrollment Criteria for Anti-Violence Experiment**

1. Forecasted as High Risk
2. No court orders to specialized supervision
3. Not eligible for Youth Violence Reduction Partnership
4. No previous High Risk forecast in last year
5. Philadelphia resident
6. Not already supervised by Anti-Violence Units
7. Has valid Police Photo Number (PPN)
8. Sentenced to at least 9 months of supervision
9. Male

Every new case of probation was screened for enrollment into the trial. During the enrollment period, an average of 433 new probationers were screened each week. There was a high degree of variability in intake; during the busiest week, the 5 working days between June 28 to July 2, 2010, 534 unique risk assessments and eligibility screens were conducted. Perhaps unsurprisingly, the fewest screenings (239) occurred in the week between the Christmas and New Year’s holidays during the same year.

**Figure 4.3: Forecasts, Screenings and Enrollment Rates**



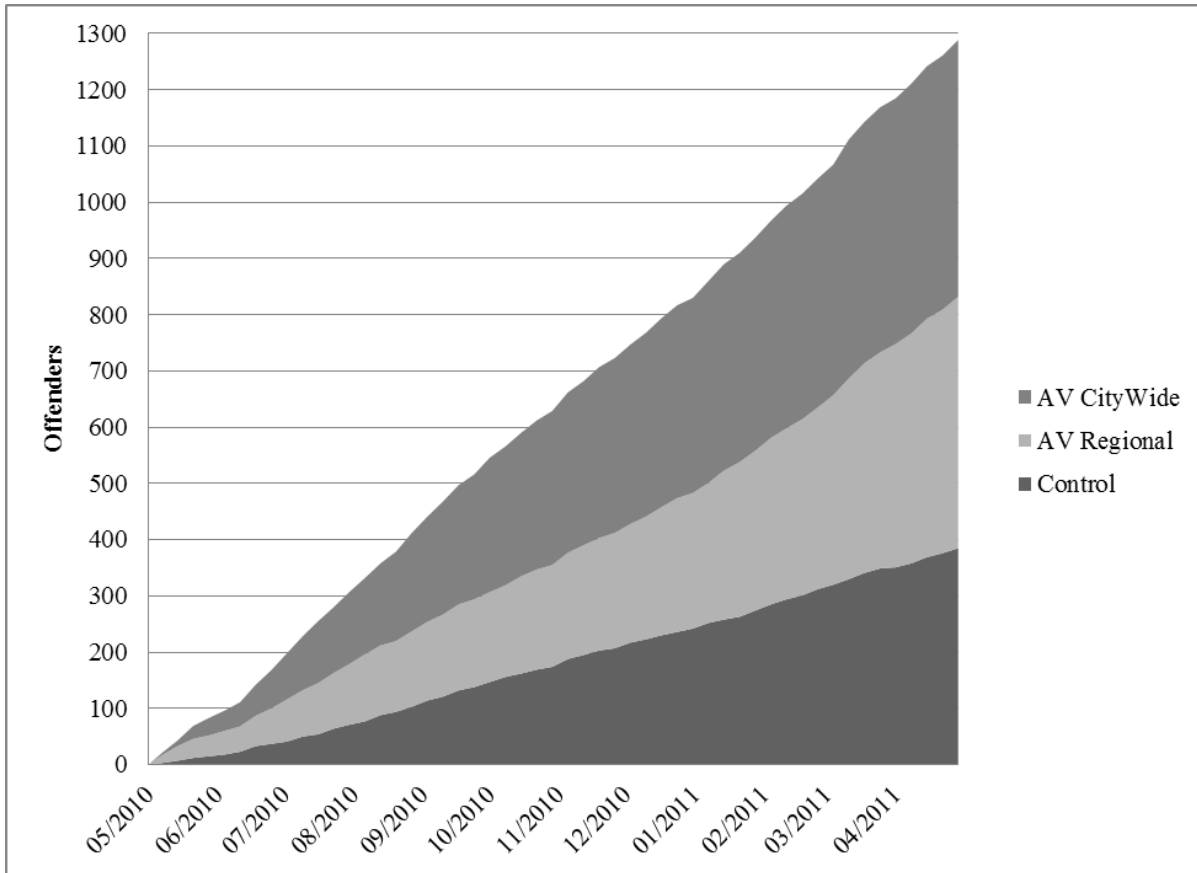


The automated, preliminary eligibility screening identified those offenders under APPD supervision who were suitable for enrollment in the trial. Of the hundreds of cases screened per week, an average of 94 cases per week (21.6%) met the most basic criteria: a high-risk score. Once the additional screening procedures took place, 24 offenders per week, on average, were eligible for random assignment. These offenders represented a relatively small proportion of the incoming APPD population. Each week, 25.4% of those offenders screened as high risk, or, considered more broadly, 5.6% of all new cases were enrolled in the trial.

## **VII. Sample Size and Power Calculations**

Even when balancing restrictive enrollment criteria with the need for a sufficiently large sample, the random assignment and screening process, detailed above, allowed for the identification of almost 1,300 eligible probationers in 12 months.

**Figure 4.4: Sample Size and Enrollment**



Sensitivity, also known as power, sets out the probability that a null hypothesis will be rejected when it is in fact true; this is known as a Type II error. As the power of a comparison increases, the probability of a Type II error decreases. This statistical test is influenced by a number of aspects of experimental design, including the  $\alpha$  used in statistical testing, the desired effect size and the sample size. Though subject to convention, Lipsey & Hurley (2009, p. 46) suggest that power should be “at least .80 to detect a reasonable departure from the null hypothesis.”

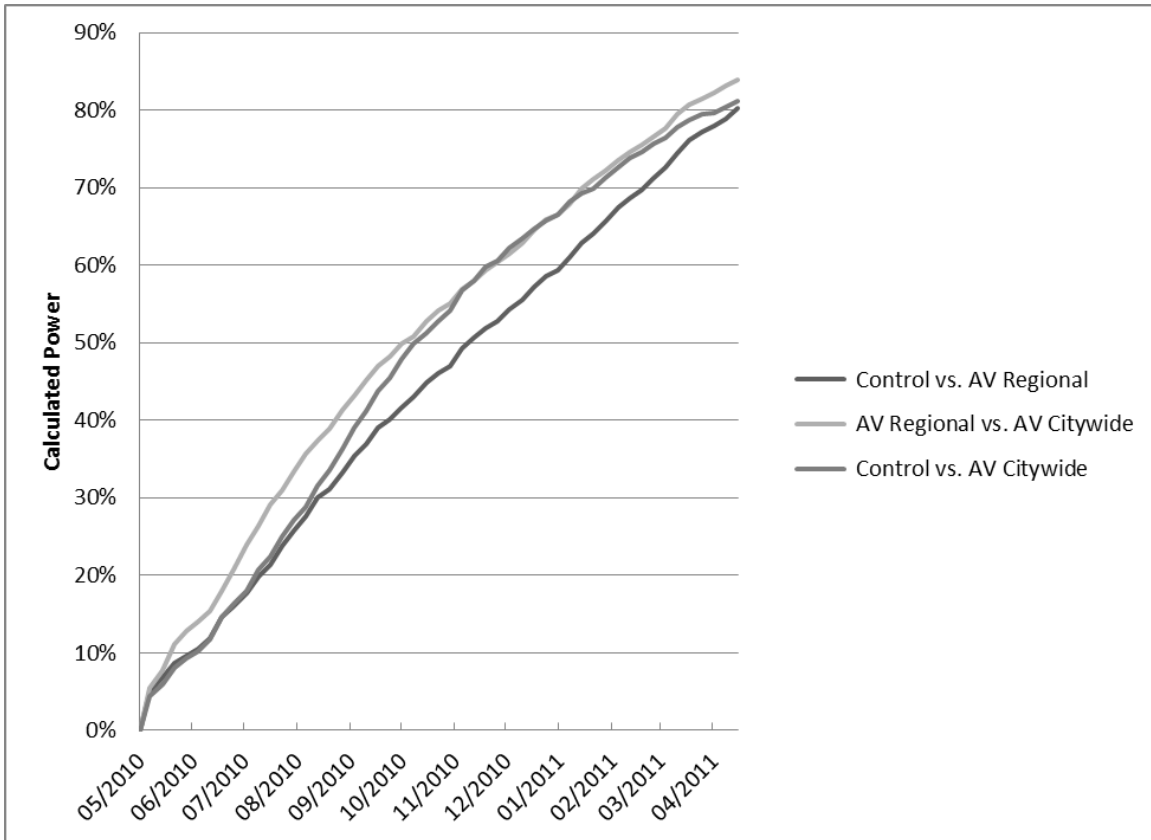
Since enrollment in the trial was variable, and the overall rate of forecasts was dependent of factors outside of the experimenters’ control, power could only be estimated

in advance. Instead, rough estimates of the sample sizes necessary to detect a minimum effect size, across each of the three comparisons, were calculated at the outset of the project. These calculations were derived from the average number of offenders per week that were assigned to the caseloads of officers in the high and moderate risk units during months prior to the experiment. However, since these estimates were based on historical intake rates and could not take into account fluctuations in the instant caseload sizes and changes in the number of available officers in each of the involved units.

The relative ratio of assignments, as detailed above, was managed by the research team. Although this tight control was necessary to ensure that caseloads were both balanced and manageable, this also provided the research team the opportunity to prospectively guarantee that each of the final comparisons would have the requisite power.

Calculations were conducted on a weekly basis to confirm that there were sufficient subjects being enrolled in each condition. When necessary, and as reflected in Figure 4.5, the ratio of offenders being assigned to a condition was adjusted to ensure adequate caseflow into each of the three experimental conditions. Power was calculated using a relatively conservative effect size of .2 for each of the comparisons below.

**Figure 4.5: Relative Power Calculations Over Time**



At the conclusion of random assignment, all three of the desired comparisons had power exceeding the .8 threshold. The high-risk comparison relevant to this analysis crossed the minimum acceptable threshold after approximately 10 months of intake. Since continued random assignment to those conditions was necessary to ensure the integrity of the assignment process, the final power calculation reached .852, or an 85.2% chance of detecting a meaningful effect, at the conclusion of random assignment.

## **VIII. Participant Recruitment Procedures**

Random assignment was conducted immediately upon the start of an individual's probation supervision. Although this meant that the greatest possible number of offenders were screened for enrollment into the experiment, there were some limitations created by this system. Notably, the automated screening and random assignment tool was limited to data that were reliably stored in the machine-readable databases maintained by or accessible to APPD at the moment the forecast took place.

During the developmental phases of the experiment, it was determined that there were a number of factors that would prevent probationers assigned to the CBT condition from being able to participate in a class. Employment posed a specific challenge, since APPD encouraged offender to seek employment and neither the research team nor the Department wanted to interfere with that process. Since it was not available in a machine-readable format, employment data was not included in the RCT enrollment screening. Even if it had been, a percentage of the participants were able to secure employment, or were already employed, during the time between trial enrollment and the start of the next Life Skills session. It was necessary, therefore, to rescreen potential participants as close to the start of the CBT class as possible.

Since space in each course was limited, only 15 probationers were enrolled for each class meeting(45 total offenders per session). This limitation was both practical and philosophical. The CBT program, since it required the development of a relationship between facilitator and student, was better delivered in smaller group settings. The 15 seat limit was also more pragmatic, as the classroom could only fit about 15 chairs in a comfortable semicircle. Scheduling was accomplished through a real-time screening and

scheduling process during the roughly three week period prior to each session. Though time consuming, the scheduling process ensured that the maximum number of treatment group offenders would participate in the intervention, a priority for the JLC team, and that the resources dedicated to the project would not go unused, a constant concern for the APPD leadership.

JLC researchers generated a list of probationers who, based upon set criteria, would be eligible for the Life Skills program. This ranked list was set up to ensure that RCT participants were given the greatest number of opportunities to be interviewed about, and scheduled for, the program. The list was distributed to all of the officers in the Citywide Unit the Friday before recruiting was to begin.

The recruiting list was prioritized in a manner that would give precedence to randomly assigned offenders who were approaching the minimum, 3 month threshold necessary to fully complete the program before the end of their probation sentence. Since all offenders enrolled had, at a minimum, 9 months under supervision, each participant had at least 2 opportunities to be recruited for the class. This hierarchy ensured that individuals approaching the end of their eligibility would appear on the list for all three weeks of the recruiting period and that appropriate attempts could be made to interview them regarding scheduling for the program. At multiple points during the evaluation, non-randomly assigned individuals were also enrolled in the Life Skills program, but they were always given a lower priority and enrolled only after it was determined that there were no RCT participants eligible for that seat in the class. Non-RCT participants were used to fill seats in the course, since the cost of operate the class was fixed and APPD wanted to maximize the impact and dispersion of the program. This was most prominent

at the outset of the experiment, as it took time for a critical mass of experimental offenders to be enrolled in the trial, and towards the end, as a large number of CBT-designated participants had either graduated from the program, completed their supervision or were no longer eligible to participate.

Once the active recruiting period began, the JLC recruiter was “on call” from approximately 8 to 5 each day. When a probationer whose name appeared on the list met with their primary officer during this period, the officer would provide them with a slip of paper to take out to the lobby waiting area and give to the front desk staff. The desk staff would then call the recruiter and advise them that an interview needed to take place. Once the recruiter had secured a space, they would call for the probationer to be admitted to the room. There was often a short delay in securing a space, as there were only 10 interview rooms, all of which were frequently occupied.

The recruiter conducted a brief interview with each potential participant. Lasting between 5 and 10 minutes, this interview was used to explain the logistics and purposes of the program, the expectations of the facilitators, and the benefits associated with completing the full class. As noted previously, these benefits included a reduction in the frequency of reporting and the satisfaction of any court-ordered anger management programming. Working with the probationer, the recruiter would then identify the class day that was best suited for that individual, record this information and provide them with a reminder stating the exact date and time of their class.

If a probationer had a valid conflict with the class, for example, concurrent employment or drug treatment, the recruiter recorded this information using a standardized form and coding scheme. A full list of these reasons is set out in Table 4.4.

A single RCT participant would have multiple entries in Table 4.4, as each time they were screened would be included as a separate incident, each of which may have had a distinct excuse.

A single recruiter was responsible for the majority of all recruiting interviews conducted during the study period. On the rare instances when they were unable to be present, another researcher took over this responsibility. A script was developed to ensure homogeneity in procedures and subjective experience, as well as to standardize the information that the enrollees would receive.

**Table 4.4: Reasons Probationers were Excused from Life Skills Participation, Post-Interview**

Reason	Event Count	Percentage
Employment	96	21.6%
School	8	1.8%
Treatment Program	2	0.4%
Unsuitable	5	1.1%
Supervision Conflict	6	1.3%
Appeal	1	0.2%
Other	2	0.4%
<b>Total</b>	<b>120</b>	<b>26.8%</b>

The interview and screening process was a necessary step, as offenders were excused from the program for a number of reasons, most of which were not reflected in the administrative data available during random assignment. Since situations changed rather frequently for enrolled probationers, each was rescreened during all sessions regardless of prior outcomes. The most common reason was for employment, as a relatively small percentage of participants remained employed for the duration of the project and a larger group had temporary employment at some point during the recruiting



period. Employment data were verified with the supervising officer who, as a matter of policy, had been asked to request copies of paystubs. Although this was a department-wide policy, only the CBT unit officers received additional reinforcement, both from the administration and through the file audits conducted by JLC researchers. This, unfortunately, prevents meaningful, between-group comparisons of employment data.

During the first 12 months of the evaluation, 68.7% of the treatment (313) group was interviewed at least one time. Multiple offenders were screened additional times, for a total of 445 distinct offender contacts. This repetition was necessary to ensure that, if an offender's situation changed and they became eligible for the class, they could be enrolled as quickly as possible. Of these 445 contacts, 73% (325) resulted in the probationer being scheduled for a Life Skills class.<sup>16</sup>

Despite the best efforts of the research team, the entire treatment group was not screened for the intervention. After 12 months, 31.5% of the treatment group was not interviewed a single time. The overall breakdown of why individuals were not enrolled in the Life Skills program is reported in Table 4.5. In this case, the count is not of individual offenders, but rather "recruiting opportunities." This construct represents the overall number of recruiting opportunities that took place and is the product of the number of probationers and the number of recruiting sessions that took place during the 12 months following each participant's enrollment date. Overall, the outcomes represented in the table below highlight the enormous amount of instability experienced

---

<sup>16</sup> The total number of enrollments exceeds the number of offenders treated as participants were given multiple opportunities to complete the class. This included situations where a participant was removed from the class for a "negative" reason (i.e. a new arrest). If the subject was released from custody and returned to APPD supervision, they would be allowed, during the next open recruiting period, to re-enroll.

by this subset of probationers. This contributed to the difficulties encountered during the recruitment and enrollment processes for the Life Skills intervention.

**Table 4.5: Reasons Participants Were Not Screened For Eligibility**

Reason	Event Count	Percent
Custody	401	51.4
Not in CityWide unit	142	18.2
Abscond	72	9.2
Non-reporting	49	6.3
All Cases Closed	34	4.4
Missed Appointment	26	3.3
State Parole	16	2.1
No Appointment	12	1.5
Not referred by PO	7	0.9
Deceased	6	0.8
“Walk off”	4	0.5
Inpatient Treatment	3	0.4
Detained at Appt.	2	0.3
Recruiting Closed	2	0.3
APPD Error	1	0.1
Case Vacated Appeal	1	0.1
Misc.	1	0.1
Not on Recruiting List	1	0.1
<b>Total</b>	<b>780</b>	<b>100</b>

During the 12 months following each offender’s random assignment into the experiment, 780 unsuccessful interview attempts were made. In the majority of the cases,

interviews did not take place because the individual was never present in the probation facility. The most common reason for this was due to some form of custody. Incarceration and immigration detention accounted for 400 (51.4%) of the missed recruiting incidents. This was unsurprising; APPD's high risk population has a relatively high rate of incarceration. At any given time, approximately 40% to 50% of probationers in the high-risk division are incarcerated or under some form of restrictive custody (Tudor, 2013).

Since this is a criminally active population, many of the RCT participants were sentenced for the commission of new offenses during the first year of their supervision. At the sentencing for these new offenses, some offenders received judicially-ordered supervision conditions that were not a part of their older, RCT-eligible cases. This resulted in the offender's transfer out of high-risk and prevented them from being interviewed, and enrolled, for the Life Skills course. 17.6% (137) of all missed recruiting opportunities occurred because the participant was no longer under the supervision of the CBT treatment unit. As a matter of policy, probationers in the specialized units were not permitted to participate in the Life Skills program. From an analytical perspective, these offenders, having received a wholly different supervision program than both high-risk treatment and control offenders, may weaken ability to detect treatment effects using direct, between-group comparison methods.

The large majority of the recruiting failures were out of the control of the research team. Overall, only 3.3% (25) of missed recruiting opportunities were missed due to the actions, or inactions, of APPD staff or the JLC recruiter. These errors included an officer's failure to refer an offender for recruiting (.9%) and other breakdowns of the

recruitment protocol. Surprisingly, 12 of these errors, approximately 1.5% of all missed opportunities, occurred when no appointments were scheduled during the full three weeks of the recruiting period. Generally, this should not have happened, as the high-risk protocol mandates weekly reporting. However, it could be the case that the line officers had information about the probationer's ability to report during that time that was not reflected in the monitor case management system.

Overall, these data suggest that the CBT recruiting protocol was able to maximize the number of individuals screened for enrollment into the program. High-risk offenders are a notoriously difficult population to manage. Under these constraints, common in field-based evaluations, the implementation of the recruitment process represents a significant accomplishment for the research team. However, as discussed in later sections, this relatively high level of treatment dilution poses a number of challenges when attempting to identify the relationship between participation in CBT and recidivism.

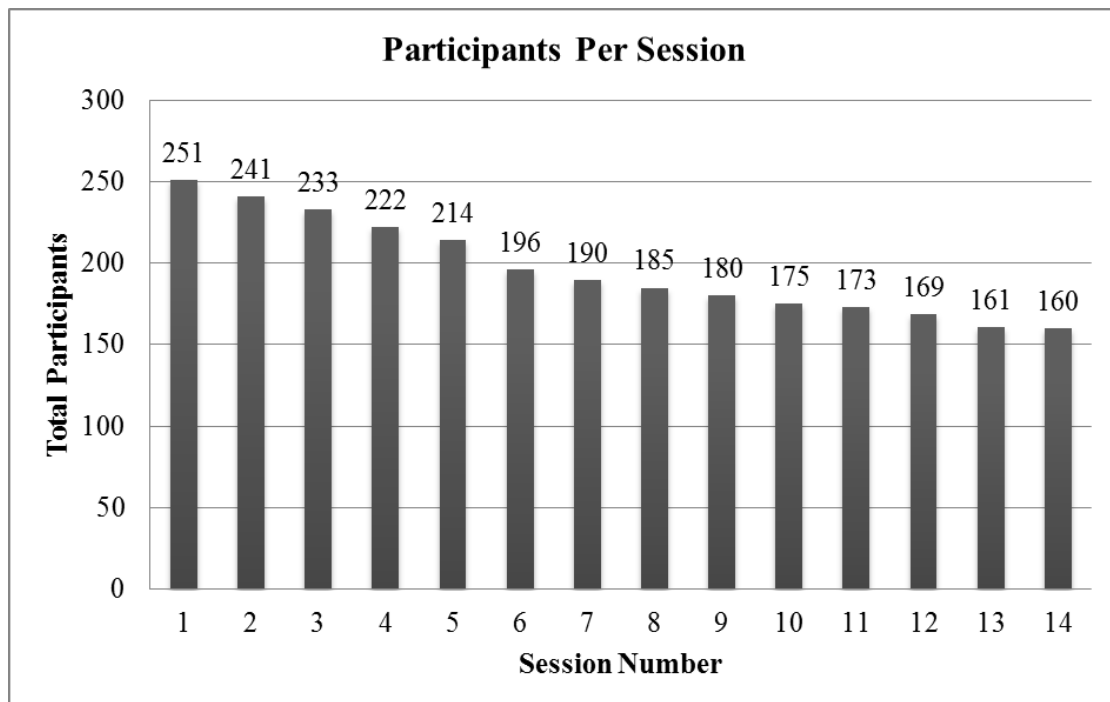
The recruiting and screening process took approximately three weeks. Active periods of enrollment were offset by almost a month due to the staggered schedule of the classes themselves. This was designed to increase the opportunities for newly enrolled probationers and those who had a change in eligibility to begin the Life Skills program as soon as was possible. For offenders who were enrolled in the trial and who attended one session of the class, the average lapse between random assignment and that class was 91.33 days. In some cases, the delay was as short as 11 days, while, in the most extreme case, the lapse was 356 days ( $sd= 76.28$ ).

## IX. Treatment Rates

Life Skills classes were run on a continuous basis throughout the evaluation period. After a 4 month “dry run”, random assignment went ‘live’ on May 1, 2010. The first CBT session that included experimental participants began on June 3, 2010. During the course of this evaluation period, 14 full CBT sessions were completed. However, as with recidivism data, only participation activities that took place within 12 months of an individual offender’s enrollment date are included in this analysis.

Treatment delivery was incomplete. As noted above, a number of offenders who were unable to participate in the Life Skills intervention for a variety of reasons, including employment and incarceration. At the conclusion of the 12 month follow-up period, 60.3% of (251) randomly assigned probationers participated in at least one Life Skills session within 12 months of their enrollment in the trial.

**Figure 4.6: Enrollment and Attrition of Life Skills Participants**



Offenders left the intervention for a number of reasons during the 14 week program. Of the offenders that began the program 160 (66.3%) met all of the intervention's requirements and graduated from the Life Skills program. Considered otherwise, 35.0% of all high- risk RCT participants who were designated to participate in the treatment program were able to complete the full curriculum within 12 months of the start of their probation case.

High risk offenders lead relatively disordered lives. Unsurprisingly, the complications that characterized the recruiting process continued throughout the program. Since it is possible for a Life Skills program participant to have been removed from multiple sessions,<sup>17</sup> Table 4.6 summarizes the total count of removal instances, not unique RCT participants. In just under one-third of all cases (32.8%), participants were removed from the program for missing too many classes. Under the high-risk protocol and Life Skills program guidelines, a participant was allowed to miss only one class. A second unexcused absence was cause for removal. Scheduling conflicts that arose when a participant secured employment or enrolled in an educational program (19.4%), a new arrest (18.3%) and technical violations (23.3%) comprised the majority of the remaining removals.

---

<sup>17</sup> In almost every case, participants who were removed from the program were, during the next open enrollment period during which they were available, targeted for recruitment. This was true for both positive reasons (i.e. securing employment) and negative (i.e. a new arrest or excessive absences).

**Table 4.6: Count of Removal Reasons**

Removal Reason	Event Count	Percent
New Arrest	33	18.3%
Excessive Absences	59	32.8%
Technical Violations	42	23.3%
Unsuitable	1	0.6%
Valid Conflict	35	19.4%
Other	10	5.6%
Total	180	100

The rate of attrition is comparable to that of similar programs. During a field-based evaluation of the R&R program, for example, approximately 60% of those assigned to the course completed the entire program (van Voorhis, Spruance, Ritchey, Listwan, & Seabrook, 2004, p. 292).

**X. Outcome measures**

The evaluation of the Life Skills program has focused on probationer's post-assignment criminal conduct. The measures used, at this stage, are all extracted or computed directly from the administrative records maintained by a variety of criminal justice agencies in Philadelphia. These include the Philadelphia Prison System (PPS), the First Judicial District (FJD), and those electronic case files maintained by APPD itself.

All criminal activity is reported at the arrest level. Conviction and sentencing data, given the influences of plea bargaining and the overall levels of case attrition in the FJD, would likely not reflect the actual distribution of offending within the sample. JLC and the APPD research team determined at the outset of the project, and as a matter of

policy, that arrests sufficiently capture underlying rates of criminal conduct.<sup>18</sup> An arrest, in itself, in accordance with APPD policy, is a violation of the general terms and conditions of probation and is classified as a technical violation. Additionally, the time delay between a new case and a trial, even for a minor offense, would prevent a vast majority of offenses from appearing in the administrative records collected during this one-year follow-up period.<sup>19</sup>

There are two general types of outcomes reported: frequency and prevalence (also known as participation). As Blumstein notes, “[p]articipation distinguishes active offenders from non-offenders within a population; frequency is a reflection of the degree of individual criminal activity by those who are active offenders” (1988, p. 4). Effective interventions should address both types or recidivism measures, though it is possible that impacts may be seen on just one of these metrics. Reductions in either area will increase public safety and contribute to the goals of probation. A decrease in the frequency of offending, for example, reduces the count of individuals victimized during a period of time, while a reduction in the proportion of a population committing criminal acts reduces the overall use of the criminal justice system in the community.

---

<sup>18</sup> The use of criminal convictions was also considered but was not employed during this stage of the analysis for two reasons. First, a criminal conviction includes the influences of the criminal justice process. Multiple factors, including plea bargaining, mandatory minimum sentences and assistance for assisting in the prosecution of co-defendants, may influence conviction rates or characteristics without regard for the actual criminal behaviors. Additionally, since the Commonwealth has, at a minimum, one year to complete the prosecution of a criminal case, the outcome variables would be censored during earlier time periods. Conviction-related outcomes, including sentencing characteristics and length, will be included in future analyses of the experiment.

<sup>19</sup> Future evaluations of this project will expand the range of inquiry to include outcome of interest to include conviction rates, sentencing outcomes and technical violations, as they become available.



## CHAPTER 5: RESULTS

This chapter begins with an explanation of the characteristics of the overall study sample. Three sets of results are reported within the subsequent sections report. For each comparison, post-randomization group equivalence and 12-month outcomes are discussed.

Outcomes regarding criminal conduct are reported for each analysis. Each set of results includes the number of overall charges filed against participants, and a categorical classification of these charges by offense type. Types of criminal charges include serious, violent, non-violent, drug and property offenses. Additional outcomes relating to the results of urinalysis screenings conducted by APPD and the number of times, and duration, of confinement in the Philadelphia County jail are also reported.

All outcomes in this analysis were calculated from data extracted at least 12 months after a participant was enrolled into the trial. As random assignment was conducted on a rolling basis, the analysis frame for each offender is not contemporaneous. However, the total amount of post-randomization analysis time remains consistent. These data were downloaded from the APPD case management system, as well as other administrative record sources within the First Judicial District on January 7, 2013. Data are right-censored for all offenders as the follow-up period was limited, by design, to the one year immediately following random assignment..

In addition to outcomes, the first two sections include a randomization check for the respective comparison groups. This brief analysis includes a post-assignment comparison of known variables that demonstrates that the assessment process was successful at “ensuring a particular probability distribution over the possible outcomes,

conditional on the truth of a given null hypothesis” (Urbach, 1985, p. 266). The administrative data available during this experiment, as well as for the creation of the random forest prediction model, allow for this to be accomplished across multiple, categorical comparisons. This is especially useful when assessing the extent to which the randomization process created treatment groups that were as similar as possible, on both measurable and unknown factors.

#### **IV. Intention to Treat (ITT) Analysis**

Intention to treat (ITT) designs are the preferred approach to analyzing most modern clinical trials (Cochrane Collaboration, 2002). Adhering to the adage “analyze as you randomize,” the ITT design requires that each case, regardless of the quality or duration of the treatment received, be included in the analysis as a member of the group to which they were assigned. A complete ITT design requires the inclusion of subjects that both fail to receive any treatment, drop out of the trial or receive an intervention other than designated through the random assignment process (Hollis & Campbell, 1999).

Under this approach, outcomes for all of the offenders assigned to the CBT unit will be compared to those in the high-risk comparison group. No distinction will be made between those probationers who participated in some, or even all, of the CBT class and those who failed to attend even one session. This is the simplest and most direct approach to the comparison.

The ITT approach is not without problems. Critics argue that this approach is too cautious and is more susceptible to Type II errors, false negatives (Sommer & Zeger, 1991). Some maintain that an efficacy approach to analysis is more valuable than the more pragmatic approach (Rubin D. , 1998). As was the case in this experiment,

randomization prior to the determination of the applicability of the intervention can create a scenario where attrition endangers the validity of an ITT analysis (Fergusson, Aaron, Guyatt, & Hebert, 2002).

Ultimately, the ITT design focuses the analysis on the impact of a department-level policy of assigning offenders to participate in the CBT program. While this is an important question, especially for practitioners seeking to understand the implications of instituting such a policy, it may fail fully to capture the impact of an intervention on participants and underreport effect sizes.

#### **A. Post-Randomization Group Comparisons**

Post-assessment randomization checks are important as they reinforce the assumption that the assignment procedure was truly random and that equivalence was achieved between the treatment and comparison groups. The randomization check should be performed at the level of randomization, rather than the level of analysis (Arceneaux, 2005). For this experiment, therefore, the appropriate unit of analysis is the individual, high-risk offender.

The overall number of participants enrolled in the treatment and comparison groups, despite the variable rate of random assignment, was roughly equal after one year. Of the 1,290 offenders enrolled in the trial, 447 were assigned to receive high-risk, intensive probation; while a distinct sample of 457 offenders were targeted with the same program with the additional cognitive-therapy component.

**i. Age**

The treatment and comparisons were equivalent on basic measures of age, both at the time of random assignment and at key points in their criminal careers. The average age of offenders, calculated on the day they began their instant probation case, was not significantly difference between those assigned to receive CBT (30.26 years old,  $sd=9.78$ ) and those in the control group (29.38 years old,  $sd= 9.48$ ) ( $p=.167$ )<sup>20</sup>. Similarly, offenders in the samples did not differ in their average ages at which the commenced offending, both for those with a juvenile record ( $p= .395$ ) or for an adult crime ( $p= .799$ ).

**ii. Race**

The post-randomization racial composition of the treatment and control groups were identical. There were no statistically significant differences in the proportion of each group that was comprised of African-Americans, Caucasians or another other racial background. All data on race was derived from the identifiers held within the First Judicial and Commonwealth databases. Ultimately, these data represent the often subjective judgments of court clerks and other agents during the criminal adjudication process. These variables, though useful for comparing the effectiveness of random assignment, fail to capture the full range of diversity and racial self-identification within the target population.

---

<sup>20</sup> Unless noted otherwise, all analyses were conducted using SPSS v. 20 (IBM Corp., 2012)

**Table 5.1: ITT Reported Race**

Race	Control	Treatment	<i>p</i>
	( <i>n</i> =447)	( <i>n</i> =457)	
Black	0.718	0.705	0.654
White	0.210	0.225	0.583
Other	0.072	0.070	0.927

Racial identification data were derived from the First Judicial District’s databases, avoiding self-identification, but not all, biases. These results were noteworthy, as they highlight the extent to which APPD’s population is disproportionately African-American, with regard to both the demographic distribution in Philadelphia County and among offenders sentenced to community corrections. Additionally, there were no missing values for racial identification variables within the sample.

**iii. Risk Scores**

The predicted risk level of participants was statistically indistinguishable between the two groups. As noted in Chapter 4, each individual risk score is a composite of 500 votes. These individual “votes” are counted in order to determine the appropriate categorical classification, in this case high, moderate or low. Quite apparently, each offender enrolled in the experiment received more high votes than any other type; this determined their final risk classification. However, the ratio of votes could vary significantly, as each of the “trees” in the forest is independent and the variables used to make decisions within those trees are selected at random (Berk R. , Forecasting methods in Crime and Justice, 2008). A comparison of the average number of each type of vote scores shows that there were no significant differences between the CBT and comparison groups. Offenders assigned to CBT had, on average, 212.8 high votes, while the

comparison group had a mean high score of 213.28 ( $p = .605$ ). Table 5.2 sets out the mean scores and significance tests for the remainder of vote comparisons.

**Table 5.2: ITT Actuarial Risk Scores**

Risk Score Profile	Control ( $n=447$ )	Treatment ( $n=457$ )	$p$
Low Votes	125.264	128.326	0.078
Moderate Votes	161.459	159.392	0.171
High Votes	213.277	212.282	0.605

#### iv. Juvenile Criminal History

Treatment and control groups were largely indistinguishable based upon juvenile offending history. Table 5.3 includes comparisons of the full range of available, juvenile-related variables. Within the total sample, 304 probationers (68%) assigned to the treatment group and 302 in the comparison group (65%) had a juvenile criminal history. Averaged across the full experimental sample, treatment group offenders had more serious juvenile charges, at 1.18, while control group participants had 0.94 charges ( $p < .10$ ). Though statistically significant, the practical implications of differences of this magnitude are limited. At traditional alpha levels, between-group differences in overall rates of offending, as well as the majority of classifications, including violent, property, drug use and drug distribution, remained insignificant.

**Table 5.3: ITT Prior Juvenile Offending History**

	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Prior Juvenile Charges			
Any Charge	9.36	9.79	.599
Serious Charges	0.94	1.18	.097
Violent Charges	2.90	3.40	.176
Sexual Charges	0.12	0.19	.238
Property Charges	2.76	2.75	.986
Weapons Charges	0.70	0.62	.512
Firearms Charges	0.61	0.50	.319
Drug Charges	1.23	1.13	.506
Drug Distribution Charges	0.49	0.47	.832

Overall, differences in the age of juvenile onset across the whole sample were insignificant ( $p < .05$ ). The age of onset for juvenile offending was also indistinguishable between the two groups within the subgroup of offenders with a juvenile record. Treatment group juvenile offenders had an average age of 14.98 years, while those in the control group had a mean age of 14.82 years old.

**v. Adult Criminal History**

The adult offending history for probationers and parolees enrolled in the experiment was statistically equivalent. As indicated in Table 5.4., there were no significant differences in the average number of prior charges, calculated across the full sample. This holds true for overall number of charges filed, as well as across each of the categorical classifications of offense type.

**Table 5.4: ITT Prior Adult Offending History**

	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Prior Adult Charges			
Any Charge	58.04	60.67	.403
Serious Charges	8.28	8.98	.228
Violent Charges	19.23	21.18	.124
Sexual Charges	0.80	0.93	.575
Property Charges	15.55	14.98	.673
Weapons Charges	5.60	6.51	.144
Firearms Charges	4.26	4.95	.181
Drug Charges	5.81	6.17	.429
Drug Distribution Charges	2.12	2.25	.524

Offenders assigned to each group had, based upon the data available, similar life course patterns of offending. Notably, there were no significant differences in the age of adult onset of offending. Treatment group offenders were, on average, 19.4 years old when charged with their first adult offense, while the mean for the comparison group was 19.3 years.

**vi. Prior Incarceration Experiences**

As indicated in Table 5.5., probationers assigned to both conditions generally had a long history of contacts with the criminal justice system. This is unsurprising, as the sample is defined by being at a high risk of committing a serious crime. These histories were statistically similar. There were no significant between-group differences in the number of times the participants had been remanded to the Philadelphia Prison System, either as part of a sentence or for pre-trial detention. Differences in incarceration histories were also not significant; members of the CBT treatment group were, on



average, incarcerated 4.82 times for 539 days, while the comparison group averaged 4.82 incarcerations totaling 518 days. Non-significant differences were also found in the prior number of judicial sentences to probation and to incarceration, as well as in the issuance of sanctions for not coming to court when ordered, also known as Failures to Appear (FTA).

**Table 5.5: ITT Prior Incarceration History**

	Control (n=447)	Treatment (n=457)	<i>p</i>
Prior Incarcerations			
Probation Sentences	2.451902	2.358862	.627
Failures To Appear	1.512304	1.654267	.445
Abscondings	0.183445	0.260394	.090
Incarcerations (count)	4.798658	4.824945	.911
Number of Days in Jail	518.12528	539.6061	.563
Judicial Sentences to Incarceration	3.387025	3.689278	.356

**vii. Instant Offense**

The two groups were comparable with regard to the characteristics of the instant offense, which is the criminal offense that resulted in the probation case under which the offender was enrolled in the trial. As was expected, there were no significant differences in the average number of charges across the majority of categorical classifications, including violent and serious offending. Though not a significant difference, and as with other comparisons, the treatment arm of the trial included individuals who, on average, exhibited slightly more serious offending patterns. The average numbers of

these charges, as well as for serious, property, firearm and drug charges, are reported in Table 5.6.

**Table 5.6: ITT Instant Offense**

Instant Offense	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Serious Charges	0.725	0.836	.267
Violent Charges	1.443	1.681	.143
Sexual Charges	0.058	0.101	.298
Property Charges	0.971	0.853	.274
Firearms Charges	0.217	0.376	.229
Drug Charges	0.774	0.702	.332

viii. **Instant Sentence**

Finally, the probationer’s instant sentences and immediate experience at the time of random assignment are set out in Table 5.7. There were no significant differences in the number of sentences to incarceration as part of the instant case, number of probation sentences, or the number of days in a row that participants were sentenced to spend in the local jail. A single case is often comprised of multiple charges, each of which can result in a distinct sanction. Therefore, on one case, a single offender can be given multiple sentences to either probation, parole, incarceration or be found not guilty. This has the potential to inflate mean levels of sanctions and make them difficult to interpret, but does not invalidate the between-group statistical tests.

**Table 5.7: ITT Instant Sanctioning Characteristics**

Instant Sanctioning	Control (n=447)	Treatment (n=457)	<i>p</i>
Number of Instant Probation Sentences	0.62	0.75	.084
Concurrent Days on Probation	369.74	380.20	.798
Number of Instant Incarceration Sentences	0.43	0.53	.132
Concurrent Days Incarcerated	187.13	216.34	.274

**ix. Experience while on Supervision**

The intensive anti-violence supervision protocol being used in Philadelphia was designed to be applied to all offenders who were assessed as high risk and who were not assigned to a Court-ordered or otherwise specialized unit. Although there were some exceptions, including the Youth Violence Reduction program (YVRP) noted previously, which resulted in the exclusion from the RCT of a small number of younger high-risk offenders from certain geographic areas of the city, the protocol was uniformly applied to participants in the experiment.

The baseline level of probation was designed to mirror the intent of other Intensive Supervisory Probation (ISP) programs developed for high-risk offenders (See Petersilia and Turner, 1990). In Philadelphia, this was characterized by a weekly

reporting schedule, weekly drug tests and a monthly field visit to the offender's home.<sup>21</sup> Quite obviously, a fully compliant group of offenders would, within the one year timeframe reported here, have 52 scheduled appointment, 52 drug tests and 12 field contacts. As with many other aspects of supervision, high-risk offenders fail to meet that benchmark. Relatively high levels of absconding, scheduled holiday breaks and a high proportion of incarcerated participants suppresses success rates. However, with the exception of the cognitive-behavioral therapy component, each of the high-risk units should have been operating under identical supervision programs and with the same logistical constraints.

The treatment and control units, when considered from a delivered, rather than intended, treatment perspective, had significantly different experiences on a number of key dimensions. Table 5.8 includes all measured comparisons of the delivery of supervision during the 12 month study period. Notably, treatment group offenders had both more contacts scheduled with their probation officers and completed more face-to-face meetings in the office. Although not a mandatory part of the supervision program, treatment condition officers also scheduled more phone calls with their clients. With the exception of field visits, both attempted and successful, all of these differences were significant at the  $p < .05$  level.

---

<sup>21</sup> There were a small number of offenders who, despite having been forecasted as "high-risk" at APPD were sentenced to less restrictive supervision programs by a judge. These sentences included, but were not limited to, non-reporting or phone-only reporting requirements. APPD, despite assigning these offenders to the Anti-Violence units, could not contravene the Order and increase supervision intensity.

**Table 5.8: ITT Supervision Contacts within 12 months**

Contact Type	Control	Treatment	<i>p</i>
	(n=447)	(n=457)	
Scheduled Office Contacts	21.47	25.22	.001
Successful Office Contacts	18.67	21.45	.005
Scheduled Phone Calls	8.45	10.56	.002
Completed Phone Calls	5.50	6.60	.007
Attempted Field Visits	8.82	9.64	.148
Successful Field Visits	5.32	5.52	.631

The variation in intensity of in-person supervision is also reflected in the rates at which CBT treatment group probationers were subjected to urinalysis screenings. On average, treatment group offenders were given an average of 8.57 tests per year, while comparison group participants were given only 6.60 screenings ( $p=.000$ ). It is also worth noting that, though this difference in urinalysis testing rates is significant, the disparity is only 3 tests over a span of 12 months. Interestingly, while a higher proportion of the treatment group (81.8%) than the control group (79.8%) were screened at least one time, the difference is not significant ( $p=.452$ ). This suggests that, since incarceration rates<sup>22</sup> were equal over the same period, the treatment group participants who were “on the street” and reporting to APPD were subject to a more rigorous and frequent drug testing schedule than their counterparts in the other units.

---

<sup>22</sup> In addition to incarceration rates, the distribution of absconding and transfers out of the RCT-assigned units was consistent between the two groups. Since there was only a single urinalysis station available for the duration of the experiment, there should have been little variation in the type and subjective experience of participating in the screening process.

The differential rates of reporting, as reflected in these data, could be due to the implementation and data management strategies employed during the evaluation. The CBT component of the evaluation, due to its classroom-based nature, required more intensive record keeping, including taking regular attendance and ensuring that participants received the appropriate forms from their primary officer. Since all case management takes place within the same computer system, classroom involvement may have increased the extent to which successful meetings or classes were recorded in only the treatment unit.

Similarly, the CBT classroom was on the same floor as the urinalysis lab and class participants were given time in the middle of the session to complete the test, encouraging compliance. Additionally, due to issues of capacity unrelated to this project, the urinalysis lab was often closed for several hours and on a daily basis. CBT-enrolled offenders had multiple opportunities, given the length of their appointments and scheduled breaks in the curriculum, to visit the lab. Non-participating offenders were simply given a form and told to report to the lab at the conclusion of their meetings. If the lab was closed at that time, they were often not required to submit a sample at that meeting and could depart APPD. The lack of post-urinalysis reporting may also have allowed probationers to leave without completing a test at all, reducing the number of tests in the control condition.

The differences could also be attributed to the culture within each of the units and the extent to which line officers received direct oversight and monitoring from their Supervisors. For most of the trial, a single supervisor was responsible for the treatment unit, while three different supervisions oversaw the geographically-organized units that

made up the comparison group. Although the CBT unit supervisor was promoted in April of 2011, he was a part of the APPD-JLC oversight team and regularly participated in meetings regarding experimental design and treatment delivery. In addition to a strong, 'by-the-book' mentality, he was aware of the need to adhere to the protocols in a manner not replicated by his peers. Due to the double-blinding mechanism, the supervisors of the control group offenders were not aware that their units were involved in the project. This too may have disproportionately encouraged compliance in the treatment unit.

Data regarding scheduled meetings, of all types, and success rates were extracted from APPD's administrative records. These data were entered by the officers who, in many cases, did not use the mandated classifications, selected from drop-down menus when entering appointments into the database. This was verified by researchers during the experiment. However, officers in the treatment unit were aware that their probationers were participants in an experimental evaluation and that their files were being reviewed regularly by the research team. The supervisor assigned to that unit was also included in planning and status meetings and stressed, on multiple occasions, the importance of consistent record keeping to his officers. The comparison group, having been blinded to both officers and APPD staff, had no information regarding the participation of offenders on those caseloads and, in fact, may not have even been aware of the project at all. The lack of reinforcement of record keeping and data entry for those officers may potentially account for the significant differences in the types and outcomes of contacts reported above.

## **B. Results: Intention to Treat**

Successful randomization and relatively complete and reliable data allow for the confident comparison of a number of different outcomes using independent samples t-tests. Below, overall rates of offending, in both frequency and prevalence, are compared between the full control and treatment groups. Crime-related outcomes are compared across a number of categories, as well as differences in absconding and drug test results after 12 months.

### **i. Absconding**

An offender, under the high-risk protocol, was considered to have absconded from supervision when he missed two consecutive, scheduled appointments and was not reachable by their probation officer. A warning letter, most often sent to the offender's address of record after the second missed appointment, preceded a change in status. At that time, a warrant was issued for the probationer's arrest for failing to report in accordance with the APPD rules and the case was transferred to the Operations unit for management by the Warrant Squad. This transfer was the triggering event for classifying an offender as having absconded during the evaluation. For measures of prevalence, the percentage reported reflects the proportion of each group that was charged with at least one absconding event during the follow up period.

Assignment to cognitive-behavioral therapy did not result in a significant reduction in absconding during the first 12-months after random assignment. Offenders assigned to the treatment condition absconded, on average, .36 times, while control group offenders averaged the exact same number of events ( $p < .05$ ). Similarly, an almost equal



proportion of probationers in each group were charged with a post-random assignment absconding event, as reported in Table 5.9.

**Table 5.9: ITT Results, Absconding**

	Control (n=447)	Treatment (n=457)	p
Percent Absconded	0.27	0.26	0.617
Number of Absconding Incidents	0.36	0.36	0.939

**ii. Drug Use**

Abstinence from the use of controlled substances is a key component of the anti-violence supervision strategy. As noted above, treatment group offenders were screened even more frequently. The analysis below reports the prevalence and frequency of positive urinalysis screenings. Each screening is comprised of tests for multiple controlled substances. These drugs include: alcohol, cocaine, marijuana, benzodiazepine, methamphetamines and phencyclidine (PCP). Included in this analysis are the results for overall drug use, marijuana and PCP. Marijuana was selected as it is the most frequently abused substance in the panel that APPD considers to be a violation of probation.<sup>23</sup> PCP is included because APPD considers PCP use to be a serious offense that often is a correlate or predicate to violence. Accordingly, there is a “zero-tolerance” policy for

---

<sup>23</sup> Alcohol is not included for two primary reasons. First, the sensitivity of the test employed is relatively low and alcohol metabolites are not detectable several hours after ingestion. More importantly, all of the offenders enrolled in the RCT are under *Adult* Probation’s supervision and are past the legal age to consume alcohol. Therefore, even if a positive test was returned, the probationer would have not broken a law and, accordingly, would not have violated the conditions of their probation or parole.

PCP use; all offenders testing positive for PCP should be taken into custody and brought before a judge for a violation hearing.

A policy of attempting to deliver Cognitive-Behavioral Therapy did have a significant impact on the number of positive drug tests returned during the first 12 months after random assignment. However, the direction of these results was not as was expected. As indicated in Table 5.10, treatment group offenders, on average had approximately 1 more positive test for any controlled substance ( $p < .000$ ).

Though not significant, the average number of positive screenings for PCP was less within the treatment group.

**Table 5.10: ITT Results, Count of Positive Urinalysis Screenings**

Number of Tests	Control ( $n=447$ )	Treatment ( $n=457$ )	$p$
Any Positive Test	1.94	2.93	.000
Positive Marijuana Test	1.22	1.79	.004
Positive PCP Test	.26	.23	.653

Treatment group offenders were, as described above, subject to a more frequent drug testing regimen. It not surprising, that offenders who are screened more often will have more positive tests. Quite simply, assuming an equal distribution of drug use across the entire sample, the treatment group was given more opportunities within the same time period to provide a urine sample that tested positive for a controlled substance. As Table 5.11 indicates, the prevalence numbers were less discouraging. The percentage of each group that tested positive, at least once, for each of the measured substances was equal after one year. This suggests that there may be an equivalent number of drug-involved

participants in each group of the experiment and that the increase in the frequency of positive tests is an artifact of the differential rates of testing.

**Table 5.11: ITT Results, Prevalence of Positive Urinalysis Screenings**

	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Percent with any Positive Test	54.1%	59.5%	.103
Percent w/ any Positive Marijuana Test	37.1%	40.7%	.272
Percent w/ any Positive PCP Test	11.4%	9.0%	.226

The results of the drug screening process are not all discouraging. Conceptualizing the dependent variable as the proportion of all drug tests that were positive avoids the complications created by the differential testing rates. Additionally, this measure captures any possible changes in the regularity of drug use that the binary prevalence measures do not. In this regard, the Life Skills intervention had a significant impact on the rate of PCP positive tests. For treatment group offenders, 2.8% of all PCP tests were positive, while 5.1% of the control group tests were positive for PCP use ( $p=.03$ ). Finally, the differential proportions of overall positive tests and positive screenings for marijuana were not significantly different between the two groups, as indicated in Figure 5.12.

**Table 5.12: ITT Results, Proportion of Screenings with a Positive Result**

Type of Test	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Any Test	33.5%	37.0%	.172
Marijuana Test	20.7%	22.5%	.437
PCP Test	5.1%	2.9%	.030

**iii. Offending**

The probationers and parolees enrolled in the experiment were, based on the statistical forecasting model, likely to commit a serious offense within the first two years of their supervision. Criminal activity and offending are captured here as the number of new charges filed against probationers enrolled in the study for which the offense date falls within the 12 month period following their enrollment in the study.

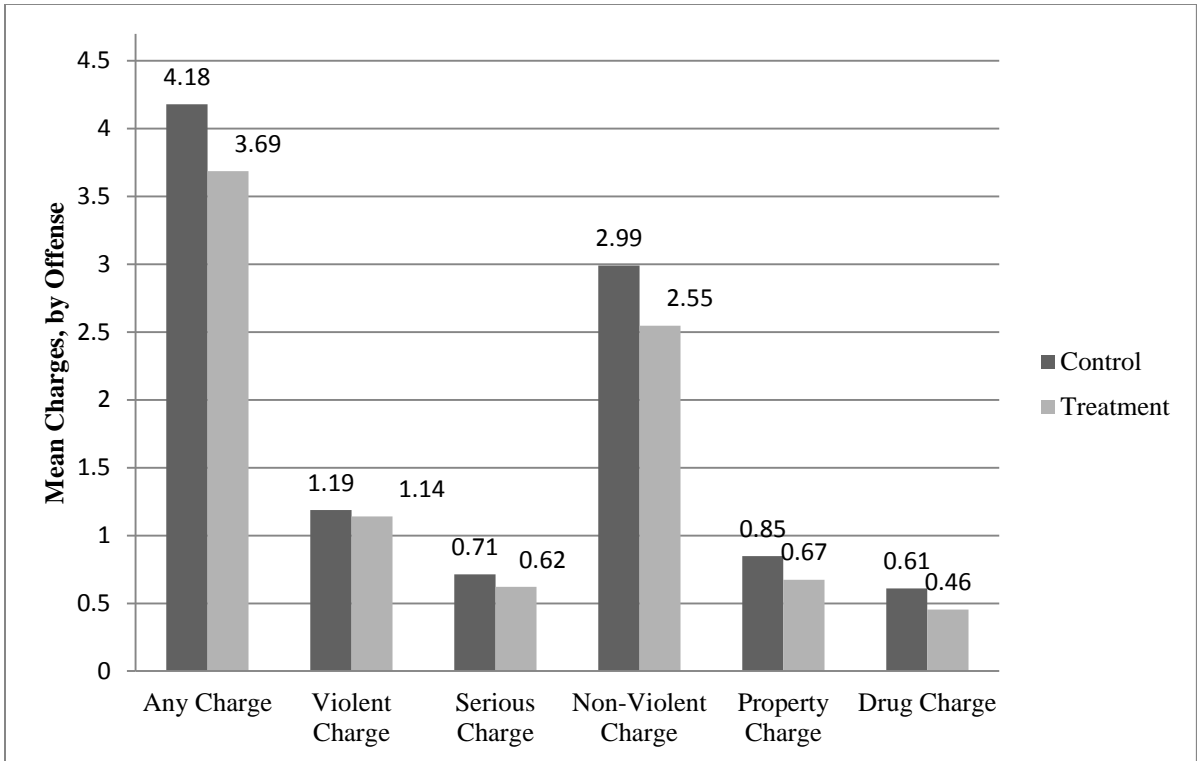
A policy of delivering Cognitive-Behavioral Therapy, after one year of post-random assignment conduct, had a mixed effect on criminal offending. When considering the frequency of offending, that is the average number of charges committed by offenders in each group, treatment group participants had lower numbers of any type of charge, violent charges, serious charges, non-violent charges, property charges and drug charges filed against them. However, as indicated in Table 5.13, none of these differences were statistically significant. The consistent direction of the results, all of which favor the treatment program, is slightly encouraging.

**Table 5.13: ITT Results, Average Number of Charges**

Charge Type	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Any Charge	4.18	3.69	.509
Violent Charge	1.19	1.14	.880
Serious Charge	0.71	0.62	.658
Non-Violent Charge	2.99	2.55	.344
Property Charge	0.85	0.67	.333
Drug Charge	0.61	0.46	.154

Although these comparisons fail to reach significance after 12 months of observed conduct, treatment group members had fewer average charges than their control group counterparts, as shown in Figure 5.1. Though the mean differences are, in many cases, slight, the consistency of the results suggests that future follow-up analyses may be warranted.

**Figure 5.1: ITT Results, Average Number of Charges**



Despite having been designed as an explicit, violence prevention program, a policy of delivering the intervention did not result in any significant differences in violence related outcomes after 12 months. In fact, of all frequency and prevalence comparisons, only when comparing the percentage of each group charged with a violent crime does the treatment group perform slightly worse, as shown in Table 5.14.

**Table 5.14: ITT Results, Proportion of Each Group Charged, by Offense**

Charge Type	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =457)	<i>p</i>
Any Charge	40.5%	33.9%	.041
Violent Charge	13.4%	13.8%	.874
Serious Charge	10.3%	11.6%	.530
Non-Violent Charge	40.5%	33.9%	.041
Property Charge	16.6%	15.1%	.549
Drug Charge	16.1%	13.3%	.242

There were some significant and meaningful differences within the remaining measures of the prevalence of offending. Fewer offenders assigned to the treatment group (33.9%) than comparison (40.5%) were charged with an offense of any kind ( $p=.041$ ). Assignment to the Life Skills program caused a 7.5% decrease in the number of offenders committing non-violent crimes. During the same period, there was also an identical, significant reduction in the overall percentage of offenders in the CBT unit who were charged with a non-violent offense. Although, in all likelihood, the reduction in overall offending is driven almost entirely by the decrease in non-violent offending, this represents a positive impact of the interventional policy. At the same time, the majority of the other comparisons, with the notable exception discussed above, uniformly favor offenders in the CBT condition.

#### **iv. Incarceration**

Within one year of each probationer's random assignment date, there were no significant differences between the treatment and control groups with regard to incarceration in Philadelphia's local jail system. Though the treatment group spent on

average just over two less days incarcerated in Philadelphia’s prison system, the difference was not statistically significant. Similarly, the treatment (.98) and control (.96) had, on average, almost the exact same number of incidents of incarceration per person ( $p=.816$ ).

**Table 5.15: ITT Results, Incarceration Characteristics**

Incarceration Type	Control ( $n=447$ )	Treatment ( $n=457$ )	$p$
Proportion Incarcerated	67.6%	65.9%	.589
Average Number of Distinct Incarcerations	0.969	0.982	.816
Days Spent in Jail	87.19	84.96	.747
Average Jail Stay Length	69.27	63.17	.306

**v. Time to Failure**

Offenders assigned to the CBT treatment group, on average, demonstrated some significant differences with regard to their time to failure. In this regard, time to failure is defined as the number of days that passed after an individual’s random assignment into the evaluation and the first instance on which they were charged with an offense or violation. Between-group differences for the survival analysis are reported in Table 5.16.

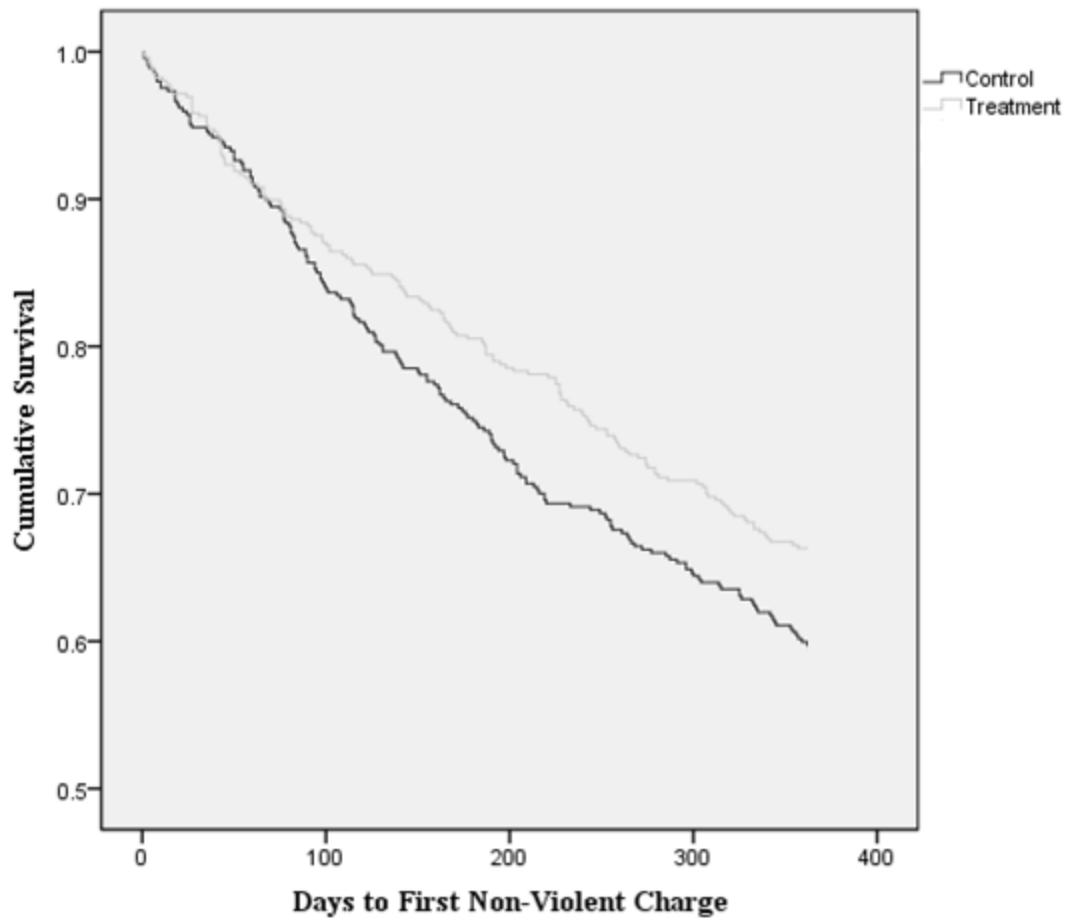


**Table 5.16: ITT Results, Cumulative Probability of Failure (Log Rank Test)**

	$\chi^2$	$df$	$p$
First Absconding Event	.303	1	.582
First Incarceration	.075	1	.784
First Charge of Any Type	4.341	1	.037
First Serious Charge	.382	1	.536
First Non-Violent Charge	4.293	1	.038
First Violent Charge	.025	1	.873
First Property Charge	.391	1	.532
First Drug Charge	1.451	1	.228

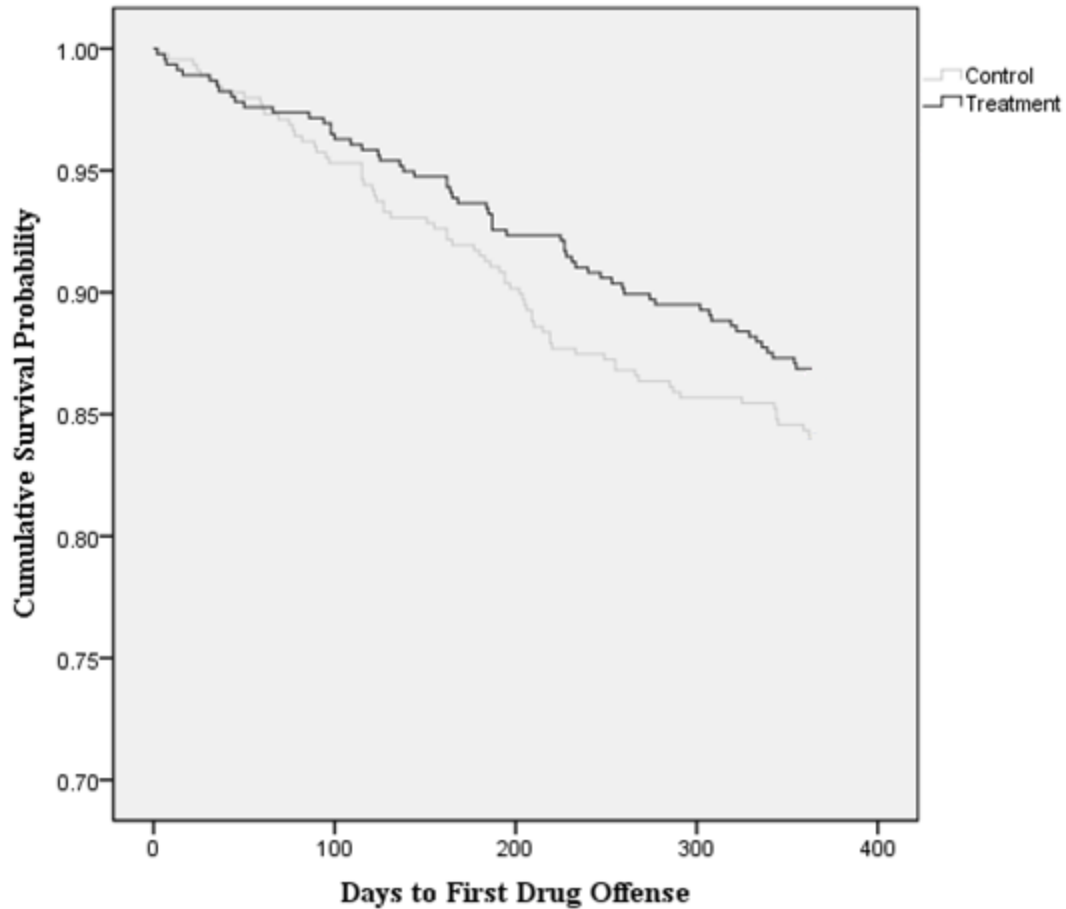
Assignment to the Life Skills intervention caused a reduction in the survival rates (that is, a lack of failure through re-arrest) for non-violent offending. On average, treatment group offenders were charged with a non-violent crime after 294.7 days, while comparison group participants were charged 15.4 days sooner, or after 279.3 days. This difference is significant using the Log Rank (Mantel-Cox) test ( $p=.038$ ). This statistic tests the null hypothesis that the population survival curves, that is the cumulative probability of failure occurring at any time point, are drawn from the same distribution. Additionally, and as was the case with overall offending, similar survival functions, also significantly different, are observed when considering the time to failure overall offending rates ( $p=.037$ ). Analyses for serious ( $p=.536$ ), violent ( $p=.873$ ), drug ( $p=.228$ ) and property ( $p=.532$ ) offending, as well for incarceration ( $p=.784$ ) and absconding ( $p=.582$ ) failed to reach significance.

**Figure 5.2: ITT Results, Survival Functions for Non-Violent Offending**



The survival functions for non-violent offending illustrates the relationship between CBT intervention and non-violent. As shown in Figure 5.2, both treatment and control groups had nearly identical survival functions for the first three months of the observation periods and then begin to diverge (and never cross again) after that point. As noted in Chapter 3, the average time to enrollment into the intervention was nearly simultaneous to this diversion point, occurring approximately 91 days after random assignment.

**Figure 5.3: ITT Results, Survival Functions for Drug Offending**



Though failing to reach significance within 12 months, plotting the cumulative probability of failure for drug offending does offer some encouragement for future waves of analysis. Figure 5.3 illustrates these relationships. As was the case with non-violent offending, the survival curves for the treatment and control group are nearly identical for the first 100 days of the observation period. After that point the probability of failure remains consistently lower for the treatment group, the two survival functions do not

cross. It may be that case that, in future analyses, these differences may reach significance.

As detailed in Table 15.17, within a subgroup limited to offenders who were charged with certain offenses, treatment group participants generally took longer to recidivate than the comparison group. Though encouraging, and not causal, the differential delays remain relatively small after 12 months of follow-up.

**Table 5.17: ITT Results, Average Time to First Incident Among Failures**

Type of Failure	Control ( <i>n</i> )	Treatment ( <i>n</i> )
First Absconding Event	134.31 (122)	145.75 (118)
First Serious Charge	143.67 (46)	151.26 (53)
First Non-Violent Charge	153.36 (181)	157.82 (153)
First Incarceration	145.39 (233)	141.39 (242)
First Charge of Any Type	152.54 (181)	157.82 (155)
First Violent Charge	146.37 (60)	145.78 (63)
First Property Charge	148.24 (74)	161.36 (69)
First Drug Charge	169.06 (72)	182.87 (61)

### C. Summary

A policy of delivering (or attempting to deliver) Cognitive-Behavioral Therapy to high-risk probationers resulted in some reductions in offending characteristics within 12 months. Notably, a lower percentage of offenders who were targeted with the intervention committed a non-violent crime than the comparison group. Additionally, the rate at which urinalysis screenings were positive for PCP was lower, suggesting that the

program may have had an impact of participant's long-term patterns of drug use. Finally, assignment to the Life Skills program resulted in a significant increase in the time to rearrest for non-violent offenses.

It is worth noting that, while the overall differences in mean levels of frequency and prevalence are small, and some do not reach statistical significance, the direction of the majority of the comparisons reported consistently favors the treated group. As discussed previously in Chapter 3, relatively low levels of treatment delivery may suppress the effects of the program on those that received it. The analytical framework employed in this section, the traditional Intention to Treat (ITT) design, considers the offending of all probationers assigned to the treatment arm of the program, regardless of whether or not they ever participated in the CBT class. A consideration of other analytical approaches is necessary to fully parse out the impact of the treatment program on those probationers who participated in the program.

## V. Treatment on the Treated (TOT) Analysis

Delivering treatment to all of the participants designated to receive it can be difficult under the best of circumstances; working with high-risk offenders in the community is far from ideal. As discussed in Chapter 3, almost 30% of the sample that was assigned to participate in the CBT intervention did not, during the entire evaluation period, have contact with the program recruiter. In some cases, and often due to incarceration, certain offenders did not have a single successful contact with the Department at all during the evaluation period. This is not uncommon in field experiments; many clinical trials include patients who fail to adhere to their assigned therapy. These losses threaten the generalizability of the conclusions (Schulz & Grimes, 2002) and may introduce bias into the results (Hollis & Campbell, 1999). Differential treatment delivery, as was the case during this project, can also result in the underestimation of the actual treatment effect. One approach is to consider those individuals receiving the treatment as part of the experimental condition, removing those who were never eligible. Although injecting a selection bias into an experiment where great pains were taken to control for such issues, this approach is a first step towards the isolation of an actual treatment effect.

More commonly referred to as a “treatment on the treated” analysis, this approach may better approximate the effect of a fully implemented program (Bloom, 2006). This framework may indicate the impact of an intervention when program compliance or integrity is at issue. An example from medical literature illustrates these potential pitfalls. Treanor and colleagues (2000) conducted a RCT to evaluate the efficacy of an anti-influenza drug. Though 649 patients were randomized to receive the drug, 40%

were later found to not have influenza; they were clearly ineligible for the trial. An ITT analysis showed a significant, 22% reduction in infection ( $p<.004$ ). However, when the researchers reran their analysis on only those patients who should have received the drug, the results shifted to a 30% reduction ( $p<0.001$ ). This approach is appropriate here, as approximately 40% of the treatment group was unable, for a variety of reasons discussed in Chapter 3, to attend even one class session.

TOT-derived results, standing alone, are rarely convincing, and certainly not causal, evidence. Gross and Fogg (2004) suggest conducting both an ITT and TOT analysis. Further support for results is found when both analyses agree. When the results differ, both ITT and efficacy subset analyses should be conducted but the results should be compared for the high and low adherence groups in the experimental condition (Feinstein, 1991). Therefore, though this analytical framework cannot convincingly stand on its own, it does provide additional evidence on the effect of CBT on offending, especially given the relatively weak, but encouraging, findings found in the ITT analysis.

For the purposes of this TOT analysis, a group consisting of “treated” participants can be constructed in multiple ways. First, the treated group can include all offenders who were enrolled in the Life Skills program and attended at least one class. This approach is markedly more liberal than an ITT analysis, as it includes enrolled participants who dropped out of or were removed from the program. Alternately, the classification of “treated” could be limited to those offenders who completed the entire 14 course curriculum and, for the duration of the program, remained arrest and technical violation free. The analysis below employs the latter approach; only those probationers who graduated from the course are considered to have been fully “treated.”

## **A. Treatment on the Treated Sample Construction**

As noted in Chapter 4, 457 high-risk probationers were assigned to participate in the CBT program. From this group, 251 (60.3%) participated in, at a minimum, one class session and were exposed to some, limited measure of the treatment. The remaining offenders were either excused from participation due to a valid conflict (e.g. employment), never reported to probation or were arrested or absconded in the relatively short period of time between the appointment at which they were scheduled for the Life Skills class and that class date. After 12 months of post-random assignment observation 160 probationers (35%) from those assigned to receive the intervention had completed the Life Skills Program. These offenders represent the treated sub-sample.

## **B. Post-Randomization Group Comparisons**

Although the treated group is a systematically defined sub-population, it remained strikingly similar to the full control sample on a number of metrics. The same, detailed and systematic randomization checks were performed on these comparison groups as with the full, ITT sample.

### **i. Risk & Age at Assignment**

The relative risk profiles of the offenders in both groups were statistically indistinguishable. The treated group, for example, had on average 211.73 high votes, while the comparison group had 213.28 high votes ( $p=.172$ ). The groups were also indistinguishable by age at the time they began their instant case, with the CBT group averaging only a non-significant .574 years, or 209 days, older ( $p=.514$ ). The very slight



difference in age of adult onset (at 19.3 years old for the control sample and 19.2 years old for program graduates) was also failed to reach significance ( $p=.723$ ).

**Table 5.18: TOT Actuarial Risk Scores**

Risk Score Profile	Control ( $n=447$ )	Treatment ( $n=160$ )	$p$
Votes Low	125.26	128.52	.172
Votes Moderate	161.46	159.76	.422
Votes High	213.28	211.73	.581

**ii. Offending History and Onset**

Criminal histories, on both the adult and juvenile levels, were also markedly similar. Within the treated group, 111 offenders had a juvenile record, while 304 probationers in the comparison group had at least one recorded juvenile offense. Across the full TOT sample, the total number of prior juvenile charges, violent charges and serious charges were statistically indistinguishable. As indicated in Table 5.19, across the remaining variables measuring pre-random assignment juvenile conduct there were no significant differences in juvenile offending histories.

**Table 5.19: TOT Prior Juvenile Offending History**

Prior Juvenile Charges	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Any Charge	9.36	9.83	.650
Serious Charges	.94	1.09	.336
Violent Charges	2.90	3.46	.207
Sexual Charges	.12	.19	.387
Property Charges	2.76	2.66	.826
Weapons Charges	.70	.66	.810
Firearms Charges	.61	.53	.586
Drug Charges	1.23	1.24	.991
Drug Distribution Charges	.49	.56	.596

Overall, the adult offending histories of the two groups were generally indistinguishable. Of all the categorical predictors, only the count of prior property offenses was significantly different. In that case, the control group, on average, committed 4.45 more of these offenses ( $p = .004$ ). Table 5.20 includes the comparisons for the remainder of the prior adult offenses.

**Table 5.20: TOT Prior Adult Offending**

Prior Adult Charges	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Any Charge	58.04	50.29	.066
Serious Charges	8.28	7.07	.103
Violent Charges	19.23	17.53	.299
Sexual Charges	.80	.54	.344
Property Charges	15.55	11.10	.004
Weapons Charges	5.60	5.45	.844
Firearms Charges	4.26	4.07	.754
Drug Charges	5.81	6.15	.576
Drug Distribution Charges	2.12	2.34	.463

Overall, program graduates had fewer overall offenses in their criminal histories, but had more drug-based offenses, though none of these differences reached statistical significance.

Probationers assigned to both groups began their criminal activity at approximately the same age. As shown in Table 5.21, the ages at which first adult and juvenile charges of any type were filed are also similar. The same holds true for the ages at which the first violent charges, as both an adult and a juvenile, were filed against members of each group

**Table 5.21: TOT Ages of Onset for Any Charge and for Violent Charges**

	Control ( <i>n</i> )	Treatment ( <i>n</i> )	<i>p</i>
Age at First Juvenile Charge	14.83 (304)	14.84 (111)	.974
Age at First Juvenile Violent Charge	14.98 (205)	14.90 (79)	.757
Age at First Adult Charge	19.39 (447)	19.29 (160)	.725
Age at First Adult Violent Charge	20.68 (414)	20.49 (140)	.623

**iii. Prior Sanctioning History**

Much like their criminal records, participants in the two groups had similar prior incarceration and sanctioning histories. As noted in Table 5.22, treated participants had been, on average, sentenced to any type of incarceration prior to the experiment less often. The difference of .968 sentences was significant ( $p = .005$ ). At traditional alpha levels, the remainder of the differences failed to reach significance.

**Table 5.22: TOT Prior Sanctioning History**

Prior Sanction	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Probation Sentences	2.45	2.02	.082
Failures To Appear	1.51	1.18	.084
Abscondings	.18	.23	.428
Incarcerations (count)	4.80	4.06	.012
Number of Days in Jail	518.13	441.14	.124
Judicial Sentences to Incarceration	3.39	2.42	.005

**iv. Instant Offense and Sentence Characteristics**

The instant offenses, the set of charges associated with the probation case that enrolled the offender in the trial, were statistically similar between the groups. As shown in Table 5.23, treated group offenders began the experiment on cases with slightly more serious, violent, firearms and drug charges.

**Table 5.23: TOT Instant Offense**

Offense Type	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Serious Charges	.72	.83	.459
Violent Charges	1.44	1.63	.386
Sexual Charges	.06	.08	.739
Property Charges	.97	.76	.148
Firearms Charges	.22	.28	.496
Drug Charges	.77	.89	.262

The similarities with regard to charges on the instant case translated to a similarity in the sentences on those dockets. There were no measured, significant differences between the sentences given to the group of treated offenders and the full comparison group. Table 5.24 sets out mean values and significance tests for these comparisons; differences were not significant.

**Table 5.24: TOT Instant Sentences and Sentence Characteristics**

Sentence Type	Control	Treatment	<i>p</i>
	( <i>n</i> =447)	( <i>n</i> =160)	
Number of Instant Probation Sentences	.62	.63	.938
Concurrent Days on Probation	369.74	307.06	.264
Number of Instant Incarceration Sentences	.43	.47	.659
Concurrent Days Incarcerated	187.13	157.13	.342

**v. Experience while on Supervision**

The group of offenders that were both enrolled in and completed the CBT program had significantly different experiences on supervision than those in the comparison group. This is unsurprising, as the sample identification process removed, from only the treatment group, those offenders who did not report to the probation department for the entire duration of the evaluation period, as well as those who failed to comply with APPD’s rules. Whether due to absconding or incarceration, those removed offenders had almost no appointments, drawing down the mean number of contacts within only the comparison group. Additionally, since the treated group is characterized

by higher levels of compliance, the successful contact rates for those participants in the treated group is even higher than those within the larger ITT sample.

Overall, treated group participants had an average of 69.64 scheduled contacts, 78% of which were successful (54.33). On the other hand, offenders in the control group were scheduled for 39.22 contacts of any kind, 76.3% (29.95) of which were successful ( $p = .000$ ). Table 5.25 breaks down these contact rates by type, including in-office meetings, phone calls and targeted home visits. Each of these differences, as would be expected, is significant.

**Table 5.25: TOT Supervision Contacts within 12 months**

	Control ( $n=447$ )	Treatment ( $n=160$ )	$p$
Scheduled Office Contacts	21.47	40.01	.000
Successful Office Contacts	18.67	36.03	.000
Scheduled Phone Calls	8.45	12.41	.000
Completed Phone Calls	5.50	8.11	.000
Attempted Field Visits	8.82	16.98	.000
Successful Field Visits	5.32	9.94	.000

### C. Results: Treatment on the Treated

Below, overall rates of offending, in both frequency and prevalence, are compared between graduates of the Life Skills program and the full comparison sample. These results compare outcomes across a number of categories, including for absconding and drug test results.

**i. Absconding**

Graduating from the Life Skills intervention was associated with a significant reduction in absconding during the first 12-months after random assignment. The Life Skills group had .16 incidents per offender, while the comparison group averaged .36 incidents ( $p < .000$ ). At the same time, only 13.75% of the treatment group had at least one absconding incident, while 27.29% of the comparison group had an incident during the follow-up period ( $p < .000$ ).

**Table 5.26: TOT Results, Absconding**

	Control ( $n=447$ )	Treatment ( $n=160$ )	$p$
Percent Absconded	27.29%	13.75%	.000
Number of Absconding Incidents	.36	.16	.000

**ii. Drug Use**

As was the case with the full treatment sample, graduates of the CBT program had higher rates of post-urinalysis screenings. As shown in Table 5.27., these results are unsurprising. Additionally, the sub-sample of treated offenders is *only* comprised of individuals who attended all requisite Life Skills sessions, each of which included time to complete the urinalysis screening. The average program graduate was screened 5.8 times more per year than the mean number of tests in the full comparison sample (8.57 tests per year).



**Table 5.27: TOT Results, Urinalysis Screening Rates**

	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Number of Drug Tests	6.61	14.41	.000
Percent with any Drug Test	79.9%	99.4%	.000

Since the treated group remained active in their assigned high-risk supervision unit for longer and so received more regular urinalysis screenings, the overall higher numbers of positive drug tests is not surprising. This discrepancy may be, in fact, magnifying the significant between-group difference noted in the ITT analysis. Program graduates, on average, had 2.1 more positive tests overall and 1.5 more positive marijuana tests than the full comparison sample. The mean number of PCP positive tests, however, was nearly identical. Counts of positive drug tests were also, unsurprisingly, higher than those for all positive tests and positive marijuana tests reported for the full ITT sample (2.93 and 1.79 tests, respectively).

**Table 5.28: TOT Results, Count of Positive Urinalysis Screenings**

Test Type	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Any Positive Test	1.94	4.14	.000
Positive Marijuana Tests	1.22	2.73	.000
Positive PCP Tests	.26	.26	.972

The prevalence of positive drug tests was also distinctly different between the graduated sub-group and the comparison group. As noted in Table 5.28, 16.4% more of

the treated group had at least one positive test ( $p < .000$ ) and 12.2% more positive marijuana tests ( $p < .05$ ). The prevalence of PCP positive tests was not significantly different.

**Table 5.29: TOT Results, Prevalence of Positive Urinalysis Screenings**

Test Type	Control ( $n=447$ )	Treatment ( $n=160$ )	$p$
Any Positive Test	54.1%	70.6%	.000
Positive Marijuana Tests	37.1%	49.4%	.008
Positive PCP Tests	11.4%	12.5%	.713

Completion of the full Life Skills intervention correlated with a change in the rate of positive tests. Though an imperfect measure, the difference in rates suggests that some component of the program (or the associated characteristics of supervision) is related to a decrease in drug use over time. As Table 5.29 indicates, differences were significant for overall positive tests ( $p < .1$ ) and for PCP positive tests ( $p < .05$ ).

**Table 5.30: TOT Results, Proportion of Screenings with a Positive Result**

Test Type	Control ( $n=447$ )	Treatment ( $n=160$ )	$p$
Proportion of All Tests Positive	33.5%	26.1%	.010
Proportion of Tests Positive for Marijuana	20.7%	17.0%	.149
Proportion of Tests Positive for PCP	5.1%	1.7%	.001

### iii. Offending

Graduates of the Life Skills program, when compared to the full control sample, committed significantly fewer crimes within the 12 months after their random assignment into the experiment. These differences, persisting across categorical outcomes, as well as measures of frequency and prevalence provide evidence, albeit weak, of the intervention's crime reduction effects.

Program graduates were charged with significantly less offenses, on average, than their counterparts in the comparison group. A typical offender who completed the program had 1.54 charges, of any kind, filed against them. Control group offenders, on the other hand, were charged with 4.18 offenses ( $p=.007$ ). Notably, graduates were charged with, on average, 1.9 fewer non-violent crimes ( $p=.001$ ) and .7 fewer violent charges ( $p=.094$ ). Full comparisons can be found in Table 5.30.

**Table 5.31: TOT Results, Average Number of Charges**

Charge Type	Control ( $n=447$ )	Treatment ( $n=160$ )	$p$
Any Charge	4.18	1.54	.007
Violent Charge	1.19	.46	.094
Serious Charge	.71	.31	.178
Non-Violent Charge	2.99	1.09	.001
Property Charge	.85	.31	.026
Drug Charge	.61	.26	.014

These effects largely hold up when considering the prevalence of offending in the comparison groups, with the exception of drug offenses. Importantly, there were statistically significant differences in the proportion of each group charged with a crime

overall ( $p=.001$ ), as well as across both violent ( $p=.027$ ), serious ( $p=.044$ ) and non-violent ( $p=.001$ ) offenses.

**Table 5.32: TOT Results, Proportion of Each Group Charged, by Offense**

Charge Type	Control	Treatment	<i>p</i>
	( <i>n</i> =447)	( <i>n</i> =160)	
Any Charge	40.5%	26.3%	.001
Violent Charge	13.4%	6.9%	.027
Serious Charge	10.3%	5.0%	.044
Non-Violent Charge	40.5%	26.3%	.001
Property Charge	16.6%	10.6%	.072
Drug Charge	16.1%	11.3%	.138

As was the case with the frequency comparison, the majority of the comparisons reflected a consistent difference in offending characteristics that almost exclusively favored the program graduates.

**iv. Incarceration**

Graduates of the Life Skills session exhibited lower rates of post-randomization incarceration. Offenders in the treated group, on average, spent 46 fewer days incarcerated in the local jail system ( $p=.000$ ) and entered the prison system on .2 fewer occasions ( $p=.000$ ). Just over 18% less of the treated group had any contact with the correctional system ( $p=.000$ ).

**Table 5.33: TOT Results, Incarceration Length and Characteristics**

	Control ( <i>n</i> =447)	Treatment ( <i>n</i> =160)	<i>p</i>
Proportion Incarcerated	67.6%	49.4%	.000
Average Number of Distinct Incarcerations	.97	.68	.000
Days Spent in Jail	87.19	29.07	.000
Average Jail Stay Length	69.27	23.27	.000

Finally, the subgroup of offenders who spent any time incarcerated at the county level was shorter for program graduates. The average stay in jail for treated probationers was 55 days less, a reduction from 102.5 to 47.2 days.

**v. Time to Failure**

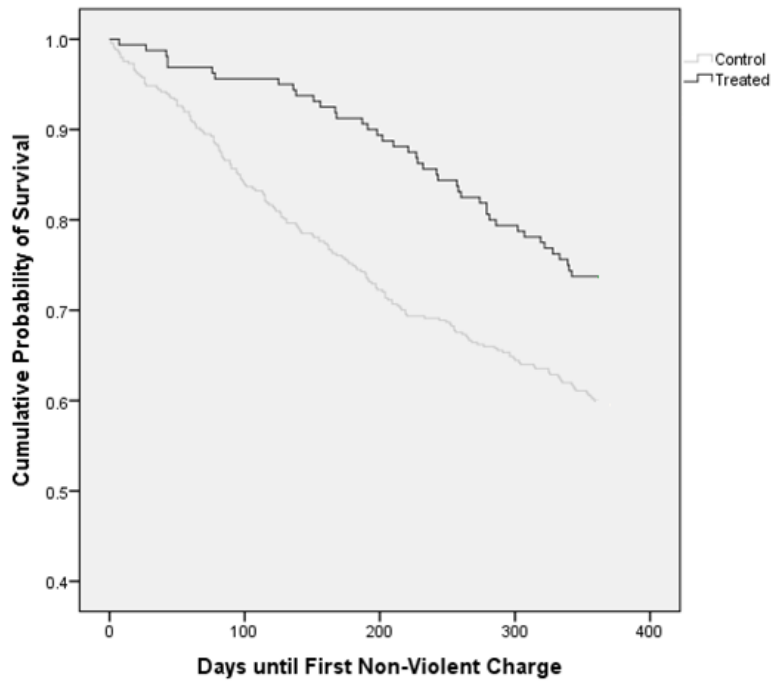
Program graduates had significantly different rates of failure within the first 12 months after random assignment. In keeping with results reported above, results of Kaplan-Meier survival analyses universally and significantly favored program graduates.

**Table 5.34: TOT Results, Cumulative Probability of Failure (Log Rank Test)**

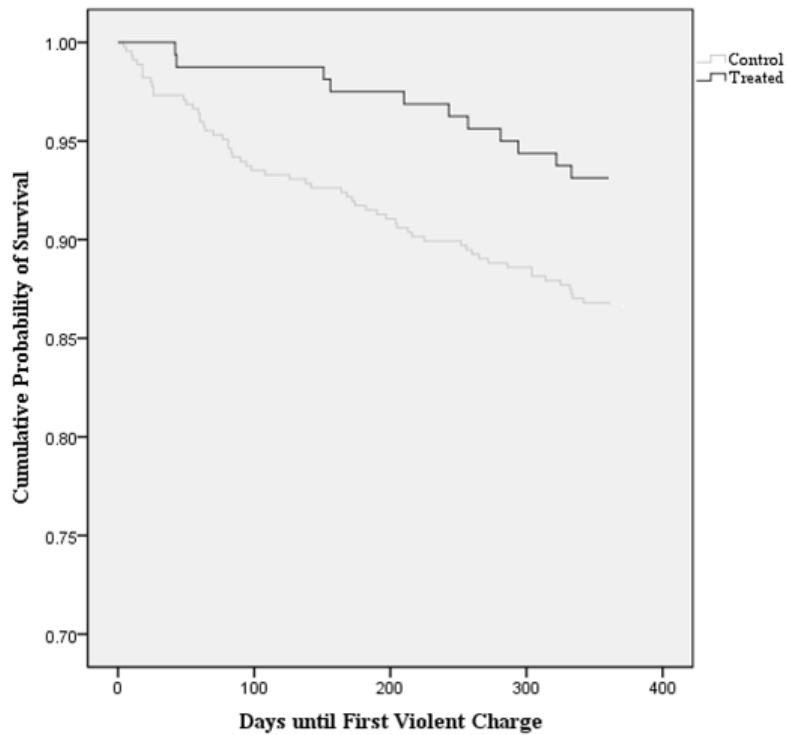
	$\chi^2$	$df$	$p$
First Absconding Event	12.852	1	.000
First Incarceration	17.701	1	.000
First Charge of Any Type	11.825	1	.001
First Serious Charge	4.208	1	.040
First Non-Violent Charge	11.810	1	.001
First Violent Charge	4.986	1	.026
First Property Charge	3.738	1	.053
First Drug Charge	2.414	1	.120

As can be seen in Figures 5.4, 5.5 and 5.6, the differences in the cumulative survival curves between Life Skills graduates and the full comparison sample are much more pronounced (as would be expected) than those in the ITT analysis.

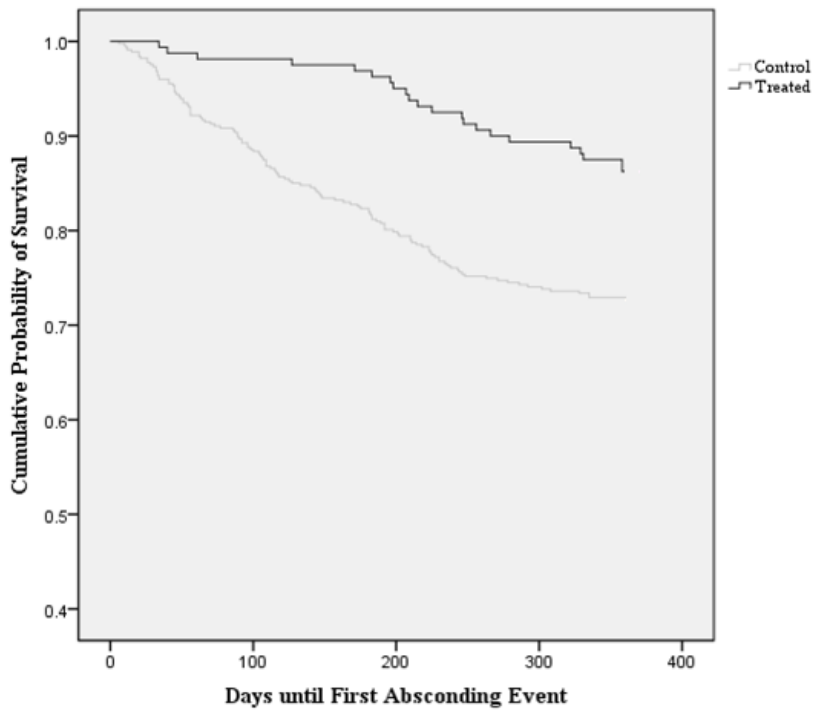
**Figure 5.4: TOT Results, Survival Functions for Non-Violent Offending**



**Figure 5.5: TOT Results, Survival Functions for Violent Offending**



**Figure 5.6: TOT Results, Survival Functions for Absconding**



With regard to the subgroup of participants who committed a new crime, the reductions in time to failure were most pronounced in overall offending, a difference of almost 107 days and in the 124 day difference in the filing of a first property charge. As Table 5.34 indicates, these differences were favored program graduates for other classifications of offending, as well as for absconding and incarceration.



**Table 5.35: TOT Results, Average Time to First Incident Among Failures**

Failure Type	Control ( <i>n</i> )	Treatment ( <i>n</i> )
First Absconding Event	134.31 (122)	220.82 (22)
First Incarceration	145.39 (233)	200.25 (67)
First Charge of Any Type	152.54 (181)	210.57 (42)
First Serious Charge	143.67 (46)	250.50 (8)
First Non-Violent Charge	153.36 (181)	210.57 (42)
First Violent Charge	146.37 (60)	212.00 (11)
First Property Charge	148.24 (74)	272.12 (17)
First Drug Charge	169.06 (72)	226.61 (18)

**IV. Summary**

After 12 months of post-random assignment observation, the Treatment on the Treated (TOT) analysis provides some additional evidence regarding the potential impact of the Life Skills intervention. There were significant reductions observed in both the frequency and prevalence of absconding. Given the increased treatment intensity that accompanied enrollment in the intervention, as well as supervision by the experimental,

CityWide unit, program graduates were screened for drug use more often. However, despite the associated increase in the raw number of positive results, a lower proportion of all drug screens and specific tests for PCP use were positive within the treated group. With regard to offending and local incarceration, reductions in frequency and prevalence were universal. Finally, when graduates of the program did commit a new offense, they took longer to do so than when compared to the full control sample.

Although the results of the Treatment on the Treated analysis are not evidence of a causal relationship, the consistent direction of these results suggests that, for those offenders who complete the program, there may be an effect on their subsequent patterns of behaviors. The relatively high levels of both implementation failure (those offenders who were never enrolled in the class) and treatment attrition (those offenders who began the program but who, after 12 months, were unable to successfully complete it) make actual treatment effects difficult.

It is possible, within a regression framework, to reapportion the effect of the program across all participants, regardless of their actual treatment status. Discussed in the next section, this approach, relying on an instrument variables (IV) approach more common in econometrics, allows for the estimation of the average treatment effect, an estimate of program effectiveness unencumbered by the implementation issues noted here.

## **VI. Instrumental Variable (IV) Analysis**

Comparing graduates of the Life Skills program to the full control sample eliminates the benefits of randomization. This is because this approach creates systematic differences between group participants. For example, program graduates were generally more compliant with the terms of probation and were under active supervision for a longer period of time. This imbalance prevents causal inference, though it may be suggestive of the direction of effects.

Experimentation and randomization can resolve many of the hurdles to isolating causal relationships, but there are still limitations. Even well designed experiments can suffer from pragmatic limitations that may inhibit clear analysis. At issue in this evaluation, and as discussed in Chapter 3, is treatment dilution. This occurs when participants who are assigned to a treatment fail to receive it during the evaluation (Gartin, 1995). Intention to Treat designs get around this limitation by ignoring whether or not treatment was actually delivered to the sample. As noted above, this is useful when assessing the impact of a policy of delivering a particular intervention to the target population. This approach predicts the outcomes when the program is delivered to a similar population, under similar constraints; however, it often underestimates the magnitude of effects. As evidenced above, simply dropping the non-compliers will bias the results in favor of the treatment (Sheiner & Rubin, 1995).

Instrumental variable analysis provides an alternative framework for causal inference that avoids some of the problems associated with the ITT approach. Effects can be estimated, in light of treatment dilution, by employing randomization as an

instrumental variable (IV). This approach fills in a gap within the experimental framework: an understanding of potential outcomes for non-treatment group participants, a necessity when attempting to support causal inferences within sub-populations (Angrist J. , 2006). Generally, an instrument is a variable that is related to treatment assignment but is not correlated with the outcome variable conditional on the other covariates. An instrument must have a causal effect on an intermediate variable (such as treatment) in the causal chain. In this case, the intermediate variable is receipt of the Life Skills intervention, which was only possible for offenders randomized into the treatment group. Assumptions about monotonicity are also satisfied through limitations on group crossover. Finally, the instrument and outcome cannot have a common, confounding cause (Angrist & Krueger, 2001).

In general, randomization provides an effective instrument. Random assignment (instrumental variable) is correlated with treatment delivery (endogenous variable). Furthermore, random assignment is only correlated with the outcome variable through its effect on the endogenous variable. At the same time, the IV distinguishes participants who may receive an intervention (the treatment group) from those who certainly will not (the control group) (Heckman, 1995). As Imbens and Rosenbaum note, “treatment effect is a function of treatment that is actually received, and, once that effect has been removed from responses the responses are independent of the treatment that was randomly assigned” (2005, p. 11). In an experiment, as here, treatment assignment is random. Therefore, it influences the probability of treatment but does not determine outcomes; it is a “strong” instrument (Imbens & Rosenbaum, 2005).

IV approaches are not a solution for weakly designed experiments. The use of instrumental variables is problematic when an IV is correlated with an omitted or unknown variable. An association between the instrumental variable and omitted variables can lead to a bias in the resulting estimates that is much greater than the bias in ordinary least squares estimates (Angrist & Krueger, 2001). However, in trials with true randomization, this concern is unnecessary. When the integrity of the randomization process is maintained, group assignment should not correlate with any known or unknown factors. In this experiment, the random number generator was automated and there was no evidence to suggest exogenous influences that may have led to possible correlation with some unobserved variables.

IV methods capture the average causal effect of an intervention, even when there is not a randomized comparison group that had no chance to receive treatment. This is most useful in observational research or when attrition may bias standard experimental comparisons (Imbens & Angrist, 1994). In an experiment where treatment is randomized and fidelity to those assignments is strong, every individual is a complier (Abadie, 2002). The local average treatment effect (LATE) is the average treatment effect for individuals whose treatment status is influenced by changing an exogenous regressor that satisfies an exclusion restriction (Imbens & Angrist, Identification and Estimation of Local Average Treatment Effects, 1994), here randomization.<sup>24</sup> As Abadie notes, after “estimate[ing] the

---

<sup>24</sup> Angrist (2006) also notes that LATE is not the same as ATET, the average causal effect of treatment on the treated. ATET differs from LATE because it is a weighted average of two effects: on always-takers and one on compliers. However,

an important special case when LATE equals ATET is when  $D_{0i}$  equals zero for everybody, i.e., there are no always-takers. This occurs in randomized trials with one-sided non-compliance, a scenario that typically arises because no one in the control group receives treatment. If no one in the control group receives treatment, then by definition there can be no always-takers. Hence, all treated

cumulative distribution functions of the potential outcomes for compliers” the distributions of all potential outcomes can be compared. This comparison indicates how the intervention would affect different parts of the distribution of the outcome variable for all compliers (Abadie, 2002, p. 286).

The 2-stage least squares method is the most frequently used technique for instrumental variable analysis (Kennedy, 2003). The 2SLS method decomposes each covariate into two components: a portion correlated with errors in that regression model and a second, error-free portion that can then be used, in the second stage, to estimate effect sizes (Stock & Watson, 2002, pp. 331-335). In models without covariates, the two stage least squares model (2SLS) estimator uses a dummy instrument and is the same as the Wald estimator<sup>25</sup> (Angrist J. , 2006). This measure answers the question of how much the average outcomes would be affected if participation in the program were universal, assuming no general equilibrium effects (Heckman, 1995, p. 3).

Using randomization as an IV provides a better estimation of how treated participants performed compared to how they would have without the intervention in situations where causal inference is not straightforward (Heckman, 1995). Using this mechanism, this approach, “obtains confidence statements relating the treatment received to the magnitude of the effects observed (Greevy, Silber, Cnaan, & Rosenbaum, 2004).” As (Angrist & Krueger, 2001, p. 81) note,

---

subjects must be compliers. (p. 31).  
In this case, therefore, LATE and ATET are identical and interchangeable.

<sup>25</sup> See **Invalid source specified.** for a description of this approach to fitting straight lines, which was developed to overcome errors-in-variables problems.

[i]nstrumental variables are useful in experiments when, either because of practical or ethical considerations, there is incomplete compliance in the treatment or control groups. In randomized evaluations of training programs, for example, some treatment group members may decline training while some control group members may avail themselves of training through channels outside the experiment.

Though most popular with economists, an IV approach has been used to estimate the magnitude of experimental results where dilution was an issue. Angrist (2006) employed an IV approach to conduct a re-analysis of the Minneapolis Domestic Violence Experiment (MDVE). The MDVE (1984), conducted by Sherman and Berk, was a randomized evaluation of police-based strategies to prevent recidivism after a domestic disturbance. Each of the three police responses (arrest, removal and advice) were randomized through the use of a multicolor incident pad, with each sheet correlating with the type of response appropriate for that particular event. A number of factors contributed to a break-down of this system and the resultant differential attrition (Sherman & Berk, The specific deterrent effects of arrest for domestic assault, 1984, p. 264). After 6 months and using an ITT analytical framework, MDVE participants who were arrested at the incident had significantly lower recidivism rates than those receiving the other two responses. Angrist's re-evaluation of the MDVE data used randomization as an IV to overcome the limitations associated with the treatment dilution. This analysis returned an effect size for the arrest response that was about one-third larger than the original intention-to-treat effect (2006).

Given the popularity of both the IV and 2SLS framework, it is unsurprising that there are multiple approaches. The analysis below employs a linear probability model, as

suggested by Angrist & Pischke (2009, pp. 198-204).<sup>26</sup> In IV models, OLS estimates of standard errors are not consistent in the presence of endogenous variables (Guan, 2003). Bootstrapping, a nonparametric approach for evaluating the distribution of a statistic based on random resampling from within the sample, is used to estimate the standard errors intervals (Hesterberg, Moore, Monaghan, Clipson, & Epstein , 1995). The bootstrap standard error ( $SE_{boot}$ ) is reported for each statistic, in addition to standard coefficients and confidence.

Ultimately, this regression framework is used within an experimental context in order to better estimate the effect sizes in light of methodological challenges. It is not an approach that can be used to find a meaningful effect when traditional tests fail to find a significant relationship. In fact, an instrumental variable analysis will always agree with an ITT analysis regarding the plausibility of a null-effect hypothesis (Greevy, Silber, Cnaan, & Rosenbaum, 2004). Therefore, in addition to sidestepping criticisms of “significance shopping,” the IV analysis offers an opportunity to obtain more accurate estimates of the magnitude of actual programmatic effects.

#### **A. Absconding (and an example)**

Completion of the Life Skills program was associated with a non-significant reduction in measures of absconding after 12 months. As indicated in Table 5.34, program graduates, on average, were charged with slightly less absconding incidents. At the same time, approximately 4% more of the comparison sample absconded within the follow-up period.

---

<sup>26</sup> The IV analysis was conducted using the `ivregress` command in Stata 12SE (StataCorp, 2011).



**Table 5.36: IV Results, Absconding**

	$\beta$	SE <sub>boot</sub>	<i>p</i>	95CImin	95CImax
Number of Absconding Incidents	-0.010	0.136	0.940	-0.277	0.257
Percent Absconded	-0.042	0.084	0.617	-0.207	0.123

These coefficients were obtained, as discussed above, in two steps. In the first stage, the instrument is regressed on the explanatory variables and a predicted value is obtained. Here, we first regresses the actual receipt of treatment on a variable that indicates whether the individual was assigned to the treatment or control group (i.e. randomization) to get estimates for the probability of treatment. As noted above, randomization is an effective instrument because it is correlated with the receipt of treatment but not with unmeasured and/or unobserved variables that influence receipt of treatment. A second stage of the regression model can then be run, where the forecasts from the first regression are used as the independent variable in another regression. This second equation regresses the outcomes observed on the forecasts. In this case, we regress the first-stage estimates for absconding on the variable indicating participation in the CBT program in order to obtain unbiased estimates of treatment effects. This estimate takes into account the imperfect compliance rates observed during the evaluation period due to uncontrolled (or uncontrollable) treatment dilution.

### **B. Drug Testing and Use**

As reflected in urinalysis results, graduates of the Life Skills program, on average, were more drug-involved than those in the comparison group. As indicated in Table

5.35, treatment group probationers had a positive results on 2.8 more tests for any substance ( $p= .000$ ) and 1.6 more times for marijuana ( $p< .05$ ) than comparable offenders. However, graduates of the program had slightly fewer positive PCP tests, despite having significantly higher positive counts in other categories. The 2SLS analysis, like the ITT and TOT approaches, cannot be used to distinguish the overall increased levels of drug testing found in the treatment group from the impact of the CBT program alone.

**Table 5.37: IV Results, Count of Positive Urinalysis Screenings**

	$\beta$	$SE_{boot}$	$p$	95CImin	95CImax
Overall Positive Test Count	2.816	0.650	0.000	1.541	4.091
Positive Marijuana Test Count	1.624	0.538	0.003	0.569	2.678
Positive PCP Test Count	-0.079	0.185	0.671	-0.441	0.284

Despite higher rates of positive drug tests, the percentage of each group that submitted at least one positive drug test was not meaningfully different. As shown in Table 5.36, measures of prevalence for any positive test, any positive marijuana test and any positive PCP test failed to reach traditional levels of significance. Mirroring the directionality of post-assignment frequency measures, only the prevalence of PCP testing favored the treatment group.

**Table 5.38: IV Results, Prevalence of Positive Urinalysis Screenings**

	$\beta$	$SE_{boot}$	$p$	95CImin	95CImax
Percent with any Positive Test **	0.154	0.095	0.104	-0.032	0.339
Percent w/ any Positive Marijuana Test	0.102	0.094	0.278	-0.082	0.286
Percent w/ any Positive PCP Test	-0.070	0.057	0.223	-0.182	0.042

Finally, there was a significant reduction in the overall proportion of administered urinalysis tests that were positive for PCP use ( $p < .05$ ). Since, as noted above, there were significant differences in the overall number of tests administered to the treatment group, irrespective of involvement in the Life Skills program, it is difficult to directly compare the number of positive tests results. These rate-based measures, however, suggest that the program had an impact on PCP use over time. Similar comparisons for overall drug use and marijuana, as shown in Table 5.37, neither favored the treatment group nor were significant.

**Table 5.39: IV Results, Proportion of Screenings with a Positive Result**

	$\beta$	$SE_{boot}$	$p$	95CImin	95CImax
Overall Proportion of Positive Tests	0.082	0.059	0.165	-0.034	0.199
Proportion of Tests Positive for Marijuana	0.043	0.054	0.429	-0.063	0.149
Proportion of Tests Positive for PCP	-0.053	0.025	0.032	-0.101	-0.005

As with prior analyses, these results should be interpreted with caution. Graduation from the Life Skills program was also associated with 5.6 more drug tests within the first year ( $p = .000$ ). The increased rates and prevalence of positive urinalysis results underscore higher rates of testing, not a meaningful difference in drug usage.

**C. Offending**

Within the IV framework, graduates of the Life Skills program had lower average numbers of subsequent offenses across the majority of measures. Despite consistently favoring the treatment group, none of these differences reached significance. As indicated in Table 5.38, the treated probationers were charged with, on average, fewer charges of any kind, serious charges, non-violent charges, property charges and drug charges.

**Table 5.40: IV Results, Average Number of Charges**

	$\beta$	SE <sub>boot</sub>	<i>p</i>	95CImin	95CImax
Any Charge	-1.405	2.176	0.518	-5.670	2.860
Serious Charge	-0.263	0.581	0.650	-1.402	0.875
Violent Charge	-1.268	1.414	0.370	-4.039	1.503
Non-Violent Charge	-0.137	0.926	0.883	-1.953	1.679
Property Charge	-0.497	0.528	0.347	-1.532	0.539
Drug Charge	-0.444	0.308	0.149	-1.048	0.159

The Life Skills intervention had a meaningful impact on the prevalence of offenders charged with a new, non-violent offense. After 12 months, 18.8% more of the control group was charged with a non-violent crime ( $p < .05$ ). The reduction in the prevalence of overall offending was also significant ( $p < .05$ ), though it is driven almost exclusively by the change in non-violent offending. There were also non-significant reductions in the prevalence of property and drug charges, approximately 4% and 7%, respectively, as shown in Table 5.39.

**Table 5.41: IV Results, Proportion of Each Group Charged, by Offense**

	$\beta$	SE <sub>boot</sub>	<i>p</i>	95CImin	95CImax
Any Charge *	-0.188	0.088	0.032	-0.359	-0.016
Serious Charge	0.037	0.061	0.542	-0.082	0.157
Violent Charge	0.010	0.065	0.873	-0.116	0.137
Non-Violent Charge *	-0.188	0.093	0.044	-0.370	-0.005
Property Charge	-0.042	0.069	0.547	-0.177	0.094
Drug Charge	-0.079	0.070	0.263	-0.217	0.059

**D. Incarceration**

After 12 months, there were no meaningful differences in incarceration rates or characteristics between the treated and comparison probationers. As shown in Table 5.40., graduates of the program spent fewer overall days in Philadelphia’s jail system and approximately 4% less of the treatment sample was incarcerated at least one time. Additionally, the average length of stay in the county jail for treated offenders was just over 25 days shorter than for comparable offenders. These differences, though relatively consistent, failed to reach significance.

**Table 5.42: IV Results, Incarceration Length and Characteristics**

	$\beta$	$SE_{boot}$	$p$	95CImin	95CImax
Average Number of Distinct Incarcerations	0.039	0.170	0.816	-0.294	0.373
Days Spent in Jail	-6.355	19.901	0.749	-45.361	32.650
Proportion Incarcerated	-0.048	0.087	0.576	-0.219	0.122
Average Jail Stay Length	-25.193	29.503	0.393	-83.017	32.631

**E. Conclusion**

Using an instrumental variable analysis, the effects of CBT on offending characteristics within 12 months was better specified. Notably, the reduction in the percentage of offenders who committed a non-violent crime was estimated at almost 19%. This is noticeably larger than then the 7.5% reduction observed in the ITT analysis. Additionally, a 5% reduction in the rate at which urinalysis screenings were positive for PCP was identified using the IV approach. A comparison of these results to the more conservative ITT analysis suggests that indications of the overall effectiveness of the intervention may be suppressed by issues surrounding treatment delivery during this evaluation.

## **CHAPTER 6: SUMMARY AND IMPLICATIONS**

This chapter summarizes the first round of findings from a multi-year randomized trial conducted by Philadelphia's Adult Probation and Parole Department (APPD) and the Jerry Lee Center of Criminology (JLC). The implications, for both policy and research, are also discussed. The primary purpose of this evaluation was to assess the impact of a cognitive-behavioral therapy program that had been developed specifically to reduce violence among APPD's high-risk population. The results for the first 12 months of the trial are reported here. The broader aims of the project were to add to the limited stock of methodologically rigorous, experimental knowledge on the efficacy of CBT in a community correctional setting, as well as to provide additional theoretical links between CBT and broader criminology.

### **I. Summary**

The Life Skills program had a meaningful effect on the criminal conduct of high-risk offenders, though the impact was not as broad as was intended. The magnitude of the results varied dependent on the analytical method employed.

Using an Intention to Treat (ITT) approach, 7.5% fewer members of the treatment group were charged with a non-violent crime (or any crime) and 2.3% fewer of PCP tests administered indicated its use, differences were significant at  $p < .05$ . Though failing to reach traditional levels of significance, for all other offenses the frequency of offending was lower for offenders assigned to attend CBT. Treatment group probationers were charged with, on average, fewer numbers of violent, serious, non-violent, property and



drug offenses. Prevalence rates were also generally lower, though not significant, with a smaller percentage of treatment group participants having been charged with drug and property crimes, in addition to those noted above. The treatment group had lower, non-significant measures of incarceration, including number of stays in jail, fewer days spent in jail overall and a lower average length of incarceration. Measures of absconding were near identical. Treatment group probationers, since they were screened at significantly higher rates, also tested positive for both any substance and marijuana at increased frequency and prevalence. Despite this, measures of PCP frequency and prevalence were lower, in addition to the significant differences noted above. Finally, time-to failure measures consistently favored the treatment group, with the difference in survival curves reaching significance for non-violent offending.

When comparing those probationers who completed the full 14 week program to the complete control group produced, unsurprisingly, results favored the intervention. Using this Treatment on the Treated (TOT) approach, program graduates were charged with, on average, fewer violent, non-violent drug and property charges, all of which were significant at  $p < .05$ . Prevalence measures were similar, with significant reductions in the proportion of the graduate group charged with violent, serious, non-violent and property offenses. Treated probationers performed significantly better on each measure relating to absconding, incarceration and time to failure.

Lastly, randomization was used as an instrument variable to better estimate the magnitude of effect sizes in light of treatment dilution. The IV results indicated that relationship were significant only where the more conservative ITT approach did as well (Greevy, Silber, Cnaan, & Rosenbaum, 2004). However, the 7.5% reduction in the

prevalence of any offending and non-violent offending identified in the ITT analysis was estimated at 18.7%. The bootstrap standard error for the “Any Change” coefficient was slightly smaller. The reduction in the proportion of urinalysis tests positive for PCP use also was estimated at 6.9%. Though failing to reach traditional levels of significance, the IV estimation for the majority of the remaining coefficients indicated slightly larger estimates in reductions for the average number of charges, proportion of each group charged with categorical offenses, characteristics of incarceration and time to failure.

A comparison of each of the sets of results highlights some concerns about the use of a treatment on the treated approach. The same cautions would hold true for any post-hoc analytical method that encourages the systematic removal of certain classifications of participants from a randomized sample. The TOT results, looking only at significances, are radically different from both the ITT and IV results. With only two exceptions, every single crime-related outcome was significantly different and favored the treatment group. Results in both the ITT and IV results, on the other hand, reached significance on *only* the same three measures. For example, traditional between-group comparisons found a non-significant reduction of about .5 charges, overall, associated with the CBT program. Under the IV approach, the difference was calculated as a larger, but still non-significant, reduction of about 1.5 charges per person. The TOT results were not only significant, but, with a 2.6 charge reduction, represented a 420% increase over the ITT findings and a 73% over the instrumentally-adjusted effect. Where all three approaches agree on the statistical likelihood of the findings, the magnitude of effects also varied significantly, indicating additional concerns regarding the TOT approach. For example, the difference in the overall prevalence of offending favored the treatment group by 7.5% and 18.7%

under the ITT and IV approaches, respectively. The TOT results indicate a reduction of 14.5%. In this sense, since the comparison sample is much larger, the TOT results may, depending on the underlying distribution of offending, underestimate effect sizes.

## **II. Discussion of Results**

As a classroom-based program with a specific aim to reduce serious recidivism, and delivered as part of an “anti-violence” initiative for high-risk offenders, the first round of results were less than ideal. Notably, there were no clear reductions in violent and serious offending within the first 12 months after individual random assignment. Overall, the results suggest that the Life Skills intervention is effective at reducing non-violent offending and certain types of drug use after 12 months.

The Life Skills intervention, despite some limitations, demonstrated an impact on recidivism comparable to other, established CBT-based programs. In one field randomized trial, participation in a life skills program based upon CBT reduced the proportion of parolees convicted of any new offense by approximately 18% (Ross, Fabiano, & Ewles, 1988). For example, offenders released on community supervision, after receiving a similar CBT-based program while in custody, demonstrated a 5% reduction in post-release offending (Robinson D. , 1995). The Georgia Cognitive Skills Experiment, a similar program and evaluation, reported that after 9 months parolees receiving the R&R program had slightly lower rates of arrest, returns to prison, revocations or in employment, but that the differences failed to reach significance. Calculations of time-to-failure were similarly insignificant, as in Philadelphia (van Voorhis P. S., 2004).

The results reported here are, by and large, directionally consistent and most often favored the treatment group. This leaves room for optimism with regard to the analyses of later waves of data collection. There are also a number of potential explanations for the lack of significant effects during the time frame reported here. Therefore, these results should be considered in context, both of the current evaluation and of the intervention itself.

The most apparent source of these muted results is derived from the offender's treatment status during the one year follow up period. For probationers who were both assigned to participate in the class and were successfully enrolled in at least one class within one year, the mean time to treatment was 91.3 days (251 offenders<sup>27</sup>). The Life Skills program was designed to take 14 weeks (approximately 98 days) to complete. It took, on average, almost 190 days for an enrolled participant to complete the course. Therefore, almost 50% of the 12 month follow-up period for the average offender was comprised of pre- and during-treatment conduct. This makes an accurate assessment of the effects of program completion difficult. Fortunately, this limitation will, by default, be overcome during future waves of data collection and analysis.

The lack of significance, but consistent directionality, may also result from the relatively short follow up period, regardless of treatment status. The time constraints may also limit the opportunity to capture meaningful differences in conduct. Polaschek et al. (2005), for example, employed a follow-up period of over 1,300 days to capture the effects of a CBT-based program delivered in a prison-community structure. Similarly,

---

<sup>27</sup> This count is slightly lower than the total number of treated individuals, as probationers who were removed from one session were recruited for, and often re-enrolled, in a later session. The count here includes only the lapse between random assignment and the first instance of treatment, not the enrollment that resulted in graduation from the Life Skills program.

Dowden, Blanchette, and Serin (1999) used a 3 year follow up period when evaluating a low-intensity anger management program. In their 2007 review, Lipsey, Landenberger and Wilson found that, though 50% of the included studies used a similar follow up period, 23% included longer observation periods (2007). Data support the use of longer observation windows. During the course of this evaluation, and especially in the early stages, regular checks were conducted to ensure that the assignment and supervision mechanisms were functioning as designed. Data extracted at these relatively short intervals suggest that there is a delayed onset of any treatment effect. Between-group differences during the first months of the evaluation were almost non-existent. Only as treatment delivery rates increased, along with sample sizes and observed time, were differences in conduct observed. This observation, taken in conjunction with the encouraging directionality of the results, suggests that a longer observation period is warranted.

Recruitment and participation in the Life Skills program has remained ongoing. Probationers who were enrolled in the RCT between May of 2010 and April of 2011, and who remain under active supervision, are given priority during the recruitment process.<sup>28</sup> Therefore, the data reported here may also underestimate the ultimate proportion of the treatment group that could successfully complete the program.<sup>29</sup> Meta-analytic evidence has shown that treatment attrition rates have a negative relationship on effect sizes ( $\beta = -.28, p < .05$ ) (Lipsey, Landenberger, & Wilson, *Effects of Cognitive-Behavioral Programs for Criminal Offenders*, 2007). Therefore, any reductions in this area should encourage

---

<sup>28</sup> As of May 2013, APPD has continued to support the Life Skills program, even in light of a decreasing proportion of RCT-enrolled probationers in each session.

<sup>29</sup> Admittedly, any improvements in binary treatment outcomes will likely be small. However at the 12 month cutoff point, there were RCT participants actively involved in the ongoing classes and more who remained eligible for enrollment in future classes.

more positive results. Within the first year of the evaluation period, 63.7% of participants who began the course were able to complete it. The .363 attrition rate, calculated at that point, is higher than the large majority of the studies included in the Lipsey, et al., review. In that analysis, only 22% of all included studies had attrition rates higher than .30, though this figure includes multiple contexts and methodological approaches. Continued enrollment and a longer follow-up period will, hopefully, remedy at least some portion of the deficit.

The relatively weak, though still promising, results reported here may not be an artifact of the time frames used, but rather due to unique and identifiable characteristics of the program itself. Treatment length, instructor training/ background and the quality of implementation have been shown to relate to the magnitude of effects. Lipsey, et al. (2007), found that the length of treatment was significantly related to total hours ( $r=.51$ ) as were the number of sessions per week ( $r=.58$ ) and number of treatment hours per week ( $r=.75$ ). In that meta-analysis, 45% of the studies included had between 11 and 20 months of treatment. This study would have fallen at the low end of that range. At the same time, just over 40% of these same studies met more often than the once weekly meetings employed here, increasing the overall dosage.<sup>30</sup> When controlling for methodology, the number of sessions per week ( $\beta=.32, p<.05$ ), hours per week (logged) ( $\beta=.23, p<.1$ ), and total hours of treatment (logged) ( $\beta=.38, p<.05$ ), were significantly associated with larger effect sizes. It may have been the case that the dosage of the CBT

---

<sup>30</sup> As noted in Chapter 4, the practical limitations of high-risk, community-based supervision prevented meaningful increases in the amount of treatment hours that could be delivered each week. Similarly, the 14 week curriculum represented the compromises necessary to develop and implement a CBT-based program in the “field.” It may, therefore, not be a completely level playing field, as just over half of the studies in the Lipsey, et. al (2007) study were completed in a custodial environment where time management and attendance are much more easily managed.

intervention was not sufficient to effect behavioral change, at least with regard to recidivism.

A number of additional influences, though measurable, are difficult to quantify in terms of their influence on outcomes measured after 12 months. For example, during the interventions, there were four different facilitators employed during the intervention and each of these individuals was a trained probation officer, not a clinician or researcher. Both of these factors have been shown to correlate negatively with decreases in recidivism (Lipsey, Chapman, & Landenberger, 2001). The fact that the Life Skills program delivered here significantly reduced some measures of non-violent crime is perhaps more meaningful in light of these constraints. However, the structure of the program was designed to be both replicable and sustainable. Increased levels of researcher-driven oversight, enrollment of less dangerous offenders or the employment of psychologists would have severely hampered, if not eliminated, the potential for the sustained use of the program. The current program can, and is, being used as one measure to reduce crime in Philadelphia. After all, when scaled up, even modest effects can have a meaningful impact on community life when the treated population is sufficiently large.

The accomplishments of the Anti-Violence Experiment go beyond the observable differences in offending rates. As Sherman (2006) notes, experimental evaluations can be used to both evaluate theory and to develop effective harm-reduction programs. It is difficult, however, to avoid making “basic mistakes when conducting field experiments, since experiments require a very different set of skills and methods than... [in] observational criminology” (2010, p. 399). The innovations in experimental design

employed during this project offer some useful additions to the “experimental criminology playbook.”

First, the random forest risk forecasting model, developed by Dr. Richard Berk, allowed for a more accurate and nuanced population identification strategy. The selection of offenders who were at a high risk of committing a serious crime, the group APPD was most concerned about, was paramount during this project. Though many CBT programs focus on “high-risk” offenders, the use of a less accurate prediction instrument may increase the number of false positives and, in the aggregate, make the treatment group more amenable to reform. Although the random forest procedure also results in false positives, the overall accuracy of the model is high (Barnes & Hyatt, 2012). The integration of this forecasting model into the case management software used at APPD was also noteworthy, as it ensured that the data used to make the prediction, as well as the outcomes, were not influenced by the individuals running the assessment program. These automated processes ensured that biases, either intentional or subliminal, did not influence the forecasting and, ultimately, the assignment procedures.

Secondly, the random assignment procedures were integrated into the management software and operated completely without user input. This allowed for a near perfect rate of assignment compliance at the outset. Once an individual was identified as being eligible for the experiment every single offender was assigned to one of the experimental units, another salient characteristic of the design. The research team can, with great confidence, note that, during the enrollment period and based upon all of the administrative data available, every single eligible participant was enrolled in the trial. This was crucial and allowed for the inclusion of almost 1,300 high-risk



probationers within just 12 months of “active” random assignment, approximately 900 of whom were included in the CBT-focused analysis. This sample size is relatively large for an intervention of this nature, especially when compared to the small sample sizes employed in many methodologically similar, field-focused analyses (Ross, Fabiano, & Ewles, 1988).

Next, through the use of a regular and in-depth auditing and reporting process, the essential requirements of the Stable Unit Treatment Value Assumption (SUTVA) are well met. SUTVA is one of the core assumptions that underlies all randomized experiments and requires that the handling of any subject, while dependent on their own randomly-assigned treatment, is independent of the assignments of all other participants (Sampson, 2010). Though it is nearly impossible to ensure that treatment and control offenders never come in contact with each other, access to the intervention was consistently well controlled. There were no cases where an individual who was assigned to the control condition participated in the Life Skills course. On very few occasions were RCT participants were transferred out of their assigned unit, most often through the operation of APPD rules outside of the control of the research team, this process ensured that a request was filed to have them returned within 4 business days. The same procedures ensured that, if offenders who had completed their randomly-assigned sentence were sentenced to APPD’s authority a second time during the evaluation period, they were also supervised by their RCT-assigned officer and unit. This process, though labor intensive, allowed for the maintenance of group autonomy and treatment integrity.

Finally, the management of the Life Skills enrollment process allowed for the greatest number of possible offenders to participate in the program. Within the first 12

months of the project, 73% of all attempts to enroll an offender in the intervention conducted during the first year of the trial resulted in successful registration for the Life Skills program. JLC and APPD worked hard to ensure that no opportunity to screened treatment group participants was missed. In fact, only 3.3% of recruiting opportunities that were missed were due to the actions, or inactions, of program staff. Anecdotally (since no measured counterfactuals were available) these rates exceeded many other local programs offered in conjunction with community-based supervision. High-risk offenders are not known for their compliance with rules and conditions; these rates indicate the effort and commitment of the involved parties in the evaluation process.

The Philadelphia Anti-Violence Experiment was, like many field trials, the result of a balancing act between the desire to ensure valid results and the necessities of conducting real world research. The findings reported here represent the outcomes of a well-designed and well-implemented experiment, conducted with a difficult to manage population. Through intensive tracking and oversight, the trial includes a relatively large sample size and evaluates an innovative approach to recidivism reduction. After one year, there are some promising effects, even with a weak instrument, and the overall direction of the findings remains encouraging.

### **III. Integration of Findings**

The general direction of the results suggests that the CBT program may have a different effect across types of offending. These outcomes encompass a broad range of recidivism and post-treatment conduct. In this case, violence was largely unaffected by the Life Skills intervention. Notably, the treatment group had lower average number of

charges filed across all categories in the ITT analysis, with the exception of serious and violent crimes. The prevalence of serious and violent offending was, however, slightly higher in the treatment group. These results are replicated within the IV analysis, where graduation from the Life Skills program was associated with a non-significant reduction in all outcome measures, with the notable exception of the prevalence of serious and violent offending.

General cognitive-behavioral research on anger, as well as the literature more broadly, may offer insight into these results. According to Beck, situations involving anger and violence are treated differently from a cognitive belief perspective than other circumstances. Beck suggests that, when aroused, “the offender’s information processing shifts to the primal mode [and] his thinking about the incident is biased and highly exaggerated” (Beck A. , 1999, p. 127). Importantly, this anger acts as a mediator between an instigating circumstance and violent reactions, thereby limiting the individual’s ability to avoid violent and reactionary behaviors (Betancourt & Blair, 1992). Quite simply, in the situations where non-premeditated violent conduct is likely to occur, graduates of the program do not have time to access their new skills and so conduct remains unaffected. Alternately, the CBT program delivered here was of insufficient quality or duration to change deeply ingrained cognitive belief systems or automatic patterns of behaviors.

Research suggests that self-control and anger, when considered through the cognitive-behavioral lens, have similar relationships with socially undesirable behaviors, including crime (Beck & Fernandez, 1998). Gottfredson and Hirschi (1990) suggest that self-control is developed at a young age and, by adulthood, is invariant. The significant

and negative effects, across only specific types of crimes, challenge this assumption. If the relationship between self-control and crime is homogeneous, that is, the same for all types of crime, then these results pose a challenge for theoretical integration. These results are instead supportive of a conceptualization of self-control that, while a predicate of criminal behavior, is amenable to development and change over time (Na & Paternoster, 2012).

The results of this analysis could be considered supportive of the hypothesis, advanced by many life course criminologists, that the propensity to commit crimes changes during the social development and maturation process. All individuals have some probability of engaging in a criminal act that is conditioned on the cumulative development and social history of that person (Sampson & Laub, 2005). Self-control predicates engagement in activities and conditional states that may decrease the likelihood of deviance or alter criminal trajectories (Sampson, Laub, & Wimer, 2006). Changes in levels of self-control brought about by participating in the CBT program may not be apparent in individuals who remain criminally active. However, for those individuals who are on a desistance path, but have not yet reached the threshold that would result in a drop in crime, increased self-control and cognitive skills would not result in an immediate or apparent decrease in offending. Instead, participation in the Life Skills program would result in differences in desistance rates, a calculation that cannot be available for some time.

The Philadelphia Anti-Violence Experiment can also add value to the increasingly large number of meta-analyses that are being conducted on the effect of CBT on recidivism. Despite the clear value of experimental evidence, these studies are

underrepresented in the current reviews. This is due, in part, to the scarcity of randomized research in this area. For example, only 20% of Wilson's (2005) review of structured CBT programs was comprised of experimental evidence. One-third of the evaluations in Lipsey, et al.'s (2007) meta-analysis were of similar rigor, while Pearson and colleagues (2002) found only 7 similar studies, representing 10% of the sample. Meta-analytical results have been, as a whole, supportive of CBT's application as a recidivism reduction measure. However, these estimates are drawn from a number of studies (both quasi-experimental and weakly constructed experiments) in addition to randomized evidence. Randomization and implementation characteristics, as well as CBT methodology, have been shown to have significant impacts on effect sizes (Lipsey, Landenberger, & Wilson, *Effects of Cognitive-Behavioral Programs for Criminal Offenders*, 2007). Therefore, even though the outcomes of the evaluation reported here may have failed to reach significance on some metrics, the integration of these randomized and large sample results into the broader, meta-analytic literature may have a meaningful impact of aggregate effect sizes. Though the shifts may be incremental, they will ensure that the resulting effect sizes are more reflective of research methods designed to both "develop and test" emerging evidence on crime prevention (Sherman L., 2006).

#### **IV. Limitations**

Like most field trials, the Philadelphia Anti-Violence Experiment was conducted in an environment that posed challenges for the implementation and analysis of results. These compromises and challenges are necessary when working within the limitations

and needs of practitioner partners or, more generally, outside of a controlled laboratory setting. Randomization and associated processes ensure high degrees of internal validity. The unique characteristics of Philadelphia's APPD in size, organization and population, may inhibit the generalizability of these findings. Similarly, conclusions can only be applied to high-risk, male probationers and parolees under APPD's supervision. A recognized limitation in many experiments, this can be overcome through replication across contexts and sites.

The current analysis is also limited to outcome measures that can be captured using administrative data that is already available. Any systematic errors in these data should be evenly distributed between groups, due to a strong randomization strategy. However, these measures may over or under-estimate actual rates or fail to capture meaningful changes in behavioral patterns. Additional data, as discussed above, will be necessary to evaluate the extent of this limitation and ameliorate its impact.

The analysis reported here may also fail to accurately reflect the impact of the program. As discussed above, the 12 month time frame includes a significant amount of pre- and during- treatment time, as well as fails to capture the full, possible extent of treatment delivery. Though it is not possible to control for these limitations using statistical methods, patience and a second wave of exploration will lessen its effects.

Under the structure of the current evaluation, it is not possible to disentangle the effect of the CBT program from the distinct supervision characteristics associated with the treatment unit. As noted above, treatment group participants reported to APPD's central office more often, were called by their officers on the phone more often and were

drug tested at increased levels. These differences between the groups, all of which were significant and favored the treatment group, were not associated with the Life Skills program itself, but impacted only one experimental group. In addition to individual officer and supervisor personalities, these differences could have been driven, in part, by the project. A single unit, responsible for coverage of the entire city, supervised all of the treatment group offenders. On the other hand, the comparison group was supervised in multiple, regionally-oriented units. This allows for more variability in the intensity of supervision actually delivered, despite the fact that all of the experimental units discussed here operated under the same protocol. It is difficult to quantify the effect this may have had on outcomes.

Alternatively, the relative difference in intensity could have been the result of a form of observation bias. More common in qualitative and ethnographic research, the presence of researchers can potentially contaminate any kind of observational data and undermine its reliability and validity (Spano, 2002). Experimental interaction is not a guarantee of bias; some research has shown that, over time, observed behaviors tend to return to normal (Gottfredson 1996). In this case, although officers in the treatment unit, due to the blinding procedures, were not aware *which* of their offenders were enrolled in the trial, they were aware that an evaluation was being conducted and that there was some probability that their cases would be included and examined. The potential bias, therefore, comes not from the impact of the observation on the treated probation population, but rather on the conduct of the involved, but not studied, officers. Here, simply knowing that the conduct of probationers under their charge was being tracked

and analyzed could have caused the officers in the treatment unit to “dot their i’s” with increased regularity.

The impact of the increased supervision on the treatment unit is reflected in the significantly higher rates of office contacts, as well as in urinalysis screenings. As an additional example, officers assigned to the treatment unit were asked, in order to facilitate the CBT enrollment process, to update employment records regularly. This should have been completed, in accordance with APPD policy, but the control group officers received no similar reminder. Since the control condition, for the duration of the experiment, was double-blinded, the officers managing the comparison group were not even aware that an evaluation was taking place. No communication with those officers was permitted, from a research standpoint, in order to preserve the hard-won benefits of the clean, automated randomization process.

Regardless of the cause of the increase in supervision intensity associated with the treatment unit, the between-group differences will remain a potentially confounding variable in this analysis. It is worth noting that previous research has shown that increases in probation intensity have either little effect on recidivism (LaTessa & Vito, 1988) or, in some instances, have been shown to cause subsequent crime rates to increase (Turner & Petersilia, 1992). Therefore, it is possible that potential reductions in criminal behaviors encouraged by the Life Skills program could be suppressed or offset by the more severe levels of supervision delivered to participants during the periods in which they were not enrolled in the program. Under the current analysis, it is not possible to separate the effects of the relatively increased form of ISP from those of the CBT program itself and further research is required.



## **V. Implications for Future Interventions and Public Policy**

Cognitive-behavioral therapy, when delivered in a classroom setting, is a relatively efficient and cost-effective intervention. When probation officers, already employed by an agency, are used as facilitators, the startup and maintenance costs are even more manageable. Agencies that wish to integrate a treatment component into an ISP supervision framework in the future may use the protocols developed in Philadelphia as a model. In addition to the potential benefits derived from the CBT program, the model employed in Philadelphia allowed for flexibility in supervision and permitted, within reasonable limitations, the balancing of the high-risk caseloads between the classroom program and standard management strategies.

Substantively, these results qualify the suggestion that CBT programs can be broadly implemented to reduce many types of recidivism (Wilson, Bouffard, & Mackenzie, 2005). Instead, policy makers may want to consider the relationship between offender characteristics, programmatic goals and support within the literature, prior to the selection of an appropriate CBT-based program. If nothing else, this project provides one more option, tailored to the needs of a male, urban population, which can be considered.

Young, high-risk, male offenders, including those on probation or parole, are a notoriously difficult population to work with (Gynes, 1988). This research challenges the notion that CBT can be used, in the absence of other support and control, as a panacea for reducing crime rates among this group. However, the identification of some significant reductions after 12 months tests the long-standing view that “nothing works” to reduce recidivism (Martinson, 1974). Additionally, these findings are incompatible

with the notion that ISP, broadly construed, is ineffective (Sherman, et al., 1997). These findings suggest that policymakers should consider the integration of treatment into increasingly high levels of supervision for dangerous offenders. When political necessities require the use of ISP, even in the face of a skeptical body of research, agency heads may rely on these findings to both offset potential increases in recidivism and to capitalize on the opportunities created through more frequent offender contacts.

The analyses reported here also advance the menu of analytical options for experimental criminologists. Taking up Angrist's (2006) challenge, randomization is used as an IV to overcome practical limitations in an otherwise well-designed experiment. Despite their theoretical appeal, valid IVs are uncommon outside of econometrics, with the most common application being medical trials that have issues with treatment contamination (Sussman & Hayward, 2010). The use of the IV approach in the social sciences can be used to bridge the gap between the policy focused question answered by ITT analyses and the outcome-focused question of programmatic impact.

## **VI. Directions for Forthcoming Research**

The administrative and operational data collected by APPD as a matter of course and the implementation and monitoring data gathered by JLC during the project will provide a rich source of data for future analyses.

Given the limitations inherent in the 12 months window, as discussed above, the first steps will include the extraction of data on all experimental participants for 18 and 24 month post-random assignment periods. These timeframes should better capture the

extent of CBT treatment delivery during the project and allow for the inclusion of much-needed post-treatment time in each analysis. Additional data should also become available during these later analyses. For example, outcome data regarding technical violations, mortality and supervision compliance need to be extracted, hand-coded and validated from APPD and FJD data systems. This is a time-consuming and labor intensive process that has not yet been completed. Similarly, data on incarceration should become available, as arrests that took place (and are reported here) during the evaluation period should have sufficiently progressed through the judicial process.

Even with these additional data, the full impact of the program may not be reflected. In order to capture changes in cognition that may not be reflected in recidivism data, a qualitative survey will be delivered to all RCT participants still under supervision after 18 months.<sup>31</sup> The instrument, developed specifically for this project, includes measures of criminogenic attitudes (Tangney, Mashek, & Stuewig, 2007), self-control and response bias (MacDonald, Morral, & Piquero, 2011). The survey itself was designed to capture any changes in thought patterns that could be attributed to the intervention, even in the absence of immediate and observable changes in recidivism.

The principles of responsivity and effective intervention (Taxman, Thanner, & Weisburd, 2006) suggest that programs should be tailored to meet the needs of the targeted population. Further work will be necessary to identify any subgroups for which the Life Skills class was exceptionally effective and the calculation of possible dose-response relationships. As noted in evaluations of similar programs, adults between 20

---

<sup>31</sup> The rollout of the survey began in late March, 2013, and is expected to continue through May of the same year.

and 30, the target population here, are often the least impacted by CBT methods (Robinson D. , 1995). Johnson & Hunter (1992) reported their program to be least effective among offenders less than 30 years of age. Similarly, research has shown that high-risk offenders are particularly difficult to reform (Guynes, 1988). The same may hold true for the high-risk offenders enrolled in the Philadelphia. In fact, the offenses for which there was a significant effect, notably non-violent and drug offenses, are of the type that characterize the *moderate* offenders in Philadelphia (Barnes & Hyatt, 2012). Both of these factors suggest that a replication of the experiment focusing on younger and less serious offenders may result in larger effect sizes. This second phase of evaluation, though not in keeping with APPD's mission and justification for the project, would also allow the Life Skills program to be refined to reflect lessons learned during this evaluation.

## **VII. Conclusions**

Despite limitations due to attrition and other methodological factors, this study was able to provide evaluative research on a new approach to delivering cognitive-behavioral therapy to high-risk offenders. Using a traditional experimental analytical approach, the Life Skills program significantly reduced the prevalence of non-violent offending and the rate of positive PCP tests. These effect sizes were both confirmed and refined using randomization as an instrumental variable within a regression framework.

With implications for criminological theory and related evaluation research, this project advances, if only incrementally, our understanding of possible psychologically-based approaches to reducing crime within a population that is both dangerous and prone

to recidivism. Since the intervention used here is both innovative and, aside from this research, untested, there remains ample opportunity to build upon these findings.

## **APPENDICIES**

Appendix A: Complete Life Skills Course Schedule

Appendix B: Judicial Letter and Revised Conditions of Supervision

Appendix C: APPD Departmental Organization and Hierarchy

Appendix D: Risk Prediction Variables

**APPENDIX A: COMPLETE LIFE SKILLS COURSE SCHEDULE**

<b>Class ID</b>	<b>Session Number</b>	<b>Recruiting Start Date</b>	<b>Recruiting End Date</b>	<b>Class Start Date</b>	<b>Class End Date</b>	<b>Class Time of Day</b>
1	Session 1	5/17/2010	6/3/2010	6/7/2010	9/6/2010	10:00 AM
2	Session 2	7/12/2010	7/28/2010	8/2/2010	11/1/2010	2:00 PM
3	Session 3	8/30/2010	9/15/2010	9/20/2010	12/20/2010	10:00 AM
4	Session 4	10/25/2010	11/1/2010	11/15/2010	2/14/2011	2:00 PM
5	Session 5	12/6/2010	12/14/2010	1/3/2011	4/4/2011	10:00 AM
6	Session 6	2/7/2011	2/18/2011	2/28/2011	5/30/2011	2:00 PM
7	Session 7	3/29/2011	4/14/2011	4/18/2011	7/19/2011	10:00 AM
8	Session 8	5/23/2011	6/9/2011	6/13/2011	9/12/2011	2:00 PM
9	Session 9	7/13/2011	7/29/2011	8/1/2011	10/31/2011	10:00 AM
10	Session 10	9/5/2011	9/15/2011	9/26/2011	12/26/2011	2:00 PM
11	Session 11	10/24/2011	11/4/2011	11/14/2011	2/13/2012	10:00 AM
12	Session 12	12/27/2011	1/4/2012	1/9/2012	4/9/2012	2:00 PM
13	Session 13	2/6/2012	2/22/2012	2/27/2012	5/28/2012	10:00 AM
14	Session 14	4/3/2012	4/16/2012	4/23/2012	7/23/2012	2:00 PM

## **APPENDIX B: JUDICIAL LETTER AND REVISED CONDITIONS OF SUPERVISION**

The text of the letter provided to each probationer at the outset of their sentence read as follows:

Dear [insert name of probationer/parolee here] – [insert date here]

You were recently sentenced to a term of probation or released on parole by Judge [insert name of sentencing Judge here]. According to the rules and regulations of your supervision, you are required to complete a Life Skills class that takes place at APPD.

I am writing to let you know that failure to attend and complete this Life Skills class will constitute a violation of your probation/parole and will result in a violation hearing in my courtroom. Upon successful completion of the 14 week class, your reporting requirements will be reduced.

I wish you all the best as you complete your term of probation/parole. I hope that you will successfully complete your sentence without any violations.

Yours truly,

Judge Joan Brown

Every probationer, regardless of their enrollment status in the RCT, was provided a copy of the Rules and Regulations. As indicated under Rule 4, all offenders under APPD supervision were required, if asked, to attend the Life Skills program. Failing to comply with this request (which, during the evaluation period, was determined solely by random assignment and the recruitment protocol) was treated the same as any other violation of supervision and could result in referral to a judicial authority for a violation hearing and/or additional sanctioning.



Sample Copy of Rules and Regulations of Probation and Parole

**RULES OF PROBATION & PAROLE**

Name of Probationer/Parolee

PD

SD

Docket Number(s)

Case Judge

You have been placed on probation and/or parole and are expected to comply with the following rules:

1. You may **NOT**:
  - Possess Firearms or any other deadly weapons.
  - Unlawfully possess, use, sell or distribute controlled substances of any kind.
  - Leave Philadelphia without permission from your Probation/Parole Officer.
2. You must:
  - Report to your Probation/Parole Officer as directed.
  - Respond promptly to any summons to appear in Court. Failures to appear in court, result in bench warrants.
  - Permit your Officer to visit you at your home or place of employment.
  - Make every effort to seek and maintain employment.
  - Obey all Federal, State and County criminal laws and city ordinances.
3. You must notify your Officer within **72 hours** if any of the following occurs:
  - A new arrest
  - A change of address
  - Change of employment
4. You are subject to:
  - Pay supervision fee to the Philadelphia Adult Probation Department by ACT NO.35-1991 of the Pennsylvania General Assembly unless the fee is waived by the Court.
  - A personal search and/or property search, including vehicle and seizure of any contraband found, if there is a reasonable suspicion that you are in violation of any of the conditions of supervision in accordance with ACT NO. 35 of the Pennsylvania General Assembly, Special Session NO.1 of 1995.
  - Attend a Life Skills Training Course
  - If you have accepted Accelerated Rehabilitative Disposition (ARD) as a result of a violation of 18 P.A.C.S. Chapter 331 (relating to sexual offenses) you will also be subject to the same search and seizure if the Court has determined that you shall be subject to personal and/or property search as a condition of participation in the ARD program.
5. You must also comply with the following special conditions of Probation/Parole:

**Fines/Costs/Court Fees, Probation**

**PROBATION WARRANT AND VIOLATION HEARING PROCEDURES**

If you are arrested, if you violate any of the rules or conditions of your Probation/Parole while under supervision, the Court has the authority to place a Probation Warrant against you which will keep you in detention until a hearing is held. If the Probation Warrant is held in force at the hearing, you will remain in detention pending disposition of your new charges or the lifting of a Violation of Probation/Parole Hearing.

**EXTRADITION TO PENNSYLVANIA**

If you are arrested in another state your Officer can place a Probation Warrant against you and arrange for you to be returned to Philadelphia for a violation hearing if deemed necessary.

**FINANCIAL PAYMENTS**

Payment of any court-imposed fee shall not be paid in cash. A money order or certified bank check is the only acceptable form of payment. Payment must be made directly at the Probation Department collection window or sent to the Probation Department financial collection section by U.S. Mail.

**GRIEVANCE PROCEDURES**

If you have any questions regarding your Probation/Parole, please consult with your Probation/Parole Officer immediately. If you have a grievance concerning the supervision of your case, you may consult the team supervisor.

**ACKNOWLEDGMENT OF PROBATIONER/PAROLEE**

I have read, or have had read to me the information above in its entirety.

\_\_\_\_\_  
Signature of Probationer/Parolee

08/18/2011

Date

Note: If signed by a coach, two witnesses must execute this instrument.

\_\_\_\_\_  
Signature of Probation/Parole Officer

08/18/2011

Date



## **APPENDIX D: RISK PREDICTION VARIABLES**

Throughout the JLC-APPD partnership, there have been three different risk forecasting models used for the forecasting of “live” offenders. There were many, many more versions created during the development process, but they were never put into practice.<sup>32</sup>

In the course of developing the three forecasting models that were used (known as Models A, B and C), 53 different predictor variables were used to predict future offending. Different types, numbers, and combinations of these predictors were featured in each of the three models. Some predictors were used only once and then discarded, while others have played a role in each model developed since the beginning of the project. The components and predictors for each model can be seen in the table below.

---

<sup>32</sup> This section is adapted from *Classifying Adult Probationers by Forecasting Future Offending: Final Technical Report* (Barnes & Hyatt, 2012).

**Predictor variables used to construct the three live forecasting models**

<b>Predictor Variable</b>	<b>Model A</b>	<b>Model B</b>	<b>Model C</b>
ProbationStartAge	✓	✓	✓
CalculatedGender	✓	✓	
ZipBase5Top29			✓
ZipPopulation	✓	✓	
ZipHouseholdIncome	✓	✓	
ZipHouseValue	✓	✓	
ZipPersonsPerHousehold	✓	✓	
ZipCityLimitDistance	✓	✓	
ZipOutsideCityLimits	✓	✓	
FirstAdultAnyChargeAge	✓	✓	✓
FirstAdultViolenceChargeAge	✓	✓	✓
FirstJuvAnyChargeAge		✓	✓
FirstJuvViolenceChargeAge		✓	
InstantMurderChargeCount	✓		
InstantSeriousChargeCount		✓	✓
InstantViolenceChargeCount	✓	✓	
InstantSexualChargeCount	✓	✓	
InstantPropertyChargeCount	✓	✓	
InstantFirearmChargeCount	✓	✓	
InstantDrugChargeCount	✓	✓	
InstantProbationSentenceCount		✓	
InstantProbationDaysConcurrent		✓	
InstantIncarcerationSentenceCount		✓	
InstantIncarcerationDaysConcurrent		✓	
PriorAdultAnyChargeCount	✓	✓	✓
PriorAdultUcrPersChargeCount	✓		

**Predictor variables used to construct the three live forecasting models (con't)**

Predictor Variable	Model A	Model B	Model C
PriorAdultSeriousChargeCount		✓	
PriorAdultViolenceChargeCount	✓	✓	✓
PriorAdultSexualChargeCount	✓	✓	✓
PriorAdultSexRegChargeCount	✓	✓	
PriorAdultPropertyChargeCount	✓	✓	
PriorAdultWeaponChargeCount	✓	✓	
PriorAdultFirearmChargeCount	✓	✓	
PriorAdultDrugChargeCount	✓	✓	
PriorAdultDrugDistChargeCount	✓	✓	
PriorJuvAnyChargeCount		✓	
PriorJuvSeriousChargeCount		✓	
PriorJuvViolenceChargeCount		✓	
PriorJuvSexualChargeCount		✓	
PriorJuvPropertyChargeCount		✓	
PriorJuvWeaponChargeCount		✓	
PriorJuvFirearmChargeCount		✓	
PriorJuvDrugChargeCount		✓	
PriorJuvDrugDistChargeCount		✓	
PriorAdultSeriousChargeLatestYears	✓		
PriorSeriousChargeLatestYears		✓	✓
PriorProbationCount	✓	✓	
PriorFailureToAppearCount	✓	✓	
PriorAbsconderCount	✓	✓	
PriorJailStays	✓	✓	✓
PriorJailDays	✓	✓	✓
PriorConfinementSentenceCount	✓		
PriorIncarcerationSentenceCount	✓	✓	

In the tables above, the names of each predictors has been abbreviated somewhat. To better understand what each of these predictors really represents, the following descriptions are likely to be helpful:

**ProbationStartAge.** The offender's age at the start of the new probation case.

**CalculatedGender.** The offender's gender, as calculated from all available data sources. This value is available from more than one of the databases used to produce predictors the model. Most of the time, these sources all agree on whether the offender is male or female. When disagreement occurs, or when some of these values are missing, this value

is calculated by using the gender value from the criminal records data (if available), and the value from the probation case management system where the criminal records value is missing.

**ZipBase5Top29.** This variable forms a categorical list of 31 distinct values to indicate the 5-digit zip code where the offender was residing at the time that the instant probation case began. These values are made up of the 29 most prevalent valid zip code values among probation case starts, along with 2 other coded values to indicate whether the offender was residing some other valid zip code. If the offender was living in one of the 29 most-frequent zip codes – all of which are located within the city limits of Philadelphia – this variable is coded with that offender’s five-digit zip code value. When the offender did not reside in any of these 29 specific zip codes, the value is coded as “99998” when the offender lived elsewhere inside the city limits, and “99999” when the offender lived elsewhere outside the city limits. Offenders with missing or invalid zip code values are excluded from the model construction data.

**ZipPopulation.** The total population, based on 2000 census data, in the zip code where the offender was residing at the start of the new probation case.

**ZipHouseholdIncome.** The average household income in the offender’s home zip code.

**ZipHouseValue.** The average house value in the offender’s home zip code.

**ZipPersonsPerHousehold.** The average number of persons residing in each household in the offender’s home zip code.

**ZipCityLimitDistance.** The number of statute miles between the offender’s home zip code and the Philadelphia city limits. Coded as zero for all observations where the offender resided within the city.

**ZipOutsideCityLimits.** A binary variable which indicates whether the offender’s home zip code is outside of the Philadelphia city limits.

**FirstAdultAnyChargeAge.** The offender’s age at the time of the first offense which resulted in charges in adult criminal court.

**FirstAdultViolenceChargeAge.** The offender’s age at the time of the first violent offense which resulted in charges in adult criminal court. When the offender has never been charged as an adult with a violent offense, this value is coded as 100 years.

**FirstJuvAnyChargeAge.** The offender’s age at the time of the first offense which resulted in charges in juvenile court. When the offender no record of juvenile offending,

this value is coded as 100 years. This variable is used only in the Models B and C, and reflects the addition of juvenile predictors to the model.

**FirstJuvViolenceChargeAge.** The offender's age at the time of the first violent offense which resulted in charges in juvenile court. When the offender no record of violent juvenile offending, this value is coded as 100 years. This variable is used only in the Model B.

**InstantMurderChargeCount.** The total number of charges for murder or attempted murder that appear in the court records for the instant case. The instant court case is the one that resulted in the offender being placed on APPD supervision for this instance of probation or parole. This variable is used only in the Model A. It was dropped from the later models because only a very small number of cases which involve charges this serious result in the offender being placed on APPD supervision.

**InstantSeriousChargeCount.** The total number of charges for serious offenses – defined as murder, attempted murder, aggravated assault, robbery, and sexual crimes – in the instant case. This variable is used only in the forecasting models from Model B onward. It replaces the number of instant charges for murder or attempted murder.

**InstantViolenceChargeCount.** The total number of charges for violent offenses in the instant case. Violent offenses include all serious offenses, as well lesser crimes such as simple assault.

**InstantSexualChargeCount.** The total number of charges for sexual offenses in the instant case.

**InstantPropertyChargeCount.** The total number of charges for property offenses in the instant case.

**InstantFirearmChargeCount.** The total number of charges for firearm offenses in the instant case.

**InstantDrugChargeCount.** The total number of charges for drug offenses in the instant case.

**InstantProbationSentenceCount.** The total number of sentences to probation that appear in the court records as a result of the instant case. This variable is used only in the new forecasting model. It was added, along with the other instant sentencing variables, to provide an indication of how dangerous the sentencing judge thought the offender to be.

**InstantProbationDaysConcurrent.** The maximum number of days sentenced to probation as a result of the instant case, assuming that all sentences are to be served concurrently. This variable is used only in the new forecasting model.

**InstantIncarcerationSentenceCount.** The total number of sentences to incarceration as a result of the instant case. This variable is used only in the new forecasting model.

**InstantIncarcerationDaysConcurrent.** The maximum number of days sentenced to incarceration as a result of the instant case. This variable is used only in the new forecasting model.

**PriorAdultAnyChargeCount.** The total number of charges for offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultUcrPersChargeCount.** The total number of charges for Uniform Crime Report (UCR) Part I Personal offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case. These offenses include murder, aggravated assault, robbery, and forcible rape. This variable was used only in Model A. It was dropped from later models because it did not include some non-forcible sexual offenses, such as statutory rape, that are included in our definition of serious crime.

**PriorAdultSeriousChargeCount.** The total number of charges for serious offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case. This variable is used only in Model B, where it replaced the number of prior charges for UCR personal offenses.

**PriorAdultViolenceChargeCount.** The total number of charges for violent offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultSexualChargeCount.** The total number of charges for sexual offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultSexRegChargeCount.** The total number of charges for sex offender registration offenses (i.e., violations of the registration requirements in Megan's Law) which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.



**PriorAdultPropertyChargeCount.** The total number of charges for property offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultWeaponChargeCount.** The total number of charges for weapon offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultFirearmChargeCount.** The total number of charges for firearm offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultDrugChargeCount.** The total number of charges for drug offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorAdultDrugDistChargeCount.** The total number of charges for drug distribution offenses which were dealt with in adult criminal court, and which took place prior to the start of the new probation case.

**PriorJuvAnyChargeCount.** The total number of charges for offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case. This variable is used only in the new forecasting model, and reflects the addition of juvenile predictors to the model.

**PriorJuvSeriousChargeCount.** The total number of charges for serious offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvViolenceChargeCount.** The total number of charges for violent offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvSexualChargeCount.** The total number of charges for sexual offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvPropertyChargeCount.** The total number of charges for property offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvWeaponChargeCount.** The total number of charges for weapon offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvFirearmChargeCount.** The total number of charges for firearm offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvDrugChargeCount.** The total number of charges for drug offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorJuvDrugDistChargeCount.** The total number of charges for drug distribution offenses which were dealt with in juvenile court, and which took place prior to the start of the new probation case.

**PriorAdultSeriousChargeLatestYears.** The number of years since the offender's most recent serious offense which resulted in charges in adult criminal court. When the offender has never been charged as an adult with a serious offense, this value is coded at 100 years. This variable is used only in the Model A, and was amended in later models to include juvenile offending information.

**PriorSeriousChargeLatestYears.** The number of years since the offender's most recent serious offense, regardless of whether that offense was dealt with juvenile or adult criminal court. When the offender has never been charged with a serious offense, this value is coded as 100 years. This variable was not used until Model B, and reflects the addition of juvenile predictors to the model.

**PriorProbationCount.** The total number of cases which were placed under APPD supervision prior to the start of the new probation case.

**PriorFailureToAppearCount.** The total number of bench warrants taken out against the offender, prior to the start of the new probation case, due to a failure to appear in court.

**PriorAbsconderCount.** The total number of arrest warrants taken out against the offender, prior to the start of the new probation case, due to absconding from supervision.

**PriorJailStays.** The total number of entries into the Philadelphia county prison system which took place prior the start of the new probation case.

**PriorJailDays.** The total number of days spent incarcerated in the Philadelphia county prison system prior to the start of the new probation case.

**PriorConfinementSentenceCount.** The total number of sentences to confinement – which includes both incarceration, house arrest, and electronic monitoring – that the offender received prior to the start of the new probation case. This variable is used only in the January 2009 forecasting model. It was dropped from later models because it strongly mirrors the incarceration sentence count variable, discussed below, and added little unique information.

**PriorIncarcerationSentenceCount.** The total number of sentences to incarceration that the offender received prior to the start of the new probation case.

## BIBLIOGRAPHY

- Abadie, A. (2002). Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models. *Journal of the American Statistical Association*, 97(457), 284-292.
- Adult Probation and Parole Department. (2012). *2011 Annual Report*. Philadelphia: First Judicial District.
- Ainslie, G. (1975). Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control. *Psychological Bulletin*, 82(4), 463-496.
- Akers, R. (1991). Self-control as a general theory of crime. *Journal of Quantitative Criminology*, 7(2), 201-211.
- Akers, R., Krohn, M. D., Lanza-Kaduce, L., & Radosevich, M. J. (1979). Social learning and deviant behavior. *American Sociological Review*, 44, 636-655.
- Alfred, B., Cohen, J., & Farrington, D. (1988). Criminal career research: Its value for criminology. *Criminology*, 26, 1-35.
- Andrews, D. A., Bonta, J., & Hogue, R. D. (1990). Classification for effective rehabilitation: rediscovering psychology. *Criminal Justice & Behavior*, 17, 19-52.
- Andrews, D. A., Bonta, J., & Wormoth, S. (2006). The Recent Past and Near Future of Risk and/or Need Assessment. *Crime and Delinquency*.
- Angrist, J. (2006). Instrumental variables methods in experimental criminological research: what, why and how. *Journal of Experimental Criminology*, 2(1), 23-44.
- Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4), 69-85.
- Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton, NJ: Princeton University Press.
- Arceneaux, K. (2005). Using Cluster Randomized Field Experiments to Study Voting Behavior. *Annals of the American Academy of Political and Social Science*, 601, 169-179 .
- Arkowitz, H., & Hannah, M. T. (1989). Cognitive, Behavioral and Psychodynamic Therapies: Converging or Diverging Pathways to Change. In A. Freeman, K. M.

- Simon, L. E. Beutler, & H. Arkowitz, *Comprehensive handbook of Cognitive Therapy* (pp. 143-170). New York: Plenum Press.
- Armstrong, J. S. (2007). Significance Tests Harm Progress in Forecasting. *International Journal of Forecasting*, 23, 321-327.
- Armstrong, T. (2003). The Effect of Moral Reconciliation Therapy on the Recidivism of Youthful Offenders : A Randomized Experiment. *Criminal Justice and Behavior*, 30(6), 668.
- Armstrong, T. A. (2003). The Effect of Moral Reconciliation Therapy on the Recidivism of Youthful Offenders : A Randomized Experiment. *Criminal Justice and Behavior*(30), 668.
- Austin, J. (2010). Reducing America's Correctional Populations: A Strategic Plan. *Justice Research and Policy*, 12(1), 32-35.
- Baker, T. B., McFall, R. M., & Shoham, V. (2009). Current Status and Future Prospects of Clinical Psychology: Toward a Scientifically Principled Approach to Mental and Behavioral Health Care. *Psychological Science in the Public Interest*, 9(2), 67–103.
- Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the Predictive Accuracy of Six Risk Assessment Instruments for Adult Sex Offenders. *Criminal Justice and Behavior*, 28(4), 490-521.
- Barnes, G. C., Ahlman, L., Gill, C., Sherman, L. W., Kurtz, E., & Malvestuto, R. (2009). Low-intensity community Supervision for low-risk offenders: a randomized, controlled trial. *Journal of Experimental Criminology*.
- Barnes, G. C., Ahlman, L., Gill, C., Sherman, L. W., Kurtz, E., & Malvestuto, R. (2010). Low-intensity community Supervision for low-risk offenders: a randomized, controlled trial. *Journal of Experimental Criminology*.
- Barnes, G. C., Hyatt, J. M., Ahlman, L. C., & Kent, D. T. (2012). The Long Term Effects of Low Intensity Supervision for Lower Risk Probationers: Updated Results from a Randomized Controlled Trial. *Journal of Criminal Justice*, 35(2), 200-220.
- Barnes, G., & Hyatt, J. (2012). *Classifying Adult Probationers by Forecasting Future Offending*. Washinton, DC: National Institute of Justice.
- Baro, A. L. (1999). Effects of a Cognitive Restructuring Program on Inmate Institutional Behavior. *Criminal Justice and Behavior*, 26, 466.

- Beck, A. (1999). *Prisoners of Hate: The Cognitive Basis of Anger, Hostility and Violence*. New York, NY: Harper Collins.
- Beck, A. (1999). *Prisoners of Hate: The Cognitive Basis of Anger, Hostility and Violence*. New York, NY: Harper Collins.
- Beck, A. T. (1964). Thinking and depression II: Theory and therapy. *Archives of General Psychiatry*, *10*, 561–571.
- Beck, A. T. (1993). Cognitive therapy: Past, present, and future. *Journal of Consulting and Clinical Psychology*, *61*, 194-198.
- Beck, A. T. (1995). Cognitive therapy: Past, present and future. In M. J. Mahoney, *Cognitive and constructive psychotherapies: Theory, research and practice*. New York: Springer.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1987). *Cognitive Therapy of Depression*. New York: Guilford Press.
- Beck, J., & Beck, A. (2011). *Cognitive Behavior Therapy: Basics and Beyond*. New York: Guilford Press.
- Beck, R., & Fernandez, E. (1998). Cognitive-Behavioral Therapy in the Treatment of Anger: A Meta-Analysis. *Cognitive Therapy and Research*, *22*(1), 63-74.
- Becker, G. (1968). Crime and Punishment: An Economic Approach. *The Journal of Political Economy*, *76*, 169-217.
- Berk, R. (2005). An Introduction to Ensemble Methods for Data Analysis. UCLA, UC Los Angeles.
- Berk, R. (2008). Forecasting methods in Crime and Justice. *Annual Review of Law and Social Science*, *4*, 219–238.
- Berk, R. (2012). *Criminal Justice Forecasts of Risk: A Machine Learning Approach*. New York: Springer.
- Berk, R. A. (2008). Forecasting methods in Crime and Justice. *Annual Review of Law and Social Science*, *4*, 219–238, 236.
- Berk, R. A. (2008). Forecasting methods in Crime and Justice. *Annual Review of Law and Social Science*, *4*, 219–238, 236.
- Berk, R. A., & de Leeuw, J. (1999). An Evaluation of California's Inmate Classification System Using a Generalized Regression Discontinuity Design. *Journal of the American Statistical Association*, *94*, 1045-1052.

- Berk, R. A., Sherman, L. W., Barnes, G. C., Kurtz, E., & Ahlman, L. (2009). Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning. *Journal of the Royal Statistical Society, Serise A: Statistics in Society*, 172(1), 191-211.
- Berk, R. A., Sherman, L. W., Barnes, G. C., Kurtz, E., & Ahlman, L. (2009). Forecasting Murder Within a Population of Probationers and Parolees: A High Stakes Application of Statistical Learning. *Journal of the Royal Statistical Society, Serise A: Statistics in Society*, 172(1), 191-211.
- Berk, R. A., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). Forecasting Murder withing a Popilation of Probationers and Parolees: A High Stakes Application of Statstical Learning. *Journal of the Royal Statistical Society (Series A)*, 172, 971-1008.
- Berk, R., Azusa, L., & Hickman, L. J. (2005). Statistical Difficulties in Determining the Role of Race in Capital Cases: A Re-analysis of Data from the State of Maryland. *Journal of Quantitative Criminology*, 21(4).
- Betancourt, H., & Blair, I. (1992). A cognition (attribution)-emotion model of violence in conflict situations. *Personality and Social Psychology Bulletin*, 18(3), 343-350.
- Beto, D., Corbett, R., & Hinzman, G. (1999). Embracing Key Strategies for a Rational Probation System . 3-6.
- Bloom, H. (2006, August). The Core Analytics of Randomized Experiments for Social Research. *MDRC Working Papers on Research Methodology*, 6.
- Bryant, R. A., Moulds, M. L., & Nixon, R. V. (2003). Cognitive behaviour therapy of acute stress disorder: a four-year follow-up. *Behaviour Research and Therapy*, 41(4), 489-494.
- Burgess, E. W. (1928). Factors Determing Success on Parole. In A. A. Bruce , A. J. Hargo, E. W. Burgess, & J. Landesco, *The Workings Of The Indeterminate Sentencing Law And The Parole System In Illinois*. Springfield, Il: State Board of Parole.
- Burt, C. H., Simons, R. L., & Simons, L. G. (2006). A Longitudinal Test of the Effects of Parenting and the General Stability of Self-Control: Negative Evidence for the General Theory of Crime. *Criminology*, 44(2), 353-396.
- Bush, J., Glick, B., Taymans, J., & Guevara, M. (2011). *Thinking for a change: Integrated cognitive behavior change program, Version 3.1*. Washington, DC: National Institute of Corrections.

- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*, 17-31.
- Carter, R., & Wilkins, L. (1976). Some conceptual models. In R. Carter, & L. Wilkins, *Probation, parole and community corrections* (pp. 391-401). New York: Wiley & Sons.
- Clark, D. A., Beck, A., & Alford, B. A. (1999). *Scientific foundations of cognitive theory and therapy of depression*. New York, NY: John Wiley & Sons.
- Clarke, R. V. (1997). Introduction. In R. V. Clarke (Ed.), *Situational Crime Prevention: Successful Case Studies* (pp. 1-43). Albany: Harrow and Heston.
- Clements, C. B. (1996). Offender Classification : Two Decades of Progress. *Criminal Justice and Behavior, 23*(1), 121-143.
- Cochrane Collaboration. (2002). *Further issues in meta-analysis*. Retrieved 01 25, 2012, from <http://www.cochrane-net.org/openlearning/html/mod14-4.htm>
- Cohen, R. L. (1995). *Probation and Parole Violators in State Prison*. Washington, D.C.: Bureau of Justice Statistics, U.S. Department of Justice.
- Copas, J. B., & Loeber, R. (1990). Relative Improvement over Chance (ROIC) for 2x2 Tables. *British Journal of Mathematical and Statistical Psychology, 43*, 293-307.
- Cottle, C. C., Lee, R. J., & Heilbrun, K. (2001). The Prediction of Criminal Recidivism in Juveniles : A Meta-Analysis. *Criminal Justice and Behavior, 28*, 367.
- Cullen, F., & Gendreau, P. (2000). Assessing correctional rehabilitation: Policy, practice, and prospects. In J. Horny, *Criminal Justice 2000* (pp. 109–175). Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.
- Dean, A., & Voss, D. (1999). *Design and Analsis of Experiments*. New York, New York: Springer.
- deRidder, D. T., Lensvelt-Mulders, G., Finkenauer, C., Stolk, F. M., & Baumeister, R. F. (2011). Taking Stock of Self-Control A Meta-Analysis of How Trait Self-Control Relates to a Wide Range of Behaviors. *Personality and Social Psychology Review, 16*(1), 76-99.
- DeRubeis, R. J., & Beck, A. T. (1988). Cognitive Therapy. In K. S. Dobson, *Handbook of Cognitive-behavioral Therapies* (pp. 273-306). New York: The Guilford Press.



- Dobson, K. S., & Khatri, N. (2000). Cognitive therapy: Looking backward, looking forward. *Journal of Clinical Psychology, 56*, 907-923.
- Dobson, K., & Block, L. (1988). Historical and Philosophical Bases of the Cognitive-behavioral Therapies. In K. Dobson, *handbook of Cognitive-Behavioral Therapies* (pp. 3-38). New York: The Guilford Press.
- Dobson, K., & Dozois, D. (2001). Historical and Philosophical Bases of the Cognitive-Behavioral Therapies. In K. Dobson, & D. Dozios, *Handbook of Cognitive-Behavioral Therapies* (pp. 3-39). New York, NY: Guilford Press.
- Dodge, K. A., & Mandel, A. D. (2012). Building Evidence for Evidence-Based Policy. *Criminology & Public Policy, 11*(3), 525-538.
- Doizos, D. J., & Beck, A. T. (2012). Cognitive Therapy. In J. D. Herbert, & E. M. Forman, *Acceptance and Mindfulness in Cognitive Behavioral Therapy* (pp. 26-57). Hoboken, NJ: John Wiley and Sons.
- Dowden, C., Blanchette, K., & Serin, R. (1999). *Anger management programming for federal male inmates: An effective intervention*. Ottawa: Research Branch, Correctional Service of Canada.
- Durlak, J. A., & Lipsey, M. W. (1991). A Practitioner's Guide to Meta-Analysis. *American Journal of Community Psychology, 19*(3).
- Ebener, P., & Kilmer, B. (2003). *Linking Drug Users with Treatment: Admissions Counselors Describe the Barriers*. Santa Monica, CA: Drug Policy Research Center, RAND .
- Elliott-Engel, A. (2011, April 15). Shortage of Probation Officers May Imperil Reforms, FJD Leaders Say. *The Legal Intelligencer*.
- Ellis, A. (1957). Rational psychotherapy and individual psychology. *Journal of Individual Psychology, 13*, 38-44.
- Ellis, A. (1969). Rational-emotive therapy. *Journal of Contemporary Psychotherapy, 1*(2), 82-90.
- Ellis, A. (1980). Rational-Emotive Therapy and Cognitive Behavior Therapy: Similarities and Differences. *Cognitive Therapy and Research, 4*(4), 325-340.
- Emery, M., Gasswint, B., Hartman, M., & Lategan, D. (2008). *Annual Statistical Report, Pennsylvania Department of Corrections, 2007*. Harrisburg, Pennsylvania: Pennsylvania Department of Corrections.

- Evans, T. D., Burton, V. S., Dunwat, R. G., & Benson, M. L. (1997). The Social Consequences Of Self-Control: Testing The General Theory Of Crime. *Criminology*, 35(3), 475-504.
- Evans, T. D., Cullen, F. T., Butron, V. S., Dunaway, R. G., & Benson, M. L. (1997). The social consequences of self-control: testing the General Theory of Crime. *Criminology*, 35(3), 475-504.
- Fagan, J., & Forst, M. (1996). Risks, Fixers, and Zeal: Implementing Experimental Treatments for Violent Juvenile Offenders. *The Prison Journal*, 76(1), 22, 22-59.
- Fehrenback, P. A., & Thelen, M. A. (1982). Behavioral Approaches to Treatment of Aggressive Disorders. *Behavior Modification*, 6(4), 465-497.
- Feinstein, A. (1991). Intent-to-treat policy for analyzing randomized trials: Statistical distortions and neglected clinical challenges. In J. Cramer, & B. Spilker, *Patient compliance in medical practice and clinical trials* (pp. 359–370). New York, NY: Raven Press.
- Fergusson, D., Aaron, S., Guyatt, G., & Hebert, P. (2002). Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *British Journal of Medicine*, 325(7365), 652–654.
- Foa, E., Hearst-Ikeda, D., & Perry, K. (1995). Evaluation of a Brief Cognitive-Behavioral Program for the Prevention of Chronic PTSD in Recent Assault Victims. *Journal of Consulting and Clinical Psychology*, 63(6), 948-955.
- Free, M. (1999). *Cognitive therapy in groups : guidelines and resources for practice*. Chichester: John Wiley.
- Fulton, B., Stone, S., & Gendreau, P. (1994). *Restructuring intensive supervision programs: Applying what works*. Lexington, KY: American Probation and Parole Association.
- Garrett, C. J. (1985). Effects of Residential Treatment on Adjudicated Delinquents: A Meta-Analysis. *Journal of Research in Crime and Delinquency*, 22(4), 287-308 .
- Gartin, P. (1995). Dealing with design failures in randomized field experiments: Analytic issues regarding the evaluation of treatment effects. *Journal of Research in Crime and Delinquency*, 32(4), 425-445.
- Geis, G. (2000). On the absence of self-control as the basis for a general theory of crime:A critique. *Theoretical Criminology*, 4(1), 35–53.

- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the Predictors of Adult Offender Recidivism: What Works? *Criminology*, 34(4), 575-608.
- Gifford, S. L. (2002). *Justice expenditures and employment in the United States, 1999*. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs.
- Gill, C. E. (2010). *The effects of sanction intensity on criminal conduct: A randomized low-intensity probation experiment*. Philadelphia, PA: University of Pennsylvania.
- Glass, G. V. (1977). Integrating Findings: The Meta-Analysis of Research. *Review of Research in Education*, 5, 351-379.
- Glaze, L., & Bonczar, T. (2009). *Probation and parole in the United States*. Washington, DC: U.S. Department of Justice Office of Justice Programs.
- Gleuck, S., & Gleuck, E. T. (1930). *500 Criminal Careers*. New York: Knopf.
- Goldiamond, I. (2002). Toward a constructional approach to social problems: Ethical and constitutional issues raised by applied behavior analysis. *Behavior and Social Issues*, 11, 108-197.
- Goldstein, A., & Glick, B. (1994). Aggression Replacement Training: Curriculum and Evaluation. *Simulation and Gaming*, 25(1), 9-26.
- Gottfredson, D., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Gottfredson, S., & Moriarty, L. (2006). Statistical Risk Assessment: Old Problems and New Applications. *Crime & Delinquency*, 52, 180.
- Grasmick, H., Tittle, C., Bursik, R., & Arneklev, B. (1993). Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency*, 30(1), 5-29.
- Grave, J., & Blissett, J. (2004). Is cognitive behavior therapy developmentally appropriate for young children? A critical review of the evidence. *Clinical Psychology Review*, 24(4), 399-420.
- Greevy, R., Silber, J., Cnaan, A., & Rosenbaum, P. (2004). Randomization Inference with Imperfect Compliance in the ACE-Inhibitor after Anthracycline Randomized Trial. *Journal of the American Statistical Society*, 99(465), 7-15,11.
- Gross, D., & Fogg, L. (2004). A Critical Analysis of the Intent-to-Treat Principle in Prevention Research. *The Journal of Primary Prevention*, 25(4), 475-489.

- Guan, W. (2003). Bootstrapped Standard Errors. *The Stata Journal*, 3(1), 71-80.
- Guynes, R. (1988). *Difficult clients, large caseloads plague probation, parole agencies*. Washington, DC: U.S. Department of Justice, National Institute of Justice Research in Action.
- Hall, G. C. (1995). Sexual offender recidivism revisited: A meta-analysis of recent treatment studies. *Journal of Consulting and Clinical Psychology*, 63(5), 802-809.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting Relapse: A meta-Analysis of Sexual Offender Recidivism Studies. *Journal of Consulting and Clinical Psychology*, 66(2), 348-362.
- Hanson, R. K., & Morton-Bourgon, K. E. (2005). The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies. *Journal of Consulting and Clinical Psychology*, 73(6), 1154–1163.
- Hart, H. (1923). Predicting Parole Success. *Journal of the American Institute of Criminal Law and Criminology*, 14(3), 405-413.
- Hayes, S. C. (2004). Acceptance and Commitment Therapy, Relational Frame Theory, and the Third Wave of Behavioral and Cognitive Therapies. *Behavior Therapy*, 35, 639-665.
- Heckman, J. (1995). *Randomization as Instrumental Variable*. Cambridge, MA: National Bureau of Economic Research.
- Herbert, J. D., & Foreman, E. M. (2012). The Evolution of Cognitive Behavioral Therapy: The Rise of Psychological Acceptance and Mindfulness. In *Acceptance and Mindfulness in Cognitive Behavioral Therapy* (pp. 3-26). Hoboken, NJ: John C. Wiley & Sons.
- Hesterberg, T. C., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (1995). Bootstrap methods and permutation tests. In D. Moore, & G. McCabe, *Introduction to the Practice of Statistics*. W.H. Freeman & Company.
- Hirschi, T. (1986). On the Compatibility of Rational Choice and Social Control Theories of Crime. In D. Cornish, & R. Clark (Eds.), *Reasoning Criminal* (pp. 105-118). New York, NY: Springer-Verlag.
- Hoffman, S. G., Meuret, A. E., Rosenfield, D., Suvak, M. K., Barlow, D. H., Gorman, J. M., . . . Woods, S. W. (2007). Preliminary Evidence for Cognitive Mediation During Cognitive–Behavioral Therapy of Panic Disorder. *Journal of Consulting and Clinical Psychology*, 75(3), 374–379.

- Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Journal of Medicine*, 319, 670-674.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Metanalysis*. Thousand Oaks, CA: Sage Publications.
- IBM Corp. (2012). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.
- Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.
- Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-475.
- Imbens, G. W., & Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1), 109-126.
- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the Quality of Reports of Randomized Clinical Trials: Is Blinding Necessary? *Controlled Clinical Trials*, 17(1), 1-12.
- Johnson, G., & Hunter, R. (1992). *Evaluation of the Specialized Drug Offender Program*. Boulder, CO: University of Colorado, Center for Action Research.
- Jüni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of controlled clinical trials. *British Journal of Medicine*, 323(7303), 42-46.
- Kazemian, L. (2007). Desistance From Crime : Theoretical, Empirical, Methodological, and Policy Considerations. *Journal of Contemporary Criminal Justice*, 23(5), 7.
- Kennedy, P. (2003). *A Guide to Econometrics* (Vol. 5th Ed.). Cambridge, MA: MIT Press.
- Killen, J., & al., e. (2008). Extended cognitive behavior therapy for cigarette smoking cessation. *Addiction*, 103(8), 1381-1390.
- Landenberger, N. N., & Lipsey, M. W. (2005). The Positive Effects of Cognitive-Behavioral Programs for Offenders: A Meta-Analysis of Factors Associated with Effective Treatment. *Journal of Experimental Criminology*.
- LaTessa, E., & Vito, G. F. (1988). The effects of intensive supervision with shock probationers. *Journal of Criminal Justice*, 16(4), 319 – 330.

- Liddle, H., Dakof, G., Turner, R., Craig, H., & Paul, G. (2008). Treating adolescent drug abuse: a randomized trial comparing multidimensional family therapy and cognitive behavior therapy. *Addiction, 103*, 1160-1670.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. L. Light, . . . F. Mostellar, *Meta-analysis for Explanation* (pp. 83-127). New York: Russel Sage Foundation.
- Lipsey, M. W. (1995). What do We Learn from 400 Research Studies on the Effectiveness of Treatment with Juvenile Delinquents. In J. McGuire, *What Works: Reducing Reoffending* (pp. 62-78). Chichester: John Wiley & Sons.
- Lipsey, M. W. (2008). *The Arizona Standard Program Evaluation Protocol (SPEP) for assessing the effectiveness of programs for juvenile probationers: SPEP ratings and relative recidivism reduction for the initial SPEP sample*. Nashville, TN: Vanderbilt University Center for Evaluation Research and Methodology.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A metaanalytic overview. *Victims and Offenders, 4*, 124–47.
- Lipsey, M. W., & Landenberger, N. A. (2005). Cognitive-behavioral interventions: A meta-analysis of randomized controlled studies. In B. C. Walsh, & D. P. Farrington , *Preventing crime: What works for children, offenders, victims, and places* (pp. 57-71). Berlin: Springer.
- Lipsey, M. W., & Wilson, D. B. (1993). The Efficacy of Psychological, Educational and Behavioral Treatment: Confrimation from Metaanalysis. *American Psychologist, 48*(12), 1181-1209.
- Lipsey, M. W., Chapman, G. L., & Landenberger, N. A. (2001). Cognitive-Behaviorial Programs for Offenders. *Annals of the American Academy of Political and Social Science, 144*-157.
- Lipsey, M. W., Howell, J. C., & Tidd, S. T. (2007). *The Standardized Program Evaluation Protocol (SPEP): A practical approach to evaluating and improving juvenile justice programs in North Carolina. Final evaluation report*. Nashville, TN: Center for Evaluation Research and Methodology.
- Lipsey, M. W., Howell, J. C., Kelly, M. R., Chapman, G., & Carver, D. (2010). *Improving the Effectiveness of Juvenile Justice Programs: A New Perspective on Evidence-Based Practice*. Nashville, TN: Peabody Research Institute,.

- Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). *Effects of Cognitive-Behavioral Programs for Criminal Offenders*. Oslo, Norway: Campbell Systematic Reviews.
- Lipsey, M., & Hurly, S. M. (2009). Design Sensitivity: Statistical Power for Applied Experimental Research. In L. Bickman, & D. J. Rogg, *The SAGE Handbook of Applied Social Research Methods* (pp. 44-76). Thousand Oaks, CA: SAGE.
- Lipsey, M., Chapman, G., & Landenberger, N. (2001). Cognitive behavioral programs for offenders. *The Annals of the American Academy of Political and Social Science*, 578, 144-157.
- Little, G. (2000). Cognitive-behavioral treatment of offenders: A comprehensive review of MRT outcome research. *Addictive Behaviors Treatment Review*, 2(1), 12-21.
- Little, G., & Robinson, K. (1993). Cognitive behavioral treatment of felony drug offenders: A five-year recidivism report. *Psychological Reports*, 73(3), 1089-1090.
- Lösel, F. (1995). The Efficacy of Correctional Treatment: A Review and Synthesis of meta-evaluations. In J. McGuire, *What Works: Reducing Reoffending* (pp. 79-103). Chichester: John Wiley and Sons.
- Lowencamp, C. T., Hubbard, D., Makarios, M. D., & Latessa, E. J. (2009). A Quasi-Experimental Evaluation of Thinking for a Change : A "Real-World" Application. *Criminal Justice and Behavior*, 36, 137.
- Lowencamp, C. T., Latessa, E. T., & Smith, P. (2006). Does Correctional Program Quality Really Matter?: The Impact of Adhering to the Principles of Effective Intervention. *Criminology and Public Policy*, 5(3), 575-594.
- Lowenkamp, C., Latessa, E., & Holsinger, A. (2006). The Risk Principal in Action: What Have We Learned From 13,676 Offenders and 97 Correctional Programs. *Crime and Delinquency*, 52(1), 77-93.
- MacDonald, J., Morral, A., & Piquero, A. (2011). Socially Desirable Response Bias in Criminology: An Example of its Effect in Testing the Effects of Self-Control. In J. M. MacDonald (Ed.), *Measuring Crime and Criminality* (Vol. 17, pp. 21-36). New Brunswick, NJ: Transaction Publishers.
- MacKenzie, D. (1997). Criminal Justice and Crime Prevention. In L. Sherman, D. Gottfredson, D. Mackenzie, P. Eck, & S. Bushway, *Preventing Crime: What Works, What Doesn't, What's Promising* (p. 447). Washington, DC: Office of Justice Programs.

- MacKinnon, D. P., & Fairchild, A. J. (2009). Current Directions in Mediation Analysis. *Current Directions in Psychological Science, 18*(1), 16–20.
- Mahoney, M., & Arnkoff, D. (1978). Cognitive and self-control therapies. In S. Garfield, & A. Bergin, *Handbook of psychotherapy and behavior change* (pp. 689-722). New York: Wiley.
- Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest, 35*, 22-54.
- Matthews, A. (1997). Information processing biases in emotional disorders. In D. M. Clark, & C. G. Fairburn, *Science and Practice of Cognitive Behavioral Therapy* (pp. 47-67). Oxford University Press: New York, NY.
- Meichenbaum, D. (1977). *Cognitive-Behavior Modification: An Integrative Approach*. New York: Plenum Press.
- Meichenbaum, D. (1977). *Cognitive-Behavior Modification: An Integrative Approach*. New York: Plenum Press.
- Meichenbaum, D., & Cameron, R. (1971). Training impulsive children to talk to themselves: A means of developing self-control. *Journal of Abnormal Psychology, 77*, 115-126.
- Muraven, M., Baumeister, R. Y., & Dianne, M. T. (1999). Longitudinal Improvement of Self-Regulation Through Practice: Building Self-Control Strength Through Exercise. *Journal of Social Psychology, 139*(4), 446-457.
- Na, C., & Paternoster, R. (2012). Can Self-Control Change Substantially Over Time? Rethinking the Relationship between Self- and Social-Control. *Criminology, 50*(2), 427-462.
- Nagin, D. S., & Paternoster, R. (1993). Enduring Individual Differences and Rational Choice Theories of Crime. *Law & Society Review, 27*(3), 467-496.
- Noble, R. (2012). Cognitive-Skills Program Development: Narrative Report.
- Noble, R., & Hyatt, J. (2010). Developing A Cognitive-Behavioral Therapy Targeted to Young, Urban Offenders. *American Society of Criminology Annual Meeting*.
- Nurius, P. S., & Yeaton, W. H. (1987). Research synthesis reviews: An illustrated critique of Research synthesis reviews: An illustrated critique of. *Clinical Psychology Review, 7*, 695-714.



- Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The Effects of Behavioral/Cognitive-Behavioral Programs on Recidivism. *Crime and Delinquency*, 48(3), 476-496.
- Petersilia, J. (1985). Probation and felony offenders. *Federal Probation*, 49, 4-9.
- Petersilia, J., & Deschenes, E. (1992). Evaluating Intensive Supervision Probation/Parole (ISP) for Drug Offenders. *Crime and Delinquency*, 38(4), 539-556.
- Petersilia, J., & Turner, S. (1990). *Intensive supervision for high-risk probationers*. The Rand Corporation: Santa Monica, CA.
- Petersilia, J., & Turner, S. (1993). Intensive Probation and Parole. In M. Tonry (Ed.). Chicago, Illinois: University of Chicago Press.
- Pew Center on the States. (2008). *One in One Hundred: Behind Bars in America*.
- Pew Center on the States. (2009). *One in Thirty One: The Long Reach of American Corrections*. Pew Center on the States. Washington, DC: Pew Center on the States.
- Pilliavin, I., Thornton, C., Gartner, R., & Matsueda, R. L. (1986). Crime, Deterrence, and Rational Choice. *American Sociological Review*, 51(Feb.), 101-119.
- Platt, J. J., & Prout, M. F. (1987). Cognitive-Behavioral Theory and Interventions for Crime and Delinquency. In E. K. Morris, & C. J. Braukmann, *Behavioral Approaches to Crime and Delinquency* (pp. 477-498). New York: Plenum Press.
- Polaschek, D. L., Wilson, N. J., Townsend, M. R., & Daly, L. R. (2005). Cognitive-Behavioral Rehabilitation for High-Risk Violent Offenders : An Outcome Evaluation of the Violence Prevention Unit. *Journal of Interpersonal Violence*, 20(12), 1611-1627.
- Pratt, T. C., & Cullen, F. T. (2000). The empirical status of Gottfredson and Hirschi's General Theory of Crime: A meta-analysis. *Criminology*, 38(3), 931-963.
- Redondo, S., Sanchez-Meca, J., & Garrido, V. (1999). The influen of treatmnet programmes on the recidivism of juveile and adult offenders: An european meta-analytic review. *Psychology, Crime and Law*, 5(3), 251-278.
- Rezmovic, E. L. (1984). Assessing Treatment Implementation Amid the Slings and Arrows of Reality. *Evaluation Review*, 8(2), 187-204.
- Robinson, D. (1995). *The impact of cognitive skills training on post-release recidivism among Canadian federal offenders*. Ottawa: Correctional Service of Canada.

- Robinson, D. (1995). *The Impact of Cognitive Skills Training on Post-Release Recidivism among Canadian Federal Offenders*. Ottawa, Ontario: Correctional Research and Development, Correctional Service Canada.
- Rosenthal, R., & Rubin, D. B. (1982). A Simple, General Purpose Display of Magnitude of Experimental Effect. *Journal of Educational Psychology*, 47(2), 166-169.
- Ross, R. R., Fabiano, E. A., & Ewles, C. D. (1988). Reasoning and rehabilitation. *International Journal of Offender Therapy and Comparative Criminology*, 32, 29-36.
- Rubin, D. (1998). More powerful randomization-based p-values in double-blind trials with non-compliance. *Statistical Medicine*, 17, 371-385.
- Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *Annals of Statistics*, 6(1), 34-58.
- Sampson, R. J. (2010). Gold Standard Myths: Observations on the Experimental Turn in Quantitative Criminology. *Journal of Quantitative Criminology*, 26, 489-500.
- Sampson, R. J., & Laub, J. H. (1990). Crime and Deviance over the Life Course: The Salience of Adult Social Bonds. *American Sociological Review*, 55(5), 609-627.
- Sampson, R. J., & Laub, J. H. (2005). A Life-Course View of the Development of Crime. *ANNALS of the American Academy of Political and Social Science*, 602(12), 73-79.
- Sampson, R. J., Laub, J. H., & Wimer, C. (2006). Does Marriage Reduce Crime? A Counterfactual Approach to Within Individual Causal Effects. *Criminology*, 44(3), 65-109.
- Sanchez-Meca, J. (1997). Methodological Issues in the Meta-analysis of Correctional Treatment. In S. Redondo, *Advances in Psychology and Law: International Contributions* (pp. 487-498). Berlin: Walter de Gruyter.
- Schmidt, F., & Hunter, J. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. Harlow, S. Mulaik, & J. Steiger, *What if there were no Significance Tests?* (pp. 37-64). London: Lawrence Erlbaum.
- Schulz, K., & Grimes, D. (2002). Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet*, 359, 781-785.
- Schwartz, R. M. (1982). Cognitive-behavior modification: A conceptual review. *Clinical Psychology Review*, 2(3), 267-293.

- Sekhon, J. S. (2007). The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods. In J. Box-Steffensmeier, H. Brady, & D. Collier, *The Oxford Handbook of Political Methodology* (pp. 271-299). Oxford, UK: Oxford University Press.
- Sheiner, L. B., & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics*, 57, 6-15.
- Sheldon, J. (1987). Legal and Ethical Issues in the Behavioral Treatment of Adult and Juvenile Offenders. In E. Morris, & C. Braukman, *Behavioral Approaches to Crime and Delinquency* (pp. 543-573). New York: Plenum Press.
- Sherman, L. (2006). To Develop and Test: The Inventive Difference Between Evaluation and Experimentation. *Journal of Experimental Criminology*, 2(3), 393-406.
- Sherman, L. (2010). An Introduction of Experimental Criminology. In A. R. Piquero , & D. Weisburd, *Handbook of Quantitative Criminology* (pp. 399-437). New York, NY: Springer.
- Sherman, L. W. (1997). *Preventing crime: What works, what doesn't, what's promising*.
- Sherman, L. W. (2006). Preventing Homicide With Probation And Parole: A Randomized, Controlled Trial, Phase One (Predicting Homicide and Serious Crime).
- Sherman, L. W. (2007). The Power Few Hypothesis: Experimental Criminology and the Reduction of Harm. *Journal of Experimental Criminology*, 3(4).
- Sherman, L. W., & Berk, R. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review*, 49(2), 261-272.
- Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (2002 ). Preventing Crime. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie, *Evidence-based Crime Prevention*. London: Routledge.
- Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising*.
- Shusik-Richards, C. (2010). Philadelphia's Jail Population: A Research Initiative. *American Jails*, 41-48.
- Silver, E., & Miller, L. (2002). A Cautionary Note on the Use of Actuarial Risk Assessment Tools for Social Control. *Crime and Delinquency*, 48(132).

- Skeem, J., & Manchak, S. (2008). Back to the Future: From Klockars' Model of Effective Supervision to Evidence-Based Practice in Probation. *Journal of Offender Rehabilitation, 47*(3), 220-247.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Smith, M. L., & Glass, G. V. (1977). Meta-Analysis of Psychotherapy Outcome Studies. *American Psychologist, 752-770*.
- Sommer, A., & Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistical Medicine, 10*, 45–52.
- Spano, R. (2002). *Potential Sources of Observer Bias in Observational Studies of Police*. Washington, DC: National Institute of Justice.
- Spruance, L. M., van Voorhis, P., Listwin, S. J., Pealer, J., & Seabrook, R. (n.d.). *The Georgia Cognitive Skills Experiment: Phase II Process Evaluation*.
- StataCorp. (2011). *Stata Statistical Software: Release 12 SE*. College Station, TX: StataCorp LP.
- Stewart, C., Quinn, A., Pleyer, S., & Emmerson, B. (2009). Comparing Cognitive Behavior Therapy, Problem Solving Therapy, and Treatment as Usual in a High Risk Population. *Suicide & Life-Threatening Behavior, 39*(5), 538-547.
- Stock, J. H., & Watson, M. W. (2002). *Introduction to Econometrics*. Boston, MA: Addison-Wesley.
- Sussman, J. B., & Hayward, R. A. (2010). An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. *British Medical Journal, 340*, c2073.
- Sutherland, E. (1939). *Principles of Criminology*. Philadelphia, PA: Lippincott.
- Sykes, G., & Matza, D. (1957). Techniques of Neutralization: A Theory of Delinquency. *American Sociological review, 22*, 664-670.
- Tangney, J. P., Mashek, D., & Stuewig, J. (2007). Working at the Social-Clinical-Community-Criminology Interface: The George Mason University Inmate Study. *Journal of Social and Clinical Psychology, 26*(1), 1-21.
- Taxman, F. (2002). Supervision: Exploring the dimensions of effectiveness. *Federal probation, 66*(2), 14-27.
- Taxman, F. S., & Marlowe, D. (2006). Risk, Needs, Responsivity: In Action or Inaction? *Crime & Delinquency, 52*(3), 3-6.

- Taxman, F., Thanner, M., & Weisburd, D. (2006). Risk, Need, And Responsivity (RNR): It All Depends. *Crime and Delinquency*, 52(1), 28-51.
- Teasdale, J. D. (1997). The relationship between cognition and emotion: the mind-in-place in mood disorders. In D. M. Clark, & C. G. Fairburn, *Science and practice of Cognitive Behavioral Therapy* (pp. 67-93). Oxford: Oxford Press.
- Thanner, M., & Taxman, F. (2003). Responsivity: the value of providing intensive services to high-risk offenders. *Journal of Substance Abuse Treatment*, 24(2), 137-147.
- Tittle, C. R., Ward, D. A., & Grasmick, H. G. (2003). Gender, Age, and Crime/Deviance: A Challenge to Self-Control Theory. *Journal of Research in Crime and Delinquency*, 40(4), 426-453.
- Tong, L. J., & Farrington, D. P. (2006). Effectiveness of the “Reasoning and Rehabilitation” programme. *Psychology, Crime & Law*, 12(1), 3-24.
- Treanor, J., Hayden, F., Vrooman, P., Barbarash, R., Bettis, R., & Riff, D. (2000). Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza A randomized controlled trial. *Journal of the American Medical Association*, 283, 1016-1024.
- Tudor, A. (2013). *APPD Population Census, January 2013*. Philadelphia, PA: Philadelphia Adult Probation and Parole Department.
- Turner, S., & Petersilia, J. (1992). Focusing on High-Risk Parolees: An Experiment To Reduce Commitments to the Texas Department of Corrections. *The Journal of Research in Crime and Delinquency*, 29(1), 34-61.
- Urbach, P. (1985). Randomization and the Design of Experiments. *Philosophy of Science*, 52(2), 256-273.
- van Voorhis, P. S. (2004). The Georgia cognitive skills experiment: A replication of Reasoning and Rehabilitation. *Criminal Justice and Behavior*(31), 282- 305.
- van Voorhis, P., Spruance, L. M., Ritchey, P. N., Listwan, S. J., & Seabrook, R. (2004). The Georgia cognitive skills experiment: A replication of Reasoning and Rehabilitation. *Criminal Justice and Behavior*(31), 282- 305.
- van Voorhis, P., Spruance, L. M., Ritchie, P. N., Johnson-Listwan, S., Seabrook, R., & Pealer, J. (n.d.). *The Georgia Cognitive Skills Experiment: Outcome Evaluation*.
- van Voorhis, P. (1999).

- Wanberg, K. W., & Milkman, H. B. (2006). *Criminal Conduct and Substance Abuse Treatment: Strategies for Self-Improvement and Change; The participant's workbook*. Thousand Oaks, CA: Sage.
- Ward, T., Hudson, S., Johnston, L., & Marshall, W. (1997). Cognitive distortions in sex offenders: An integrative review. *Clinical Psychology Review, 17*, 797-507.
- Warr, M. (1998). Life-Course Transitions and Desistance from Crime. *Criminology, 36*(2), 183–216.
- Welsh, B. C., & Farrington, D. P. (2000). Monetary Costs and Benefits of Crime Prevention Programs. *Crime and Justice, 27*, 305-361.
- Wikström, P.-O. H., & Treiber, K. (2007). The Role of Self-Control in Crime Causation. *European Journal of Criminology, 4*(2), 237–264.
- Wilson, D. B., & Lipsey, M. W. (1993). The Efficacy of Behavioral, Education and Psychological Treatment : Confirmation from Meta-analysis. *American Psychologist, 48*(12), 1181-12-09.
- Wilson, D., Bouffard, L. A., & Mackenzie, D. (2005). A Quantitative Review of Structured, Group-Oriented, Cognitive-Behavioral Programs for Offenders. *Criminal Justice and Behavior, 32*, 172.
- Wilson, D., Bouffard, L., & Mackenzie, D. (2005). Quantitative review of cognitive behavioral programs. *Criminal Justice and Behavior, 32*(2), 172-204.
- Wright, B. R., Caspi, A., Moffitt, T. E., & Paternoster, R. (2004). Does the Percieved Risk of Punishment Deter Criminall Prone Individuals? Rational Choice, Self Contol and Crime. *Journal of Research in Crime and Delinquency, 41*(2), 180-213.