Screening for Violent Recidivism in the Juvenile Justice System: Establishing Accuracy

by

Larry Michael Kalb

B.A., State University of New York at Stony Brook, 1995

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Larry Michael Kalb

It was defended on

June 9th, 2006

and approved by

Rolf Loeber, Ph.D.
Committee Chair

Michael Pogue-Geile, Ph.D.

Daniel Shaw, Ph.D.

Robert McCall, Ph.D.

Jacqueline Cohen, Ph.D.

Screening for Violent Recidivism in the Juvenile Justice System: Establishing Accuracy

Larry Michael Kalb, B.A.

University of Pittsburgh, 2006

Abstract

Recently there has been increased interest among researchers and those working in the Juvenile Justice System to develop risk assessment instruments (RAIs) to predict violence among juveniles. Although there are violence prediction instruments designed for use with adults that have been widely studied, few such instruments exist that have been specifically designed for use with juvenile populations, and none have been sufficiently evaluated for predictive accuracy. The purpose of this study was to evaluate whether a RAI could be constructed to accurately predict a specific type of recidivism, violent recidivism, among juvenile offenders. Two RAIs were created using separate methods of instrument construction, the Burgess method and predictive attribute analysis. Both instruments were created and tested for predictive accuracy using a sub-sample from the Pittsburgh Youth Study. Results showed that both RAIs were able to predict violent recidivism to a degree, however, neither instrument performed well when tested on the validation sample. The limitations of commonly used measures of predictive accuracy are discussed in relation to the findings reported here and in other validation studies with similar instruments. Recommendations for constructing RAIs to predict violence among juveniles are outlined, and some cautionary remarks are made with regard to their use.

Acknowledgements

First, I would like to thank my advisor, Dr. Rolf Loeber, for all the guidance and support he has given me. He continuously provided research opportunities for me throughout my graduate career, and I am extremely grateful he allowed me to use data from the Pittsburgh Youth Study for this dissertation. I would also like to thank Dr. Magda Stouthamer-Loeber for all of her personal encouragement, and assistance with the PYS data. I would like to thank my committee for all of their encouragement and contributions to the project. Additionally, I would like to thank all of my colleagues in the Department of Psychology for their words of encouragement and support. On a personal note, I would like to thank my colleague and close personal friend Judy Thompson for all her patience, and personal and professional support she has given me. She was always there for me whenever I needed guidance, and never failed to provide a positive outlook in times when it was most needed. I would like to thank my friend Michelle Stocker for supporting me in many ways and for all her words of wisdom. I would like to thank my parents who have always provided encouragement and unconditional support. I would have achieved few of my goals without their unwavering assistance. Finally, I would also like to thank Donna Marie Drewelow for all her understanding and insights.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Introduction

Risk assessment instruments (RAIs) are frequently used by decision makers in the U.S. Juvenile Justice System (JJS) to guide judicial action and treatment referral to various mental health agencies. Using RAIs to assess/measure the degree of risk for future serious offending provides the opportunity to employ more intensive intervention for those at highest risk for recidivism, an idea that has received more and more attention as the effectiveness of treatment programs aimed at reducing delinquency has increasingly been established (Lipsey & Derzon, 1998). In the past decade researchers have begun to stress the importance of using RAIs to aid decision makers within the juvenile justice system who must evaluate risk, particularly during the early stages of processing (Dembo & Brown, 1994). Along these lines, the Office of Juvenile Justice and Delinquency Prevention has developed a comprehensive strategy designed to stop juvenile delinquents from proceeding further along established delinquent developmental pathways (Loeber and Hay, 1994; Loeber et al., 1993). This strategy emphasizes the use of effective RAIs to identify juveniles most at risk for future serious offending (Wilson & Howell, 1993).

Risk assessment instruments typically attempt to measure a variety of child, family, peer, school, and neighborhood characteristics, along with delinquency history factors such as type of offense, that have been found to be associated with serious adolescent delinquency (Hawkins et al., 1998; Lipsey & Derzon, 1998; Loeber & Dishion, 1983). A youth's level of risk can be used to inform disposition and placement decisions, with the hope of preventing crime by matching a youth's level of risk with appropriate treatment programs and interventions designed to both rehabilitate the delinquent youth and to protect the general public from further victimization (Le Blanc, 1998). While interest in the design and implementation of RAIs has increased, research in this area is still in its initial stages (Howell, 2001). Although recent studies validating the predictive utility of RAI's have increased for adults (Andrews, 1996; Monahan et al, 2001; Steadman et al, 2000), the same cannot be said for RAIs designed for use with juvenile offenders.

The purpose of this study was to evaluate whether a RAI could be constructed to accurately predict a specific type of recidivism, violent recidivism, among juvenile offenders. More specifically, using a series of variables associated with violence, two actuarial models are compared and contrasted for their utility as screening devices to predict future violence among 1^{st} arrest and 2^{nd} arrest juvenile offenders. Prior to addressing the primary question of this study, a brief overview of the procedures involved in processing a delinquent through the juvenile court is presented. As a way of highlighting issues specific to RAIs used within the JJS, a review of current risk assessment practices, such as how RAIs aid decision makers within the JJS, and what those decisions are will be explored. RAIs are used by juvenile courts in a number of ways that are unique to screening instruments in general. For instance, results from RAIs are often used to directly influence decisions about placement and intervention of delinquents, and there are also a unique set of variables or predictors that are available for measurement to court workers. Court workers and judges have a limited amount of time and resources to evaluate risk factors for recidivism and are constrained by having access to only a limited number of informants who can supply information regarding the offender's characteristics and behavior. A review of research evaluating procedures for designing, constructing, and evaluating the effectiveness of RAI's will also be presented. This review will focus specifically on the predictors of delinquency and violence included in RAIs, and review clinical and actuarial methods of risk prediction. From this review a set of guidelines for creating a RAI will be presented, and subsequently used to shape the analyses in this study to test the primary question: can a prediction equation that makes use of variables associated with violent recidivism be constructed to predict which juvenile offenders will commit violent acts? Using this brief review as a guide, a number of design and application recommendations for RAIs are presented.

For this study, violent recidivism is the outcome of interest. This choice of outcome as opposed to all recidivism for example, is based on many discussions with court administrators and probation officers within the Allegheny County Juvenile Court system. From these discussions it was determined that what was most needed was an instrument that could identify a small group of individuals from among the entire juvenile offending population who are most at risk for committing violent acts in the future. This information would suggest to judges which individuals are most likely to commit the most serious crimes following release, and also the level of security that might be needed to confine such high risk individuals. It is important to the courts to know who is most likely to commit the most serious crimes so that attempts can be made to reduce the cost of such crimes to society. When making placement decisions the courts will also benefit from knowing who is most likely to become violent in placement, because then these individuals can be placed in the more secure and heavily monitored environment. It is also important to the courts to know who is most at risk for future violence and thus requires the more intensive interventions and treatments that are specifically designed to curb violent behavior.

To evaluate whether a prediction equation can be constructed to predict which juvenile offenders will later be adjudicated violent, a subset of subjects from the Pittsburgh Youth Study (PYS), a prospective longitudinal study including approximately 1500 boys will be examined. This group of subjects is an ideal sample by which to evaluate these questions, given that boys make up the majority of juveniles referred to courts for delinquent offenses (Snyder, 1996), and that official court records were accessible on youth in the PYS.

Overview of Juvenile Justice System Procedures

A youth can come into contact with the juvenile justice system in one of two ways, because the court suspects they may have committed a crime, or because the court suspects the child is being abused or neglected in some way. It is possible for a child to have simultaneous delinquent and dependent status. If a youth is determined to be in neglect, meaning that his or her current caretaker is not appropriately caring and providing for them in some way such as not providing adequate supervision, shelter, food, or a safe environment, he or she is typically placed either with a relative when possible or a shelter, group home, or foster home. Figure 1 shows the processing of youths who enter the JJS, and how they may exit the JJS by either having their charges dropped, their case dropped, or completing some type of intervention. A youth's contact with the juvenile system for delinquency begins with an arrest by a member of a law enforcement agency such as city or school police. Once a youth has been arrested, he either is quickly released or charged with some type of delinquency and then either released or detained. The type and severity of his charge(s) along with other factors, such as criminal history, adult support, and quality of home environment, will determine how he is processed. For serious (i.e., felony class) crimes, the youth most likely will be fingerprinted and then taken to a detention facility for further processing. For less serious (i.e., misdemeanor) crimes, the youth will probably not be fingerprinted and only may be taken to the police zone office for processing. During processing, the police determine whether the youth has any papers outstanding against him such as warrants or orders to place the youth in detention. If the youth has an outstanding paper against him or if the police believe him to be sufficiently dangerous to others, he will be placed in a detention center until a formal detention hearing can be held to determine the youth’s

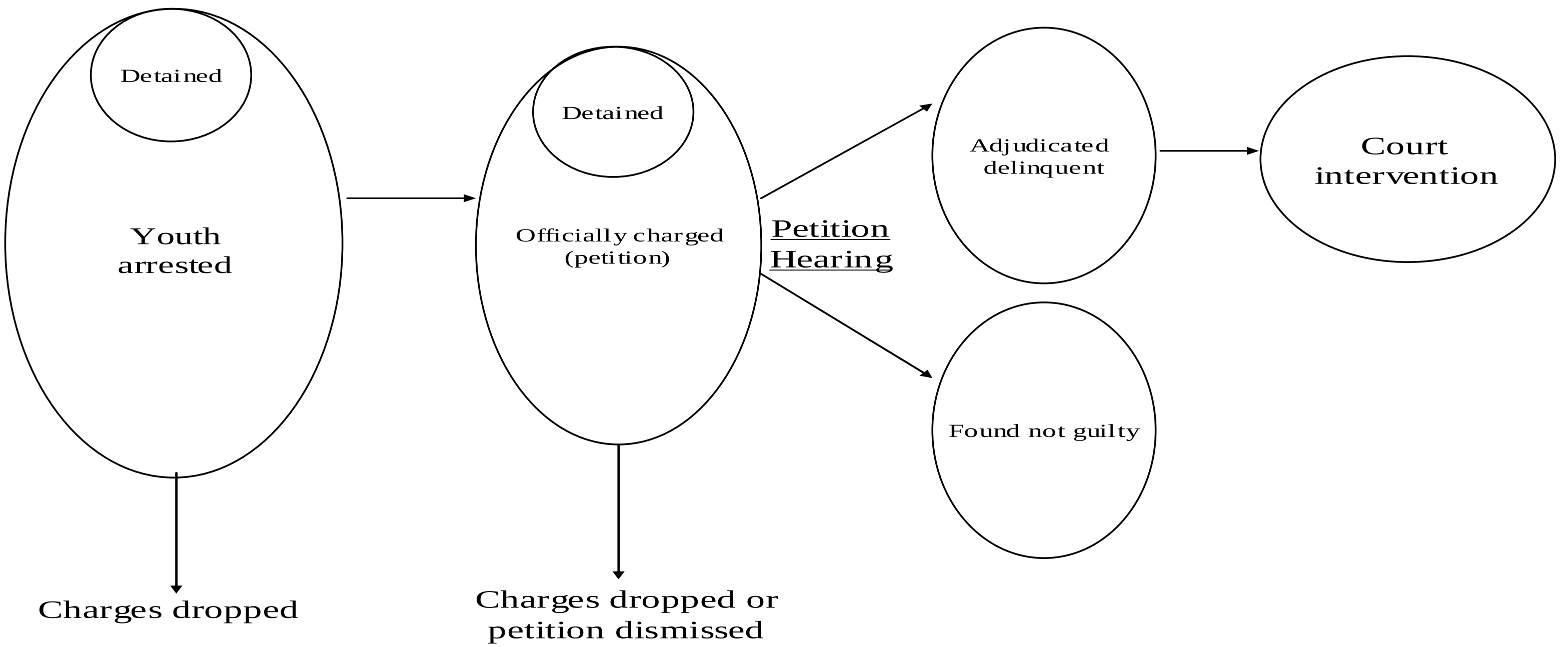

Figure 1. Schematic model of the juvenile justice system. Youth can exit the system in different ways with and without court intervention.

level of risk, which by law must occur within 72 hours of the arrest.

Following an arrest, a police report is filed by the arresting officer(s). The report includes who was arrested and charged, and others who were present at the time of the criminal incident. It also includes a listing of the charges as well as a description of what took place, and where and when it happened. Once a youth has been charged, the court then makes a decision about whether to hold the youth in detention until a hearing on the charges against him can be held (usually within 30 days) or to release him to the custody of his parents. Typically, those youth with previous court contact and those who have committed more serious crimes are held in detention. At the detention hearing, a judge decides if the youth should continue to be held in detention until he is to appear before the court for the charges against him, or if he is to be released from detention possibly with other restrictions or instructions from the court.

The juvenile court staff conducts a series of interviews with parents, victims, the youth, and other parties and then they create a social background and an investigative report on the youth and the alleged incident. This may include a psychological assessment by a qualified professional, in depth interviews with the offender’s caretaker, and collection of school records. At this point the case may be closed or a petition may be prepared on the incident. Preparation of a petition means that a formal hearing will be held at which a judge will decide whether to adjudicate the youth on the charges named in the petition. The petition is prepared by the juvenile court staff on the basis of the police report and information from various informants. The petition provides a brief summary of the incident including the date, time, and description of the alleged event; the name(s), address, and age of all perpetrators involved; and the charges arising from the incident. In Allegheny County, all original documents such as petitions and other official court orders are contained in the Prothonotary's office, where each participant’s arrest and detention information was obtained for this study.

At the hearing, the judge may take a number of actions on the specific charges named in a petition, or on the petition as a whole. The judge evaluates information about the arrestee including school records, past criminal history, psychological assessments and consultations with the youth's case worker(s). Consultations allow the judge to hear case worker’s recommendations for how to handle the youth and to determine if additional petitions are being prepared and the specifics of those petitions At the hearing, the judge may 1) dismiss an entire petition with or without permission to re-file the petition again; 2) discontinue action on an entire petition; 3) continue action on a petition until a later date; 4) adjudicate the youth on one or more charges named in an incident on a petition; or 5) adjudicate the youth on charges named on a petition but discontinue the petition in a subsequent hearing. If the youth is adjudicated, the judge may dispose of the case by placing the youth on probation, imposing a fine or restitution, or placing the youth in one or more programs that are expected to be beneficial to the youth and to protect the population at large. Placement decisions are made by the judge from a short list of available options and recommendations compiled by the youth's case worker. A record of all placement and intervention decisions are kept in the Prothonotary's office.

Risk Assessment Practices, Design, and Goals

RAIs have been used in JJS for the main purpose of informing decision makers which juveniles are most at risk for repeat and/or serious offending. RAIs are based on the assumption that knowing a youth’s risk status will allow decision makers to select the most appropriate and effective intervention, thereby reducing the chances that an offender will recidivate. RAIs are used by decision makers such as intake officers, probation officers, prosecutors, and judges working in the JJS who make or influence decisions regarding the placement, supervision, and treatment of juvenile offenders. RAIs are currently used in many state JJS's, but historically risk assessment has been done informally by decision makers who may vary considerably in their particular philosophies, personal experiences with offenders, and their beliefs about what interventions are most effective and for whom (Wiebush, Baird, Krisberg, & Onek, 1995). RAIs typically consist of a brief questionnaire or checklist designed to assess specific characteristics and behaviors of an offender and their environment. Data can be gathered from a variety of sources, such as police reports, court records, school report cards, evaluations from past corrections agencies, and informants including parents, teachers, corrections officers, and mental health workers. Data from these sources are then combined in some way to create a numeric expression of risk that is intended to represent an individual offender's likelihood of engaging in various unwanted behaviors, including reoffending. Statistical means of expressing the risk of an offender is typically presented in the form of a number, but usually risk scores range from low to high risk. This risk score is then used to group offenders into empirically or arbitrarily derived categories thought to posses different levels of risk for reoffending. Level of risk for an individual offender is determined by comparing his characteristics to other predetermined groups of offenders with different levels of demonstrated risk and placing him into the group whom he most resembles. This level of risk is then used to make decisions concerning the amount and types of interventions and court supervision a youth receives, based on what is known to work with other offenders in his risk group. This approach allows decision makers to incorporate information from a variety of sources and environments for the purpose of anchoring decisions on a limited number of salient factors that are predictive of recidivism or specific types of recidivism.

RAIs are designed to serve a number of purposes within the JJS, the most primary of which is to predict either recidivism in general or specific types of recidivism such as chronic, serious, or violent offending. They can also be used to determine the likelihood of an individual succeeding in any one of a number of available treatment options. Essentially all decision points in a JJS require some type of prediction of future behavior or offending; however, because the behaviors that need to be predicted are different for each decision point, most also have features that are idiosyncratic to the purpose of the instrument, such as the inclusion of specific variables known to predict reoffending, violent behavior, or escape from custody. These factors usually reflect the purpose and theoretical grounds on which the instrument is based. In other words, predicting different behaviors or types of offending requires using different sets of predictor variables.

Each decision point in a JJS also typically involves a unique group of offenders for whom practitioners must make decisions. For instance, of all juveniles arrested, only about two-thirds are actually referred to a juvenile court, and of those referred only about half are formally processed, and of those formally processed only about 5% are actually detained (Snyder, 1996). Most first-time juvenile offenders are not processed, because the case is dismissed for lack of evidence, there is an unwillingness of the victim(s) to prosecute, or the offender receives some type of informal sanction, such as community service or drug rehabilitation, agreed upon by the defendant and the prosecutor (Champion, 1998). RAIs are usually created for use at one of many specific decision points within a JJS where important intervention and placement decisions are made concerning individual offenders and thus can vary in the behaviors they are designed to predict. RAIs are designed to help with decisions, such as whether to detain following arrest (detention decisions); whether to place in a facility following adjudication (placement decisions); determining the most appropriate level of in-home supervision including probation, house arrest that may include electronic home monitoring, and day treatments (level of court supervision decisions); and which if any type of mental health intervention(s), such as individual therapy, drug rehabilitation, or sex offenders programs (mental health decisions), are needed. Taken together, these facts demonstrate that prediction methods designed for a specific decision point in a JJS may have less or no usefulness at other decision points that require predicting different behaviors.

Many juvenile corrections agencies conduct a standardized assessment during intake when an offending youth is initially brought into the system to determine whether the offender should be placed in detention, on electronic home monitoring, or not detained at all, until the initial hearing for any offense(s) for which the offender is being charged occurs. Detention decisions require evaluating the risk of reoffending before the time of the petition hearing, which usually occurs within the first couple of months following arrest, and also the likelihood of the youth not showing up for his hearing if he is not detained. Hence, RAIs developed for use with detention decisions would be best served by using variables that are predictive of reoffending in the short run, as well as those that predict absenteeism during petition hearings. If it is decided that a youth will not be placed in detention, RAIs can assist in choosing from a host of other placement options, including placing the child in parental custody and the use of electronic home monitoring devices. Placement decisions that occur after a juvenile has been adjudicated delinquent require evaluating the risk of the juvenile committing offenses or behaviors while incarcerated, including absconding from a secured facility, violence including sex offenses, and suicide. Placement decisions also involve making predictions about how well a given offender might respond to various available interventions. Upon release the risk of the juvenile reoffending in general or engaging in violence must be re-evaluated which requires prediction over long period of time that is technically only limited by the life of the individual in question.

While the primary function of a RAI is to produce some type of measure of risk of recidivism, RAIs can also serve the purpose of standardizing the decision making process. RAIs can guarantee that the process by which information is gathered and used to make judicial decisions is at least somewhat uniform across decision makers. In this way RAIs are used as a sort of checklist, ensuring that for each offender who passes through the system, a variety of factors will be considered as part of any decision, and at the same time hopefully expediting this process. Using standardized measures such as RAIs also adds validity to court decisions by ensuring that decisions being made are both transparent and based on agreed upon methodology. Optimally, RAIs also add continuity to how offenders are perceived and dealt with by those having contact with offenders at different points and agencies within a JJS. In this way RAIs are not always pure prediction instruments. Statutory guidelines can also be incorporated into RAIs, requiring that certain types of offenders receive certain types of interventions and restrictions. For example, homicide offenders usually require some type of mandatory confinement and intervention that occurs without considering either an offender's risk in general or individual set of risk factors. In the Broward County Detention Risk Assessment Instrument (Baird, 1984), all capital, life, first degree felonies, and violent second degree felonies automatically receive risk scores that require detention. By indicating that certain types of offenses require a minimal level of supervision or intervention, RAIs essentially take into account society's view of the seriousness or cost of different offenses. It is also important to understand that many decisions concerning offenders within a JJS can be influenced by factors not directly related to risk of reoffending, such as the number and types of placement and intervention options available to decision makers, the expected cost of each option, and the funds available for such options.

The use of RAIs provides the opportunity to examine, analyze, and compare data for large numbers of offenders within and across JJS's using the same RAI. In making the decision process more transparent, the use of RAIs allows for a clearer picture of what actually occurs in the JJS process, provides the opportunity to determine and understand what types of information are being used to make decisions, and determines what types of information are given the most weight in specific decisions. Having a thorough understanding of the current status of the assessment process can then be used as a bench mark for policy makers considering changes to the decision process in an attempt to decrease recidivism rates in general or for specific types of crimes. This type of information can also ensure the most efficient use of available resources. Conversely, when formalized assessment procedures are not used it can be difficult to determine exactly what types of information are being used and how. Knowing how judicial decisions are being made is especially important given that certain minority groups are over represented within the JJS. For example, African Americans are currently over represented at all decision points within JJS's, sparking debate over why such over-representation exists (Snyder, 2000). Making decision processes more transparent also allows researchers to more accurately explore what accounts for minority over representation by examining the specific factors that most influence judicial decisions. It can also offer insight into existing fluctuations in the level of over representation.

The underlying premise of RAIs is that data on offender characteristics and outcomes can be collected and analyzed to identify groups or profiles of offenders with similar profiles that are associated with different categorized outcomes, including different levels of risk for recidivism. When juvenile offenders enter a JJS, they can then be compared with theoretically and empirically derived offender profiles to see who they most closely resemble. An individual offender's risk score then is based on the recidivism rates among juveniles with characteristics similar to them. Given the goal of the decision maker to equate the offender with similar individuals who have a predetermined level of risk or dangerousness, results from RAI's then must always involve case classification (Glaser, 1987), an integral element of any useful RAI. Putting offenders into various categories of risk is an essential component to state juvenile corrections systems nation wide and is often used at virtually all major decision points in the JJS as a means of determining the best type of intervention and level of security given the available options for placement.

The goal of categorization is to be able to place offenders into groups of offenders who have different rates or risks of reoffending. Most RAIs categorize offenders into a small set of risk categories, usually high, medium, and low risk. Typically offenders who are placed into a "high risk" category are four to five times more likely to recidivate than offenders in low risk categories (Wiebush et al., 1995). Investigating the validity of RAIs, Baird (1991) found that for one instrument used in the Detroit JJS, 76% of offenders rated as high-risk on an RAI reoffended as opposed to only 39% of offenders rated as medium risk and 19% for lower risk offenders. RAIs can be used to place offenders into other categories based on the specific court decisions for which they were designed. For instance RAIs used for detention decisions might have three groups send home, place on electronic home monitoring, and place in detention in contrast to RAIs used for long-term placement decisions that might group offenders into categories that reflect facilities with different levels of security and length of time committed to the institution.

Clinical and Actuarial Prediction

There are two methods by which human behavior can be predicted, clinical predictions and actuarial predictions. Clinical predictions involve having the decision-maker examine information which is thought to have some relation to the outcome being predicted, and then mentally evaluating that information to make a prediction. Actuarial prediction also makes use of information related to the outcome of interest, however, this information has been quantified and ideally been shown to be statistically related to the outcome of interest. Most decisions within the JJS for which RAIs have been developed are essentially clinical. Despite the fact that a judge may make use of none or some instruments designed to predict risk of offending to inform the decisions he or she must make, in the end the judge usually incorporates the information from instruments along with other information to make a clinical decision. However, by definition true actuarial decisions are automatic and based solely on the risk score produced by a particular instrument. Essentially all information available for clinical prediction can in theory be quantified for use with actuarial methods. Although clinical and actuarial methods can also be combined such as when a clinical decision is informed by the risk score derived from a prediction instrument, in practice a decision can only be either clinical or actuarial. For example, if a judge who is attempting to make a prediction concerning an offender's risk for violence using a RAI makes automatic decisions of placement based entirely on the score of the RAI, then that would represent an actuarial judgment. Given that most decisions made in the JJS are not automatic, most decisions made within the JJS are essentially clinical, with the results from RAI's simply providing additional information for the judge to incorporate into his or her clinical decision.

Clinicians are uniquely qualified to gather certain types of data that may be difficult or too laborious to quantify such as variations in speech and facial expressions, however, this does not imply that clinicians are better than prediction equations at using these and other types of data to make accurate predictions. In fact a number of studies have found prediction equations or actuarial methods to be more accurate than clinical methods with a variety of outcomes including psychiatric diagnoses and survival time (Dawes, Faust, & Meehl, 1989), parole violation (Carroll, 1982), and criminal recidivism (Dawes & Corrigan, 1974; Gottfredson, 1987). In a meta-analysis of the literature, Grove et al. (2000) concluded that actuarial methods are approximately 10% more accurate than clinical predictions. In a separate meta-analysis Ægisdóttir et al. (2006) concluded that actuarial predictions are approximately 17% more accurate at predicting violent behavior than clinical predictions. Actuarial methods of prediction have a number of advantages over clinical prediction. Actuarial methods always yield the same results for a given data set, unlike clinical judgments that are influenced by a host of dynamic variables that may not be statistically related to the outcome of interest. Such variables include but are not limited to recent experiences by the clinician or judge including recent or salient correct and incorrect predictions, the perceived relation of specific pieces of information or variables to the outcome of interest, and the perceived purpose of the decision itself. Ideally, actuarial methods ensure that only information that is statistically related to the outcome of interest is used in decision making.

Given their design, actuarial methods of prediction are based on cases or predictions of known outcome. This is in sharp contrast to clinical prediction that is at best based on random cases of known outcome. For example, a judge or parole officer who is making predictions about delinquency may have knowledge of the outcome of some cases, it is not likely that he or she will know what happens to most juveniles who pass through their court. This is especially true of cases who either never reoffend or who do not offend for long periods of time. Lack of outcome data is extremely detrimental to accurate prediction. Not knowing the accuracy of one’s predictions and the specific variables that most predict the desired outcome makes it impossible to establish validity of the variables used to make the prediction. Not making the decision process explicit allows information to influence the decision making process even though the judge may not be aware that this information is influencing his or her decisions. If a judge is unaware that certain information is influencing their decisions, they cannot evaluate whether using this information increases the accuracy of their predictions of risk. For example, if knowingly or unknowingly a judge evaluates offenders being more at risk when they have brown eyes, not having feedback does not allow the judge to know whether having brown eyes really is a valid predictor. Also, making use of a decision process that is not transparent does not allow the judge to determine whether the proper weight is being given to all pieces of information concerning a particular offender, or if the information should be given any weight in the decision at all. This type of judgment also allows for variables such as race that are specifically not supposed to influence a judge's decisions, to do so without the judge's knowledge.

Principles of Psychometric Test Construction

Juvenile justice screening instruments are a special case of test construction. The procedures for psychometric test construction have been identified and discussed in detail in a number of sources (Kline, 2000; Kline, 1998; Janda, 1998). In general, developers of prediction equations have used two approaches to test construction, the rational approach and the empirical approach. Using the rational approach, test items are chosen and evaluated according to theoretical grounds, and the results or final score of the test are also based in theory. The empirical approach emphasizes reason and logic for item selection and test validation. Initially, a variable pool is generated from all available variables with little concern for theory. Items are then analyzed to identify those for inclusion and how they should be weighted based on the strength of the association between each variable and the predicted outcome in relation to other variables that might be included. Cronbach and Meehl (1955) outline similar steps for test construction that include preparation of items on the basis of theory, performance of appropriate item analysis, cross validation of test, collection of reliability information, collection of evidence of construct validity, and development of norms.

More specific to RAI construction Gottfredson and Snyder (2005) have outlined procedures to follow when creating a prediction instrument designed to be used in the Juvenile Justice System. These steps are: (1) defining the specific outcome criterion or behavior that is to be predicted, (2) selecting appropriate predictors of the outcome being predicted, (3) assessing the association between the predictors and the outcome criterion categories, (4) validating the relations between predictors and outcome criterion on a new sample, and (5) applying the RAI to situations for which it was designed. This study will make use of the elements of both the rational and empirical approach to construct and compare two actuarial models or prediction procedures, the Burgess method and an extension of predictive attribute analysis termed iterative classification (Steadman et al., 2000). These methods will be employed to identify those juvenile offenders most at risk for violent recidivism, for the purpose of evaluating the potential accuracy of such an instrument. In doing so the procedures outlined by Gottfredson and Snyder (2005) will be followed with the exception of the fifth step, applying the RAI to a situation for which it was designed. Following these procedures, one half of the available data will initially be used to construct the actuarial models (construction sample), and then each RAI will be evaluated for predictive accuracy on the other half of the sample (validation sample).

An initial step to test construction involves choosing items that are to be included in the RAI. Cronbach and Meehl (1955) suggest that this should be accomplished by choosing items on the basis of theoretical grounds, and on the specific application of the test. How predictors have typically selected for inclusion in RAIs will be discussed in the next section. Given that RAIs are designed for use within the JJS, only variables that are likely to be available to court workers will be discussed. The items selected for possible inclusion should then be examined empirically to assess their association to the predetermined risk outcome and whether they are appropriate for inclusion in the risk equation. This is done by examining the distribution of each item across outcome groups (violent and not violent) to determine which items distinguish between outcome groups. This procedure demonstrates the predictive utility of each individual variable for inclusion in the risk prediction equation. Not all variables that distinguish between outcome groups are always included in the final equation, however, because some variables that do distinguish between outcome groups may be highly prevalent in both outcome groups, and should therefore be eliminated from the equation. For example, if lifetime use of alcohol as a variable were to distinguish between outcome groups, it still should not be included in the equation if over 50% of the non-recidivists reported using alcohol during their lifetime.

Once the items for the prediction equation have been chosen, and the equation is constructed, the next step is to cross validate the instrument. This is often done by creating construction and validation samples, or using one half of a sample to perform the steps needed to select variables for inclusion in the prediction equation, and then making use of the other half of the sample to determine the predictive accuracy of the measure. In doing there is a certain loss in the predictive accuracy that is expected, a notion known as “shrinkage”. Ways to measure accuracy and to predict this “shinkage” will be discussed in relation to RAIs in the design, validation and accuracy section of this introduction. Evidence of construct validity can be assessed by examining both convergent and divergent validity. This is done by testing the relation between the RAI and other measures or constructs that would be expected to be either positively or negatively related to the construct that the equation or test purports to represent, in this case, risk of violent recidivism.

Selection of Predictors

An initial step in RAI design involves selecting predictor variables or domains expected to add predictive utility to the instrument. RAIs differ in the domains they measure and the number of items used to assess each domain, the wording that is used to assess each item, and how each item or domain is weighted in comparison to other items or domains included in the measure. Predictor variables considered more important or predictive of offending than others are typically given more weight than other variables. Variables that initially appear easy to assess can often be difficult to measure reliably and accurately across raters. For instance many instruments assess past criminal behavior, but this can be done using a variety of sources such as charges on a police report, petition charges, adjudications, and self-reports. There can also be variation in the way in that the number of criminal incidents are calculated. RAIs may count the number of arrests, number of times officially charged on a petition, or number of times adjudicated delinquent. Of course adding items to a RAI or any other prediction equation in general will result in some increase in predictive efficiency, however, this improvement in prediction must be weighed against the goal of keeping the RAI easy and quick to administer. Variables to be used in a RAI should be grounded in delinquency prediction models, and empirically demonstrated to independently predict offending. In regards to variable domains, the possibilities are limited by what is available to court workers, so predictors must be available for easy and simple measurement that can be assessed both reliably and accurately.

While there is considerable variation across RAIs in terms of the specific variables that are measured (Wiebush et al., 1995), there are some common areas or domains that are included in most RAIs. Domains that have consistently been found to be predictive of recidivism in the delinquency research literature including criminal history, drug use, and school behavior and performance (Farrington & Hawkins, 1991: Hawkins et al., 1998) are commonly used in RAIs (Baird, 1984; Le Blanc, 1998). These variables can be grouped together as child characteristics, school factors, peer factors, caretaker factors, family factors, neighborhood factors, and criminal history factors.

Most RAIs include some type of criminal history domain. Criminal history factors include the number and type of offenses the individual has been adjudicated or found guilty of in the past and those he is currently being charged with, the most serious offense committed, dollar value of offenses, age at first arrest or adjudication, age at first commitment, length of time since last offense, longest time served on any commitment, number of prior arrests, number of prior convictions, number of prior convictions for specific types of crimes, number of prior sentences, number of parole violations, length of probation, number of prior commitments, escape history while confined, and failure on prior conditional release. Child characteristics include past drug use, low impulse control, delinquent peer group, physical aggression and fighting, degree of victimization, and psychiatric diagnoses such as antisocial personality disorder or psychopathy. School factors include those related to school performance, such as highest grade completed, number of missed days from school, being held back a grade, suspensions, and grade reports. A number of studies have shown that crime tends to runs in families (Farrington, Jolliffe, Loeber, Stouthamer-Loeber, & Kalb, 2001; Ferguson, 1952; Gendreau, Little, & Goggin, 1996). Family factors often included in RAIs are parental criminal history and substance use, parental attitudes towards crime, and variables describing parenting practices such as amount of parental supervision and discipline techniques. Neighborhood variables are also frequently used in RAIs and they tend to describe factors outside the home that might increase risk for recidivism such as type of neighborhood, the prevalence of gangs, availability of drugs and firearms, and available social resources in an offenders community.

Researchers in the area of criminal prediction have begun to stress the importance of assessing dynamic factors or those that can change over time, such as peer affiliation (Andrews, Bonta, & Hoge, 1990). In their meta-analysis Lipsey & Derzon (1998) found dynamic factors, such as lack of social ties and delinquent peers, to be some of the strongest predictors of serious or violent delinquency for juveniles age 12-14. Unlike many predictors of recidivism that cannot be manipulated such as race and gender, dynamic factors are not only susceptible to possible interventions, but can also be viewed as potential indicators of intervention success resulting in a reduction in risk. However, the importance of giving dynamic variables more weight in risk assessment procedures has yet to take place (Corbett & Harris, 1996).

This research suggests that when considering variables for inclusion in a RAI, variables should be selected that have been shown empirically to be predictive of specific types of offending and can be accurately and reliably measured by administrators. It is also important to consider correlations between possible variables. Variables that are similar or are part of the same domain should be evaluated against each other, and only those that uniquely add to the predictive utility of the instrument should be included. It is also important to remember that variables that may predict violent offending in the general population, such as a diagnosis of antisocial personality, probably would not add much to the predictive utility of a RAI given that this diagnosis characterizes such a large percentage of the offending population. The next section will review the prediction equations used to create risk scores in RAIs for the purpose of delineating which variables might be best to include in a RAI.

RAI Design, Accuracy and Validation

Design

There are a number of ways in which variables may be combined to create a prediction equation to predict a designated outcome such as recidivism or violent recidivism. The most basic of prediction methods is the simple additive method. Originally developed by Burgess (1928), this additive method uses non-weighted values for each of a series of questions or predictors, the values are then simply added up to produce a risk score. Ideally, as the number of points or point level increases, so does the probability of offending. All variables are given equal weighting in the scale and therefore also have equal importance in the equation.

The second method commonly used in RAIs is similar to the Burgess method, with the exception that predictor variables are given different weights based on either theoretical grounds or on the statistically predetermined strength of the variable as a predictor to the outcome, in this case violent recidivism. This method was initially put forth by Glueck and Glueck (1950) who advocated selecting predictors that are mutually exclusive and weighting each predictor variable category according to the percentage of those in that category who were determined to be delinquent. A third method of combining factors that is similar to the Glueck method, in which predictor variables are weighted according to their association to the outcome variable, are a number of different regression approaches. Unlike the Glueck or Burgess methods, regression approaches allow intercorrelations between predictors to be taken into account. The most common means of determining the weight of each variable is through one of various multivariate statistical techniques, most commonly multiple, log-linear, or logistic regression.

A fourth type of method for combining predictor variables in a RAI is known as predictive attribute analysis, also called recursive partitioning or automatic interaction detection.

This method typically relies on computer software such as Chi-squared Automated Interaction Detector (CHAID; SPSS 14.0) to conduct discriminate analysis in order to partition a sample into smaller and smaller subgroups based on contingent associations between a selected risk factor and the outcome being studied. This method can be used when the predictor variables are dichotomous. The goal of this method is to make use of a classification tree procedure to separate the sample into subgroups that consist entirely of cases either entirely with or without the studied outcome. Figure 2 shows an example of the type of classification tree that is typically produced using predictive attribute analysis. As can be seen, all offenders are initially split into two groups based on the dichotomous predictor most closely related to the outcome variable. Once two groups have been formed, the strongest predictor of the outcome for each of the two groups is determined, and four groups are formed. This analysis continues until no further partitioning is possible. In this study adjudicated violence is the outcome variable, therefore here the goal of this method is to partition the sample into groups that all or almost all cases were either adjudicated or convicted of a violent offense or were not adjudicated or convicted of a violent offense. Similar to regression approaches this method allows for the identification of interactions among variables in relation to the outcome. Regression approaches can also be used to account for interactions among variables, however, researchers have done so very infrequently.

Farrington (1985) evaluated each of these methods using data from the Cambridge Study in Delinquent Development to predict self-reported and official delinquency. Farrington found that the simpler Glueck and Burgess methods performed at least as well and in some cases better than the more sophisticated regression approaches and predictive attribute analysis. Farrington and Tarling (1985) reviewed several studies employing these methods and also concluded that

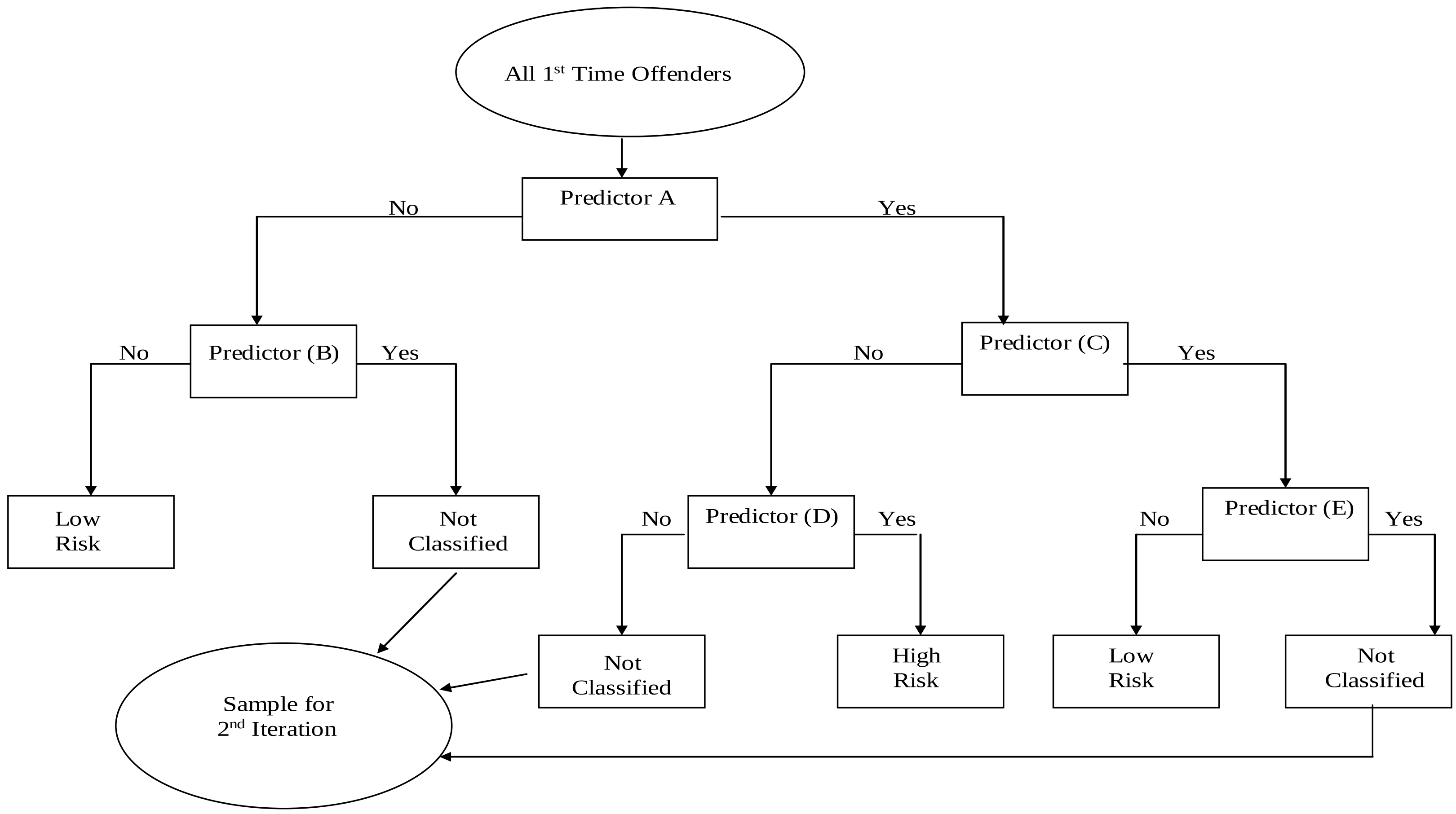

Figure *2.* Model of a classification tree (1st Iteration). Predictors most closely associated with the outcome are chosen to divide the sample into smaller and smaller sub-samples.

they were similar in predictive accuracy (Farrington & Tarling, 1985). Gottfredson and Gottfredson (1980) conducted an extensive evaluation to compare the predictive efficiency of the Burgess method, multiple regression approaches, and predictive attribute analysis. These authors concluded that they could find no clear advantage of one method over the other and that the particular method selected actually makes little difference in terms of predictive efficiency. They also tested the robustness of each of these methods in the face of unreliability by introducing 10% and 30% random error into the predictor item pool. These researchers found the Burgess method more resistant to this type of error due to its simplicity in design, suggesting a possible advantage of using more simple prediction equations. They also caution that when choosing a method, researchers should select a method that is not solely based on its statistical properties but rather one that can be used most practically given the specific constraints under which the RAI is administered, such as available data and skill level of those administering and interpreting the results of the RAI. More recently, Caulkins, Cohen, Gorr and Wei (1996) compared regression methods of constructing a RAI to neural network models at predicting recidivism. These researchers concluded that there seemed to be no advantage to using the more sophisticated neural network model, in that both methods demonstrated similar predictive accuracy.

Recently, an extension of predictive attribute analysis termed iterative classification (Steadman et al., 2000) has received attention from researchers studying violence prediction. This approach is essentially identical to predictive attribute analysis with the exception that once the sample has been partitioned as much as possible, groups that have not been classified as being either high or low-risk using a predetermined criteria, are re-analyzed in a second iteration to form a second classification tree. The hope is that by conducting discriminate analysis on those cases that in the first classification were determined to be “average risk”, those cases can be classified into high or low-risk in a second classification tree. This process is continued until remaining cases can no longer be partitioned into groups of high or low-risk. This process of re-analyzing cases that were not able to be classified as high or low-risk using an initial classification scheme is of course not unique to predictive attribute analysis, and can easily be applied to regression and additive approaches as well. However, this method has mostly been employed using predictive attribute analysis to perform each iteration and has only been used with RAIs designed to predict violence more recently. Silver, Smith, and Banks (2000) compared this method to standard predictive analysis, the Burgess approach, and logistic regression at predicting one-year and five-year arrest and imprisonment in a large adult sample. These authors found that the iterative classification procedure was able to classify a larger percentage of cases as either high or low risk than the other methods. However, in regards to overall predictive utility, all methods performed similarly. Monahan et al. (2001), making use of the MacArthur Violence Risk Assessment study, compared the iterative classification method to a basic predictive attribute approach and a logistic regression method. These authors also found the iterative classification method was able to classify a larger percentage of cases as either high or low risk, but that it performed similarly to other methods in regards to predictive ability.

Predictive Accuracy

There are a variety of ways of testing the predictive accuracy of RAIs. First, correlation coefficients provide a single statistic reflective of the strength of the relation between the results produced by the prediction equation and the specified outcome. Other methods require understanding the two types of errors that can occur when using a RAI, false positives and false negatives. Individuals who are identified to become violent who do eventually commit violent acts are valid positives, those who are predicted to desist in violent behavior and do are valid negatives. A false positive describes an error that occurs when a case that is predicted to reoffend does not. On the other hand a false negative occurs when it is predicted that an individual will not reoffend, but in fact they do reoffend. An easy to understand method of expressing the association between false negatives and false positives is the percentage of cases correctly identified. The percentage correctly identified is simply the total number of true positives plus the total number of true negatives, divided by the entire sample. However, this statistic has been shown to be dependent on the base rate of the outcome being examined and the selection ratio (total number of predicted failures).

To address this issue, Loeber and Dishion (1983) presented a statistic representing the relation between what is predicted and what actually occurs as the proportional reduction in prediction error that is gained by using an instrument over chance. Two ways of expressing this association are the Relative Improvement Over Chance (RIOC) presented by Loeber and Dishion (1983), and the Mean Cost Rating (MCR; Duncan & Ohlin, 1949). Each of these methods compares false negatives to false positives. The RIOC corrects for chance while taking into account selection ratio and the base rate. The RIOC has as an advantage over the MCR that the RIOC can be used to calculate confidence intervals, which can be used for comparisons across studies (Copas & Loeber, 1990). Analysis related to the RIOC and MCR is receiver operation characteristics (ROC) analysis. An ROC analysis plots the sensitivity and specificity pairs that result as a single decision threshold is moved from the lowest to the highest possible value. The ROC analysis is independent of the base rate of the outcome being examined, in this case violence. The ROC analysis produces the Area Under the Curve (AUC) statistic which ranges from .5 (accuracy is equivalent to chance) to 1. (accuracy is perfect). Over a decade ago Mossman (1994) suggested that ROCs should be used to evaluate the accuracy of predictions about violence, and they have since been increasingly used by researchers studying violence prediction. In their examination of prediction methods Farrington and Tarling (1985) concluded "there are no widely accepted methods of measuring predictive accuracy" (p. 20). Similarly, Gottfredson (1987) concluded that the current methods for measuring predictive accuracy cannot truly address the accuracy of a prediction instrument.

In considering the optimal proportion of false positives to false negatives, close attention must be paid to the choice of a cut-point that determines which offenders will be considered high risk and which will be considered low risk. For example, say a RAI yields a risk score that ranges from 0 to 10 and has a cut-point of 8 meaning that everyone receiving a score of 8 or more will be considered high risk. Obviously if a cut-point of 7 is chosen instead this will result in more individuals being considered high risk. In other words, changing the cut-off point changes the selection ratio, or the percentage of individuals predicted to become violent by the measure. Therefore, different cut-off points must be considered and examined for an optimal balance between sensitivity or hit rate (the likelihood of a randomly chosen violent offender having been identified by the measure as high risk) and specificity (the likelihood of a randomly chosen non-violent offender having been rated as low risk). Related to sensitivity and specificity are false negatives and false positives discussed previously. The optimal ratio of false positives to false negatives is not just an empirical question but also a moral question. It should depend on what is being predicted, and the expected financial and social costs of false negatives and false positives. Consequently, part of choosing an optimal cut-point involves deciding how big a group is useful to identify as high risk for decision makers. This not only involves evaluating the cost of false positives and false negatives, but also the amount of resources that are available to treat those identified as high risk and the ability of an intervention to affect outcome. For example, if it is found that based on RAI scores 90% of those identified as high risk already receive the most intensive treatment available, then there might be little to be gained from using the RAI. A well-designed RAI, then, should consist of an optimal balance between appropriate cut-points that take into account the base rate of violent offending, the cost of false positive and false negative errors, and available interventions.

Validity

Basic principles of research methodology assert that for any type of prediction instrument to be considered sound it must be empirically shown to be valid. RAIs need to have face validity to ensure those working with the instrument that it makes use of information that would be expected to be predictive of the expected outcome. Given that RAIs are designed to identify those youth most at risk for general recidivism or specific types of recidivism such as violence, empirical validation of a RAI involves demonstrating that the instrument can predict reoffending in individuals who have been adjudicated delinquent. In order to be informative, this involves waiting a considerable amount of time between when individuals are first identified as offenders and when recidivism is measured. Of course the longer the time frame between initial court contact and when recidivism is measured, the clearer the picture of which youth eventually reoffend.

A common way of validating a RAI involves splitting a single sample in half, and using the first sub-sample to construct the instrument, and then using the other half of the sample to test the predictive accuracy of the newly created instrument. However, due to the fact that both samples are drawn from the same pool of offenders in the same geographic area, caution must be used with this type of validation when making estimations about how well the instrument might predict recidivism in other geographic areas with different JJS's in different time periods.

Differences in time and place are expected to be important, because crime rates for different types of crimes will vary across geographic areas and time periods. Available resources and interventions as well as the process by which individuals are placed also will all vary over time. In the adult criminal justice system, a number of RAIs have been developed to predict risk (Gottfredson & Gottfredson, 1980), but as Clements (1996) points out, similar to the RAI's developed for juveniles, few have been thoroughly validated. Several researchers have concluded that the predictive accuracy of RAI's is somewhat weak. Gottfredson and Gottfredson (1985) concluded that while more sophisticated methods had been developed and employed, RAIs have not increased in their predictive efficiency. It has also been noted that workers using RAIs remain extremely limited in their ability to predict an individual's future offending behavior, and that the best RAI's lead to substantial prediction errors (Wiebush et al., 1995).

To overcome these road blocks to prediction, Andrews (1996) delineates several principals for ensuring maximum predictive accuracy of RAI's, including using RAIs that are standardized and structured, making use of multiple informants, and presenting more in depth predictions to specific types of criminal behavior rather than a simple composite measure of risk. Andrews also encourages intensive staff training on RAIs with ongoing monitoring of their use, with offenders being assessed for longer follow-up periods that include multiple time points to ensure that changes in an individual offenders' risk will be included in decision making.

Shrinkage

It has been noted that when a predictive instrument is created on a given sample and then later validated on another sample, there will always be a reduction in the predictive accuracy of the instrument. This reduction in accuracy has been termed "shrinkage," and validation studies suggest shrinkage can be substantial (Copas, 1985). Given the nature of their design, all prediction equations are created using a certain data set and random effects within that data set will influence the prediction equation that best fits the construction sample. Consequently, a RAI's predictive efficiency will be stronger on the sample on which it was created than for any new sample that it is used. To estimate the shrinkage of a RAI, Copas (1985) demonstrated how the degree of shrinkage can be approximated using an equation that uses information about the total number of predictor variables and the size of the construction sample. This expected shrinkage can then be used to calibrate the prediction equation and thus reduce the true shrinkage that occurs when the equation is employed on a new sample, thereby increasing the RAI's performance. Empirical selection of variables leads to more shrinkage than does selecting variables on some predetermined and or theoretical criteria. Scientific principal suggests that the more complicated the method of design the more places there are for error to be introduced into the data, making more complicated methods more susceptible to shrinkage than simpler methods.

Base Rate Problem

Much of the difficulty in predicting recidivism stems from the fact that so many first time offenders, many of whom would be considered high risk, do not go on to re-offend (low-base rate). Similarly, of those juveniles who do reoffend, not all would have been categorized as high risk. Most RAIs are developed to predict recidivism in general and not to predict specific types of reoffending (Wiebush et al., 1995). It is more difficult to predict specific types of crimes rather than recidivism in general in part because of the low base rate for specific types of crime. For example, data from the Juveniles Taken Into Custody statistical reporting program designed to collect data from 29 states on the characteristics of youth's entering state juvenile corrections programs, indicated that only about 14% of youth taken into custody were charged/adjudicated as a serious or violent offender and only approximately 10% of recidivists were serious or violent offenders (Austin, Krisberg, DeComo, Del Rossario, Rudenstine, & Elms 1994). This suggests that most incarcerated juvenile offenders are not adjudicated on serious or violent offences but rather adjudicated on less serious crimes. Due to the low base rate problem it is statistically difficult to identify factors unique to those offenders who are serious or violent recidivists. Given that most RAIs are designed to predict recidivism in general, this means that many offenders rated as high risk will most likely not be arrested later for a serious or violent offense.

Criminological data suggests that the base rate of serious violent juvenile offending is around 6% in the general population (boys), 20% of arrestees and 45% of adjudicated boys (Gottfredson, 1987). Gottfredson has highlighted that prediction becomes increasingly more difficult as the base rate differs from 0.5. He proposes two ways of dealing with the base rate problem, use of continuous predictor variables and sequential prediction (predicting outcomes for a homogenous group, picked from a larger group). Another way to deal with the base rate problem has been proposed by Le Blanc (1998) who proposes using a multistage screening strategy as an appropriate solution for dealing with a low base rate problem. This emphasizes the need to include number of prior contacts with the JJS into the risk assessment as well as ensuring that multiple assessments are conducted while the offender remains under court supervision.

Related to the base rate problem is the misclassification of offenders. RAIs developed for juveniles who have already offended have in general been shown to have high rates of false positive errors (offenders who were classified as high risk but who did not go on to reoffend) that typically fall within the 40% to 50% ranges (Wiebush et al., 1995), meaning that almost half of all those offenders who are identified as high risk will not go on to reoffend, at least using official records rather than self-reports as a criterion. The issues are similar when predicting other specific types of offending with lower base rates such as property, drug, and violent offenses. This highlights the problem of labeling offenders as “high risk” in general or for specific crimes especially when the impact of the low base rate on the false positive problem may not be completely understood by decision makers, and why extreme caution should be used when incorporating results from RAIs into decision making.

Effectiveness of RAIs and Classification

Knowing that a RAI has been shown to be valid or accurate certainly adds support for their use, however, accuracy only demonstrates predictive efficiency and not necessarily usefulness. It is possible for valid RAIs to be of only minimal use to decision makers. In fact, it is possible for an RAI to lead to a less efficient decision making resulting in an increased risk to public safety, improper use of available resources, and racial or other types of biases in decision making (Wiebush et al., 1995). RAIs can serve multiple purposes but most produce some type of risk score that reflects the risk of an individual offender engaging in some unwanted behavior, usually offending. This assumption that knowing an offender’s level of risk will lead to better decision-making is flawed in two ways. First, because an offender has some characteristics that are similar to those of recidivists, means less and less to the extent that recidivists as a group are heterogeneous. In fact, recidivists do make up a very heterogeneous group of individuals that includes offenders who have committed all different types of crimes requiring different types of interventions. Offenders also differ in other ways such as living environment, SES, and available social support. Recidivists by definition have already had at least minimal contact with a JJS, and many already have been exposed to any number of interventions that presumably did not work. Secondly, knowing that an offender is at high risk can be of little use to the decision maker who must choose appropriate interventions from multiple treatment options, many of which are designed for high risk offenders.

This point can be made clearer by taking a closer look at the actual decisions that take place in JJS. Figure 3 shows an overview of a typical JJS and includes some of the main elements of a system. As can be seen from the figure, as offenders pass through various decision points in a JJS, decision makers must choose from a number of treatment options. In Figure 3 these options are labeled as “no treatment” meaning not even minimal court supervision, “probation,” “mental health treatment,” “home monitoring,” and two levels of “residential confinement.” Offenders who pass through a JJS eventually either desist in their criminal activity or are arrested at a later time, thereby becoming a recidivist. Also included at the bottom of Figure 3 is an arrow representing the current method by which offenders coming into the system are predicted to become recidivist.

What is clear from examining Figure 3 is that recidivists as a group are very heterogeneous in a number of ways. First, as stated earlier, recidivists differ greatly in the kinds of offenses (serious and non-serious) they have committed during their initial and subsequent criminal actions. Secondly, recidivists consist of individuals who have been exposed to various types and levels of intervention ranging from no intervention at all to the most intense available. Individuals categorized as a recidivist by definition have passed through some JJS at least once but often have entered the JJS multiple times, perhaps receiving different interventions on each occasion. What can also be seen from Figure 3 is that another aim and perhaps true goal of RAIs is not actually to predict who will reoffend, but rather to predict for an individual offender which available treatment options will most likely result in a successful intervention. In this way then, the usefulness of an RAI is measured by the extent that it will result in better decision making,

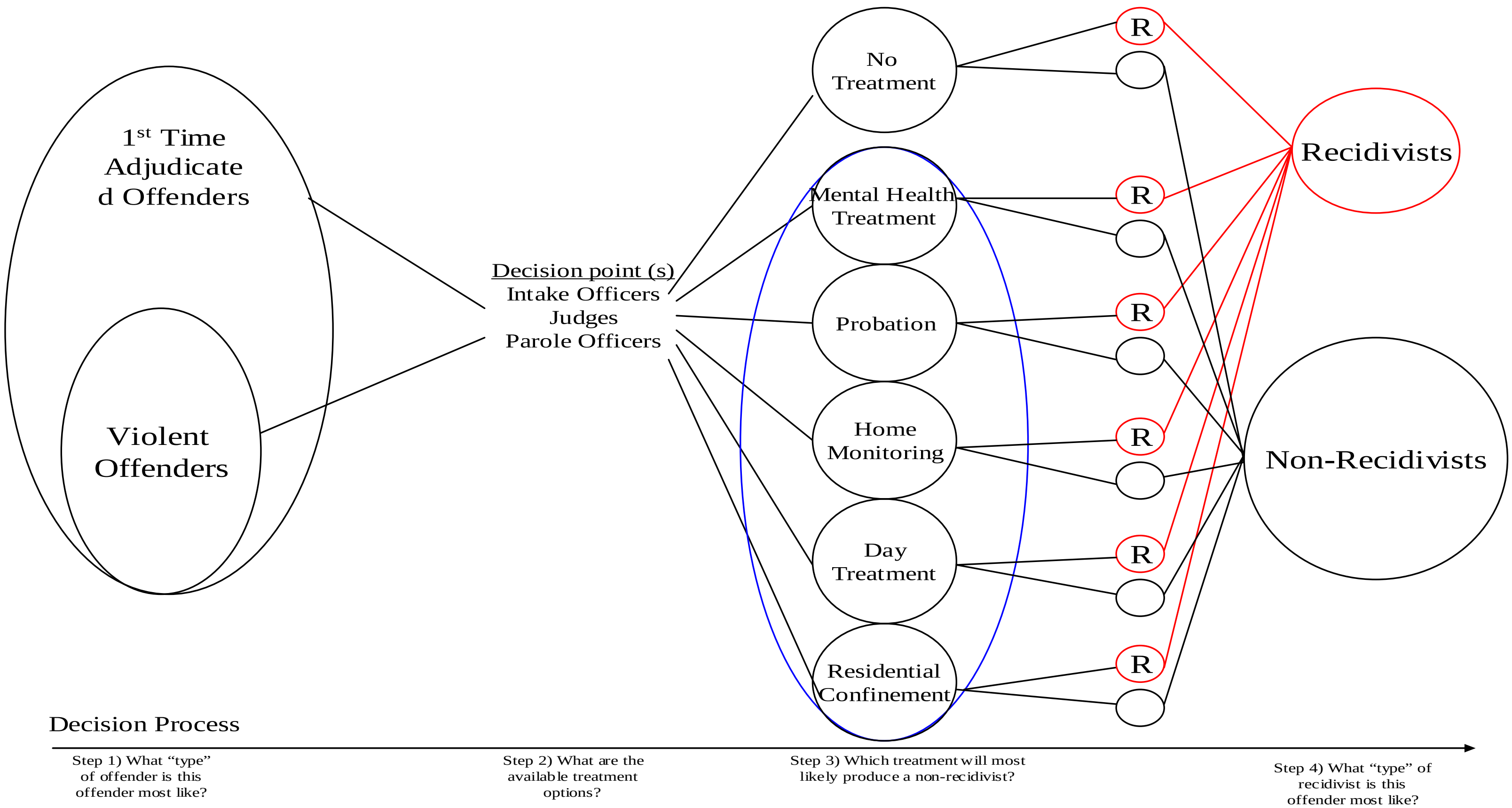

Figure 3. Model of interventions within a juvenile justice system designed to reduce all recidivism and serious offending.

with better decision making being defined as an overall decrease in recidivism rates in those individuals passing through the system. RAIs then can be as beneficial to the JJS decision maker only to the extent they result in better decision making. This also highlights the point that if RAIs are to be used to make appropriate detention, placement, probation, and other important decisions, it is key that the instrument be linked to an offender's needs for incarceration, probation, and mental health services. In other words, results from the instrument need to direct decision makers to specific interventions that are most likely to bring about the desired outcome, desistence. This, in turn, means that a RAI also needs to be linked to available agencies in charge of applying specific types of interventions. However a closer look at the actual decisions being made by the judges and their staff, highlights the difficulty of using some type of risk score to help with real world decisions. In this way, identifying recidivists as a group can also be defined as identifying those offenders that have previously passed though a JJS and the system did not work for them. In other words, some option(s) previously selected by a JJS to keep the individual from offending again failed. Creating a RAI by identifying factors associated with recidivists or offenders who have been arrested at least twice also assumes that persons coming into the system are representative of recidivists in the population. However this also is not necessarily true, given that not all juvenile offenses result in someone being caught, arrested, or charged. There is also the assumption that knowing the crimes that brought an offender into a JJS for which they were adjudicated tells you what type of offender they are, however, this is clearly not the case. Some criminal acts, especially those that are more serious, can have different risks of arrest, and because serious offenders are also known to engage in higher rates of less serious offending, serious offenders will often be arrested for less serious crimes than the ones they are committing (Huizinga & Elliot, 1987), allowing many serious offenders to be considered non-serious offenders. This point is further highlighted by self-reported delinquency statistics showing that only about 1/4 of juveniles committing offenses are actually arrested, and of those arrested only about 2/3 are processed by the court (Huizinga & Elliot, 1987).

Using RAIs in this way to determine who will most likely reoffend from a group of offenders who have received various types of interventions from least to most intensive can be compared to decisions that medical doctors frequently must make concerning appropriate treatments. Knowing that a particular patient is not likely to be successfully treated does suggest that perhaps the problems that the patient is having are more resistant to treatment. However, as is the case with recidivists, knowing a patient in general might be less responsive to treatment provides little to no information regarding the types of treatments that might work better or worse for them. In other words, it’s not enough information to know that a first time offender is likely to reoffend. To make an informed decision one must also know which if any interventions seem to work best for those first time offenders who are most at risk.

This point is exemplified in a RAI designed for use with juveniles that is currently in use in many JJS’s, the WSJCA-RA (Washington State Institute for Public Policy, 1998). The WSJCA-RA resource manual states that the purpose of the instrument is identify high-risk youth for the purpose of linking these offenders with intervention programs thought to be most effective for such youth. High-risk youth are targeted for more intensive efforts allowing administrators to spend scarce resources where they are most needed. The WSJCA has undergone a number of revisions since its introduction into the Washington State JJS. To date no studies have specifically addressed the validity of this instrument to predict violence. However, research examining felony convictions has been conducted. Reporting on the validity of the instrument, Barnoski and Markussen (2005) note that in two separate studies that included large samples (Barnoski, 1998; Barnoski & Matson, 1997) the instrument was able to identify high and low-risk groups. These authors report that the 6-month felony recidivism rates for the low-risk group was 6.6%, and the high-risk group 31.6%. The 18-month felony recidivism rates for the low-risk group was less than 10% (with 29% of the sample classified as low risk) and over 30% for the high risk group (with 42% of the sample being identified as high risk). While this instrument was found to be predictive of recidivism, with higher risk scores being associated with increased recidivism rates, it is difficult to see how this information would help a decision maker choose one of any number of treatment options over another, because the instrument provides little information about the differences in expected outcome for different types of confinement/treatment. These authors compared recidivism rates for youth who were granted a diversion (no intervention) to those who had been placed on probation. They found that for any given criminal history score, the 18-month recidivism rates for the diversion and probation groups were nearly identical. From this they accurately concluded that the criminal history domain predicts recidivism rates equally well for diversion and probation populations. Again, considering the options that decision makers have available to them, it is difficult to see how knowing that an offender falls into one of three large groups of offender categories would provide useful information as to what types of treatments might best serve the offender. Clearly grouping offenders into large and heterogeneous categories provides little information that can be used to make better decisions. The result is that those offenders most at risk for future serious or violent offending may not get the level of intervention required to protect public safety, while other, lower risk offenders or offenders committing non-violent crimes may receive overly intensive and expensive interventions.

Creators of the WSJCA-RA also presented criminal history scores that were calculated post-hoc for recidivists who had their initial court contact before the instrument was being used. Not surprisingly, they found offenders with lower criminal history scores to be more often granted diversion than high risk offenders. However, in regards to actual decision making, only 2% of those offenders with high risk scores were even diverted. This means that without the use of the instrument, 98% of offenders who were later categorized as high risk on the RAI were given probation. This suggests that decision makers are already quite good at identifying very high-risk offenders or at least identifying those offenders who would be rated as high risk using a RAI. So in terms of high-risk youth, with a decision to either divert or place on probation, it would seem that the instrument has little to offer in terms of useful information that will lead to better decisions. This highlights the problem that court personal often express concerning RAIs, which is that RAIs tend to identify offenders who can easily be identified by decision makers without the use of a structured instrument (Schneider, Ervin, & Snyder-Joy, 1996) and that predictions are always easier at the extremes, so what matters is how well a RAI can predict accurately from the intermediate group.

It follows from this discussion that any reported risk score from a RAI is only as useful as the categories in which it places offenders, and any categorization system will only be useful to the extent to which the categories of offenders identified have distinctly different probabilities of future offending in general or specific types of offending such as violent offending. It is also key that the categories in which offenders are placed be used by decision makers to make specific meaningful projections about how the different groups of offenders will respond to various treatments regimes. RAIs that group offenders into a few global categories such as low, medium, and high risk for general recidivism, can do little to inform decisions that require choosing one or more interventions from a host of options It is true that ideally any classification system should have a minimal number of classification groups for simplicity's sake, with an attempt to maximize between-group variance while minimizing within-group variance (Austin, 1986). Megargee (1984), in discussing classification systems, suggested that a classification system among other things should allow most offenders to be classified, should produce categories that can be reliably categorized across raters, and should direct treatment and placement decisions. In other words, results from an RAI should be meaningful to those working with the instrument in that offenders are placed into groups that have significantly different degrees of risk of reoffending and that the results have direct implications for intervention.

In his in-depth review of 20 years of offender classification research, Clements (1996) discusses the importance of classification for the purpose of differential treatments specific to offender needs. He proposes that the underlying premise of classification is that offenders are a markedly heterogeneous group and consequently it is useful to divide offenders into meaningful categories. In further discussing this issue Clements (1996) writes “not only are decisions made daily with respect to case-by-case offender management, but so too are long-range, policy-level decisions about the types of programs, the need for additional institutions, and the need for correctional personnel. These are influenced heavily by beliefs about the aggregate profiles of offenders received into the justice system (p.124),” further highlighting the discrepancy between the decisions that court practitioners face and the information they are provided using a RAI that predicts only general recidivism.

Due to the fact that the characteristics of first-time offenders and recidivists will vary across geographic areas and time periods, we would also expect those offenders who are not responding to treatment and those who do respond to change as the types of offenders coming into the system change and as the treatments change. Constant fluctuation in crime trends means that the types of offenders entering into the system are changing, and will theoretically result in fluctuations in RAI risk scores. So, for example, if drug usage in a jurisdiction increases, then it would also be expected that there would be an increase in drug related arrests. If during that same period of time other offenses remained consistent in their prevalence, then those crimes would account for less of the total crimes committed, highlighting the fact that as crime trends fluctuate, so will the group of offenders who are arrested, and consequently who will eventually become a recidivist, and the types of crimes they would be expected to commit.

It is also important to understand that use of any particular RAI that results in the inappropriate placement of offenders also represents inefficiencies in resource allocation. Appropriate placement can be defined on an individual level as placement that results in the most reduced likelihood of reoffending possible. However, across a whole system, optimal RAI efficiency can be defined as one that results in the largest overall reduction in either recidivism in general or specific types of reoffending, such as violent offending. In other words, the most efficient placement strategy is the one that results in the fewest number of serious or violent re-offenses by first time offenders. So, because different crimes have different levels of cost to society, the best placement system is the one that results in the fewest number of serious or violent offenses rather than less serious offenses.

Using a RAI to influence judicial decisions in a JJS will result in a change in the characteristics of offenders who eventually make up the populations at different corrections agencies. Likewise, offenders as a group who go on to reoffend will also change. For example, in Sacramento County, California, a RAI designed for detention decisions was applied post hoc to a group of 396 offenders not originally evaluated using the RAI to determine the impact the RAI would have made on detention admissions. Results showed that if the RAI had guided the detention decisions, there would have been a 45% decrease in the number of detention placements (National Council on Crime and Delinquency, 1993). Similarly, when Broward County, Florida began using an RAI that it designed for detention decisions, they noted a considerable drop in juveniles who were placed in detention, with the average daily population dropping nearly 2/3 from 166 to 53 offenders over the course of the first four years of use (Bazemore, Dicker, & Nyhan, 1994). The fact that using a RAI resulted in fewer offenders being detained of course also meant a considerable reduction in resources being spent on pre-trial detention as well as improved conditions for staff and residence of the detention facility. Given that overcrowding is such a wide-spread problem with secure juvenile facilities (Wiebush et al., 1995), this suggests that proper RAI use might be a partial solution to this problem. However, no effort was made by these researchers to evaluate the possible increase in pretrial offending due to less offenders being detained. Thus, employing the instrument may have resulted in a reduced number of individuals being detained, however, it is possible that it came at the cost of an increase in general and violent offending.

To return to the discussion of the best predictors to include in a RAI, the most optimal set of predictor variables then are those that can be used in such a way as to classify offenders into groups that have distinct rates of specific types of offending. However, knowing a certain offender is high risk is only useful if decision makers, such as judges within the JJS, are informed about differences in recidivism rates across independent interventions, allowing them to evaluate secondary options, an important point given that interventions have different levels of availability and cost. The hope, then, is that decision makers would administer the most optimal intervention, the intervention with the lowest recidivism rates for that type of offender, and be more informed about secondary options when the most desirable option is not available.

Specific variables within an RAI that are intended to drive decision making will not only affect the overall recidivism rates within a JJS but also recidivism rates for each treatment type as well. Changing theoretical approaches that suggest specific ways of treating offenders, and the addition of new interventions to a JJS and the removal of older less valued ones will also be reflected in a changing recidivist population. Changes at the local level, such as staff turnover and fluctuating levels of funding over time, as well can also increase or decrease a program's effectiveness differentially for different groups of offenders, thereby manipulating the group of offenders who go on to recidivate.

RAIs Predicting Violence

Recently, there has been increased attention focused on the development of RAIs that can be used with juvenile populations to predict violent offending in both adolescence and adulthood. Although there are a number of violence prediction instruments that have been designed for use with adults that have been researched, there are very few such instruments to date that have been specifically designed for use with juvenile populations. For example, the most widely used and researched measure of violence is the Violence Risk Assessment Guide (VRAG; Harris, Rice, & Quinsey, 1993). The VRAG is made up of the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991; Hare, 2003) and 11 other questions regarding demographic factors, psychiatric diagnosis, childhood history, and adult criminal history. Numerous studies have found the VRAG to be predictive of violent behavior in adults (for review see Quinsey, Harris, Rice & Cormier, 2006). Although this instrument has been widely studied with adults, no studies have been conducted to examine the extent to which this instrument might work with juveniles.

In recent years, however, there has been an increase in research designing similar instruments for use with adolescents, primarily to aid those working in the JJS to identify young offenders who are most likely to commit serious or violent offenses. In response to a drastic increase in juvenile violence during the 1980's and early 1990's, most states developed ways to transfer more serious juvenile offenders to adult courts. During this time juvenile courts were put under increased pressure by the public to identify and contain those offenders who were most likely to commit violent acts either as a juvenile or as an adult. The hope is that by identifying adolescents at high risk for future violence, juvenile courts will be able to more effectively allocate available intervention resources with the overall goal of reducing the amount of violence in the community. These newly developed instruments primarily stem from instruments originally created for use with adults, and are either designed specifically to predict violence and antisocial behavior, or to identify a group of traits associated with psychopathy rather than to specifically predict violent behavior.

During the last two decades researchers have increasingly focused on identifying offenders that could be characterized as being psychopaths (Seagrave & Grisso, 2002). Psychopathy, is a term first described by Cleckley (1976) to label individuals who are characterized as manipulative, shallow, grandiose, and forceful. Psychopaths are typically sensation seekers and risk takers. The primary characteristic of psychopathy is a lack of empathy and remorse for behaviors that negatively affect others. Psychopathy as a predictor of future violence has received extensive attention in the adult literature. In this regard a small set of measures that are intended to identify psychopathy in adults have been shown to predict both recidivism in general and also violence in adults (Hart & Hare, 1997) are now being researched with juvenile populations. Three such measures are the Childhood Psychopathy Scale (CPS; Lynam, 1997), the Hare Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003), and the Psychopathy Screening Device (PSD; Frick, O'Brien, Wootton, &

McBurnett, 1994). Although to date there have been few validation studies examining the ability of these measures to predict violence, studies examining the predictive utility of the PCL-YV have begun to emerge.

One of the most widely studied measures of psychopathy is the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991; Hare, 2003) that has been shown to predict violent offending in adults (Serin & Amos, 1995). Adapted from this measure, the Hare Psychopathy Check List: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003) is perhaps the most researched instrument designed to assess psychopathic traits and behaviors in adolescents. The PCL-YV makes use of a structured clinical rating approach. The PCL-R, from which the juvenile version was taken, has been shown to be predictive of official violence in late adolescence (Forth, Hart, and Hare, 1990). Although validation studies on the PCL:YV have just begun to emerge, a few studies have shown promising results. Catchpole and Gretton (2003) using retrospective data found results from the PCL:YV in juveniles to be associated with violence. In a study that examined the prospective validity of the PCL:YV, Murrie, Cornell, Kaplan, McConville, and Levy-Elkon (2004) found the PCL:YV to predict to multiple types of violence. The PCL:YV has also been shown to predict institutional violence among juveniles in the short term (Dolan & Rennie, 2005).

Two other instruments designed to assess risk for violence and antisocial behavior and not specifically psychopathy have also been studied with juveniles. The Early Assessment Risk Lists for Boys (EARL-20B; Augimeri, Koegl, Webster, & Levene, 2001) is designed to predict violent behavior in boys displaying disruptive behavior under the age of 12. Modeled after the Historical/Clinical/Risk Management instrument (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) designed to assess violence in adults, the EARL-20B contains 20 items covering a number of domains used to assess risk. No studies to date have been conducted to examine how well this measure predicts violent behavior, however, Augimeri (2005) reported that when a median split on the total EARL-20B scores of 379 boys was used as a cut-point, 38% of those youth categorized as low risk were adjudicated on a charge prior to their 18th birthday as opposed to 49% of the high risk boys, with boys in the high-risk sample also having significantly more court appearances and being adjudicated on more charges than the low risk group. Clearly this measure offers promise for predicting violence, however, more research is needed that specifically addresses this instruments ability to predict violence.

The Structured Assessment of Violence Risk in Youth (SAVRY; Bartel, Bartum, & Forth, 2000) is an RAI designed to aid in the assessment of violence risk with adolescents. This instrument is based on the "structured professional judgment" model that makes use of structured professional clinical appraisals of relevant risk factors. Retrospective studies looking at how well SAVRY total score's perform at predicting violent recidivism have found promising results showing that the instrument was able to predict violent behavior among juveniles (Catchpole, & Gretton, 2003; Borum, Bartel, & Forth, 2005). To date only two prospective studies examining the predictive utility of this instrument have been conducted. Gretton and Abramowitz (2002) found violent recidivism rates of young offenders to be 5.7% for those categorized has low risk, 13.1% for those categorized as moderate risk, and 40.4% for those categorized has high risk. In a similar study, Catchpole and Gretton, (2003) found violent recidivism rates at one year follow-up to be 6%, 14% and 40% for juveniles categorized as low, moderate, and high-risk respectively.

Childhood and Juvenile Predictors of Delinquency and Violence

In adult populations the predictors of recidivism in general are similar to those that specifically predict violence (Gendreau et al., 1996). Many factors measured in childhood and adolescence have been found to be predictive of delinquency, including criminal history, age, gender, factors associated with the family, and delinquent peers (Andrews & Bonta, 1994). A number of studies have specifically examined the influence of predictors measured in childhood and adolescence on violence. For the purpose of this study these predictors are grouped into child factors, developmental factors, peer factors, school factors, caretaker factors, family factors, neighborhood factors, and arrest (criminal history) factors. A brief review of research finding associations between the predictors in each of these categories and violence will be presented here, but for a more in depth review see Hawkins et al. (1998) and Lipsey and Derzon (1998).

A number of individual or child factors have been found to be predictive of later delinquency and violence. Physical aggression in childhood and adolescence is perhaps one of the more widely studied predictors of later violence. Farrington (1989) found high aggressiveness to be associated with self-reported violence in adolescence, self-reported violence at age 32, and official violent convictions between ages 10 and 32 in males. Similarly, Stattin and Magnusson (1989) found teacher-rated aggression at both age 10 and at age 13 to be related to official violence by age 26 in both males and females. A number of studies have examined the association between ADHD symptoms and later violence. Childhood hyperactivity has been shown to be associated with both official violence (Mannuzza, Klein, Konig, & Giampino, 1989) and self-reported violence (Hechtman & Weiss, 1986) in adulthood. Other symptoms that make up the ADHD disorder measured in youths have also been found to be related to later official and self-reported violence, such as concentration problems and impulsivity (Farrington, 1989). It has been proposed that youth with hyperactivity, impulsivity, and attention problems, and conduct problems are at special risk for becoming chronic offenders as adults (Lynam, 1996), and conduct disorder is the most frequently occurring diagnosis among juvenile delinquents

(McManus, Alessi, Grapentine, & Brickman, 1984). Farrington (1989) found both peer-rated dishonesty at age 10 and hostility to police at age 14-16 to be related to both self-reported juvenile and adult violence and official violence from ages 10 to 32. Mitchell and Rosa (1979) also found both parent and teacher reports of dishonesty in youths to be associated with official violence in adulthood.

Developmental factors here refer primarily to variables identifiable in very young children that are biological in nature. It is thought that such variables may make children more vulnerable to environmental and social influences at later points in development. Few studies have evaluated the impact of developmental factors on delinquency or violence compared to most of the other predictor domains reviewed here, however, there is evidence to suggest that they offer some predictive utility in predicting violence. Raine, Brennan, and Mednick (1994) found birth and delivery complications along with maternal rejection to be correlated with officially reported juvenile violence in males. Similarly, Kendel and Mednick (1991) also found an association between minor physical abnormalities measured at age 11-13 that are thought to signal pregnancy complications, and officially measured violence in both males and females Despite these findings it should be kept in mind that the strength of the association between variables in this domain and later violence does not appear to be as strong as compared to other predictors of violence, for example past aggressive behavior and antisocial peers. Findings related to pregnancy and delivery complications are also less consistent compared to other domains. At least two large longitudinal studies failed to find an association between pregnancy and delivery complications, and later violence (Denno, 1990; Farrington, 1997). Also, early physical trauma including birth and delivery complications seems to be particularly dependent upon environmental influences. Raine, Brennan, and Mednick (1994) did find indications of pregnancy complications to be related to future violence, but only when paired with maternal rejection. There is also evidence to suggest that prenatal trauma is associated with violence only in unstable home environments (Mednick & Kendel, 1988).

Associating with delinquent peers is thought to play an important role in the development of violent behavior and is one of the strongest predictors of youth violence (Hawkins et al, 1998; Lipsey & Derzon, 1998). Ageton (1983) found that exposure to delinquent peers in adolescence to be correlated with self-reported sexual assaults in adolescences. In a sample of both male and female adolescents, Herrenkohl et al. (2000) found gang membership to be related to self-reported violence at age 18. Both Farrington (1989) and Herrenkohl et al. (2000) found involvement with delinquent peers to be predictive of self-reported violence. Farrington (1989) also found involvement with delinquent peers to be associated with official violence in adolescence and adulthood. Lipsey and Derzon (1998) concluded from their meta-analysis of prospective longitudinal studies that among juveniles age 12-14 one of the best predictors of later violence was antisocial peers.

Poor academic achievement has consistently been shown to be predictive of later delinquent behavior (Maguin & Loeber, 1996). Children who consistently perform poorly in school and demonstrate little interest in school are at greater risk for school failure and dropout (Rumberger, 1995), and are at increased risk for engaging in violence. Farrington (1989) and Herrenkohl et al. (2000) found that academic failure at multiple ages in adolescence to be related to increases in self-reported adult violence, with Farrington also finding that academic failure predicted official violence. In the same study, Farrington also found truancy in the early teen years and dropping out of school prior to age 15 to be related to increased levels of both self-reported and official violence. Denno (1990) reported higher levels of academic achievement age ages 7 and 13-14 in males to be negatively correlated with official juvenile violence.

Parenting practices exemplify parental attempts to socialize their children and can lead to both positive and negative developmental outcomes. Farrington (1989) found a number of aspects of parenting as reported by parents to be related to violent criminal convictions including an authoritarian parenting style, low parental involvement in education, poor child rearing practices and disagreements among caretakers at age 8, and father's involvement in leisure activities at age 12. Farrington (1989) also found that parental disharmony at age 14 predicted juvenile and adult self-reported and official violence. Straus (1991), examining the effects of physical punishment, found the more parental physical punishment a person received as a child, the more likely they were to self report assaulting a spouse, commit a violent crime as a juvenile, and engage in physical aggression as an adult. Coughlin and Vuchinich (1996) found less effective parental discipline techniques at age 10 to be predictive of police arrest by age 17. Several studies have found that an association exists between parental attitudes and teen behavior problems, such as drug and alcohol use (Peterson, Hawkins, Abbott, & Catalano, 1994), although the association between parental attitudes towards violence and adolescent violence has been less studied (Hawkins et al., 1998). Using data from the Seattle Social Development Project Herrenkohl et al. (2000) found that parent's attitudes favorable to violence at age 10 was positively related to self-reported violence at age 18 in both males and females. These authors also found that poor family management practices at ages 14 and 16 were predictive of self-reported violence at age 18.

Aside from parental practices and attitudes, family characteristics have also been examined to determine their association with violent behavior. Analyzing data from the Rochester Youth Development Study, Kelly, Thornberry, and Smith (1997) found that maltreated youth, including youth physically neglected, exposed to illegal activities, and lacking in supervision, to give higher self-reports of having been involved in all types of delinquency including serious and violent delinquency. Using data from the Cambridge-Somerville Youth Study, McCord, McCord, and Zola, (1959) found that males living in high-conflict families were more likely to be convicted of violent crimes. Being a member of a low income family at age 8 has been shown to be predictive of self-reported and official violence in adolescence and adulthood (Farrington, 1989). Coughlin and Vuchinich (1996) found that living in a step-family or single-parent household more than doubled the likelihood of arrest by age 17. Examining youth in the National Youth Survey, Elliott, Huizinga, and Menard (1989) found that the prevalence of self-reported assault and robbery to be higher in youths from urban neighborhoods and low-income families.

Perhaps the most established association between a juvenile criminal history variable and later violence is an early onset of delinquency and violence. Using self-report data, Farrington (1989) found delinquency at age 14 to be related to violent crime convictions between ages 10 and 32. Tolan and Thomas (1995) found delinquent behavior prior to the age of 12 to be related to official and self-reported juvenile violence in both males and females. Moffitt, Mednick and Gabrielli (1989) found that as the age of first arrest increases the likelihood of violence prior to the age of 30 decreases.

Data from the PYS has also provided a wealth of information regarding what predicts delinquency and violence. Loeber, Farrington, Stouthamer-Loeber, Moffitt, and Caspi (2001) using data from the PYS found the best predictors of either reported violence by teachers, parents, and the participant, or court referred violence to be low guilt, low achievement, young mother, broken family, single mother, low SES, family on welfare, and bad neighborhood.

Poorly educated mother and large family were predictive specifically of court referred violence. Thornberry, Huizinga, and Loeber (1995) also reported that violent offenders are frequently involved in other types of crimes, especially property and drug crimes. Examining chronic violent offending, these researchers also found chronic violent offenders to have higher school drop out rates, lower attachments to parents and teachers, and more delinquent peers. They also reported that chronic offenders are more likely to experience poor parental monitoring, belong to a gang, and to reside in high crime neighborhoods. Taken as a whole these studies contribute extensively to our knowledge of what best predicts future serious delinquency and indicates that there are a host of specific factors and domains that might be appropriate for inclusion in a RAI designed to predict violence amongst juveniles. Next, it must be determined which subset of variables available to decision makers for assessment will be most predictive of violent reoffending and useful to those making decisions about juvenile offenders. When choosing predictors, it should be kept in mind that some variables such as past arrest or criminal history will be easier for those working in the JJS to assess than others such as parental attitude towards violence. In determining which factors are most appropriate for inclusion in a RAI and to minimize the total number of predictors, Copas (1985) calls for using a correlation matrix to compare the strength of predictors to the outcome, and to evaluate correlations between predictors to determine which are most highly correlated for the purpose of identifying variables that can be eliminated without reducing the overall utility of the instrument. Significant interactions that are found can also be explored for possible inclusion in the equation. However, others have recommended choosing predictors based on theoretical considerations (Farrington & Tarling, 1985; Gottfredson, 1987). Although there are many factors that can be measured in childhood and adolescence that have been shown to be related to later violence, some variables show more promise for this purpose than others. Some risk factors appear to be more strongly related to later violence than others, and some have been more thoroughly studied and consistently found to be predictive of violence. Evaluating 34 longitudinal studies on delinquency, Lipsey and Derzon (1998) conducted a meta-analysis to evaluate predictive factors of adolescent and early adult serious and violent delinquency. Examining predictors of violence in two age groups (ages 6-11 and 12-14), these researchers found the best predictors of serious and violent delinquency were having a juvenile offense or engaging in substance use prior to the age of 12, and having poor social ties and antisocial peers in early adolescence. Low family SES, having antisocial parents, and prior aggressive and violent behavior were also strongly related to future violence. The best predictors at age 6-11 were committing a general offenses and substance use. For boys age 12-14 the best predictors were lack of social ties and delinquent peers, followed by committing a general offense. Lipsey and Derzon also found that juveniles who were considered at risk based on these predictors were 3-20 times more likely to commit a serious or violent act in the future than those rated as low risk. Given that the strength of the predictors was dependent on age, Lipsey and Derzon suggest a need for RAI's developed specifically for certain age groups.

Prior to Lipsey and Derzon (1998), Loeber and Dishion (1983) also presented an extensive review of research focusing on the prediction of delinquency. They found that among a number of factors predictive of later delinquency, the strongest were problematic behaviors such as aggressiveness, family factors such as parental criminal involvement, and indices of school performance. They too found the strength of predictor variables to be dependent on age with different factors having more predictive utility at different ages. Grendreau et al. (1996) conducted a meta-analysis to evaluate the best predictors of adult recidivism. They found that socializing with other offenders to be the best predictor of adult recidivism. Adult criminal history, pre-adult antisocial history, antisocial personality, criminalgenic attitudes, and race were also highly predictive of recidivism.

Thus, it appears the best indicators of future violence that may be most useful in a RAI may be past criminal, aggressive and violent behavior, and substance use. Family factors that suggest a poor parental-child relationship, parental antisocial attitudes, and low family income are also likely to add to the accuracy of a RAI. Additionally, the relation between some variables and future violence is dependent upon or mediated by other factors, and perhaps might be less useful in prediction equations, especially those RAIs built on less complex methodology (i.e. Burgess based RAIs) that do not take into account interactions among predictors. For example, research suggests that the association between developmental factors such as birth and delivery complications and later violence may be strongly influenced by environmental factors (Raine, Brennan, and Mednick, 1994). In discussing which predictors are best, to include in RAIs, Le Blanc (1998) concluded that much more research is needed before researchers can "establish a common set of predictors of risk and need that are defined in the same way that can be applied to arrest, detention, adjudication, and placement decisions (p. 188)."

Primary Question

This study evaluated the strength of the association between a wide range of possible predictors of violence and both official and self-reported violence, for the purpose of identifying potential factors that court workers can measure and use to assess an individual's level of risk to guide them in their decision making. Two risk prediction equations designed to predict violent recidivism were constructed and compared following the unique principles of test construction described by each method. The usefulness of putting predictors together in each of these RAIs was then evaluated for the purpose of answering the primary question: can a prediction equation that makes use of variables associated with violent recidivism be constructed to predict which juvenile offenders will commit violent acts? It was hypothesized here that a prediction equation could be created to successfully or accurately predict which juvenile offenders would commit violent acts in the future. In addition, prevalence rates for property, drug, and violent offenses are presented to demonstrate the volume and scope of juvenile offending in the PYS.

Hypotheses

The primary purpose of the study was to determine whether it is possible to construct an RAI to successfully identify which juvenile offenders will become violent. Successful identification or prediction is defined here as meeting two criteria. First, the instrument must predict, with greater accuracy than would be expected by chance, which first time offenders will be adjudicated of a violent charge in the future, either as a juvenile or as an adult. Second, those offenders who are classified as high risk must have at least two times the base rate of violence as compared to all offenders, and those classified as low risk must have at 1/2 or less the rate of violence as compared to all offenders.

Main Hypotheses: It was hypothesized that indeed it would be possible to predict which first time juvenile offenders would become violent offenders in the future using the two criteria used here to define successful prediction of violence.

Secondary Hypotheses 1: It is possible to identify a group of variables typically available to court workers that are associated with future violent offending that can be used in an RAI to predict future offending.

Secondary Hypotheses 2: When two different methods (Burgess and iterative classification) for constructing a RAI are compared, the iterative classification approach would perform similarly to the Burgess method in predicting violence. However, it was hypothesized that the iterative classification method would outperform the Burgess method by classifying a larger percentage of offenders as either high or low risk.

Method

Participants

The Pittsburgh Youth Study began in 1987-88, at which time participants were randomly selected from boys in the first, fourth, and seventh grade (also called the youngest, middle, and oldest cohort), within the inner-city Pittsburgh school system. From the families selected, 84.7% agreed to participate. Following the initial selection which consisted of approximately 850 boys at each grade level, participants were screened to identify those who were at high risk for disruptive and delinquent behavior. Boys in the first and fourth grade were administered the Self-Reported Antisocial Behavior questionnaire (SRA; Loeber, Stouthamer-Loeber, Van Kammen, & Farrington, 1989), and those in the seventh grade completed the Self-Reported Delinquency questionnaire (SRD; adapted from the National Youth Survey, see Loeber, Farrington, Stouthamer-Loeber, & Van Kammen, 1998). Parents and teachers of the participants were administered the respective forms of the Child Behavior Checklist (CBCL; Achenbach, 1978; Achenbach & Edelbrock, 1983).

Using data from all informants at this screening phase, a risk score was calculated based on 21 serious antisocial behaviors. This score was then used to identify approximately 30% of the most antisocial boys in each grade (250 in each grade). A random sample of boys from the remaining 70% of each age group was also chosen (approximately 250 in each grade), yielding a total sample of approximately 500 boys in the first (n=503), fourth (n=508), and seventh (n=506) grade cohorts, with half being at high risk and half being average or low risk. Slightly over half of the boys identified themselves as African-American, and just under half identified themselves as White, a ratio similar to that found in the Pittsburgh public schools. Participants included Caucasians (*n* = 641), Asians (*n*=11), African-Americans (*n* = 837), Hispanics (*n*=4), 23 children of mixed race, and one American Indian. A more detailed description of the characteristics of the participants and the methods used to select them can be found in Loeber et al. (1998). Attrition has been quite low in this study, with an average participation rate across phases of over 95%.

For the purpose of the present study, only data from the youngest and oldest sample were examined. Each of the two cohorts was initially followed up every six months, and then every year. Because the primary purpose of this study was to examine predictors that the juvenile court could collect around the time of arrest when assessing the risk of juvenile offenders, only data prior to the age of 18 was used here to create the predictor variables. Data from follow-ups (excluding the juvenile and adult arrest and court data collection) that occurred over 13 years for the youngest sample, and 6 years for the oldest sample were used. There were a total of 14 data collection waves used for the youngest sample and 9 data waves for the oldest sample. At each assessment, the boy's primary caretaker was also interviewed and the boy's primary caretaker and teacher completed questionnaires. The term caretaker or parent is used to describe the person who claimed to have the primary responsibility for the boy in the household. In most cases (91.4%) the caretaker either was the biological, step, or adoptive mother. Approximately 4% of the respondents either were aunts, grandmothers, or foster mothers. The remainder of the caretakers were male.

The average ages of the participants at the screening phase were 6.9 for the youngest, and 13.4 for the oldest cohort. The age of the boys at the final phase used for the predictors of violence was 18. For arrest and court data the average age of participants varied depending upon when the data were collected, with the oldest participants being age 21 for the youngest cohort and age 30 for the oldest cohort at the time of the last collection of court data. Table 1 presents attrition information regarding who was eliminated from the sample and why, and the resulting N’s for the sample used for the analyses. As can be seen from the table, initially there were a total of 1009 participants in the youngest and oldest cohorts in the PYS sample used for this study. Of the 1009 participants available, 110 (63 from the youngest and 47 from the oldest cohort) were eliminated because consents could not be obtained for the participants, they died prior to the end of the study, or were no longer living in the state of Pennsylvania prior to age 16 meaning that complete juvenile data were unavailable for these participants. This left 899 (440 from the youngest and 459 from the oldest cohorts) participants remaining to be assessed for juvenile and adult offending. Of these 899 participants, 447 either had no juvenile arrest at all or were only arrested as adults and had no juvenile arrest record. This left 452 participants with at least one contact with a juvenile court somewhere within the United States. A participant was considered to have a juvenile court contact, if they had been arrested for at least one criminal charge prior to the age of 18. This was considered to be the case even if the participant was only arrested but not formally charged by the juvenile court, or if they were not adjudicated delinquent of any charge.

Table 1

Participants Eliminated From the Study and Resulting Sample Used For Analyses

Participants	Youngest	Oldest	Combined
Total Initial Sample	*503*	*506*	*1009*
Without consents	37	20	57
Died prior to 18th birthday	2	0	2
Died after 18th birthday	4	17	21
Not living in the state of PA prior to age 16.	20	10	30
Total Eliminated from analyses	*63*	*47*	*110*
No official contact with Juvenile justice system	231	216	447
Official contact with Juvenile justice system	209	243	452
Total Sample Used For Analyses	*440*	*459*	*899*

Sources of Information

In this study data were used from two categories of sources. First, data were collected from the participants, their primary caretakers (in most cases biological parents), school records, and teachers. Data from these sources were used to create the predictors of violence with the exception of the arrest variables. In this regard, an effort was made to use only data from sources of information that would most likely be available to court workers within most U.S. JJS's. These independent sources provide a large array of useful information that could be obtained by a juvenile court regarding the participant's life, including past and present experiences and behaviors, living situations, caretakers background and history, school attendance and performance, and previous criminal justice system contacts. These sources of data are often available to juvenile court workers and frequently used to assess a juvenile's level of social support, living environment, and risk of re-offending including risk of violent behavior. However, it is not uncommon for at least one of these sources of information to be unavailable to the court. Data from the participants themselves were also used to examine reported offending.

The second source of data was collected from official court records from five government sources covering both juvenile and adult data regarding arrests and court contacts from 1986 to 2001. The Pittsburgh Police Department Records were obtained in 1991 and 1993, and provided information regarding police contacts that participant's may have had in the city of Pittsburgh, Pennsylvania. The Allegheny County Juvenile Court Information Management Office Records and the Prothonotary Office Records provided additional city of Pittsburgh and Allegheny county wide juvenile data on court contacts through participants' 19th birthdays. The Juvenile Court Information Management Office Records stores and maintains files on each youth in the system, which include copies of official documents such as adjudication and placement orders, and also background and supporting documents such as psychological evaluations and placement reviews used to evaluate and monitor a youth's progress. The Prothonotary's Office Records are independent of the juvenile court, and contain the originals of all documents that are acted upon or created at a hearing before a judge for all courts, including family (custody) and juvenile. Hence, both offices needed to be searched for the collection of juvenile court records data to be as complete as possible. Three separate searches and retrieval of records were conducted at these two locations, the first in 1991, the second in 1993, and the third in 2001. The first two searches retrieved information from the oldest cohort, and in 2001 a third search was undertaken for the youngest cohort. This author, along with two employees of the PYS collected data from the records from each of these two offices. Once all the data were collected and entered into a computer data base, this author was also heavily involved in an extensive data cleaning process to ensure that all data from these sources had been entered correctly, and assisted in an detailed checking process to ensure that there were no duplications of arrests and/or charges from previous collections. With each new collection, new information on boys who previously had no record was collected, and for the 1993 and 2001 collections, any additional or new information on boys for whom data were previously collected, was also retrieved. Data that were previously collected were verified and amended if needed.

The Pennsylvania Juvenile Court Judges' Commission Records provided state wide data regarding participant's juvenile court contacts up to 1997, with the youngest sample being covered from age 10-16, and the oldest sample 10 to 18. The Pennsylvania State Police Repository provided state wide adult criminal history data on participants up to the spring of 2001 (up to age 23 for the youngest sample, and up to age 30 for the oldest sample). The U.S. Federal Bureau of Investigation provided country wide data on participants up to the spring of 2001 (age 23 for the youngest cohort and age 30 for the oldest cohort). Using data from each of these sources provided a uniquely comprehensive picture of official offending for participants from 1986-2001. Combining data from all of these sources provided the most complete picture possible of both juvenile and adult court contacts.

Collection of Records

Data from the Allegheny County Juvenile Court Information Management Office Records and the Prothonotary Office Records were collected by this author and two PYS employees, by reviewing the court records and then recording data from the records onto data collection forms covering many different areas of an offender's history through the Juvenile Court System. The Incident Record Sheet was used to collect information about offenses that a participant was charged with including those recorded on a police arrest report or an official petition of delinquency. Information collected included 1) dates of criminal incident a participant was involved in; 2) the date of the arrest and petition; 3) charges listed on the police report; 4) whether the participant was officially charged on a petition and if so what those charges were; 5) whether a referral to another agency was made at intake; 6) whether the youth was placed in detention prior to his petition hearing; and 7) whether codefendants were involved in the incident. The Petition Disposition and Hearing Record Sheet was used to record the results of a hearing before a juvenile judge on one or more petitions against the youth, including those of neglect. This form was used to collect information regarding the date of the hearing, those incidents decided upon at the hearing, the action taken on each incident of each petition (adjudicated delinquent, dismissed, etc..), and each participants specific adjudicated charges.

Data from the other four official sources were reviewed for each participant having a court record, and then these data along with data from the Allegheny County Juvenile Court Information Management Office Records and the Prothonotary’s Office Records were recorded and organized into a MS Word table for each participant that contained information regarding each participants’ delinquency history including the dates of arrests, types of crimes that participants were charged with during arrests, and charges that participants were adjudicated or found guilty of committing. Combining data from all sources in this way was done to account for any overlap in information between sources, ensuring that charges were not counted twice. Once delinquent history tables were completed for participants having a court record, information from the tables were combined for all participants onto an SPSS spreadsheet for statistical analyses.

Measures

At each data collection phase of the PYS information was gathered from the participants, their teachers, and their primary caretakers using a number of different measurements. Data were collected from participants and their primary caretakers through interviews that were mostly done privately in the participant’s home. The participant’s primary caretakers and teachers also completed questionnaires. Informed written consent was obtained from participants and their primary caretakers prior to each data collection. Data from these collections were used to construct the predictors in this study, with the exception of the arrest variables. A list of the measures used in this study, along with a brief description of each from Loeber et al. (1998) is presented below.

Child Behavior Checklist (Extended Version)

The CBCL (CBCL; Achenbach, 1978; Achenbach & Edelbrock, 1983) is a standardized behavior rating scale filled out by parents that contains 112 items designed to assess a wide range of child behavior problems, such as depression, anxiety, hyperactivity, and delinquency, over the previous six months, and includes social and academic competence questions. As was the case for similar measures, informants were initially asked about the last six months; however, for follow-ups that took place yearly, informants were asked about the previous year. The response format for the CBCL is “not true”, “somewhat or sometimes true, and “very true or often true”, yielding scores of 0, 1, and 2 respectively. This measure has been widely used and has been shown to have adequate test-retest reliability (Achenbach & Edelbrock, 1983).

The 4-year stability of the CBCL has been reported to be .66 for total problem scores using parents as informants and .37 using teachers (Verhulst, Koot, & Berden, 1990). Achenbach & Edelbrock (1987) reported overall agreement between parents, teachers and children as informants on the CBCL to be quite low, with agreement being highest between parents and teachers. They also found that agreement across these informants is highest when reporting on externalizing problems.

Youth self report. Participants were administered the 112-item Youth Self-Report (YSR; Achenbach & Edelbrock, 1987). Modeled after the CBCL, the YSR also measures child behavior problems, as well as social and academic competence over the previous six months.

Teacher report form. The Teacher Report Form (TRF; Edelbrock & Achenbach, 1984) is a measure that is also analogous to the CBCL given to caretakers. Also, 23 delinquent and concealing behavior items were added to this scale to increase its comparability with the child and parent reports. While teacher data were collected at all school age phases, the specific teacher reporting on any given student was likely to change over time, usually from year to year.

Self-Reported Antisocial Behavior Scale (SRA)

This scale was developed to examine self-reported delinquency among first and fourth graders (Loeber, Stouthamer-Loeber et al., 1989). The scale contains 33 items that asked the participant whether they had ever engaged in specific delinquent behaviors, and if so how often in the last six months. Included in the 33 items were six questions related to substance use.

Self-Reported Delinquency Scale

When participants were approximately ten years of age and thereafter, they were administered the Self-Reported Delinquency Scale (Elliott, Huizinga, & Ageton, 1985). This measure consists of 40 questions that ask participants if they have ever engaged in certain delinquent behaviors, and if so how many times in the previous six months. The measure also includes 15 questions that ask about substance use.

Supervision/Involvement Scale

The Supervision/Involvement Scale (Loeber et al., 1998) contains 43 questions regarding the parents’ knowledge of the boy’s whereabouts, amount of joint discussions, planning, and activities, and the amount of time that the boy is unsupervised. The response format scored on a 3-point scale is: ‘almost never’, ‘sometimes’, and ’often’. The boy reported on both his mother and father, however, the parent reported only on her knowledge of and interaction with the child.

Discipline Scale

This scale assessed the participant's perception of his mother's and father's discipline methods through 16 questions (8 items for each parent). Parents were also given a version of this scale. Example questions: 'If your mother warns you that you will get punished if you do not stop doing something, does she do what she says and punish you?' and 'If your mother had planned some punishment for you , could you talk her out of it?'

Counter Control Scale

This 11-item scale measures the extent to which the participant may have been in control of the family, and the extent that the parent avoids disciplining to minimize the escalations in antisocial child behavior. Example questions: 'Do you hesitate to enforce the rules with your son because you fear he might then harm someone in your household?', and 'When you are by yourself, do you have much difficulty controlling your son?'

Revised Parent-Adolescent Communication Form

This scale measures the quality of parent-child communication and display of affect. The child version asked participant's to assess both their mother and father and included 77 items. The parent version included questions for the mother only. Example questions: 'Do you have trouble believing everything your mother tells you?' and 'Are you afraid to ask your mother for something you want?' Questions were rated on a 3-point scale: 'almost never', 'Sometimes', and 'always'.

Child's Relationship with Parent/Siblings Scale (Stouthamer-Loeber, 1991)

This Scale combines 15 questions about the participant's relationship with his mother and 15 questions regarding his relationship with his father. Questions also address the participant's relationship with other children living with the participant. The parent also answered questions about their relationship with the child, the other parent, and other children living in the home.

Likelihood of Being Caught Scale

This scale measures the participant's perceptions of the likelihood of being caught for specific delinquent acts, and what would happen if they were caught by the police. The scale contains 10 items rated on a 3-point scale: 'not likely', 'somewhat likely', and 'very likely'. Example questions: 'What do you think the likelihood of being caught by the police if you……steal something worth $100?' '…..use alcohol?' and 'hit someone with the idea of hurting them?'

Attitude Toward Delinquent Behavior Scale

This 15-item scale was designed to measure the attitudes of the delinquent behaviors included in the SRD. Participants were asked how wrong it was for someone of their age to engage in each of the delinquent behaviors. Participant's rated each behavior on a 4-point scale. Response format is: 'not wrong at all', 'a little wrong', 'wrong', and 'very wrong'.

Perception of Antisocial Behavior Scale

This scale contains 12 items that measure a participant's attitude towards antisocial and rule breaking behaviors. Participants were asked how wrong it was for someone of their age to engage in each of the behaviors, such as skipping school, lying to adults, and hitting a teacher or parent. Participant's rated each behavior on a 4-point scale. Response format is: 'not wrong at all', 'a little wrong', 'wrong', and 'very wrong'.

Perception of Child Problem Behavior Scale

This scale contains 20 questions asking the parent to assess the parent's attitude regarding the participant's antisocial behavior. The scale covers a host of antisocial behaviors including aggression, conduct problems, and substance use. Example questions: 'Is it all right for your son to take a drink of alcohol?' and 'Is it all right for your son to have friends of whom you do not approve?' Questions were rated on a 3-point scale: 'not all right', 'all right sometimes', and 'all right most of the time'.

Gangs Questionnaire

This scale contains 37 questions that asked the participant about their membership in gangs, the importance of gangs they were members of, and the activities they engaged in with gangs. Example questions: 'If you are a current member of a gang: how often do you and some of the gang members get together?' and 'If you are a current member of a gang: how important to you is the gang and it's activities?'

Parents and Peers Scale

This measure consists of 11 yes or no questions measuring the number and nature of the boy's friendships, the extent to which his parent(s) approve of these friendships, as well as the behaviors that friends engaged in that caused parents of the participant to disapprove of his friends. Example questions: 'Are there any children in your group of friends of whom your parents disapproved?' and 'Have your parents said that any of your friends were a bad influence on you?'

Revised Diagnostic Interview Schedule for Children (DISC); National Institute of Mental Health (1992)

This scale was developed as a measure of child psychopathology to be administered by lay interviewers in epidemiological surveys. It was designed to assess most forms of child psychopathology contained in the DSM-III and DSM-III-R (American Psychiatric Association, 1980, 1987), as well as the age at which the problem behaviors were first noted.

Demographics Questionnaire

The Demographics Questionnaire (Loeber et al., 1998) was administered to caretakers and includes questions to gain extensive information about the household in which the study child lives, including gender, age, and employment history.

Family Health Questionnaire

This measure, administered to caretakers, contains questions about the help sought for mental health problems of the parent(s) and absent biological parents.

Subject Health Questionnaire

This measure, administered to caretakers, contains questions about the help sought for mental health problems of the participant.

Birth Questionnaire

This measure, administered to the parent contains questions regarding the mother's pregnancy with the participant, the birth of the participant, and the achievement of developmental milestones. Questions included whether there were complications at birth, a low birth weight, or if the birth was breech or Cesarean. Questions regarding developmental milestones asked about sitting, crawling, walking, talking, and using the toilet.

Predictors of Violence

In keeping with the overall goal of this study, to assess whether workers can use an RAI to predict violent recidivism, only information that court workers have or could have available to them were examined as predictor variables. Similarly, although race is often examined as a predictor of violence, because it is obviously unethical to base any decision made by the juvenile courts even in part on race, it will not be included as a potential variable for inclusion in the RAIs designed here. However, race was examined for its relation to future violence purely for scientific purposes, prior to creating the RAIs. Table 2 provides detailed information on the predictor variables assessed for their association to official violence. This table gives a brief description of each variable and a listing the measure(s) used to create them, along with the informant(s) that provided the data and the time points from which data were taken according to when each participant was arrested. Variables are grouped into child, developmental, peer, school, caretaker, family, neighborhood, and arrest factors. Aside from the official records, data were collected for most participants initially every six months and then yearly, thus, it was possible to use data that were collected fairly close in time prior to the date of a participant's first and second arrest. When possible, the data collection phase that occurred closest in time prior to a participant's arrest was used, with the exception of when data had been collected within two weeks after a participant's arrest. However, some data from which variables were created were not collected during each collection phase and therefore it was necessary to use data that were collected at time points that may have occurred more than 12 months from when a participant was arrested. These procedures were followed to mimic as closely as possible data collection procedures used by court workers assessing risk.

All child, developmental, peer, school, caretaker, family, and neighborhood constructs were previously created. With the exception of the noncompliance, codefendants, placed in detention, official charge of neglect, and bad neighborhood variables, all predictors were previously reported in Loeber et al. (1998), and descriptions of each construct and corresponding alphas given here were taken from that source. Alpha scores presented are taken from the initial data wave at which the construct was created. With the exception of the arrest variables, all predictor variables that were initially continuous were dichotomized at the one-quarter/three-

Table 2

Risk Factors Examined For Their Association With Future Violence

Risk Factor	Description	Instrument	Informant	Wave measured
	Child factors			
Physical aggression	Combines questions regarding a participant starting fights, hitting students or siblings and similar behavior.	CBCL TRF	PT	Y=Z-T O=Z-I
Cruel to people	Based on one 3-point question administered to the parent and one to the teacher.	CBCL	PT	Y=Z-T O=Z-I
Cruel to animals	Based on one 3-point question administered to the parent.	CBCL	P	Y=Z-P O=Z-I
Lack of Guilt	Based on 1 question administered to the parent and to the teacher.	CBCL TRF	PT	Y=Z-T O=Z-I
Depressed Mood	Combined from 11 questions regarding symptoms necessary for making a diagnosis of Major Depression according to DSM-III-R criteria.	CBCL TRF	PT	Y=Z-T O=Z-I
Noncompliance	Based on three questions administered to the parent and teacher regarding disobedient behavior.	CBCL TRF	PT	Y=A-R O=A-I
DSM-IIIR Conduct Disorder Diagnosis	Based on questions regarding diagnostic criteria for Conduct Disorder.	DISC-P	P	Y=A, P O=A
DSM-IIIR Attention Deficit Hyperactivity Disorder Diagnosis	Based on questions regarding diagnostic criteria for Attention Deficit Hyperactivity Disorder.	DISC-P	P	Y=A, P O=A
Covert behavior	Combines questions regarding the participant's covert behaviors such as being unaccountable to adults, manipulative, and untrustworthy behavior.	CBCL TRF YSR	PTC	Y=Z-T O=Z-I
Previous Placement in	Scored as positive if the participant had ever been	Family	PC	Y=A, G-V

Correctional Institution	placed in a correctional institution.	Screening for Violent Health Subject Health SRD		O=A-I
Psychopathy Without Delinquency	Construct modeled after Donald Lynam's (1997) Childhood Psychopathy Scale.	CBCL TRF	PT	Y=A-R O=A-I
Boy not involved	Combines 4 questions to the parent and 4 questions to the participant regarding the level at which the boy was involved in family activities such as planning family activities and joining the family in outings.	Supervision/ Involvement Scale	PC	Y=A-T O=A-I
Unlikely to get caught	Based on 11 questions regarding how likely a participant thought it was that he would be caught by the police if he engaged in specific delinquent acts.	Likelihood of Being Caught Scale	C	Y=C, E, G, J, L, N, P, R O=A, C, E, G
Positive attitude towards substance use	Based on questions regarding a participant's opinion on whether it is wrong to consume or sell various kinds of substances including alcohol, marijuana or hard drugs.	Attitude Toward Delinquent Behavior Scale	C	Y=A, C, E, G, I-V O=A, C, E, G, I, K
Positive attitude towards problem behavior	Based on questions regarding the participant's opinion on whether it is right to engage in problem behaviors such as smoking, using fists to resolve conflicts, or driving a car without a license.	Perception Of Antisocial Behavior Scale	C	Y=G-V O=A, C, E, G, I, K
Positive attitude to Delinquency	Based on questions regarding the participant's opinion on whether it is right to engage in a number of delinquent acts.	Attitude Toward Delinquent Behavior Scale	C	Y=A, C, E, G, I-V O=A, C, E, G, I, K
Gang membership	Scored as positive if the participant admitted to being a	Gang	C	Y=H-V

	member of a gang in the previous 6 months.	questionnaire		O=D-K
Weapon use	Scored as positive if a participant admitted to carrying a weapon (youngest cohort) or carrying a hidden weapon (oldest cohort).	SRA SRD	C	Y=A-T O=A-K
	Developmental factors			
Mother's alcohol use during pregnancy	Scored as positive if the mother reported having more than one drink per week during pregnancy.	Birth Questionnaire	P	C
Child premature	Scored as positive if the participant's biological mother indicated that the participant had been born three weeks or more early.	Birth Questionnaire	P	C
Prenatal problems	Scored as positive if the mother of the participant indicated problems during her pregnancy with the participant.	Birth Questionnaire	P	C
Perinatal problems	Scored as positive if the mother of the participant indicated problems during or after the birth of the participant. Questions included whether there were complications at birth, a low birth weight, or if the birth was breech or Cesarean.	Birth Questionnaire	P	C
Developmental delays	Created using questions that asked the parent about when the participant achieving developmental milestones such as walking and talking.	Birth Questionnaire	P	C
	Peer factors			
Bad friends	Combines parent and participant report summarizing information on the boy's associating with bad friends.	Parents & Peers Scale	P	Y=A-T O=A-I
Peer delinquency	Construct summarizes the proportion of friends with whom the participant reported to engage in 11 different forms of delinquency.	Peer Delinquency Scale	C	Y=A-V O=A-K
Peer substance use	Combines 4 questions administered to the participant	Peer	C	Y=G-V

	regarding the proportion of friends who use alcohol or drugs, or sells drugs.	Delinquency Scale		O=A-K
	School factors			
Low academic achievement	Combines judgments on participant's performance in seven academic subjects.	CBCL	PT	Y=Z-R O=Z-I
Truancy	Any positive response to truancy sometimes or often resulted in positive score	SRA SRD CBCL	PTC	Y=Z-T O=Z-I
Suspended	Scored as positive if the participant had been suspended from school prior to arrest.	CBCL SRA SRD	PC	Y=Z-T O=Z-I
	Caretaker factors			
Poor supervision	Based on questions regarding the parent's level of supervision in and out of the home.	Caretaker History Questionnaire	PC	Y=A-T O=A-I
Physical punishment	Scored as positive if either the parent or participant reported that the parent slapped or spanked the participant.	Discipline Scale	PC	Y=A-T O=A-I
Discipline not persistent	Combines questions from both the parent and the participant regarding the persistence of disciplining by the parent.	Discipline Scale	PC	Y=G-T O=A-I
Poor communication	Combines parent and participant report of how often they communicate directly or indirectly about emotions, disagreements, and problems.	Revised Parent-Adolescent Communication Form	Y=(A-F=P G-T-PC) O=PC	Y=A-T O=A-I
Counter control	Created from questions from the parent regarding whether the boy's behavior became worse when	Counter Control	P	Y=A-T O=A-I

	punished.	Scale		
Teenage Mother	Scored as positive if the mother was a teenager at the time the participant was born.	Demographics	P	Y=X O=X
High parental stress	Based on 14 questions administered to the parent regarding the mother's perceived stress and her perceived ability to handle problems.	Perceived Stress Scale	P	Y=A-T O=A-I
Poor relationship with caretaker	Combines 13 questions asked of the participant and 16 questions asked of the caretaker regarding the perceived relationship of the participant to the caretaker.	Child's Relationship with Parent/Siblings Scale	PC	Y=A-T O=A-I
Father behavior problems	Scored as positive if the primary caretaker reported the biological father to have had behavior problems.	Family Health Questionnaire	P	Y=A O=A
Father alcohol problems	Scored as positive if the primary caretaker reported the biological father to have had alcohol problems.	Family Health Questionnaire	P	Y=A O=A
Father drug problems	Scored as positive if the primary caretaker reported the biological father to have had drug problems.	Family Health Questionnaire	P	Y=A O=A
Father school adjustment problems	Scored as positive if the primary caretaker reported the biological father to have had school adjustment problems.	Family Health Questionnaire	P	Y=A O=A
Parent Antisocial Attitude	Construct created from 18 questions about the parent's attitude to antisocial behaviors of the boy such as whether it is all right to yell, argue, and fight.	Perception of Child Problem Behavior	P	Y=A, G O=A-G

		Scale	Screening for Violent	
	Family factors			
Family on Welfare	Scored as positive if the parent reported anyone in the household was on welfare.	Demographics	P	Y=A-T O=A-I
Official Charge of Neglect	Scored as positive if the caretaker of the participant was found to be in neglect by the juvenile court.	Court Records	O	Na
	Neighborhood factors			
Neighborhood Status	This construct is based 1990 U.S. Census tract data in which Pittsburgh neighborhoods formed the units of analysis. Variable is based on two primary factors: SES and Familism (defined in text). Each boy was matched with a neighborhood type based on his address during the first follow-up assessment.	Demographics	P	Y=A-T O=A-I
Negative Impression of Crime	This construct is created from 10 questions regarding problems that occur in neighborhoods such as racial conflict, vandalism, prostitution, gambling, burglaries and open drug use/dealing.	Demographics	P	Y=A-T O=A-I
	Arrest factors			
Placed in Detention	Scored as positive if a participant was placed in detention following arrest.	Court Records	O	Na
Co-defendants	Scored as positive if others were also charged at the time of arrest.	Court Records	O	Na
Violent Arrest Charge	Scored as positive if a participant was charged with a violent offense during arrest.	Court Records	O	Na
Property Arrest Charge	Scored as positive if a participant was arrested on a property offense charge.	Court Records	O	Na
Drug Arrest Charge	Scored as positive if a participant was arrested on a drug offense charge	Court Records	O	Na
Weapons Arrest Charge	Scored as positive if a participant was arrested on a	Court	O	Na

Screening for Violent

	weapons charge	Records		
Adjudicated Charge	Scored as positive if a participant was adjudicated delinquent on any charge.	Court Records	O	Na
Adjudicated Violent Charge	Scored as positive if a participant was adjudicated delinquent on a violent offense.	Court Records	O	Na
Adjudicated Weapons Charge	Scored as positive if a participant was adjudicated delinquent on a weapons offense.	Court Records	O	Na
Adjudicated Drug Charge	Scored as positive if a participant was adjudicated delinquent on a drug offense.	Court Records	O	Na
Juvenile Arrest Prior to Age 13	Scored as positive if a participant's first arrest occurred prior to age 13.	Court Records	O	Na
Adjudicated Charge From Arrest Prior to Age 13	Scored as positive if a participant was adjudicated delinquent prior to age 13.	Court Records	O	Na

*Note: All constructs that were created from continuous variables were dichotomized at the 75% cutoff. All constructs were taken from the data collection wave that occurred closest in time prior to or within two weeks after a participant's juvenile arrest. P=parent report, T=teacher report, S=self-report, O=official court records, Y=youngest cohort, and O=oldest cohort.

quarters split. Use of dichotomized variables allows findings to be reported in odds-ratios that tend to be more meaningful and easily understood by wider audiences than other statistical procedures (Farrington & Loeber, 2000).

Child Factors

Physical aggression. This variable summarizes the boy's physical aggression using seven questions from the participant's parent version of the CBCL and five questions from the participant's teacher using the TRF. Questions asked of the parent included: 'Hits teacher?', 'Hits parent?', and ' Hits or physically fights with siblings?'. Questions asked of the teacher included: 'Gets in many fights?', 'Physically attacks people?', and 'Hits or physically fights with students?' Ratings of questions on a 3 point scale were: 'never', 'sometimes', and 'often' (α (Youngest cohort; Y) = .63; α (Oldest cohort; O) = .75).

Cruel to people. This construct is created from one question from the parent version of the CBCL, and one from the teacher form the TRF. Questions asked if the participant was cruel, bullying, or mean to others. Questions were rated on a 3-point scale: 'never', 'sometimes,' and 'often' (α (Y) = .34; α (O) = .57).

Cruel to animals. This construct was based on one question from the parent version of the CBCL that asked the parent if the participant had been cruel to animals. Question was rated on a 3 point scale: 'never,' 'sometimes,' and 'often'.

Lack of guilt. This construct reflects the lack of guilt displayed by the boy and is measured using one question ('does not feel guilty after misbehaving') from the parent CBCL and one question from the TRF. The values from each of these questions, rated on a 3-point scale 'never," "sometimes," and "often," were then summed and dichotomized.

Depressed mood. Combined 11 questions from the parent version of the CBCL, and 11 questions (ex: ‘complains of loneliness,’ ‘cries a lot,’ and ‘feels and complains that no one loves him’) from the TRF regarding symptoms necessary for making a diagnosis of Major Depression according to DSM-III-R criteria. Questions were rated on a 3-point scale ‘never,’ ‘sometimes,’ and ‘often’ (α (Y) = .80; α (O) = .84).

Noncompliance. This construct was created using combined CBCL data from both parents and teachers. For caretakers there are three items on each measure that ask about the child’s disobedience at home. For teachers there are complementary questions about school noncompliance. Respondents are asked how often their child talks back to adults, how well they carry out assigned tasks, and how disobedient they are in each environment. For each boy a composite score was computed using the six questions pertaining to noncompliance, resulting in scores that range from 0 to 12. A boy was considered noncompliant if his composite score was greater than 7 for any single assessment period.

DSM-IIIR conduct disorder diagnosis. This construct was created using information from the DISC-P completed by the parent regarding diagnostic criteria for conduct disorder. This variable was scored as positive if a participant had five or more DSM-IIIR criteria for a diagnosis of conduct disorder. Among other areas the questions assessed the participant running away overnight, fire setting, breaking and entering, and fights with weapons.

DSM-IIIR Attention deficit hyperactivity disorder diagnosis. This construct was created using information from the DISC-P completed by the parent regarding the participant’s symptoms of ADHD. This variable was scored as positive if a participant had seven or more DSM-IIIR criteria for a diagnosis of Attention Deficit/Hyperactivity disorder, two of which had onset prior to age 7. Questions assessed if the participant was easily distracted, did not follow instructions, talked excessively, or interrupted others.

Covert behavior. This construct represents the extent to which the participant makes himself unaccountable to adults. The construct was created from 18 questions from the parent CBCL, six questions from the TRF, and three questions from the YSR concerning the level at which the participant displays concealing, manipulative, and untrustworthy behavior. Each question was rated on a 3-point scale ‘never,’ ‘sometimes,’ and ‘often’ (α (Y) = .90; α (O) = .93).

Previous placement in correctional Institution. This variable was created from both the Family Health Questionnaire and the Subject Health Questionnaire. Both the caretaker and the participant were asked if the participant had ever been placed in a correctional institution. An affirmative response by either informant led to a positive score for this variable.

Psychopathy without delinquency. This construct was modeled after Donald Lynam’s (1997) Childhood Psychopathy Scale, and it combines 18 questions from the parent CBCL and 14 questions from the TRF. Questions include ‘Cruelty, bullying and meanness to others’, ‘Destroys things belonging to his family or other children’, and ‘Teases a lot’.

Boy not involved. This construct was based on four questions from the parent version and four questions from the participant version of the Supervision Involvement Scale. This construct reflects the degree that the participant was involved in family activities such as planning family activities and joining family members on outings. Questions include ‘How often do you help to plan a family activity?’ and ‘How often do you go with members of the family to church, Synagogue, or Sunday School’ (α (Y=.61; α (O) = .73).

Unlikely to be caught. This construct was created from 10 questions asked of the participant from a scale with the same name. Questions addressed how likely the participant believed it would be to get caught if he engaged in specific delinquent acts such as going into or trying to go into a building to try and steal something, hitting someone with the idea of hurting them, and using alcohol (α (Y) = .70; α (O) = .89).

Positive attitude towards substance use. This construct was created from four questions given to the participant in the Attitude Toward Delinquent Behavior Scale asking how wrong it is to consume various kinds of substances (ex: how wrong do you think it is for someone your age to use alcohol?). Responses were scored on a 4-point scale, ‘Not wrong at all’, ‘a little wrong’, ‘wrong’, and ‘very wrong’ ((Y) = .66; α(O) = .80).

Positive attitude towards problem behavior. This construct was created from 18 questions from the Perception of Antisocial Behavior Scale asking the participant whether it is right to engage in various problem behaviors such as smoking, using fists to resolve conflicts, or driving a car without a license. Responses were scored on a 3-point scale, ‘not all right’, ‘all right sometimes’, and ‘all right most of the time’ (α(Y) = .76; α(O) = .84).

Positive attitude towards delinquency. This construct was created from 11 questions from the Attitude Toward Delinquent Behavior Scale asking the participant how wrong he judges a number of delinquent acts. Delinquent acts assessed included whether it is right to engage in various problem behaviors such as smoking, using fists to resolve conflicts, or driving a car without a license. Responses were scored on a 4-point scale, ‘not wrong at all’ ‘a little wrong’, ‘wrong’, and ‘very wrong’ (α(Y) = .82; α(O) = .87).

Gang membership. This variable was created from the Gangs Questionnaire from one question that asked the participant the number of gangs they were involved in during the previous six months. This construct was scored as positive if the participant reported having been involved in a gang.

Weapon use. This variable was created from one question on the Self Reported Antisocial

Behavior (youngest sample) and the Self-Reported Delinquency (oldest sample) scales. Participants were asked if in the last 6 months they had carried a weapon (youngest sample), or if they had carried a hidden weapon (oldest sample). An affirmative response was scored as positive.

Developmental Factors

Mother's alcohol use during pregnancy. Created from one question from the Birth Questionnaire that asked the participant's biological mother about alcohol consumption during pregnancy. This construct was scored as positive if the mother reported having more than one drink per week during pregnancy.

Child premature. Created from one question from the Birth Questionnaire that asked participant's biological mother if the participant had been born premature, and if so how early. The construct was scored as positive if the participant had been born three weeks or more early.

Prenatal problems. Created from eight questions from the Birth Questionnaire that asked the mother of the participant about prenatal problems during her pregnancy with the participant. Questions included whether the mother had high blood pressure, an infection, or used illegal drugs during pregnancy.

Perinatal problems. Created from six questions from the Birth Questionnaire that asked the mother of the participant about perinatal problems during and after the birth of the participant. Questions included whether there were complications at birth, if the participant had a low birth weight, or if the birth was breech or Cesarean.

Developmental delays. Created from five questions from the Birth Questionnaire that asked the mother of the participant about five developmental milestones; sitting, crawling, walking, talking, and using the toilet. An example question, 'at what age was (the participant) toilet trained and stayed dry’, was scored as positive if the child had not been toilet trained by the age of 4.

Peer Factors

Bad friends. Created from the Parents and Peers Scale and the Peers Scale, this construct combines five questions from the parent and five questions from the participant summarizing information on the boy’s association with friends of which the parent did not approve. Example questions: ‘Are there any children in your group of friends of whom your parents disapproved?’ and ‘Have your parents said that any of your friends were a bad influence on you?’ The original questions were dichotomous (α(Y) = .61; α(O) = .75).

Peer delinquency. For the oldest sample, the continuous version of this construct summarizes the proportion of friends who engage in 11 different forms of delinquency. The types of delinquency correspond to the items from the Self Reported Delinquency Scale. Examples of questions are: ‘How many of them [i.e., friends] have gone into or tried to go into a building to steal something?’ and ‘…attacked someone with a weapon with the idea of seriously hurting that person?’ (α(O) = .90). For the youngest sample, the construct is based on nine items and asks for the proportion of friends who engaged in delinquent behaviors (α (Y) = .82; (α(Y) = .79).

Peer substance use. The continuous version of this construct represents the proportion of friends who use alcohol or drugs or sell drugs. This construct was created from four items on the Peer Delinquency Scale. Examples of questions are: ‘How many of them [i.e., friends] used marijuana or hashish?’ and ‘…sold hard drugs such as heroin, cocaine, or LSD?’ (α (Y) = .86; (α(O) = .68).

School Factors

Low academic achievement. This construct combines parent (CBCL) and teacher (TRF) reports on each participant's academic performance in seven academic subjects. Each academic subject was rated on a four-point scale by parents, and a five-point scale by teachers that range from failing to above average. Ratings of each academic area by both informants were averaged creating the final dichotomized variable (average intercorrelations: (Y)=.56; (O)=.49).

Truancy. This construct is created from information from the boy using the SRA/SRD, YSR, the parent CBCL, and the TRF. The construct represents the lifetime estimate of whether a boy has been truant. Any positive response by any informant resulted in a positive score.

Suspended. This construct reflects whether the boy had ever been suspended from school. Data were combined from both the parent CBCL and the boy SRA/SRD. The construct was based on four questions regarding whether the boy had ever been suspended and any affirmative response led to a positive score for this construct.

Caretaker Factors

Poor supervision. This variable uses four questions each from the caretaker and child Supervision/Involvement Scale. The construct taps into how well the boy has been supervised during the previous six months. A sample question is: 'Do your parent (s) know who you are with when you are away from home?' (α (Y) = .63; α (O) = .75).

Physical punishment. This variable combines parent and child reports from the Discipline Scale, and was scored positive if either the boy or the parent reported that the parent slapped, spanked, or hit the boy during the previous six months. Example question: 'If your son does something that he is not allowed to do or that you do not like, do you…. Slap or spank him or hit him with something?'

Discipline not persistent. Created from the Discipline Scale, the questions used to create this variable reflect the degree to which a parent persisted with disciplinary actions during the previous six months, and were scored on a three-point frequency scale. For the oldest sample, this construct combines four items from the parent's report of her persistence in disciplining, and five items from the boy's report on the same behavior (α (Y) = .59). For the youngest sample, initially only the parent's report was available (α (O) = .48). A sample question in the boy's report is: 'If your mother asks you to do something and you don't do it right away, does your mother give up trying to get you to do it?'

Poor communication. This construct summarizes 28 questions from the boy and 30 questions from the parent regarding how directly or indirectly they communicate about emotions, disagreements, and problems. Created using the Revised Parent-Adolescent Communication Form, responses were coded on a 3-point frequency scale: 'almost never', 'sometimes', and 'always'. Example questions for the participant are: 'Do you tell your mother/father about your personal problems' and 'If your mother/father is angry with you, do you hear about it from someone else?' (α (Y) = .85; α (O) = .90).

Counter control. This construct is the summation of 11 items from the Counter Control Scale using parent report to assess whether the boy's behavior became worse when punished. Example questions: 'Do you think that your son will take it out on other children if you try to make him obey you?', and 'If you punish your son, does his behavior get worse?' (α (Y) = .78; α (O) = .81).

Teenage mother. This dichotomized construct was created using the Demographics Questionnaire, and divides mothers into those who were less than 20 years old at the time of the participant's birth, and those who were not.

Parental stress. Created from the Perceived Stress Scale, this construct measures the caretakers perceived stress and perceived ability to handle stress. Example question: ‘Have you found that you could not cope with all of the things you had to do?’ (α (Y) = .83; α (O) = .83).

Poor relationship with caretaker. This construct combines 13 items from the boy and 16 items from the Child’s Relationship with Parent/Siblings Scale reflecting how positively or negatively they perceived their relationship. Examples of questions for the participant are: ‘How often have you thought your mother/father bugged you a lot?’ and ‘How often have you liked being your mother/father’s child?’ For the parent example questions are: ‘How often have you thought your child was a good child?,’ and ‘How often have you wished he would just leave you alone?’ (α (Y) = .73; α (O) = .84).

Father behavior problems. This construct was created using one question from the Family Health Questionnaire and was scored as positive if the parent reported that the biological father had experienced behavior problems in the past. Question: ‘Do you know if [participant’s] biological father ever had any of the following problems? Behavior problems’.

Father alcohol problems. This construct is created from the Family Health Questionnaire and was scored as positive if the parent reported that the biological father had experienced alcohol problems in the past. Example Question: ‘Do you know if [participant’s] biological father ever had any of the following problems? Alcohol problems’.

Father drug problems. This construct is created from the Family Health Questionnaire and was scored as positive if the parent reported that the biological father had experienced problems with drugs in the past. Example Question: ‘Do you know if [participant’s] biological father ever had any of the following problems? Drug problems’.

Father school adjustment problems. This construct is created from the Family Health Questionnaire and was scored as positive if the parent reported that the biological father had experienced school adjustment problems in the past. Example Question: ‘Do you know if [participant’s] biological father ever had any of the following problems? School Adjustment problems’.

Parent antisocial attitude. This construct summarizes 18 questions asked of the parent regarding their attitude towards antisocial behaviors of the boy from the Perception of Child Problem Behavior Scale. Example questions: ‘Is it all right for your son to miss school for reasons other than being sick?’ and ‘Is it all right for your son to take a drink of alcohol?’ Questions were rated on a 3-point scale: ‘not all right’, ‘all right sometimes’ and ‘all right most of the time’ (α (Y) = .58; α (O) = .67).

Family Factors

Family on welfare. This dichotomized construct was created from the Demographics Questionnaire and was scored as positive if the participant’s caretaker reported that during the past year anyone living in the household with the boy was receiving welfare.

Official charge of neglect. Taken from data collected from the Allegheny Juvenile Court records, this variable was scored as positive if there was a petition in the court records charging a parent of the participant with neglect. If a parent was found to be in neglect; the children in their care including the participant would have been removed from the home and typically placed with another relative when possible or placed in some type of group home, temporary shelter, or foster home.

Neighborhood Factors

Bad neighborhood. This construct is based on a principal components analyses of 1990 U.S. Census tract data in which 88 Pittsburgh neighborhoods formed the units of analysis (see Wikström & Loeber, 2000, for a more detailed description). The factor analysis yielded two primary factors: 1) SES, comprised of several items including percent of families headed by single parents, median household income, and percent of households on public assistance, and 2) Familism comprised of mean household size, percent youth (ages 10-19), and percent living in the same residence for 5 or more years. Factor scores were then split at the median corresponding to lower or higher SES neighborhoods. Each boy was then matched with a neighborhood type based on his address during the first follow-up assessment.

Negative impression of crime. This construct was created from 10 questions regarding the caretakers perception of problems that occur in neighborhoods such as racial conflict, vandalism, prostitution, gambling, burglaries open drug use/dealing etc…from the Demographics questionnaire. Questions were scored on a 3-point scale: ‘not a problem’, ‘somewhat a problem’ and ‘big problem’.

Offense Factors

Given that there are hundreds of charges that an arrestee might be charged with, for the purpose of these analyses all charges were initially grouped into 30 different categories of offenses corresponding with the computerized format used by the National Center for Juvenile Justice, Pittsburgh, PA. Attempts to commit delinquent acts or charges of criminal conspiracy to commit delinquent acts were also coded as the criminal act itself. For example, if a participant was charged with attempted rape, the charge was coded as rape. Offense constructs were then created using the 30 categories. Four types of dichotomous offense constructs were created from these 30 initial categories: violent crimes, property crimes, drug crimes, and weapons crimes. For each of these four types of crimes, two separate constructs were created representing if the offender had been charged at the time of arrest with each type of crime and if the offender had been adjudicated on that type of crime. These constructs were created for first and second juvenile arrests, and recidivism after these two arrests. It should be noted that not all of the initial 30 categories of offenses were used to create the offense constructs. For example, two categories, disorderly conduct and traffic violations were not considered to fall into the four types of offenses examined here, and were therefore not used in the analyses. Table 3 provides the offense categories that were considered violent in this study and were used to construct the violence variables. Included in the table are specific offenses or acts that were included in each category. Table 4 lists the offenses that made up the property, drug, and weapons offense constructs.

All offense constructs were taken from official court data. With the exception of the Arrested/Adjudicated Prior to Age 13 variables, each of the offense variables were calculated separately for both first and second arrest. A third variable was also created for each type of offense that combined data from both first and second arrest. For example, for the variable Placed in Detention, three separate variables were created; 1) placed in detention after first arrest; 2) placed in detention after second arrest; and 3) placed in detention after either first or second arrest. This was done to ensure that when predicting violence from the second arrest, an offender’s immediate charges and entire offense history up to the point of second arrest could be evaluated for any differential predictive ability.

Table 3

Offense Categories Considered Violent and the Specific Behaviors Included

Offense category	Specific behaviors
Criminal homicide	Murder and manslaughter.
Robbery	Taking money or goods from another by violent force or threat of violence (ex: armed robbery).
Aggravated assault	Causing serious bodily injury to another intentionally, knowingly, or recklessly under circumstances manifesting extreme indifference to the value of human life (ex: shooting, cutting, or stabbing that does not result in death and was determined not to intentionally cause death).
Simple assault	Intentionally, knowingly, or recklessly causing bodily injury to another (ex: physical fights involving hitting with fists or objects).
Threats and endangering	Reckless endangerment, terroristic threats, causing or risking a catastrophe, violence with intent to terrorize another, or causing public terror (ex: death threats, bomb threats).
Car jacking	Taking a car from a driver of a car by force or threat of force.
Rape	Sexual intercourse by forcible compulsion or threat of forcible compulsion. Includes consensual sexual intercourse with persons under the age of 13.
Involuntary deviate sexual intercourse	Deviate sexual intercourse by forcible compulsion or by threat of forcible compulsion. Includes consensual deviate sexual intercourse with persons under the age of 13.
Aggravated indecent assault	Penetration of the complainant with a part of the person's body by forcible compulsion or threat of forcible compulsion. Includes such sexual acts that are consensual with persons under the age of 13.
Indecent assault	Sexual contact (ex: groping, fondling) by forcible compulsion or threat of forcible compulsion. Includes such sexual acts that are consensual with persons under the age of 13.

Note: Attempts of specific acts within each category were also coded under that category. Sex offenses not otherwise specified were not included in any of the above violent offense categories. Spousal sexual assault was also not included because although it was assessed there were no cases.

Table 4

Offenses Included in the Property, Drug, and Weapons Constructs

Construct	Offenses Included
Property Offenses	Burglary
	Larceny
	Auto theft
	Arson
	Forgery and counterfeiting
	Fraud
	Embezzlement
	Stolen property
	Vandalism
Drug Offenses	Narcotic laws (nos)
	Possession or sale of marijuana
	Possession or sale of cocaine
	Possession or sale of opiates
	Possession or sale of combo/other drugs
	Drug paraphernalia
	Misc. drug offense
	Liquor law offense
	Drunkeness
Weapons Offenses	Possession of an unregistered firearm
	Carrying a concealed weapon
	Sale of a weapon

Note: A number of offenses were not included in any of the above offense classifications including: sex offenses (not otherwise specified), voluntary deviate sexual intercourse, indecent exposure, criminal conspiracy, malicious mischief, disorderly conduct, traffic motor laws, violation of ordinances, and status offenses.

Placed in detention. This variable was scored as positive if an offender was placed in detention following arrest for any length of time. There are a number of reasons that the court may decide to place an offender in detention including that the arrest charges were more serious such as weapons or violence offenses, that the court did not feel as though the offender would receive proper supervision if returned to their home, or that the offender would fail to appear for their petition hearing if not detained. This variable was only able to be collected through the

Allegheny County Juvenile Court Information Management Office Records and Prothonotary Office Records, and was not available from the other official records sources.

Codefendants. Scored as positive if others were also charged along with the offender at the time of arrest. Other offenders may have been charged with the same or similar offense(s) or charged with completely different offenses. In the JJS, codefendants arrested along with the participant are typically listed with the primary offender on the primary offender's court petition. This variable was only able to be collected through the Allegheny County Juvenile Court Information Management Office Records and Prothonotary Office Records, and was not available from the other official records sources.

Violent arrest charge. This variable was scored as positive if a participant was charged with a violent offense as defined by those charges listed in Table 3, during arrest. An offender may not have been adjudicated on these charges for a number of reasons such as lack of evidence, evidence supporting the offenders innocence, the victim(s) failure to testify against the offender, or if the offender was adjudicated the charges may have been reduced to a less serious offense.

Weapons arrest charge. This variable was scored positive if a participant was charged with a weapons crime during arrest. Charges included carrying a concealed weapon, possession of an illegal firearm, and illegal use of a firearm.

Drug arrest charge. This variable was scored positive if an offender was charged with a drug related offense during arrest. Charges include possession of any of a number of illegal substances, sale of an illegal substance, and sale of a look-a-like substance.

Adjudicated first arrest charge. This variable was scored positive if an offender was adjudicated on at least one charge of any type associated with their first arrest.

Adjudicated violent charge. This variable was scored positive if a participant was adjudicated on a violent charge offense as defined by those charges listed in Table 3, associated with arrest.

Adjudicated weapons charge. This variable was scored positive if an offender was adjudicated on a weapons charge associated with arrest. Adjudicated charges included carrying a concealed weapon, possession of an illegal firearm, and illegal use of a firearm.

Adjudicated drug charge. This variable was scored positive if an offender was adjudicated on a drug related offense during arrest. Charges include possession of any of a number of illegal substances, sale of an illegal substance, and sale of a look-a-like substance.

Juvenile arrest prior to age 13. This construct was scored as positive if an offender was arrested on any charge prior to their 13th birthday.

Adjudicated charge from arrest prior to age 13. This construct was scored positive if an offender was arrested on any charge prior to their 13th birthday, and then later adjudicated of at least one charge from that arrest.

Outcome Variables

Official Violence

The primary outcome of interest studied here, official adjudicated violence, was defined as any adjudicated criminal act or threat of that act, the result of which is to cause bodily harm to another person. This construct was created in the same manner that the predictors measuring violence were constructed, using the offense categories listed in Table 3. However, only charges that participant's were adjudicated or found guilty of were included to examine as closely as possible criminal acts that the participant indeed committed. To assess adjudicated violence after first arrest, only adjudicated charges that were associated with participant arrests after their initial juvenile arrest were included. The adjudicated violence after second arrest construct was created in a similar manner with only adjudicated charges associated with arrests after second juvenile arrest being included.

Reported Violence

This construct, created from a reported violence construct, was previously created (Loeber et al., 2005) and included self', parent', and teacher-reported violent behavior. The construct was based on the general Delinquency Seriousness Classification (Loeber et al., 1998; Wolfgang et al., 1985) and was scored as positive if any informant reported that the participant had committed rape, robbery, or aggravated assault. The previous construct was created for phases up to age 20 for the youngest sample, and up to age 23 for the oldest sample. The construct used here was scored as positive if the construct created by Loeber et al. (2005) was scored as positive for any phase that occurred more than 6 months after the initial juvenile arrest for biannual assessments, and more than 12 months after yearly assessments. This was done to create a reported violence construct that resembled as closely as possible only reported charges that occurred after the first juvenile arrest, but did not include charges associated with the first arrest.

Analyses

Prevalence of Reported and Official Violent Offending

Prevalence rates of violence after first arrest for both the oldest and youngest cohorts were examined for the purpose of comparing how the prevalence of reported violence in the PYS compares to the prevalence of officially reported violence as measured by adjudication of a violent charge. Given that the PYS is a study of high risk youth, all prevalence rates were weighted to adjust for the over selection of participants who were considered to be high risk when selected for inclusion of the study. Prevalence rates are provided for both the construction and validation samples.

Development of Risk Assessment Instruments

To examine the usefulness of an actuarial tool to predict violence, two methods of risk assessment, the Burgess or additive method and the iterative classification method were compared to determine their ability to predict violence in a high-risk sample. Each of these methods were developed separately following procedures similar to those outlined by Gottfredson and Synder (2005) for developing risk assessment instruments. These procedures included defining outcome criterion, selecting predictors, measuring relations between predictor variables and the outcome variable, assessing the predictive utility of combining predictors using either the Burgess or iterative classification method, and cross-validation.

Prior to beginning the analyses, a construction and validation sample was created by randomly splitting both the youngest and oldest cohorts into two separate samples with approximately equal numbers of participants in each sample. This was done so that both RAIs could be cross-validated, a necessary step to validating any predictive instrument (Cronbach & Meehl, 1955). One half of the youngest cohort and one half of the oldest cohort were then combined to form a construction sample. The other two halves from the youngest and oldest cohort samples were then joined to form a validation sample.

Predictors to be examined. There were a total of 58 variables that were assessed for their relation to violence after arrest and possible inclusion in each of the RAIs. Each of these variables is listed in Table 2. Forty-five previously created predictors (Loeber et al., 1998) from data collected from the participants, their caretakers and teachers were selected based on theoretical grounds and extensive previous research described earlier demonstrating a relation between these variables and violence. Another 12 arrest variables and the official charge of neglect variable created from the collection of official court records and thought to be related to future violent behavior were also examined.

Selecting and combining predictors. Following the Burgess approach, predictors that were found to be associated with adjudicated violence were selected for inclusion in the equation. Given that this method does not take into account intercorrelations among predictors, no assessment concerning the relations between predictors was needed. Once the equation was created, risk scores based on the number of risk factors present were created for all offenders. Groups with similar risk scores and rates of violent recidivism were then combined resulting in two groups, high- and low-risk offenders.

To conduct predictive attribute analysis and any subsequent iterative classification, computer software known as Chi-squared Automated Interaction Detector (CHAID; SPSS 14.0) was used. Unlike the Burgess method, this procedure takes into account relations between predictor variables to select the best variables to be used to create a classification tree that allows for the largest number of offenders possible to be classified as low or high risk. Using Figure 2 as a template for carrying out the predictive attribute analysis, all 1st time offenders were initially split into groups based on the variable most related to adjudicated violence after first arrest. These groups were then separately analyzed with each group again being split into two more groups based on the best predictor to the outcome for each group. This process was continued until it was not possible to further classify offenders. Once this process was complete, those offenders who could not be classified as either high or low risk, were then re-analyzed in a second iteration using the same procedure in an attempt to classify these offenders. Following the work of Steadman et al. (2000), high-risk groups were defined as having a prevalence of violence at least twice that of the base rate in the construction sample, with low risk groups having a prevalence violence half or less than that of the base rate. It was expected that this iterative classification approach (Steadman et al., 2000), would lead to the identification of a different set of predictors than using those identified using the Burgess method. There are two reasons for this first, predictive attribute analysis and iterative classification identifies predictors specific to groups of offenders with similar characteristics. Second, predictive analysis and iterative classification takes into account the interrelationships among predictors and the Burgess approach does not.

Assessing validity. To assess the validity of each of the instrument building procedures on the construction sample, four separate criteria were used. First, the Burgess method and predictive attribute analysis were compared on the percentage of offenders that each procedure was able to classify as either high or low risk. The percentage of offenders correctly identified is equal to the number of true positives plus the number of true negatives, divide by the total number in the sample. The sensitivity (true positive rate) and specificity (true negative rate) of each measure is also reported here. Second, the RIOC (Loeber & Dishion, 1983) statistic discussed earlier that corrects for chance while taking into account the selection ratio and the base rate is also reported for each of the RAIs. Third, the phi (ϕ)statistic was used to assess the association between each of the classification models and future adjudicated violence. The appropriate test to compare differences with a dichotomous outcome is the chi-square statistic. Phi was used here as an estimate of correlation calculated from chi-square between risk categories created from each of the procedures and the studied outcome, future adjudicated violence. Phi represents the correlation between two dichotomous variables, and is the same as the computing Pearson's correlation for two dichotomous variables. Forth, the predictive accuracy of each of these methods was also evaluated using receiver operation characteristics (ROC). An ROC analysis plots the sensitivity and 1-specificity pairs that result as a single decision threshold is moved from the lowest to the highest possible value. The ROC analysis is optimal for evaluating the usefulness of the entire instrument, not using a specific cut-point, given that the results are independent of the base rate of the outcome being examined, in this case violence. The ROC analysis produces the Area Under the Curve (AUC) statistic which ranges from .5 (accuracy is equivalent to chance) to 1. (accuracy is perfect).

Cross-validation. The procedures described above were first evaluated on the construction sample and then the validation sample. First time offenders in the validation sample were classified into high- or low-risk groups by applying the risk assessment equations produced by the Burgess and predictive attribute methods on the construction sample. Each of the RAIs were then compared on how well they performed on the validation sample using the four criterion used to assess performance on the construction sample.

Results

Official Offending Among PYS Participants

Most youth commit some delinquent act prior to the age of 18 (Shannon, 1988), although most are never arrested. It is estimated that approximately 20% of white males and 40% of black males will be arrested prior to their 18th birthday (Blumstein, Roth & Cohen, 1986). To examine the processing of offenders during their first juvenile arrest and future juvenile and adult offending, a number of flow charts were created for both youngest and oldest cohorts, and both cohorts combined. The first of these charts, Figure 4, shows the progression of offending in the youngest cohort, with the left hand side of the chart providing information on first juvenile arrest, and the right hand side providing detailed reoffending information that occurred after first juvenile arrest. In the youngest cohort, just under half (45.9%) of participants were arrested at least once as a juvenile, with 42.6% of those arrested actually being adjudicated on some charge associated with the first arrest. A surprising number of those who were arrested at least once as a juvenile were arrested again sometime in the future. More than three out of four first time juvenile offenders were arrested again sometime before the age of 24. Whether an offender was adjudicated delinquent on their first juvenile arrest seemed to make little difference in terms of whether they would be arrested again in the future, because more than 75% of both of these groups were arrested again at a later time. Of 202 participants who were arrested at least once as a juvenile, 58 (28.7%) were later convicted of a violent offense sometime before the age of 24. Figure 5 is a schematic diagram representing the makeup of first time juvenile offenders in the youngest sample, and offenders that went on to reoffend. As can been seen from Figure 5, of all first time juvenile offenders, less than one in five was adjudicated on a violent offense associated with their first arrest. Among offenders who were arrested again at least once, a substantial percentage (74.4%) at some point after their first arrest were adjudicated or found guilty of some

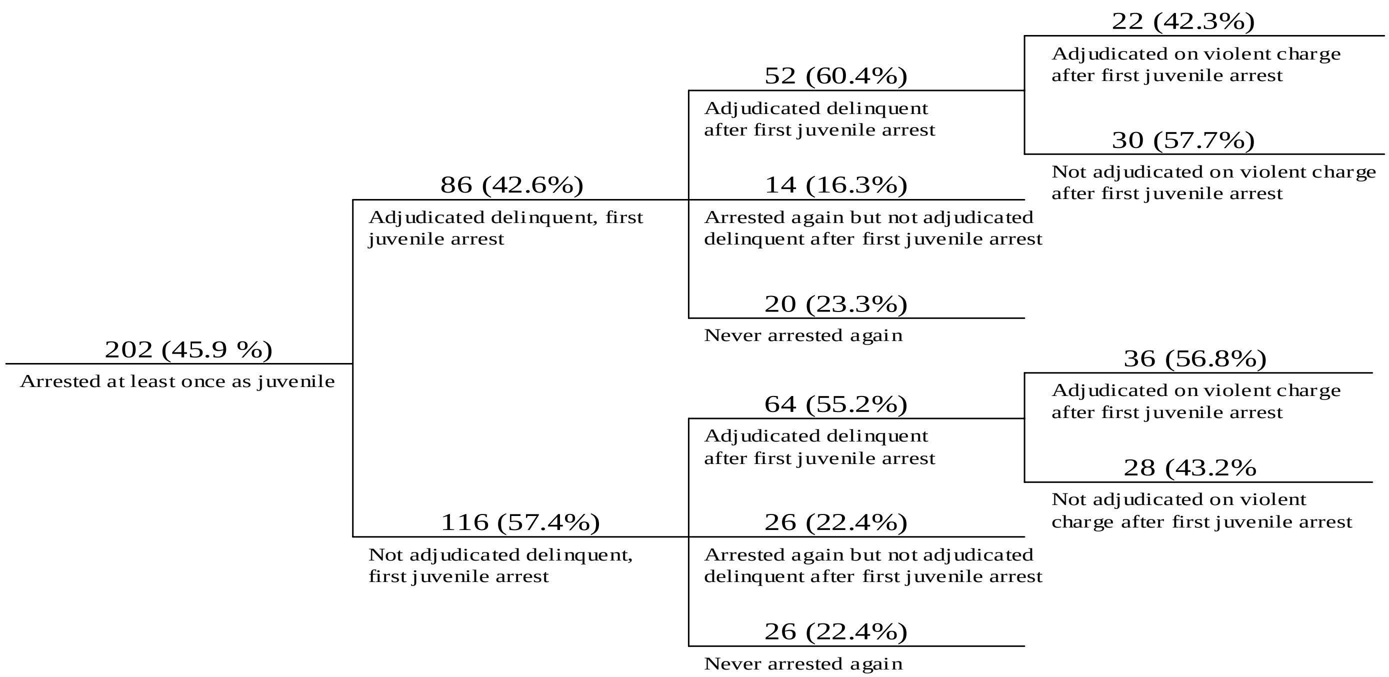

Figure 4. Juvenile justice flow chart showing the progression of offending in the youngest cohort. Of 202 participants who were arrested at least once as a juvenile, 58 (28.7%) were later adjudicated or convicted of a violent offense sometime before the age of 23.

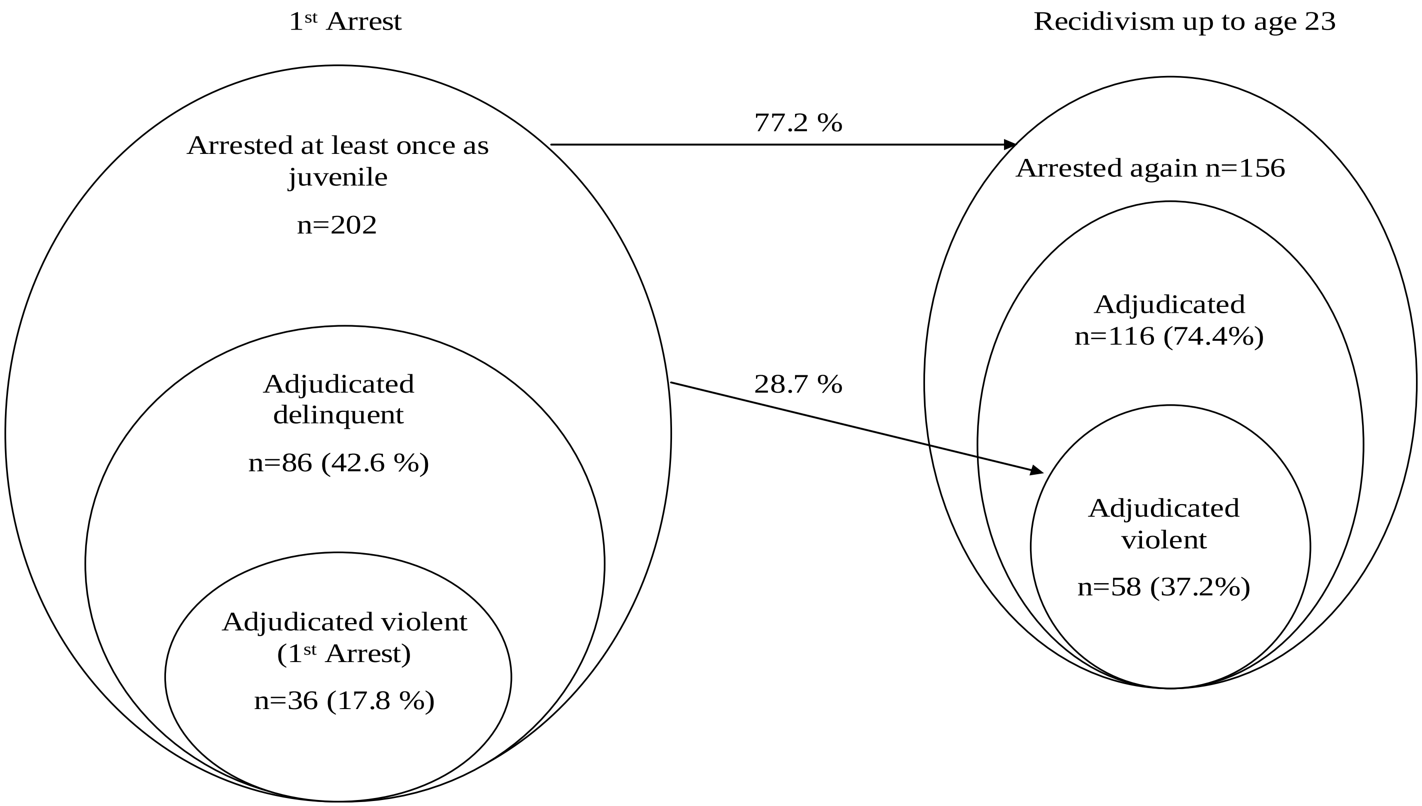

Figure 5**.** A schematic diagram representing the makeup of 1st time juvenile offenders in the youngest sample, and offenders that go on to re-offend.

charge. In regards to violent reoffending, just under one third (28.7%) of all first time juvenile offenders would later be adjudicated or found guilty of committing a violent crime.

Figures 6 and 7 provide analogous data for the oldest cohort. Figure 6 shows that of the 221 participants who had at least one juvenile arrest, 96 or 43.4% were adjudicated delinquent for at least one charge on their first arrest. A striking number of participants who were arrested at least once as a juvenile were arrested again, with 85.1% of all juvenile offenders being arrested again at least once before the age of 31. Figure 7 illustrates that although only 10.9% of juvenile offenders were adjudicated on a violent offense on their first juvenile arrest, just over 1/3rd (36.2%) would eventually be adjudicated or found guilty of a violent offense in the future, a percentage that was just somewhat higher than that found for the youngest cohort (28.7%).

Figures 8 and 9 are similar flow charts for both the youngest and oldest cohorts combined. Of all participants included in this study having at least one juvenile arrest (N=423), 43% were adjudicated delinquent of at least one charge associated with their first arrest. The total number of juvenile offenders who were arrested at least twice was 344 or 81.3% of all juvenile offenders. A total of 60 or 14.2% of all juvenile offenders were adjudicated on a violent charge associated with their first arrest. The rate of adjudicated violence among recidivists were much higher, with 40.1% of offenders arrested at least twice having been adjudicated or found guilty of a violent offense after their first arrest.

Data presented in Figure 10 show the percentage of offenders adjudicated on a violent charge associated with first and second juvenile arrests and recidivism after second arrest. Although the outcome windows are somewhat different (recidivism included arrest data up to age 23 for youngest cohort, and age 30 for oldest cohort), the figure shows that the percentage of offenders who were adjudicated of a violent offense were similar for both cohorts. However,

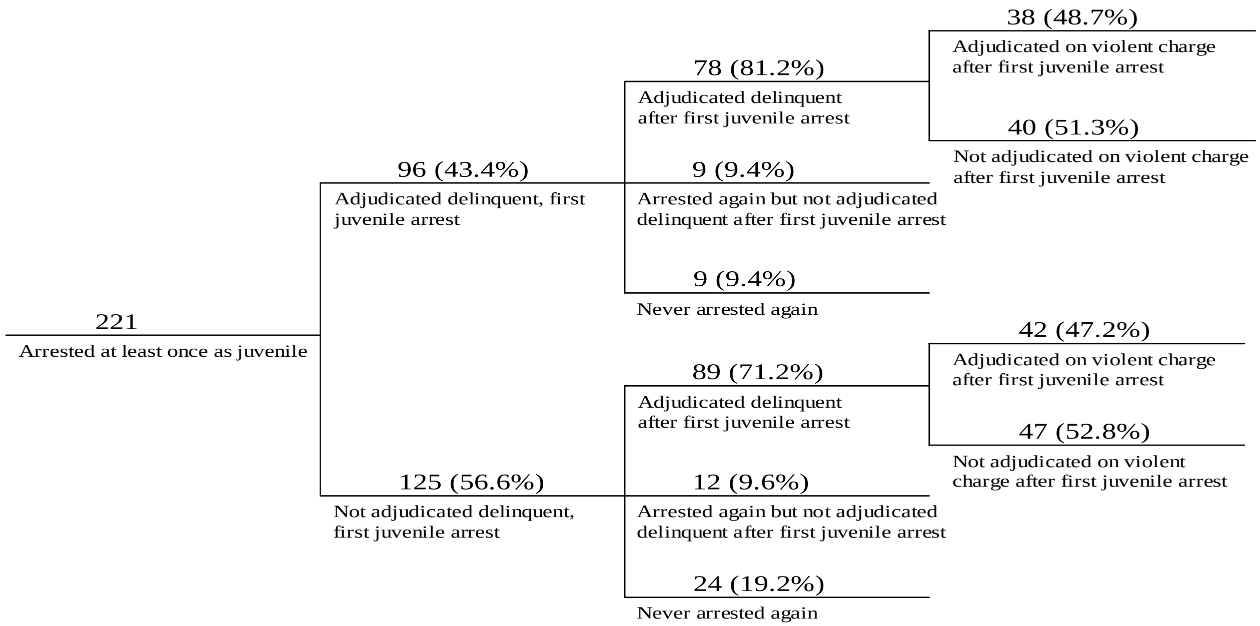

Figure 6. Juvenile justice flow chart showing the progression of offending in the oldest cohort. Out of the 221 participants who were arrested at least once as a juvenile, 80 (36.1%) were later adjudicated or convicted of a violent offense sometime before the age of 31.

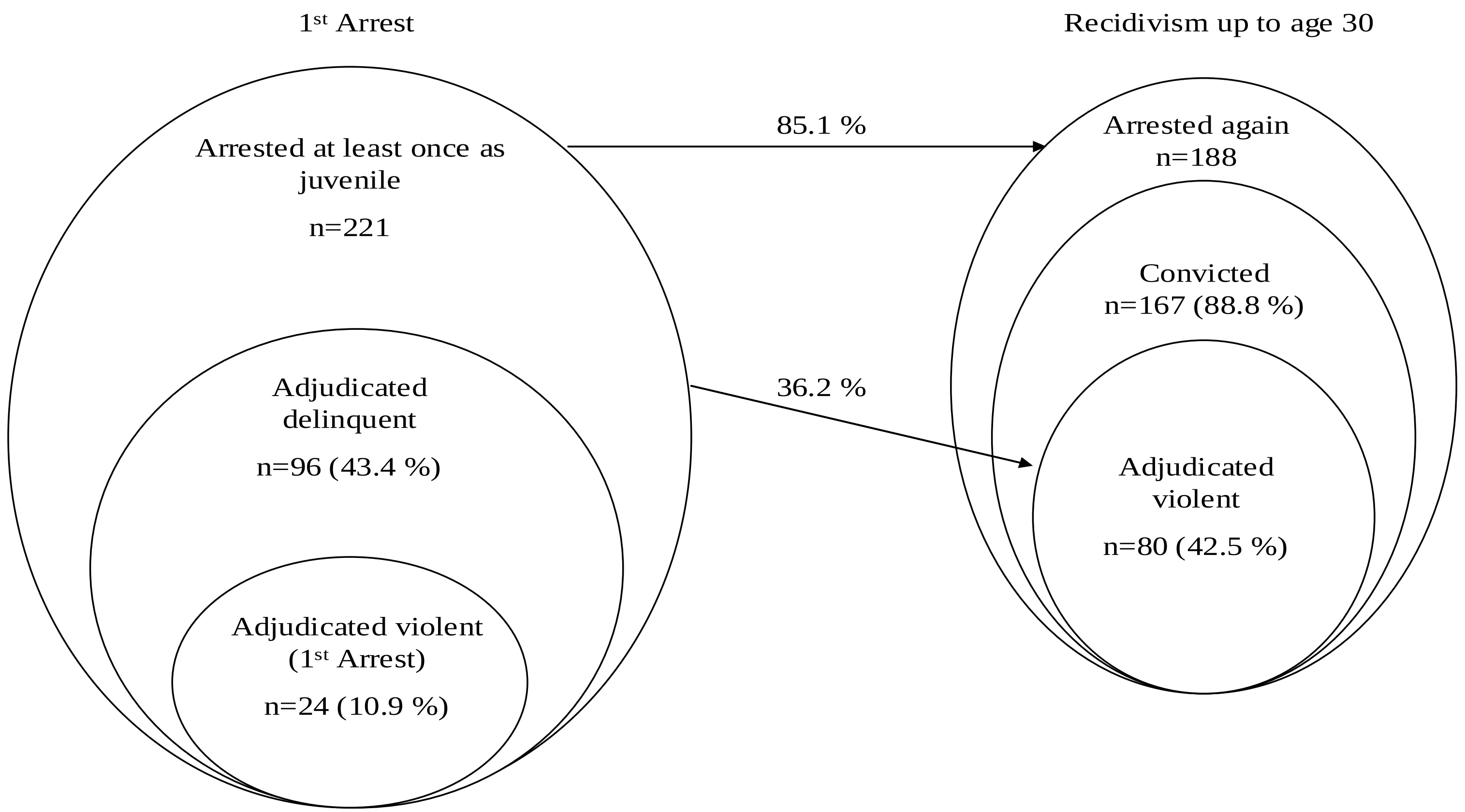

Figure 7. A schematic diagram representing the makeup of 1st time juvenile offenders in the oldest cohort, and offenders that go on to reoffend.

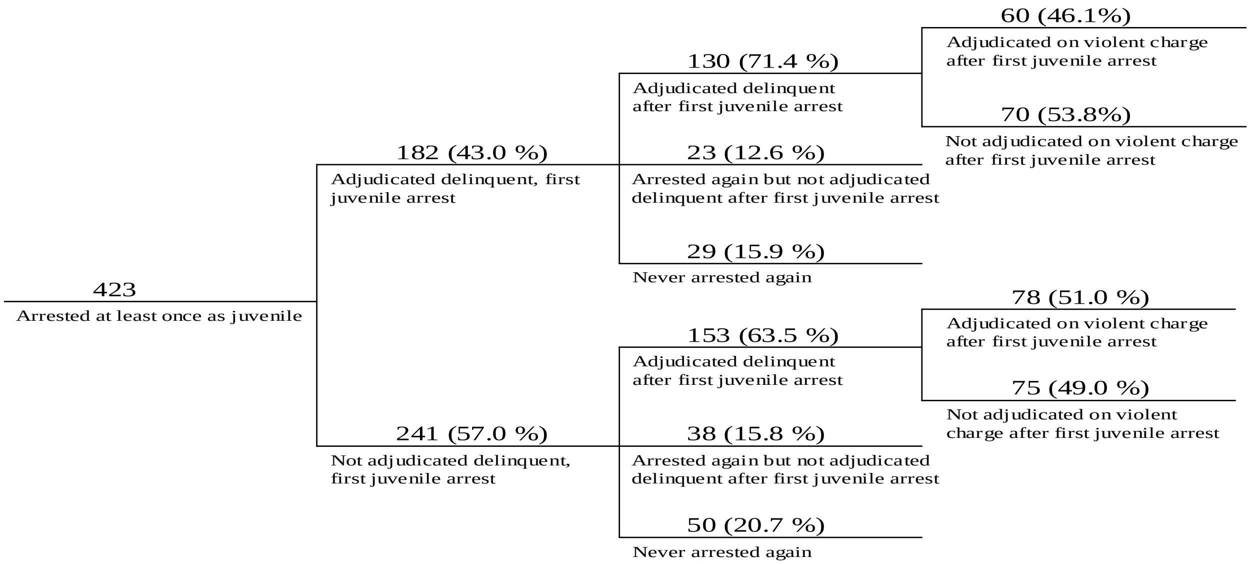

Figure 8. Juvenile justice flow chart showing the progression of offending in the youngest and oldest cohorts combined. Of 423 participants who were arrested at least once as a juvenile, 138 (32.6%) were later adjudicated or convicted of a violent offense.

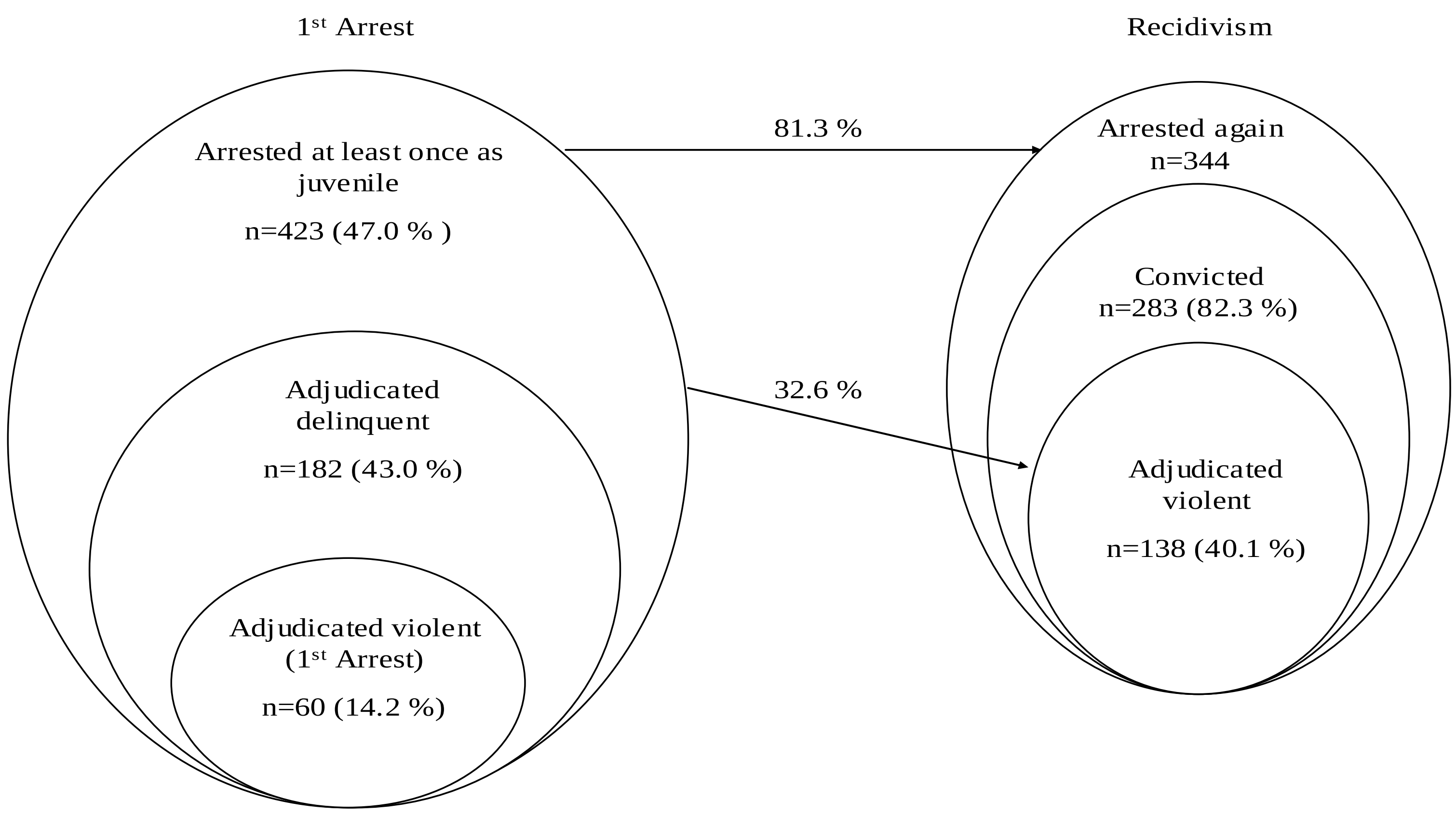

Figure 9. A schematic diagram representing the makeup of 1st time juvenile offenders in the oldest and youngest cohorts combined, and offenders that go on to re-offend.

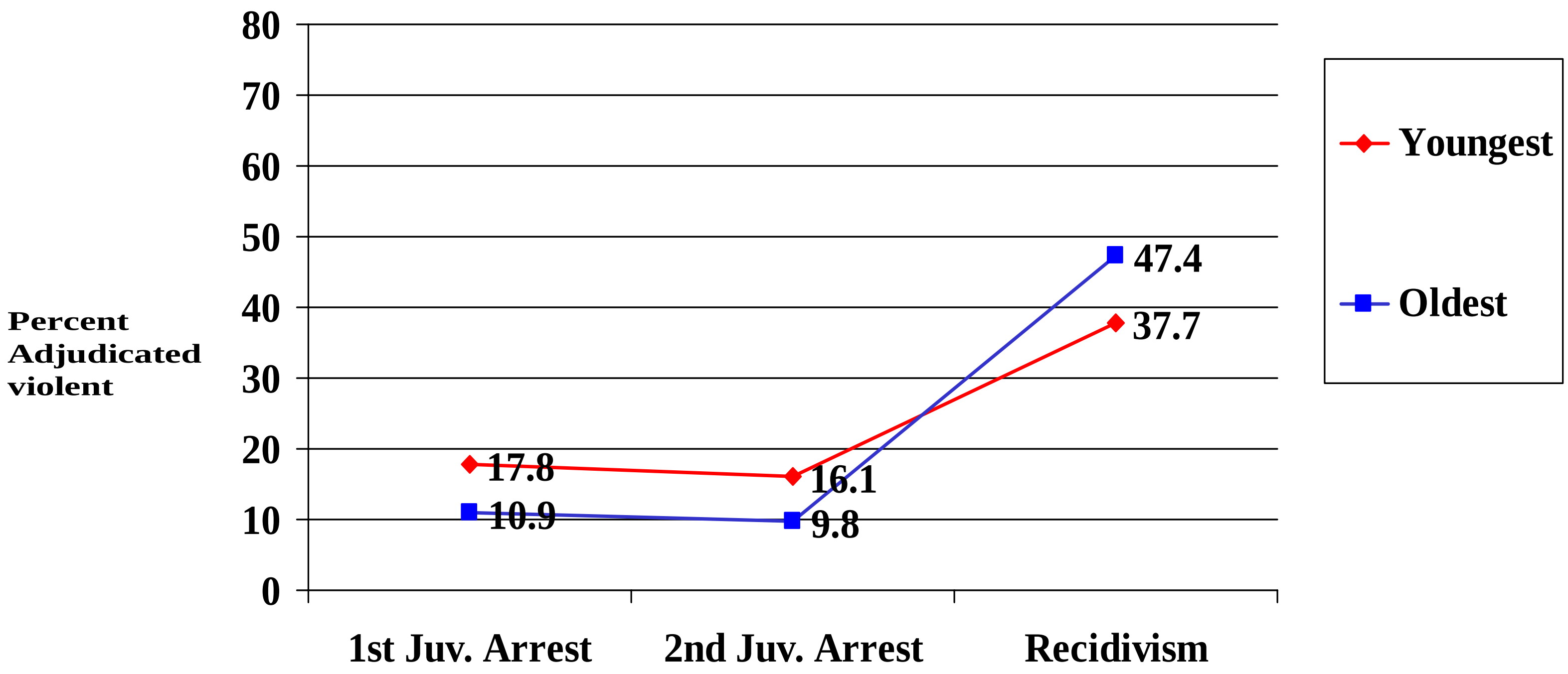

Figure 10. Percentage of offenders adjudicated on a violent charge associated with first and second juvenile arrest, and recidivism after second juvenile arrest. Recidivism included arrest data up to age 23 for youngest cohort, and age 30 for oldest cohort.

fewer numbers of juvenile offenders in the oldest cohort were adjudicated delinquent on their first and second juvenile arrest than offenders in the youngest cohort despite the fact that more juvenile offenders in the oldest cohort went on to be adjudicated or found guilty of a violent offense than in the youngest cohort.

Table 5 presents the percentage of juvenile offenders who were adjudicated on violent, property, drug, and weapons charges. Included in the table are data for first and second juvenile arrests, and recidivism after each arrest first and second juvenile arrest. Data are provided for the youngest and oldest cohorts separately and for both cohorts combined. For the youngest cohort, more juvenile offenders were adjudicated on at least one violent offense than any of the other offense types at first juvenile arrest. For adjudicated charges associated with the second juvenile arrest and recidivism measured after first and second juvenile arrest, property charges were the most common, with the same number of offenders receiving a violent charge after second juvenile arrest. However, for the oldest sample, more juvenile offenders were adjudicated on a property charge for their first juvenile arrest than any other offense type including a violent offense. This also held when both cohorts were combined, with 21.8% of juvenile offenders being adjudicated on a property offense on their first arrest, and 14.2% being adjudicated on a violent offense. For those offenders who were later arrested again as a juvenile, the most frequent offense type for which they were adjudicated delinquent was also some type of property crime, with violence being the next most frequent. Across all four points of assessment, weapons charges were the least common type of charge that offenders were adjudicated or found guilty of committing. Figure 11 provides a graphic representation of the percentage of offenders who were adjudicated of four types of specific offenses studied here (violence, property, drug, and weapons offenses). Figure 11 combines data from the youngest and oldest cohorts, for first juvenile arrest

Table 5

Prevalence of Specific Types of Adjudicated Offending For First and Second Juvenile Arrests and Recidivism

Offense Type	Youngest Cohort *n*	%	Oldest Cohort *N*	%	Combined Cohorts *N*	%
First juvenile arrest						
Arrested	202		221		423	
Violent	36	17.8	24	10.9	60	14.2
Property	30	14.8	62	28.0	92	21.8
Drug	18	8.9	11	5.1	29	6.9
Weapon	7	3.5	3	1.2	10	2.4
Second juvenile arrest						
Arrested	134		141		275	
Violent	22	16.1	14	9.8	36	13.1
Property	25	18.7	52	36.9	77	28.0
Drug	12	9.0	9	6.4	21	7.6
Weapon	2	1.5	4	2.8	6	2.2
Recidivism after first juvenile arrest						
Arrested	156		188		344	
Violent	58	37.5	80	42.5	138	40.2
Property	64	41.1	95	50.3	159	46.1
Drug	36	22.9	83	44.2	119	34.5
Weapon	8	5.4	37	19.5	45	13.1
Recidivism after first two juvenile arrests						
Arrested	114		135		249	
Violent	43	37.7	64	47.4	107	43.0
Property	43	37.7	61	45.2	104	42.0
Drug	28	24.6	66	48.9	94	37.8
Weapon	8	7.0	30	22.2	38	15.3

Note. The n's and percentages reported have been weighted for risk status at start of study.

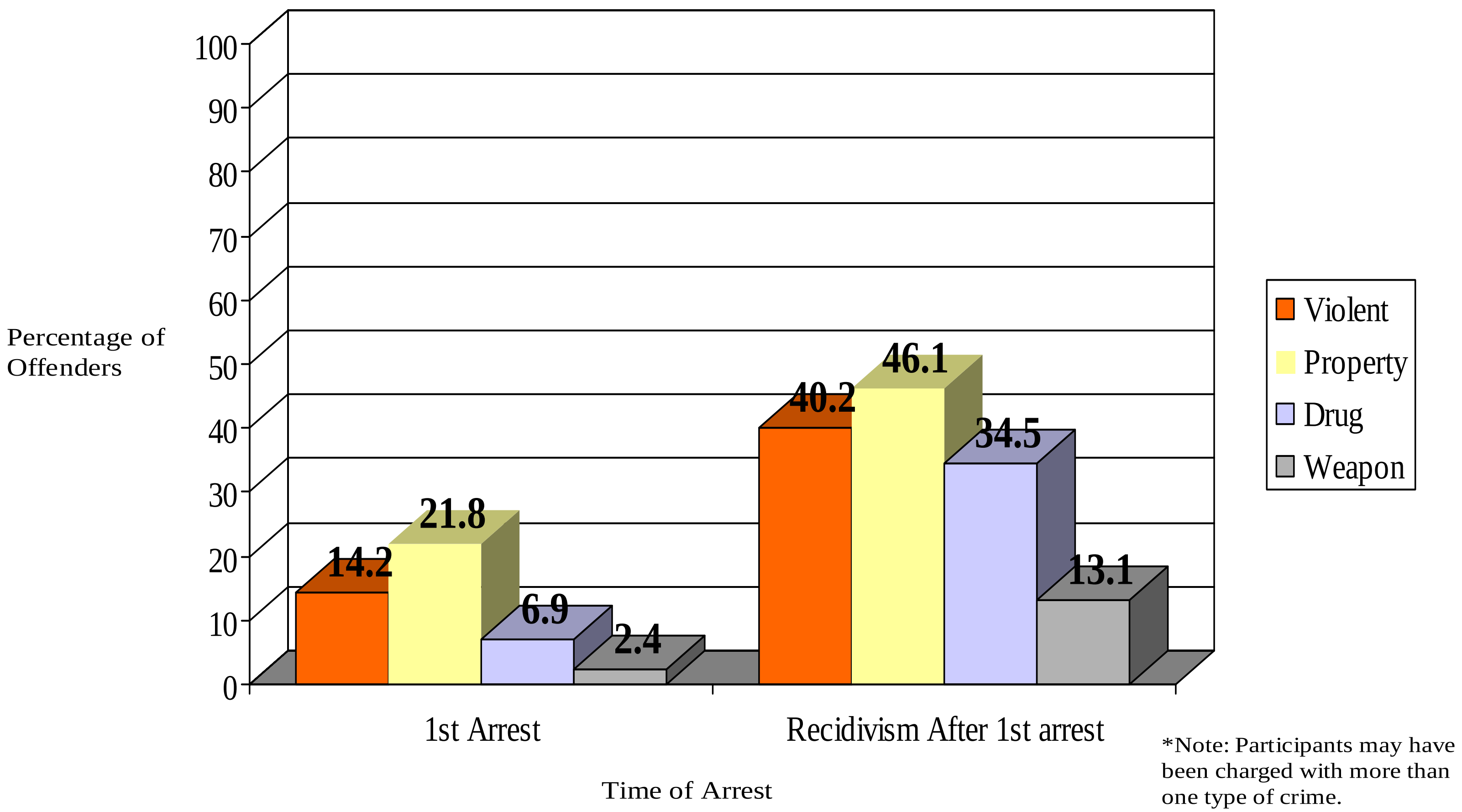

Figure 11. Percentage of Offenders Adjudicated on Specific Types of Offense for 1st Arrest and Future Recidivism.

and recidivism after first arrest. As can be seen from the Figure, property offenses were the most frequent type of offense that juveniles were adjudicated of on their first juvenile arrest, and future recidivism after the first juvenile arrest. Violent offenses were the next most frequent type of offense for which offenders were adjudicated. Just over 14% of juvenile offenders were adjudicated delinquent of a violent offense associated with their first arrest, and 40.2% of juvenile offenders were adjudicated or found guilty of a violent offense after their first juvenile arrest. Weapon offenses were the least common type of offense for which offenders were adjudicated or found guilty of committing. For first juvenile arrest, being adjudicated delinquent on a weapons charge was very rare with only 2.4% of first time juvenile offenders being adjudicated on a weapons charge. Although the least common offense type that offenders were adjudicated or found guilty of committing, among recidivists 13.1% at some point were adjudicated or found guilty of a weapons charge.

Construction and Validation Sample

Prior to forming a construction and validation sample, a comparison between the youngest and oldest cohorts was conducted. To examine possible differences between cohorts, the youngest and oldest cohorts were compared on a number of important variables including race (African American), neighborhood status (bad vs. average neighborhood), high risk at the start of the study, adjudicated on a violent offense after first arrest, and arrested prior to the age of 13. Chi-square analysis was conducted to test for any differences between groups on these variables. As can be seen from Table 6, the youngest and oldest cohorts were similar in terms of the variables listed above. A t-test that was also performed to examine differences in socio-economic status between the construction and validation samples was not significant. Although there were more first time offenders that were later adjudicated of a violent offense in the oldest

Table 6

Comparison of Youngest and Oldest Cohorts

Variable	Youngest cohort %	Youngest cohort N	Oldest cohort %	Oldest cohort N
African American	71.3	144	67.4	149
Bad neighborhood	31.0	63	28.0	61
High risk at start of study	48.0	97	45.2	100
Adjudicated violent after 1st arrest	28.7	58	36.2	80
Any juvenile arrest prior to age 13	27.1	55	14.5	32

Note. No significant differences were found between the youngest (n=209) and oldest (n=243) cohorts for any of the variables examined.

than in the youngest cohort, this difference was not significant. This finding was perhaps not surprising given that the outcome window for the oldest sample was somewhat larger for the oldest than for the youngest cohort, giving those offenders in the oldest sample additional years to be adjudicated on a violent offense than those in the youngest cohort.

Following the procedures outlined in the method section, prior to constructing a risk assessment equation using either the Burgess method or predictive attribute analysis, construction and validation samples were formed. This was done by combining the youngest and oldest cohorts into a single data file, and then randomly assigning cases to either the construction sample or the validation sample using SPSS to generate a variable with randomly assigned values of either 0 or 1 to all cases. Similar to the comparison that was done between the youngest and oldest cohorts, any possible differences between the construction sample and the validation sample were examined by comparing these groups on the same variables. Table 7 shows the number and percentage of offenders possessing each of these characteristics in the construction and validation samples. Chi-square analyses were conducted to test for any differences between

Table 7

Comparison of Construction and Validation Samples

Variable	Construction sample %	Construction sample N	Validation sample %	Validation sample N
African American	69.2	146	69.5	148
Bad Neighborhood	30.5	64	28.1	59
High Risk at Start of Study	47.4	100	45.8	97
Adjudicated Violent After 1st Arrest	34.4	73	31.0	66
Any Juvenile Arrest Prior to Age 13	23.7	50	17.5	37

Note. There were no significant differences between the construction sample (n=225) and validation sample (n=227) for the variables tested above, suggesting that no selection bias occurred when creating the construction and validation samples.

groups on these variables; no significant differences were found. No significant differences were found between the construction and validation samples, indicating that that no selection bias occurred when the two samples were formed.

Missing Data Analysis

Prior to selecting the predictors, a missing data analysis was conducted to ensure that none of the predictor variables had high levels of missing data. It was decided that any predictor having over 15% of the data missing would not be used in either risk prediction equation, although all variables would be examined initially to test their association to violent offending after first juvenile arrest. It is important to note that an important difference between the Burgess method and predictive attribute analysis, is how each method is able to deal with missing data. The Burgess method allows a risk score to be created even when there are some data missing for a given offender. For example, if there are 10 predictor variables in a Burgess based RAI and a juvenile court can only gather data for 7 out of the 10 variables, a “pro-rated” risk score can be calculated. In this example with the offender having 3 missing predictors, if out of the remaining 7 available predictors an offender was positive for 4 predictors, their level of risk would be:

5.71=((4/10-3)x10) or 5.71=((4/7)x10). Using predictive attribute analysis, however, if a variable used in the classification tree is missing, there is no accurate way to properly classify the offender. Thus, the reason for limiting the variables used to those with less than 15% missing was in part driven by the fact that if a variable with more than that amount of missing data was used in the final classification scheme produced by the predictive attribute analysis, a large portion of offenders would not be able to be classified with this method. Listed in Table 8 are each of the predictor variables examined and the amount of missing data for each variable. Both the number of cases and the percentage missing for the construction and validation samples is provided. As can be seen from the table, overall there was very little missing data; however, there were a number of variables that did have over 15% of the data missing and thus were not included in the final prediction equations. These variables included gang membership, father behavior problems, father school adjustment problems, placed in detention, codefendants, and developmental delays. Although these variables were not used in constructing the risk prediction equations, they were still examined for their association to future violent offending.

Table 8

Amount of Missing Data For Each Predictor of Violence

Risk factor	Construction sample N	Construction sample %	Validation sample N	Validation sample %	Total sample N	Total sample %
			Child factors			
Physical aggression	0	0	0	0	0	0
Cruel to people	0	0	0	0	0	0
Cruel to animals	0	0	0	0	0	0
Lack of guilt	0	0	2	.9	2	.4
Depressed mood	0	0	0	0	0	0
Noncompliance	7	3.1	14	6.2	21	4.6
DSM-IIIR Conduct Disorder diagnosis	7	3.1	12	5.3	19	4.2
DSM-IIIR Attention Deficit Hyperactivity Disorder diagnosis	24	10.7	13	5.7	37	8.2
Covert behavior	1	.4	2	.9	3	.7
Past correctional institution	0	0	0	0	0	0
Psychopathy without delinquency	0	0	3	1.3	3	.7
Boy not involved	1	.4	3	1.3	4	.9
Unlikely to get caught	1	.4	1	.4	2	.4
Positive attitude towards substance use	0	0	1	.4	1	.2
Positive attitude towards problem behavior	0	0	2	.9	2	.4
Positive attitude to delinquency	0	0	1	.4	1	.2
Weapon use	10	4.4	11	4.8	21	4.6
			Developmental factors			
Mother's alcohol use during pregnancy	26	11.5	31	13.6	57	12.6
Child premature	29	12.9	30	13.2	59	13.1
Prenatal problems	27	12.0	32	14.1	59	13.1
Perinatal problems	28	12.4	31	13.7	59	13.1
			Peer factors			
Bad friends	1	.4	4	1.8	5	1.1
Peer delinquency	4	1.8	9	4.0	13	2.9

			Screening for Violent			
Peer substance use	12	5.3	6	2.6	18	4.0
	School factors					
Low academic achievement	0	0	0	0	0	0
Truancy	0	0	0	0	0	0
Suspended	1	.4	2	.9	3	.7
	Caretaker factors					
Poor supervision	1	.4	0	0	1	.2
Physical punishment	1	.4	1	.4	2	.4
Discipline not persistent	11	4.9	5	2.2	16	3.5
Poor communication	0	0	0	0	0	0
Counter control	1	.4	2	.9	3	.7
Teenage Mother	10	4.4	11	4.8	21	4.7
Father alcohol problems	26	11.6	26	11.5	52	11.5
Father drug problems	24	10.7	29	12.8	53	11.7
High parental stress	1	.4	2	.9	3	.7
Poor relationship with caretaker	0	0	0	0	0	0
Parent antisocial attitude	1	.4	2	.9	3	.7
	Family factors					
Family on welfare	3	1.3	10	4.4	13	2.9
Official charge of neglect	1	.4	1	.4	2	.4
	Neighborhood factors					
Neighborhood status	2	.9	3	1.3	5	1.1
Negative Impression of Crime	2	.9	3	1.3	5	1.1
	Arrest factors					
Violent Arrest Charge	0	0	0	0	0	0
Weapons Arrest Charge	0	0	0	0	0	0
Drug Arrest Charge	0	0	0	0	0	0
Adjudicated Charge	0	0	0	0	0	0
Adjudicated Violent Charge	0	0	0	0	0	0
Adjudicated Weapons Charge	0	0	0	0	0	0
Adjudicated Drug Charge	0	0	0	0	0	0
Juvenile Arrest Prior to Age 13	0	0	0	0	0	0

			Screening for Violent			
Adjudicated Charge From Arrest Prior to Age 13	0	0	0	0	0	0
Variables Eliminated Due to More Than 15% Missing in Construction Sample						
Developmental delays	68	30.2	64	28.2	132	29.2
Gang membership	100	44.4	84	37.0	184	40.7
Father behavior problems	35	15.6	28	12.3	63	13.9
Father school adjustment problems	40	17.8	38	16.7	78	17.3
Placed in Detention	61	27.1	60	26.4	121	26.7
Codefendants	58	25.8	58	25.6	116	25.6

*Note: Construction sample n=225, validation sample n=227, total sample n=452.

RAI Construction: Burgess Method

Once the construction and validation samples were formed, the next step to designing each of the RAIs was to select the appropriate predictors for each of the methods. Using the Burgess method, variables that were significantly related to violent offending after first juvenile arrest were used in the prediction equation. Table 9 shows the predictor variables that were significantly related to adjudicated violent offending after the first juvenile arrest. Listed in Table 9 are the number and percentages of offenders that either had a positive score (with attribute) or a negative score (without attribute) for each of the predictors significantly related to adjudicated violence. Also presented in Table 9 are the resulting chi-squares, odds ratios, and p-values for each of the predictors. Of the 58 variables examined, 13 were found to be significantly related to future adjudicated violence after first juvenile arrest including physical aggression, truancy, cruel to people, lack of guilt, noncompliance, covert behavior, psychopathy, suspended from school, family member on welfare, codefendants, a violent offense charge associated with a participants first juvenile arrest, arrest prior to age 13, and having been adjudicated of an offense that occurred prior to the age of 13. The most significant predictor was parent and teacher rated physical aggression ($p<.001$). Along with physical aggression, six other predictors in the child characteristics category were associated with future violent offending. Two school factors, truancy and suspended from school, were also found to significantly predict violent offending after first arrest. There were four arrest variables associated with future adjudicated violence, demonstrating the importance of including arrest or criminal history variables in any RAI. Surprisingly, none of the developmental, peer, caretaker, or neighborhood factors were associated with future conviction of a violent offense.

In regards to predictors that were related to violent juvenile offending after the second juvenile arrest, only two variables, family member on welfare and arrest prior to age 13 were found to be significant. The reduction in the number of predictors that were found to be related to violence from first juvenile arrest to second juvenile arrest is most likely in part due to the reduction in the number of recidivists after first arrest to second arrest. There were 225 offenders who were arrested at least once as a juvenile and 158 who were arrested twice as a juvenile. Thus, the reduced number of offenders examined most likely reduced the ability to detect associations between the predictors and violent offending following the second juvenile arrest.

Continuing with the Burgess method, those factors found to be significantly associated with violence were then used to construct an equation or measure that yielded a risk score with higher scores indicating an increased risk for violence. However, as indicated earlier, one variable, codefendants, could not be used to create either risk prediction equation because there was more than 15% of the data missing from this variable. The primary reason for this is that codefendant data were only able to be collected through the Allegheny County Juvenile Court Information Management Office Records and Prothonotary Office Records, and were not available from the other official records sources. Given that there were 12 risk factors significantly related to future violence with less than 15% of the data missing, the equation used to construct risk scores for each of the offenders was: Risk score=((number of positive risk factors/12-number of risk factors missing) x 12). Figure 12 provides an example of what an actual Burgess based RAI using these variables might look like in practice, using the predictor variables included in the risk prediction equation here. Included at the bottom of the example RAI is a simple equation that can be used to calculate a risk score if data for some of the predictor variables is unavailable.

Table 9

Predictors of Adjudicated Violence Following First and Second Juvenile Arrest: Construction Sample

Predictor	Without attribute: Percent violent	Without attribute: N	With attribute: Percent violent	With attribute: N	X^2	OR	*P*
			First arrest				
Physical Aggression	25.7	152	56.7	60	18.32	3.79	<.001
Truancy	27.0	89	40.2	122	3.96	1.82	.047
Cruel to People	27.9	147	50.0	64	9.63	2.59	.002
Lack of Guilt	28.7	122	42.2	90	4.21	1.82	.040
Noncompliance	28.6	154	47.1	51	5.91	2.22	.002
Covert Behavior	27.4	135	46.7	75	7.94	2.32	.005
Psychopathy	29.1	148	46.0	63	5.67	2.08	.017
Suspended From School	21.7	92	43.7	119	11.13	2.79	.001
Family Member on Welfare	25.8	97	41.4	111	5.65	2.04	.017
Codefendants	27.9	61	49.5	93	7.11	2.53	.008
1st Arrest Violence Charge	27.2	136	47.3	74	8.59	2.40	.003
Arrest Prior to Age 13	28.6	161	52.0	50	9.32	2.71	.002
Adjudicated Prior to Age 13	32.7	196	60.0	15	4.61	3.09	.032
			Second Arrest				
Family Member on Welfare	27.7	65	44.9	78	4.49	2.13	.034
Arrest Prior to Age 13	30.1	103	59.5	42	10.90	3.42	.001

Note. Percentages and n's reported have been weighted for risk status at the start of the study. The weighted baseline prevalence of adjudicated violence following first arrest in the construction sample is 34.4%. In the construction sample, there a total of 225 juvenile offenders were arrested at least once and 158 juvenile offenders were arrested at least twice.

12-Item Scale to Assess Violence Risk

Add one point for each item:

Prior physical aggression	________
Cruel to others	________
Displays lack of guilt	________
Noncompliant with parents or teachers	________
Covert behavior	________
Psychopathy	________
Truant from school	________
Suspended from school	________
Family on welfare	________
1st Arrest violent charge	________
Arrested prior to age 13	________
Adjudicated prior to age 13	________

Risk score:

Classification:

Score:	Risk Group:
0-5	Low risk of violence
6-12	High risk of violence

Risk Level

* For cases with missing data:

$$\text{Risk score} = \left(\frac{\text{number of positive risk factors}}{\text{12- number of risk factors missing}} \right) \times 12$$

Figure 12. Example Burgess RAI.

Next, using the Burgess based prediction equation, risk scores were calculated for all first time offenders in the construction sample. Risk scores calculated with missing data were rounded to the nearest whole number. Risk scores were only calculated for those offenders who had less than 1/3 missing data (meaning data were available for at least 9 out of 12 predictors) for the predictors used in the RAI. Figure 13 presents the percentage of first time offenders who later went on to be adjudicated of a violent offense for each risk score. Also listed underneath each risk score is the number of first offenders having each particular risk score. As can be seen from Figure 13, a dose-response relationship is present such that as the number of positive risk factors that an offender received on the RAI increased, the more likely the offender was to later be adjudicated on a violent offense. In the extreme, of all offenders having a risk score lower than three, less than 20% later went on to become adjudicated or convicted of a violent crime. For those offenders who had a risk score of eight or more, more than 60% were later adjudicated or convicted of a violent crime. However, this dose-response relation was not uniform across all risk scores as can be seen from the percentage of offenders who became violent with risk scores in the middle (between three and seven). For example, a larger percentage of offenders with a risk score of three went on to be adjudicated of a violent offense than those offenders who scored a six on the instrument.

Once risk scores were calculated for each offender, the next step was to divide offenders into two groups which involved choosing an appropriate cut-off score to classify offenders as being at average risk and high risk for future violence. In choosing an appropriate cut-off score there are a number of important points to consider, such as the purpose for which the instrument is designed, base rate of violent offending after first arrest, selection ratio, and percentage of offenders correctly identified in the construction sample. In actual practice, court systems would

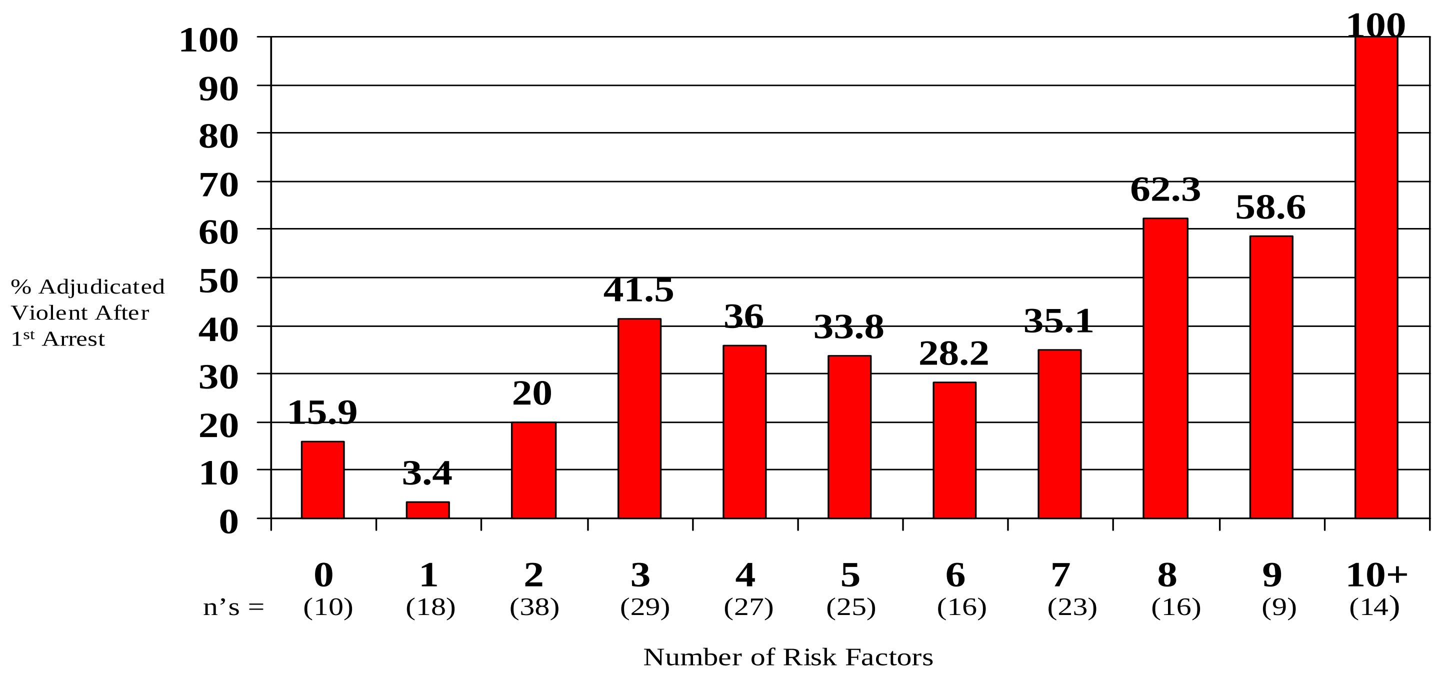

Figure 13. Burgess Method: Percentage of Adjudicated Violent Offenders After First Arrest by Number of Risk Factors: Construction Sample.

quite likely wish to choose a cut-point to identify a certain percentage of offenders most at risk for violence for which they are able to place in a high security facility. For example, administrators in a juvenile corrections system might want to identify the 10% of offenders most likely to become violent because that's the percentage of all of those arrested that their maximum security facility can hold. Similarly, if the facility can hold 25% of all first time offenders they might wish to identify the 25% of offenders most likely to become violent. Given that the intention of this study was to evaluate the predictive utility of an instrument to predict violent offending after first arrest, a choice was made that when choosing a cut-point a premium would be placed on minimizing the number of false negatives. A decision was also made to choose a cut-point that resulted in a selection ratio that was similar to the base rate of violent offending after first arrest as other previous researchers have done (Rice & Harris; 1995). Thus, given that in the construction sample 38.7% of first time offenders later became violent, it was decided to choose a cut-point that resulted in approximately 38.0% of the offenders being considered high risk. This was done because if a higher cut-point was chosen it would be impossible for the instrument to correctly classify all offenders who would later become violent. A decision was also made to give greater consideration to cut-points that maximized the percentage of offenders who were correctly classified in the construction sample. If there is a big difference between two cut-points in regards to the percentage correctly identified, then this suggests that the instrument might work much better at a certain cut-point than at another. To examine each of these factors, the percentage correct identified and the selection ratios for all cut-points were examined. Also, an ROC curve was generated for the RAI which produces specificity and sensitivity data for each particular cut-point. After examining each of these factors, a cut-off score was chosen with offenders having six or more risk factors being considered high risk. Figure 14 is identical to

Figure 13 with the exception that Figure 14 includes the specific cut-point to divide the sample into low- and high-risk groups. Table 10 shows the two-by-two table that was produced using this cut-point. Using this cut-point resulted in a selection ratio of 31.3%, with the percentage correctly identified being 66.8%.

Table 10

Two by Two Table For Burgess Method: Construction Sample

		Observed		
		Not Violent	Violent	
Predicted	Not Violent	107	38	145
	Violent	32	34	66
		139	72	211

RAI Construction: Predictive Attribute Analysis

To classify offenders using predictive attribute analysis, the (CHAID) program was run on the construction data. Prior to running the program, missing data on predictor variables were replaced using a method recommended by Breiman, Friedman, Olshen, and Stone (1984), in which missing values are replaced with the value of another variable among the predictors most closely associated with the variable for which there was missing data. As stated previously, only variables with less than 15% missing data were used. All 1st time offenders were initially split into a groups based on the variable most related to adjudicated violence after first arrest. These groups were then separately analyzed with each group again being split into two more groups based on the best predictor to the outcome for each group. This process was continued until it was not possible to further classify offenders.

The classification tree from the predictive attribute analysis that the CHAID program produced is presented in Figure 15. Six classification groups were formed, and the number of offenders and the percentage that later went on to be adjudicated on a violent charge is given for each terminal node or classification group on the tree. Group six contained the highest percentage of offenders who later went on to be adjudicated violent with approximately 3 out of 4 offenders being adjudicated on a violent charge following their first arrest. Group 1 had the lowest amount of offenders later adjudicated on a violent offense with only 10.1 percent of the group adjudicated on a violent offense following first arrest. Once this process was complete, those offenders who could not be classified as either high or low risk (groups 2 through 4, were then re-analyzed in a second iteration. Unfortunately, in this study after the first classification tree was produced CHAID was not able to identify any other predictor variables to further classify moderate risk offenders. Thus, only a classification tree using standard predictive analysis procedures was able to be constructed, making it not possible for this study to evaluate the usefulness of the iterative classification approach.

Using the same considerations used to determine an appropriate cut-point for the Burgess method, a cut-point was chosen by which offenders in groups 1 thru 4 were considered to be low or average risk, and groups 5 and 6 were considered high risk. Table 11 shows the two-by-two table that was produced using this cut-point. A selection ratio of 28.3% was produced using this cut-point, and 69.3% of the offenders were correctly classified.

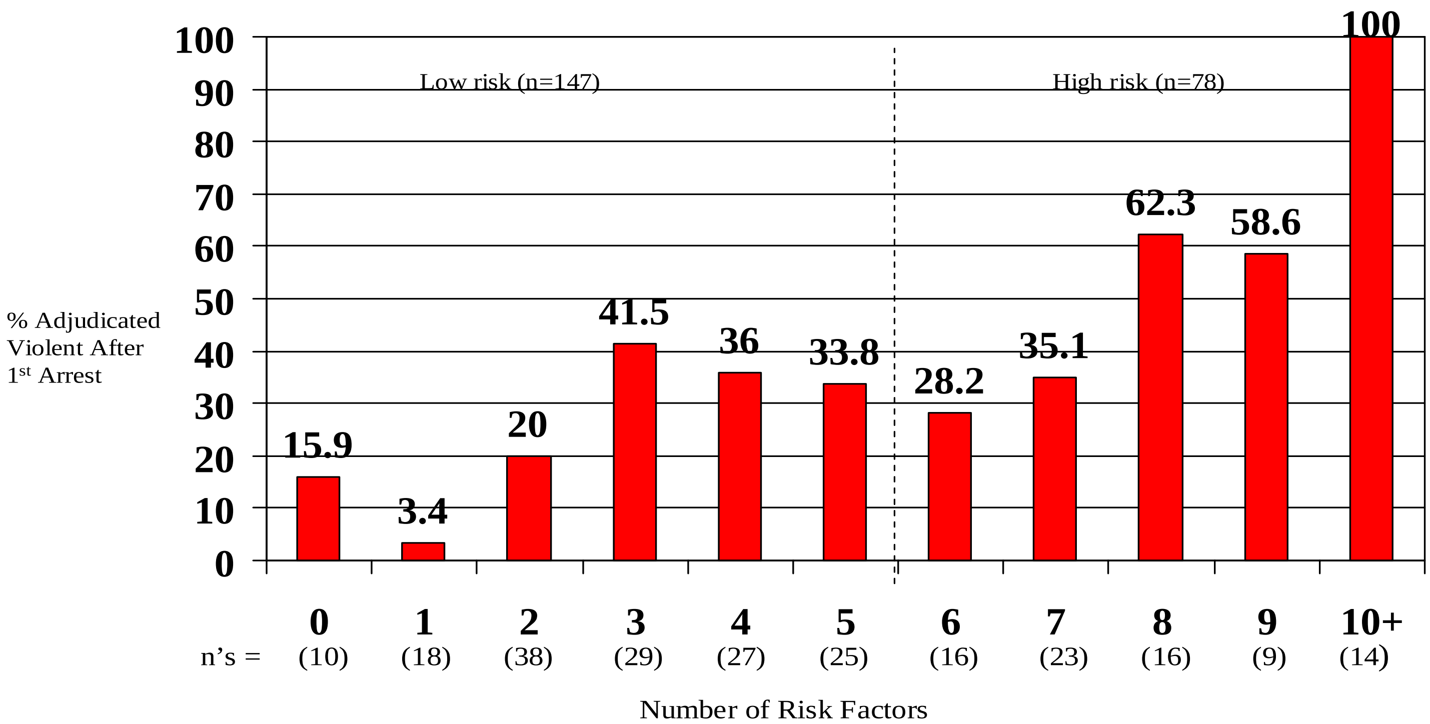

Figure 14. Burgess Method: Prevalence of Adjudicated Violent Offenders After First Arrest by Number of Risk Factors: Construction Sample.

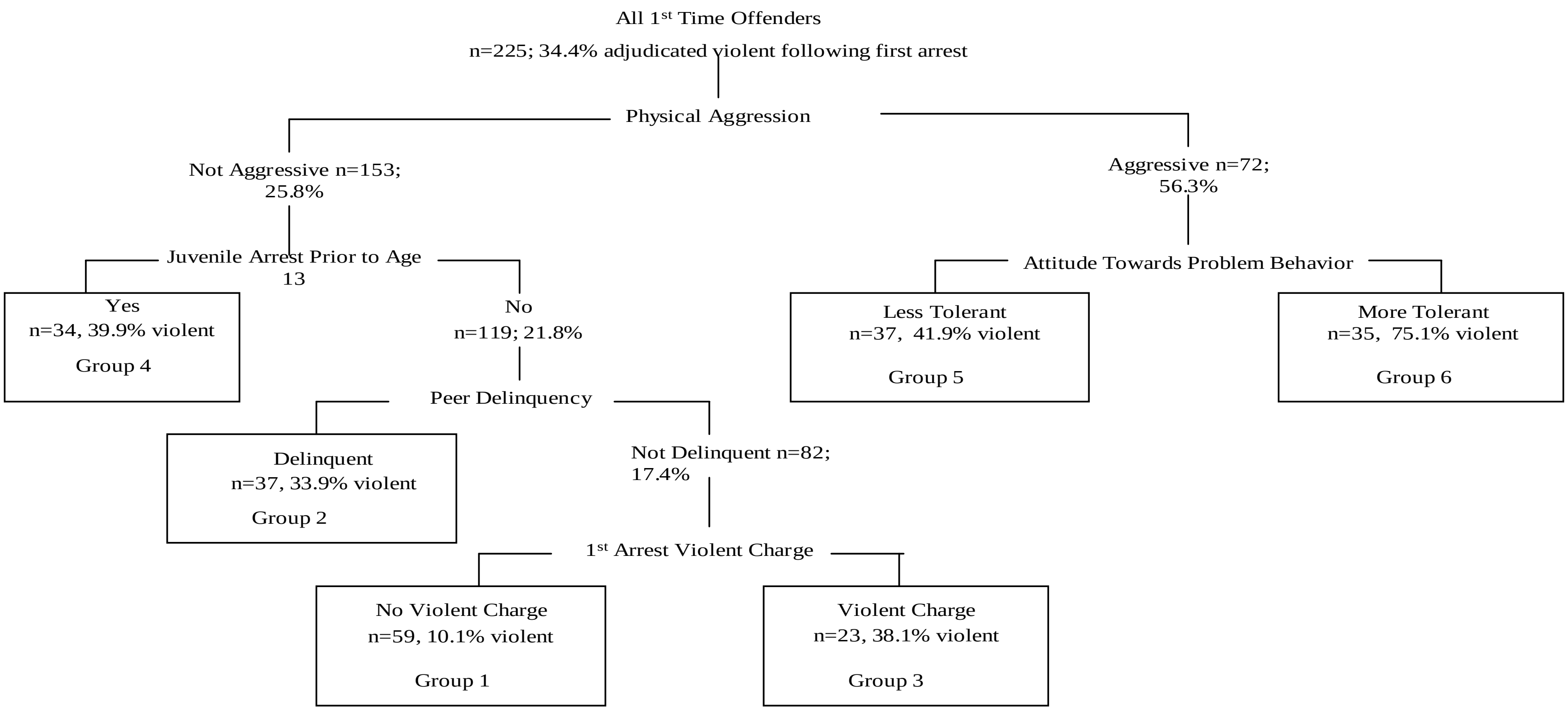

Figure 15. Predictive Attribute Analysis Classification Tree: Construction Sample. Using the classification tree approach, six different risk groups were formed with each having different percentages of offenders being adjudicated on a violent charge after their first arrest. Percentages reported are weighted for risk status at start of the study.

Table 11

Two by Two Table For Predictive Attribute Analysis: Construction Sample

		Observed: Not Violent	Observed: Violent	
Predicted	Not Violent	113	39	152
	Violent	26	34	60
		139	73	212

Table 12

Measures of Predictive Accuracy: Construction Sample

	Method	
Statistic	Burgess	Predictive attribute
% correct	66.82	69.34
Sensitivity (true positive rate)	.47	.47
Specificity (true negative rate)	.77	.81
RIOC (%)	26.44	33.93
Odds ratio	2.99***	3.79***
Chi-square	12.92***	18.32***
Phi	.25***	.29***
ROC (AUC)	.71	.73

Note. The selection ratio for the Burgess RAI (.31) and the predictive attribute analysis RAI (.28) were similar. ***p<.001.

Measures of Predictive Accuracy: Construction Sample

Table 12 provides data on each of the measures of predictive accuracy used in this study for the construction sample. The Burgess method and predictive attribute analysis performed similarly on each of the measures of predictive accuracy examined here. This of course is not surprising, given that the cut-points for each of the risk prediction equations were intentionally chosen to have similar sensitivity, specificity, and percentage of offenders correctly identified. Each of the methods also produced similar RIOC scores, and the Chi-square, Phi, and Odds-

Ratios were significant for both the Burgess and predictive attribute methods with at similar levels ($p<.001$).

Taking into account all possible cut-points, each of the measures also produced similar ROC curves. Figure 16 shows the ROC curve produced from the Burgess RAI, and Figure 17 shows the ROC curve produced by the predictive attribute analysis RAI. Both methods generated similar AUC statistics with the Burgess RAI producing an AUC=.71, and the predictive attribute analysis RAI producing an AUC=.73.

RAI Validation

Once each of the RAI's or risk prediction equations were created using the construction sample, the next step was to test their performance using the validation sample. Analogous to Figure 14 that made use of data from the construction sample, Figure 18 provides data on the prevalence of adjudicated violent offenders after first arrest by the number of risk factors for the validation sample. As can be seen from Figure 18, similar to results from the construction sample, as the number of positive risk factors that an offender possessed increased, so to did the likelihood that they would be later adjudicated on a violent offense. However, this pattern was much less pronounced than was the case in the construction sample. For example, even at the extreme high risk scores the measure failed to identify a risk group with a large percentage of offenders later being adjudicated of a violent offense, with only 38.7% of offenders with a risk score of 10 or more later becoming violent. However, this measure did appear to perform somewhat better at predicting groups of offenders who were lower at risk, although this too was less pronounced than in the construction sample.

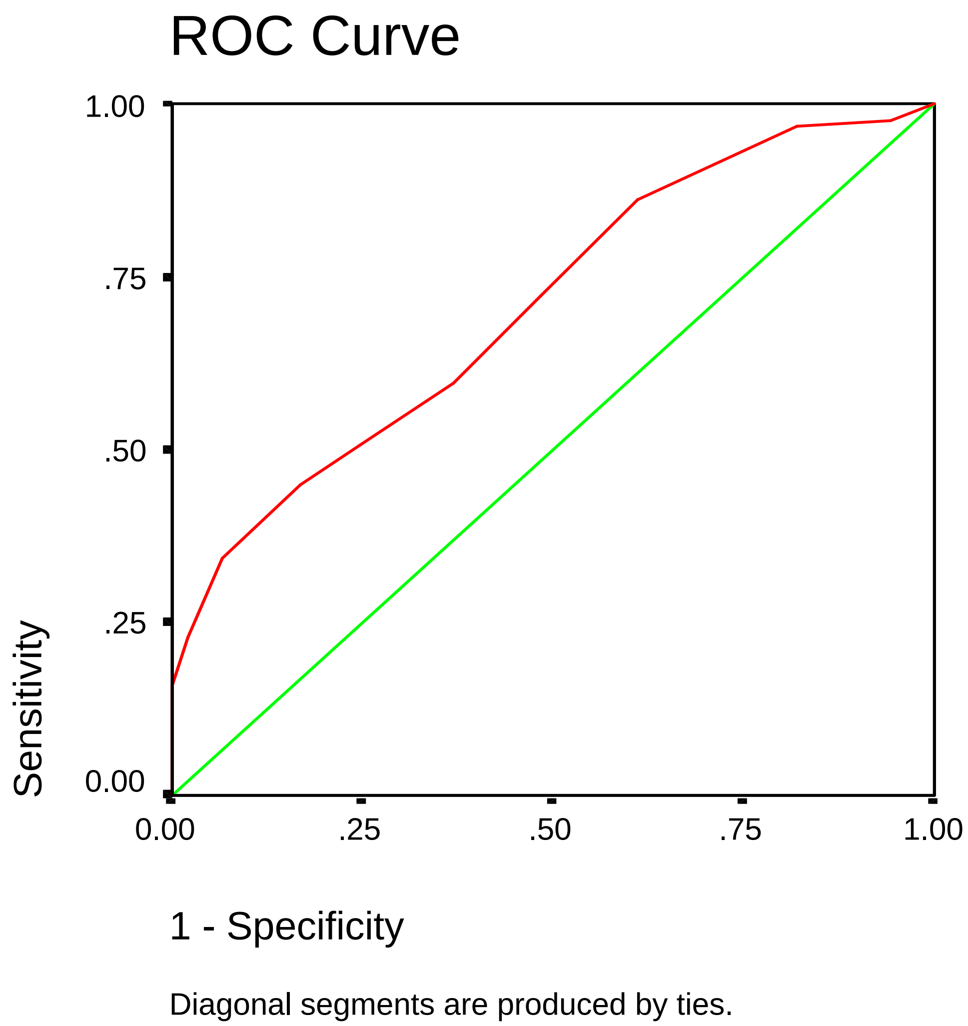

Figure 16. Roc Curve for Burgess Method – Construction Sample.

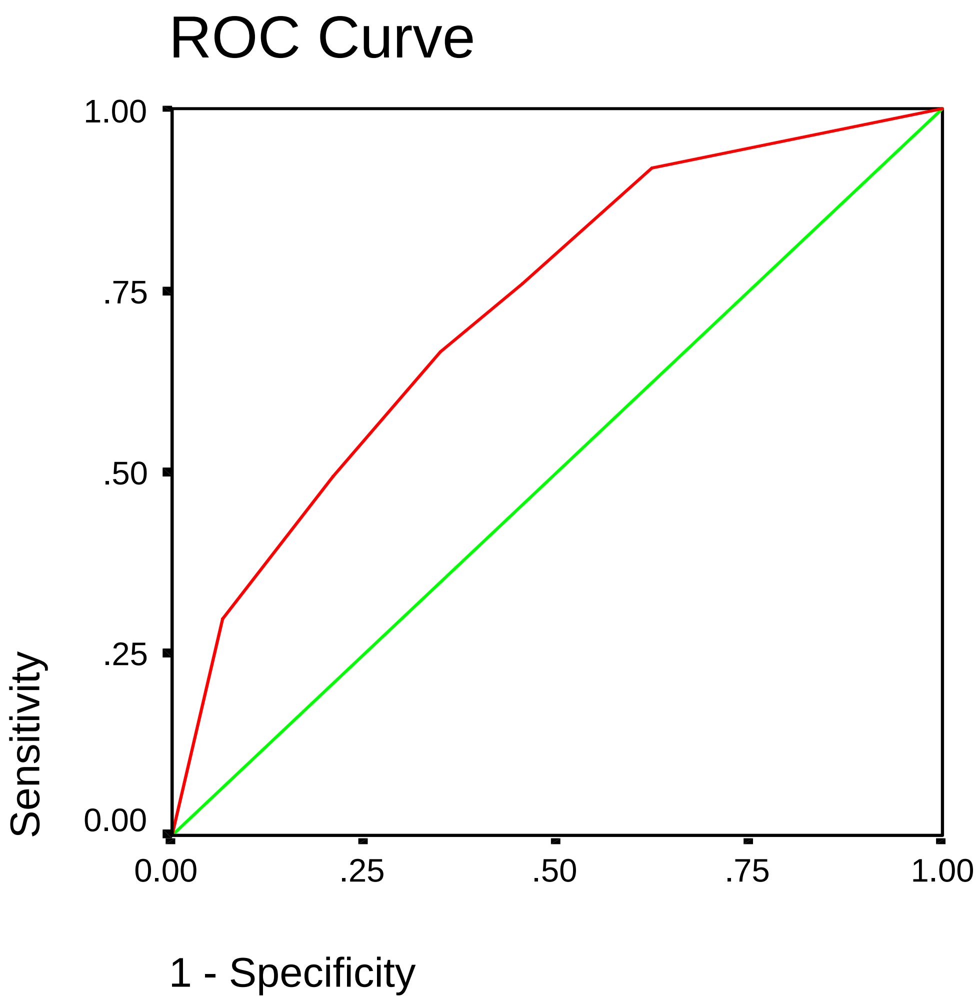

Figure 17. Roc Curve for Predictive Attribute Method – Construction Sample.

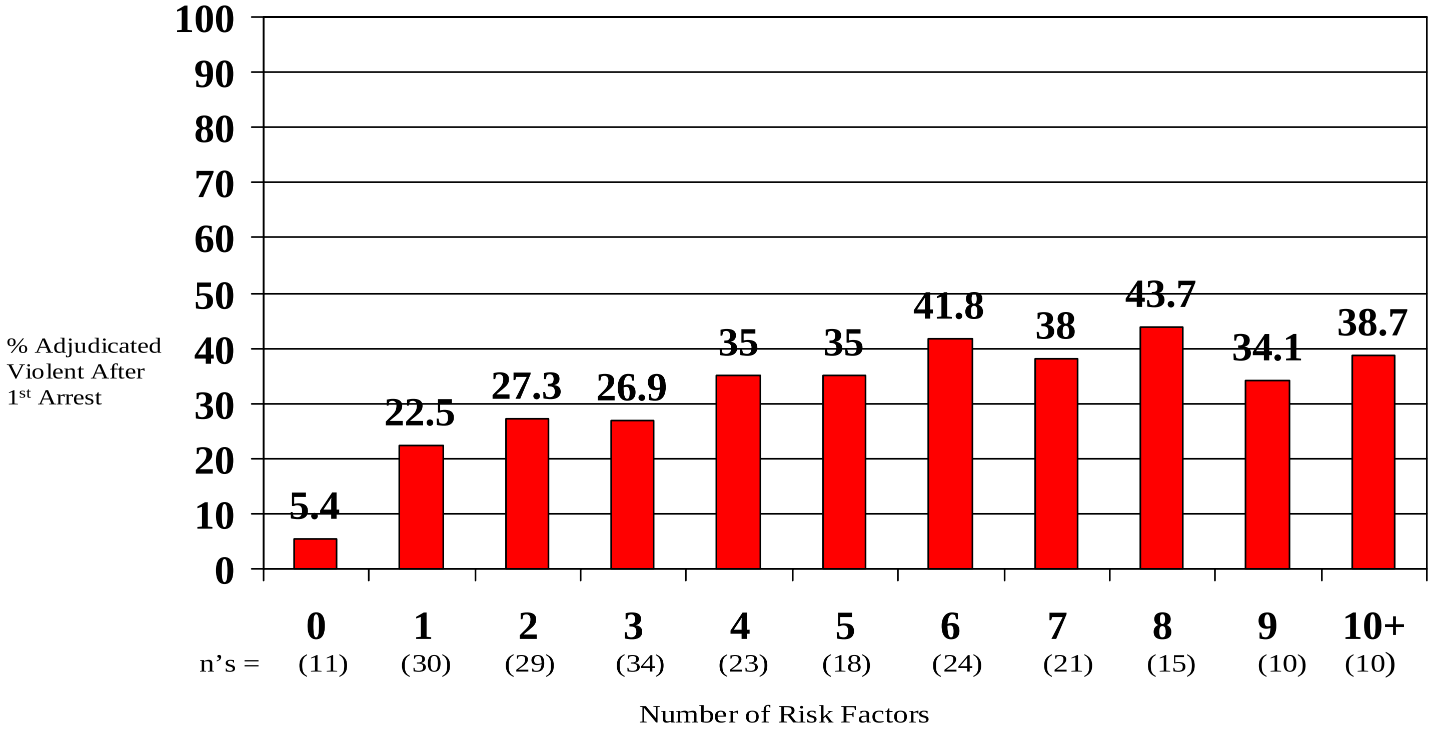

Figure 18. Burgess Method: Prevalence of Adjudicated Violent Offenders After First Arrest by Number of Risk Factors-Validation Sample. As the number of positive risk factors increases for an offender, so does the likelihood that he will be adjudicated violent after his first arrest. There were 227 juvenile offenders in the validation sample, however, two offenders had missing data for more than 1/3 of the risk factors, and were eliminated.

The two-by-two tables produced by the Burgess RAI and predictive attribute analysis RAI using the validation sample data are presented in Tables 13 and 14 respectively. Selection ratios for both methods for the validation sample were similar to those in the construction sample with the Burgess RAI having a selection ratio of .33, and the predictive attribute analysis RAI producing a selection ratio of .38. Table 15 presents the measures of predictive accuracy for both RAI construction methods on the validation sample. As can be seen, neither of the RAIs performed very well at predicting adjudicated violent offending after first arrest when employed on the validation sample. Overall, the Burgess RAI was able to correctly identify just over 60% of the offenders, and the predictive attribute RAI did just slightly less well, correctly identifying exactly 60% of the offenders. The relative improvement over chance statistic was also less than optimal with the Burgess RAI having improved roughly 15% the ability to predict future adjudicated violence above what would be expected by chance. Similarly, the predictive attribute RAI improved the prediction of violence 13% above what would have been expected by chance. The odds-ratio, chi-square, and phi statistics were only significant for the Burgess method ($p<.05$), although given that an offender was classified as high risk, they were not even twice as likely to later become violent as compared to those classified as low risk. Similarly, the AUC statistics for each of these methods was less than what one would expect to see in a RAI that can predict violence, with an AUC of .58 for the Burgess RAI, and .57 for the predictive attribute RAI.

Table 13

Two by Two Table For Burgess Method: Validation Sample

		Observed		
		Not Violent	Violent	
Predicted	Not Violent	104	37	141
	Violent	42	28	70
		146	65	211

Table 14

Two by Two Table For Predictive Attribute Analysis: Validation Sample

		Observed		
		Not Violent	Violent	
Predicted	Not Violent	95	33	128
	Violent	49	28	77
		144	61	205

Table 15

Measures of Predictive Accuracy: Validation Sample

	Method	
Statistic	Burgess	Predictive attribute
% correct	62.56	60.00
Sensitivity (true positive rate)	.43	.46
Specificity (true negative rate)	.71	.66
RIOC (%)	14.82	13.36
Odds ratio	1.87*	1.65
Chi-square	4.15*	2.58
Phi	.14*	.11
ROC (AUC)	.58	.57

Note. The selection ratio for the Burgess RAI (.33) and the predictive attribute analysis RAI (.38) were similar.
*p<.05.

Given that neither RAI performed well when tested on the validation sample, post-hoc analysis was conducted to determine which if any of the initial 58 predictor variables examined here were predictive of violence in the validation sample. Table 16 presents the eight variables that were predictive of adjudicated violence after first arrest in the validation sample. Interestingly, only three of the predictors (suspended from school, family member on welfare, and arrest prior to age 13) that were significantly associated with adjudicated violence in the construction sample were also significant in the validation sample. Although not predictive of adjudicated violence in the construction sample, the Race (African American) variable was significant in the validation sample. Four additional variables that were not predictive of adjudicated violence in the construction sample were significantly related to violence in the validation sample. This analysis sheds light on why neither of the RAIs performed well on the validation sample, and also indicates that it may be somewhat easy to “miss” identifying useful variables for inclusion in a RAI if only a single sample is used to construct the RAI. Given that the sample was randomly split to form the construction and validation samples, the variables found to be related to violence in the validation sample are equally likely to be useful in future RAIs, not taking into account the specific strength of each predictor. Finding that variables other than those identified in the construction sample were associated with later violence also highlights the need to use validation samples when constructing RAIs, and directly identifies the main source of shrinkage in this study.

Table 16

Predictors of Adjudicated Violence Following First Juvenile Arrest: Validation Sample

Predictor	Without attribute Percent violent	N	With attribute Percent violent	N	X^2	OR	*P*
African American	16.9	65	37.2	148	8.65	29.0	.003
DSM-IIIR ADHD	28.2	177	50.0	24	4.69	2.54	.030
Suspended from school	22.0	91	37.5	120	5.85	2.13	.015
Family member of welfare	19.1	115	45.5	88	16.28	3.52	<.001
Bad neighborhood	26.5	151	42.4	59	5.01	2.04	.020
Neglect Before 1st arrest	28.5	193	57.9	19	6.97	3.45	.011
Arrest prior to age 13	27.4	175	47.2	36	2.37	5.49	.018
1st arrest adjudicated property charge	26.9	160	43.1	51	4.80	2.06	.023

Note. Percentages and n's reported have been weighted for risk status at the start of the study. In the validation sample, there a total of 227 juvenile offenders were arrested at least once and 186 juvenile offenders were arrested at least twice.

Table 17

Measures of Predictive Accuracy Applied at Second Juvenile Arrest

Statistic	Method: Burgess	Method: Predictive attribute
% correct	51.09	54.82
Sensitivity (true positive rate)	.42	.39
Specificity (true negative rate)	.57	.65
RIOC (%)	-1.43	3.79
Odds ratio	.95	1.18
Chi-square	.04	.41
Phi	-.01	.04
ROC (AUC)	.50	.56

Note. The selection ratios for the Burgess method (.45) and the Predictive Attribute (.37) methods were similar.

To examine the utility of the RAIs from each of these methods to predict future adjudicated violent offending from the second juvenile arrest, each method was employed among all participants with a second juvenile arrest. Given that this greatly reduced the number of offenders, the construction and validation samples were combined (N=274). Table 17 presents the measures of predictive accuracy for the Burgess and predictive attribute RAI's when predicting from the point of second juvenile arrest. Neither of the RAIs were successfully able to predict future adjudicated violence from the second juvenile arrest. Both measures correctly identified just over 50% of the offenders, similar to the predictive ability that would be expected by chance. None of the measures of predictive accuracy were significant for either RAI.

Predicting Reported Violence

To examine the ability of each of the RAIs to predict reported violence, each method was employed on the total sample rather than on either the construction or validation samples alone. Table 18 shows the prevalence of reported and official violent offending for both the youngest and oldest cohorts, and for both cohorts combined. As can be seen, the prevalence of reported offending was greater in both the youngest and oldest cohorts when compared to official offending. Table 19 provides the measures of predictive accuracy for each RAI predicting reported violence. Interestingly, both RAIs performed similarly when predicting reported violence as opposed to official violence, with the Burgess method performing slightly better when predicting reported violence. The odds-ratio, chi-square, and phi statistics were significant for both the Burgess RAI ($p<.001$) and the predictive attribute RAI ($p<.001$). Offenders classified as high risk by the Burgess RAI were more than twice as likely to become violent than those not predicted to become violent, although the predictive attribute RAI did not meet this standard.

Table 18

Prevalence of Official and Reported Violence After First Juvenile Arrest: Youngest vs. Oldest Cohorts

Cohort	Adjudicated		Reported	
	n	%	*N*	%
Youngest (n=202)	58	28.9	66	33.0
Oldest (n=221)	80	36.1	103	46.9
Combined (n=424)	138	32.7	169	40.6

Note. The n's and percentages reported above have been weighted for risk. There were two offenders in the youngest cohort and one offender in the oldest cohort for which there was no reported violence data was available. A t-test revealed significant differences between the prevalence of adjudicated and reported violence in the both cohorts (p<.001).

Table 19

Measures of Predictive Accuracy: Reported Violence

	Method	
Statistic	Burgess	Predictive attribute
% correct	63.25	59.80
Sensitivity (true positive rate)	.44	.41
Specificity (true negative rate)	.76	.73
RIOC (%)	25.65	16.12
Odds ratio	2.52***	1.81**
Chi-square	18.97***	7.83**
Phi	.21***	.14**
ROC (AUC)	.65	.62

Note. The selection ratio for the Burgess (.32) and predictive attribute (.33) methods were similar.
**p<.01. ****p*<.001.

Discussion

The main hypothesis of this study, that it would be possible to predict which first time juvenile offenders become violent in the future as measured by official adjudication, was only minimally supported. In the construction sample, the two constructed RAIs were able to predict future official violent offending from first juvenile arrest. Also in the construction sample, both RAIs were able to predict better which offenders would later become violent than what would be expected by chance, and both were able to identify a high-risk group of first time offenders that were more than two times likely to go on to be adjudicated on a violent offense than those offenders in the low-risk group. However, neither RAI performed well when tested on the validation sample. Only the RAI constructed from the Burgess method was able to significantly improve the prediction of violence after first arrest above what would be expected by chance when tested on the validation sample. Neither of the RAIs were able to identify a group of offenders in the validation sample that were at least twice as likely as those considered low risk to be adjudicated or found guilty of a violent offense after first arrest.

There was minimal support for the secondary hypothesis in this study that it is possible to identify a group of variables typically available to court workers that are associated with future violent offending that can be used in an RAI to predict future violent offending. Following the procedures of the Burgess method in which a RAI is constructed using variables independently associated with violent offending, there were 13 variables found to be independently associated with adjudicated violent offending after first juvenile arrest. Using predictive attribute analysis that takes into account the inter-relationships between variables, five variables were identified that could be used to classify offenders into groups of high and low risk. Although there continues to be debate over which risk factors are most predictive of future violence, a set of common factors continue to prove to be predictive of violence, including arrest variables such as early entry into the JJS and parent and teacher reports of a youth's behavior. The predictor variables identified here by both RAIs as being related to future violence were similar to those used in other instruments designed to predict violence among juvenile populations. Both procedures used here to create the RAIs identified arrest variables as important for inclusion in a RAI. The PCL:YV (Forth, Kosson, & Hare, 2003) designed to detect psychopathy also includes the assessment of past criminal behavior, and the SAVRY (Bartel, Bartum, & Forth, 2000) assess for prior violence. Most RAIs designed to predict violence in adults include measures of past physical aggression or violence with the notion that past behavior is the best predictor of future behavior. Here too, the idea that the best predictor of future behavior is past behavior was supported. Several factors related to antisocial behavior were predictive of later violence including cruel to people, lack of guilt, psychopathy, and first arrest violence charge, with the strongest predictor being a prior history of physical aggression. Although truancy and being suspended from school were two variables that here were related to future adjudicated violence and used in the Burgess RAI, neither the PCL:YV or the SAVRY include these as predictors. However, both the EARL-B (Augimeri, Koegl, Webster, & Levene, 2001) and the SAVRY include assessments of academic performance. In this study an offenders family being on welfare was related to future violence, however, none of these instruments address family income directly. Whether a youth was arrested with others or codefendants in this study was also found to be related to future violence, although this variable is not included in any of the measures reviewed here. This suggests that more research examining how these and other possible variables might increase the predictive utility of instruments designed to predict violence should be conducted. In this regard, perhaps greater weight should be given to those variables, such as whether a youth is arrested with codefendants, that can be easily obtained by juvenile courts.

Although the results here shed light on which predictor variables might be important for inclusion in a RAI designed for the purpose of predicting violence after first juvenile arrest, there was less evidence that the predictor variables identified could collectively be used to predict violent offending after first juvenile arrest with a great deal of accuracy. Although each of the RAIs constructed here with it's unique set of predictor variables was able to demonstrate predictive qualities in the construction sample, the same was not true when these variables were collectively used to predict violence in the validation sample. Neither RAI was able to accurately predict violence in the validation sample using the criteria outlined at the beginning of the study. The Burgess RAI did perform somewhat better in the validation sample than did the predictive attribute RAI. Only the RAI produced from the Burgess method was able to identify a high risk group that had a significantly greater probability of being adjudicated on a violent offense in the future than the low risk group. There was also little evidence to suggest that either RAI could predict violent offending from second arrest.

The other secondary hypothesis in this study, that the predictive attribute method would outperform the Burgess method by classifying a larger percentage of offenders as either high or low risk was not supported. Rather, the RAI constructed using the methods outlined by the predictive attribute method produced high and low risk groups of similar size to those produced by the Burgess method. In this study, the optimal cut-point for both RAIs resulted in roughly one third of the sample being classified as high risk. Although often the case, the cut-points that determined high and low risk categories were chosen based on similar criteria, with the high-risk group or selection ratio being roughly the same size as the base rate of violent offenders to minimize false negatives. Thus, it is possible that using different criteria to group offenders into high- and low-risk groups may have produced slightly different results. However, prior to choosing a cut-point, neither method clearly identified a larger group of offenders in which the percentage of offenders who were later violent was at least twice that of the base-rate of violent offending for the entire sample.

Both RAIs produced from each of the two methods performed better when predicting reported violence than adjudicated violence. Both the Burgess and predictive attribute RAI were able to predict reported violence above what would be expected by chance. Also, the Burgess RAI identified a high-risk group that was more than two times more likely to later be adjudicated on a violent charge than the low-risk group. There are a number of possibilities why the RAIs were better able to predict reported violence than official violence in the validation sample. It is possible that at least in this study reported violence was in fact more representative of true violence than was being adjudicated or found guilty of a violent offense. Certainly not all violent acts will result in arrest, let alone result in adjudication or conviction for the offense. Additionally, not all violent acts are reported to police and not all violent acts that are reported result in arrest, and even fewer will result in conviction because of lack of evidence or unwillingness of a victim to prosecute or testify against the aggressor. Thus, although it is possible that at least a few of those offenders who were adjudicated delinquent were in fact innocent, it is most likely that the actual base rate of violent offending is higher than that of official violent offending.

When compared with predictive attribute analysis and other more complicated means for constructing RAIs, results from this study indicate that the Burgess method is the optimal method for designing a RAI, and has a number of advantages over predictive attribute analysis. Overall the RAI designed using the Burgess method performed at least as good and in some cases better than the RAI constructed using the predictive attribute method. Both RAIs performed similarly in the construction sample in regards to the AUC measure of predictive efficiency. In the construction sample the RAI constructed using predictive attribute analysis did performed somewhat better than the Burgess RAI with respect to the percentage of offenders correctly identified and the odds of an offender categorized as high risk becoming violent. However, when predicting violence in the validation sample and reported violence, the Burgess RAI outperformed the predictive attribute RAI, correctly categorizing a higher percentage of offenders. When predicting official violence in the validation sample and when predicting reported violence, offenders categorized as high risk by the Burgess RAI were also more likely to later become violent than those rated as high risk by the predictive attribute RAI. Thus, the RAI constructed from predictive attribute analysis suffered from more shrinkage when tested on other groups of offenders, perhaps due to the fact that the model is more complex and "over fit" the construction sample data.

Perhaps it is not surprising that the Burgess RAI performed somewhat better than did the predictive attribute RAI, given that multiple studies have demonstrated that RAIs constructed using the Burgess method do better than other more complex RAIs, especially when used on new data sets. RAIs constructed using the Burgess method appear to perform as well if not better in regards to predictive ability than other more complicated methods of instrument construction not examined here such as regression and neural network approaches (Farrington and Tarling, 1985; Caulkins et al., 1996). Also, at least one study has shown that the Burgess method may be more resistant to errors made by informants completing a RAI (Gottfredson & Gottfredson, 1985). The Burgess method also has a number of advantages over other more complicated measures because of its simplicity. A RAI constructed from the Burgess method lends itself more readily to paper and pencil measure than other more complicated methods making it more practical. For example, the Iterative Classification method and neural network models of predicting future offending would most likely require computer software to calculate the risk of an individual offender, which would result in increased monetary cost and give the instrument less face validity to those working in the JJS and using the RAI. The Burgess method also has the advantage of being able to deal with missing data more effectively than RAIs constructed from the predictive attribute analysis and other more complicated methods of RAI construction . When information used on an RAI is not able to be collected by those in the JJS on a particular offender, it can be difficult to know how to correctly classify the offender using a RAI that makes use of a classification tree or other more complicated classification system. Also, because of it's simplicity, the Burgess method might allow juvenile court systems to more easily monitor the status of high risk groups in the court system for validation purposes and to calibrate cut-points for high and low risk groups based on the offenders in a specific jurisdiction. For example, if a JJS was most interested in identifying the 10% of offenders most at risk for future offending because that is the maximum number of offenders that their maximum security facility holds, if because of increased funding they suddenly had room for additional high risk offenders, depending on the characteristics of the specific instrument, it might be possible easily adjust the cut-point to identify a larger number of offenders simply by lowering the cut-point.

Less Than Optimal Accuracy in the Validation Exercise

Neither RAI was able to predict adjudicated violent offending in the validation sample, or reported offending using the criteria stipulated in this study. Why did the instruments designed using two separate methods perform less than optimally when tested on the validation sample? As discussed in the introduction section, when ever an instrument is created on one sample and then tested or validated on another sample, it is expected that the instrument will suffer from shrinkage and not perform as well as on the sample it was constructed. With this in mind it is not surprising that both of the instruments performed less well on the validation sample, however, neither instrument was able to predict violence with the criteria used in this study. One reason they may not have been able to meet this criteria is that violence as a construct is difficult to accurately assess. Perhaps adjudicated violence is especially difficult to predict because some offenders who commit violent acts are not actually arrested, adjudicated delinquent, or convicted. Although this study had the strength that both reported and official violence were examined, it is possible that not all of those juvenile offenders who eventually did or will become violent were accurately identified. This is especially true of those offenders who may offend at a later date beyond the outcome window. It is also possible that in this study the most optimal set of predictors of violence were not identified. Numerous constructs from a host of domains have been found to be predictive of violence among juveniles (Hawkins et al., 1998; Lipsey and Derzon, 1998), however, to date no common set of predictors have been identified that predict violence among juveniles with a great deal of accuracy.

It is important to realize that the bar was set high in this study with respect to the criteria for how accurate the instrument is compared to other studies of this nature. Even though the importance of increasing true positives and true negatives and reducing false negatives and false positives is often recognized by researchers in this area, few studies examining instruments that predict violence evaluate their measures in this manner, especially on validation samples. Although neither RAI was able to predict future violence using the criterion outlined in the hypothesis, it could be argued that both RAIs designed here were able to predict future adjudicated and reported violent offending. The RAI constructed using the Burgess method significantly predicted official violent offending in the validation sample (OR=1.87, $p<.05$), and both RAIs significantly predicted reported violence with the RAI constructed from the Burgess method (OR=2.52, $p<.001$) slightly outperforming the RAI constructed using the predictive attribute method (OR=1.81, $p<01$). However, this means less to those working in the juvenile courts than the percentage of offenders who are rated as high risk that are expected to become violent in the future.

In this study each of the instruments were designed using a construction sample made up of youths from ages 10-17. It is possible that predictors that were associated with later violence only for certain age groups were not included in either RAI because all juvenile offenders were grouped together. If a factor was predictive for one age group but not another, then it is possible that they would not have been identified here as having an association to later violence. Research has demonstrated that youth that have an early onset of delinquency appear to have a different trajectory than later onset offenders, with early onset offenders having more psychopathic traits, mental health problems, substance abuse problems, and job and financial problems in adulthood (Moffitt, Avshalom, Harrington, & Milne, 2002). Youth who have an early onset of delinquency are also more likely to become violent later in life (Farrington, 1989; Moffitt, Mednick and Gabrielli, 1989; Tolan and Thomas, 1995; Moffitt, Avshalom, Harrington, & Milne, 2002). In their meta-analysis of prospective longitudinal studies, Lipsey and Derzon (1998) found clear evidence that many child and adolescent predictors of violence are more strongly related to violence at certain ages more than others. For example, these researchers found one of the best predictors of violence among youth age 6-11 to be substance use, yet this variable was only slightly related to violence among 12-14 year olds. The notion that at least some factors are more predictive of violence for certain age groups than others suggests that RAIs that are made up of different predictors for specific age groups, especially offenders under the age of 13, might prove to be more accurate. Indeed Lipsey and Derzon conclude from these results that separate RAIs should be developed for different age groups.

The idea of creating different RAIs for different subgroups of offenders is not limited to groups based on age. RAIs might also prove to be more predictive if created separately for different offender profiles, such as property, drug, and violent offenders. Similarly, it is possible that variables might be more predictive of future violence in certain racial groups than in others, and thus some benefit in accuracy might be gained from using separate RAIs for African Americans, Caucasians, and Hispanics. However, it is likely that any attempt to use different RAIs for different racial groups would be viewed as a mechanism for racial discrimination, especially given that African Americans are currently overrepresented at basically all points in the juvenile justice system. A tenet of the JJS is that all juveniles will be treated equally regardless of race, thus, basing predictions of risk on different types of information for different racial groups is not likely to occur within the JJS. To some degree it is likely that all RAIs might select certain groups more than others no matter what predictors are used in the RAI. This is because RAIs that include any variable assessing a characteristic related to the offender, whether that variable be poor neighborhood or early physical aggression, will possibly identify certain racial groups as being more at risk than others if certain racial groups possess those characteristics more than other racial groups.

It is also possible that one reason neither RAI performed as well as hoped is that the optimal set of predictors was not chosen. As Tables 9 and 16 indicate, even though the sample of offenders here was randomly split in half, many predictors were only related to future violence in either the construction or validation sample. It was therefore less surprising that neither RAI performed well on the validation sample, especially given that the strongest predictor in the construction sample, physical aggression, was not significantly associated with violence in the validation sample. Both RAIs, relied on this variable being a predictor of future violence, especially the predictive attribute based RAI that essentially divided groups into low and high risk based on this single variable. This suggests that a wider range of variables may have been more useful in the Burgess based RAI, and perhaps explains why the predictive attribute based RAI did not perform as well as the Burgess RAI in regards to predictive accuracy.

Finally, it is possible that the way in which the predictor variables were dichotomized here led to less than optimal predictive accuracy. Most of the predictors were selected from previously created constructs, many of which were created from continuous variables that were dichotomized on the one-quarter/three-quarters split using the entire sample including both offenders and non-offenders. Dichotomizing the predictors in the same manner but instead using only data from the juvenile offenders in the sample, may have increased the predictive utility of the individual predictors and in turn the RAIs that made use of these variables.

What is clear is that neither in this or previous studies has an optimal set of risk factors been identified to be used in a RAI to predict violence. It is also not clear whether a set of variables can be identified that will collectively predict violence among all juveniles offenders with any useful degree of accuracy. RAIs may demonstrate better predictive accuracy among juveniles if created separately for subgroups of juvenile offenders with similar characteristics.

Comparison with Previous Research

Although there are a number of violence prediction instruments that have been designed for use with adults that have been researched, there are very few such instruments to date that have been specifically designed for use with juvenile populations. In recent years, however, there has been an increase in attention to designing similar instruments for use with adolescents, primarily to aid those working in the JJS to identify young offenders who are most likely to commit serious or violent offenses. The hope is that by identifying adolescents at high risk for future violence, juvenile courts will be able to more effectively allocate available intervention resources with the overall goal of reducing the amount of violence in the community. These newly developed instruments primarily stem from instruments originally created for use with adults. They typically are either designed specifically to predict violence and antisocial behavior, or to identify a group of traits associated with psychopathy rather than to specifically predict violent behavior. Given that research addressing the construction, reliability, validation, and general use of these measures is in the very early stages, validation studies of these instruments are scarce. Most studies testing the properties of these instruments have focused primarily on issues of reliability and concurrent validity. Very few studies exist that make use of prospective data to assess the predictive validity and accuracy of such instruments, and of those that have almost none have evaluated any instrument using the common means of evaluating predictive accuracy used here (RIOC, sensitivity, specificity, and percentage correctly identified). Thus, no instrument designed to predict violence among juveniles has truly been evaluated for its ability to accurately predict violent behavior among adolescents even in a specific jurisdiction, let alone across different jurisdictions.

In this study the Burgess RAI correctly identified just over 62% of the offenders in the validation sample and the predictive attribute RAI correctly identified 60% of the offenders However, it is difficult to compare these results with previous studies because the few studies that have focused on predictive validity tend to use measures of validity that are more meaningful to researchers such as correlations and the AUC produced from an ROC curve that are less informative to those actually using the instruments to aid their judicial decision making. Statistics such as correlations and the AUC are less informative to those attempting to make use of the instrument because these statistics more or less measure the incremental or linear association between the overall scores obtained from an instrument and some type of violence, and do not typically require or make use of a specific cut-off score where groups are categorized as either high or low risk. An example will make this situation clearer. Imagine a given instrument measures 10 risk factors and thus produces scores that range from 0 to 10. It is unclear how a judge who knows that an individual offender received a total score of 6 might use this information differently knowing that the instrument correlates .30 with future violence than if she or he knows the instrument correlates .50 with future violence. Similarly, it is even more difficult to see how knowing that an instrument yielded an AUC of .60 might be more or less informative than if the instrument yielded an AUC of .75. In this example it is much more informative to know that a certain percentage of juvenile offenders who had a total score on the measure of 6 or more eventually became violent. Despite this fact, after thoroughly reviewing the literature this author was unable to identify a single study examining the predictive validity or accuracy of an instrument designed for use with juveniles to predict future violence that made use of more informative, simple, and easy to understand statistics such as overall percentage of offenders correctly classified despite the fact the percentage correctly identified is commonly used as a measure of predictive accuracy in studies examining similar instruments in adult populations. A few studies did include the percentage of false positives and false negatives, however, these studies also included categorizations that were neither high or low risk (average risk), making it difficult to determine the overall percentage of offenders correctly identified. It is unclear why this is the case when in the last few years at least some validation studies using these instruments have begun to appear in the literature. Related to this issue, is the fact no studies employing any of the instruments described above have empirically established appropriate cut-points for their measures. Instead researchers have chosen to use ROC curves and the resulting AUC, and correlation statistics that essentially take into account all possible cut-points as measures of accuracy. Although both of these statistics are very informative regarding the overall association of a measure to the outcome the instrument is designed to predict, it does not inform those wishing to use the measure how well it may or may not work for them. This is especially true given that a predictive instrument will be less effective at some cut-points than others, and that it is not likely that those working in a JJS will be able to establish the optimal cut-point for the particular population they wish to employ the measure without using the measure for a number of years and collecting and analyzing their own data. Indeed this is a step that some juvenile courts have begun to take. Therefore, it is difficult to compare the results obtained here with other studies, at least in regards to the most important measures of accuracy including the percentage of offenders correctly identified and the RIOC.

In regards to other indices of accuracy the RAI’s examined here did not perform very differently from other predictive measures of violence. In this study the AUCs were .71 and .73 respectively for the Burgess and predictive attribute RAIs employed on the construction sample. In the validation sample the AUCs were .58 and .57. So essentially both measures performed equally well in terms of the AUC, with both performing less well on the validations sample. These results appear to be similar to those found in other validation studies of instruments designed to predict violence among adolescents. Adapted from the Hare Psychopathy Checklist-Revised (Hare, 2003), the Hare Psychopathy Check List: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003) is perhaps the most researched instrument designed to assess psychopathic traits and behaviors in adolescents that makes use of a structured clinical rating approach. Examining violence prediction with this measure, Catchpole and Gretton (2003) found an AUC of .73 using retrospective data among 74 young violent offenders. In a study that examined the prospective validity of the PCL:YV, Murrie, Cornell, Kaplan, McConville, and Levy-Elkon (2004) found the PCL:YV to predict to multiple types of violence that included violence while incarcerated, assault with a weapon, victim required medical attention, and instrumental violence. Reporting correlations and AUCs, these authors found correlations between the PCL:YV and multiple types of violence to range from .30 to.36, and AUCs that ranged from .68 to .79. Examining the ability of the PCL:YV to predict to institutional violence in a three month prospective study, Dolan and Rennie (2005) reported an AUC of .64 for the total PCL:YV score predicting to institutional violence.

The AUCs found in these studies are somewhat stronger but similar to those found in this study. However these studies did not make use of specific cut-off scores and examine accuracy as was done in this study, making it difficult to assess how well the instruments would serve a decision maker in the JJS. A second instrument that has been examined in this fashion using cut-off scores is the Early Assessment Risk Lists for Boys (EARL-20B; Augimeri, Koegl, Webster, & Levene, 2001) designed to predict violent behavior in youth displaying disruptive behavior under the age of 12. Modeled after the Historical/Clinical/Risk Management instrument (HCR-20; Webster, Douglas, Eaves, & Hart, 1997) designed to assess violence in adults, the EARL-20B contains 20 items covering a number of domains used to assess risk. No studies to date have been conducted to examine how well this measure predicts violent behavior, however, Augimeri (2005) reported that when a median split on the total EARL-20B scores of 379 boys was used as a cut-point, 38% of those youth categorized as low risk were adjudicated on a charge prior to their 18th birthday as opposed to 49% of high risk boys, with boys in the high risk sample also having significantly more court appearances and being adjudicated on more charges than the low risk group. Clearly this measure offers promise for predicting violence among youth, however, more research is needed that specifically address this instruments ability to specifically predict violence if it is to be used for such purposes.

Research examining the validity of a third instrument the Structured Assessment of Violence Risk in Youth (SAVRY; Bartel, Bartum, & Forth, 2000) designed to aid in the assessment of violence risk with adolescents indicates that this instrument has some ability to predict future violence. This instrument is based on the “structured professional judgment” model that makes use of structured professional clinical appraisals of relevant risk factors. Retrospective studies looking at how well SAVRY total score’s perform at predicting violent recidivism have found AUCs between .73 (Catchpole, & Gretton, 2003). and .80 (Borum, Bartel, & Forth, 2005). To date only two prospective studies examining the predictive utility of this instrument have been conducted. Gretton and Abramowitz (2002) found violent recidivism rates of young offenders to be 5.7% for those categorized has low risk, 13.1% for those categorized as moderate risk, and 40.4% for those categorized has high risk. This means that most offenders categorized as high risk may not have gone on to commit violent acts. In a similar study, Catchpole and Gretton, (2003) found violent recidivism rates at one year follow-up to be 6%, 14% and 40% for juveniles categorized as low, moderate, and high risk respectively. These results are similar to those found here for both RAIs, with roughly 40% of those youth categorized as high-risk going on to become violent.

Unlike the validation studies examining the PCL:YV, the validation studies using the EARL-20B and the SAVRY have been conducted using predetermined cut-points. Although the rate of true positives from these studies is similar to those found with both RAIs here, none of the instruments discussed here have been researched with respect to how different cut-points might effect the rate of true positives or false negatives. Few studies made use of a specific cut-point or included the overall percentage of offenders correctly identified but rather they relied exclusively on the AUC statistic and correlations (Murrie et al., 20004). These studies also did not include the necessary information such as number of offenders in each classification category so this could be calculated. Catchpole and Gretton, (2003) did report this information, however, they also included a moderate risk group, making it unclear whether those juveniles categorized as moderate risk are predicted to become violent. Augimeri (2005) did include such data using a median split on the total EARL-20B score to predict general recidivism, finding similar results to those in this study. Reporting the base rate of offending in their study to be 43%, 38% of the low risk cases were found guilty of an offense at follow up as compared to 49% of the high risk group. However, no information in regards to how well the instrument predicted violence was included. Similarly, it seems perplexing that while many of the studies mentioned had the ability to assess accuracy using RIOC scores or percentage correctly classified, percentage of false negatives or false positives, none of them did so. Relying on the AUC or a simple correlation says less about how a measure would do once a specific cut-point is adapted since the AUC takes into account a variety of possible cut-points for a given instrument.

Limitations

There were a number of limitations to this study. First, with this and other studies of this nature it is difficult to obtain an exact picture of which offenders eventually go on to become violent recidivists. For reasons previously discussed, it is difficult to have a clear picture of the number of violent acts that a given offender may have committed over a specified period of time, and how serious those violent acts were regardless of how violence is assessed, especially when the offender is not incarcerated. Although this study did have the advantage in that measures of both reported and official violence were obtained, it is difficult know exactly which juvenile offenders went on to commit a violent act after their initial juvenile arrest. Secondly, although statewide and national arrest records were obtained, most offenders examined resided within a given jurisdiction for all or most of the outcome window, Allegheny County, Pennsylvania. It is unclear how each of the set of predictors used to construct the two RAI's might perform in other regions of the country that have larger percentages of minority populations who were for the most part not represented here, such as in states in the southwest where Hispanics make up a larger percentage of the juvenile population. Other jurisdictions could also have higher or lower rates of violence among their juvenile populations which might also affect the predictive utility of both instruments constructed here. Thus, to increase the validity of these results it would be optimal to test these and other such instruments in other jurisdictions to evaluate their predictive utility in other areas of the country.

Although the outcome window in this study was far better than most studies examining the predictive utility of these types of instruments even with adults, combining two different cohorts with outcome windows of slightly different length may have reduced the predictability of the RAIs constructed. It is also most likely that at least some offenders considered nonviolent will eventually be arrested and/or convicted on a violent offense beyond the outcome window in this study in their late 20s or 30s. This issue also exists in all of the studies reviewed here with juveniles and most previous research in this area assessing adult instruments designed to predict violence, because most studies have a much shorter outcome window. Of course no outcome window is totally complete unless participants are followed until all have deceased. Given the unlikely scenario that such a study would occur due to impracticality, a more optimal outcome window to be strived for might be when all participants reach the age of 40, an age when violent arrests subside. Finally, in regards to the results found here related to predictive attribute analysis and lack of ability in this study to use the iterative classification method, it is possible that the sample used here may have been too small to successfully perform such analysis. Studies in which iterative classifications were possible to predict violence (Monahan et al., 2001) made use of samples that were much larger in size than in this study. Thus, a much larger and more elaborate classification tree might have been able to be constructed in this study if the sample was larger.

Future Directions

Clearly violence prediction among juvenile delinquent populations is in the initial stages of development and a great deal of work remains before measures will be developed that can predict future violence among juvenile populations in ways that are sufficiently accurate and that can be communicated to those working in the JJS in a language that is meaningful. To this end more research identifying a common set of predictors of violence that can be measured in adolescents efficiently by those working with the juvenile courts, that can also be used across different jurisdictions needs to be conducted. Previous research has shown that less than 50% of adolescents who have committed a violent offense will develop into antisocial adults (Moffitt & Caspi, 2001). Thus, even though arrest variables that were used in the measures in this and other studies are helpful at identifying those most at risk for future violent offending, more emphasis must be placed on identifying less studied variables including variables less often used with these types of RAIs such as previous response to treatment or how often an offender misses a meeting with his probation officer. More emphasis also needs to be placed on identifying variables that may be more predictive at specific time periods in development, although assessing variables prior to the age a youth can be arrested can prove to be difficult. By identifying additional predictor variables of violence the hope is that at least some will improve the predictive accuracy of the instruments studied here over and above what the more common predictors such as arrest or criminal history variables currently allow. This in turn would reduce the percentage of cases that are misidentified or false positives and false negatives. More emphasis must be placed on evaluating the accuracy of instruments designed to predict violence, using statistics that are easily understood by those expected to use the instruments. Despite the fact that RAIs do often lack predictive efficiency, decision makers value RAIs as useful tools in the decision process, and most believe that JJSs that use RAIs are more effective at reducing recidivism (Schneider, Ervin, & Snyder-Joy, 1996).

Research in this area must continue and be more directed at creating measures that have been validated with multiple populations. Measures must be tested with more samples, including multiple validation samples in jurisdictions other than the one(s) in which they were developed. The importance of using a validation sample when assessing predictive utility of an instrument cannot be understated, and simply examining how well an instrument performs on a construction sample can mean little about how the instrument might do when tested on other populations or samples. Additionally, statistical methods such as bootstrapping that attempt to foresee how well a measure might do in new samples in no way should be viewed as a replacement for testing an instrument on real validation samples.

Optimal cut-off points that minimize false positives and negatives and maximize true positives and true negatives need to be established before those working in the JJS give a great deal of weight to the results from such instruments. When instruments to predict violence are used within the JJS, each jurisdiction should be heavily encouraged to examine how well any particular instrument and any specific cut-point used on that instrument works for the specific population(s) that are being assessed with the instrument. The ways in which different cut-points influence selection ratios, sensitivity, specificity, and the overall percentage of offenders that are correctly identified should be monitored by those using the instruments when possible. Similarly, how using such instruments to guide judicial decisions influences overall placement decisions and outcomes for offenders should also be monitored. Indeed, any juvenile court system employing such prediction measures that categorize offenders as either high or low risk must conduct ongoing research to determine the best cut-point to be used in their system, and to calibrate the cut-point based on the purpose for which the instrument is being employed. To this end continual data collection by the courts in terms of scores from such instruments and outcome data should be collected using computer software to be analyzed with some degree of regularity. To aid in the collection of data and interpretation of any analyses from such data, predictive instruments should be employed that are not overly complex in nature. Using instruments that are more straight forward increases the chances that results from such instruments will be understood by those using them and thus used properly.

There must be more research that is aimed specifically at demonstrating the accuracy of these measures in statistics that are meaningful to those working in the JJS. Although correlations and other statistics such as the AUC are helpful in demonstrating how well scores on a specific instrument relate to violence among researchers, they will have less meaning to those actually attempting to use the instrument to make appropriate placement and treatment decisions. It is minimally helpful to courts to know that there is a correlation between a given measure and violence. Further, knowing that a particular measure produced an AUC of .75 tells one little about where the best cut-off score between high and low risk might be and the types of errors that a particular cut-point might favor. More emphasis needs to be placed on examining the accuracy of RAI's meant to predict violence using statistics such as percentage correctly identified, and RIOC. Similarly, the types of errors that a particular cut-point on an instrument is meant to favor (i.e. false negatives or false positives) should be made clear by those suggesting the use of a certain cut-point. At a minimum at least some validation data should also be provided to those expected to employ the measure in order that they will have at least some knowledge as to the percentage of false positives and false negatives that are likely to result from using the instrument. Thus, to be certain that RAIs meant to predict violence are used appropriately by the courts, optimal cut-off scores need to be identified for these types of measures, and multiple validation studies using the same cut-off score must be conducted.

Finally, research in this area has failed to address and communicate to those needing to use instruments designed to predict violence in the juvenile courts, the egregious problem that many offenders will be misclassified as either high or low risk and thus result in false negatives and false positives. For example, if 30% of the sample is categorized as high risk, and 50% of the high risk group eventually becomes violent, this means that 50% of the high risk sample was incorrectly categorized resulting in many false positives. This is a problem that researchers studying the prediction of violence in juvenile populations have failed to sufficiently address. In fact most studies simply reported on the relation of the measure to future violence without ever addressing what percentage of offenders were misclassified. Misclassifying offenders as low risk who later commit violent acts most likely leads to inappropriate interventions with those offenders, including less intense interventions services than those really needed, which in turn results in more violence in the community. Perhaps less problematic are false positives, or identifying juvenile offenders as high risk who do not go on to become violent. The main cost here is the price of providing more extensive treatment than is needed. Identifying offenders as high risk who do not go on to become violent also leads to greater numbers of offenders being incarcerated for longer periods, the cost of which can be dramatic. However, knowing which offenders are false positives is very difficult to determine because it is difficult to know the impact that specific treatments are having for specific subgroups of offenders. Some offenders who are identified as high risk who would have later become violent will not go on to reoffend primarily because they appropriately were identified as high risk and received more intensive treatments and/or were incarcerated for a longer period of time. Thus, for any RAI to be used appropriately in this way each juvenile court system must monitor both the impact that using the instrument has on placement and treatment decisions, and also the impact of specific treatments and placements on specific subgroups of offenders. This issue also highlights the importance of relaying accurate information about these instruments to the courts regarding the percentage of both false positives and false negatives that are likely to result from employing any given RAI. It is highly important that more emphasis be given to increasing communication between researchers and those working in JJS's to accurately communicate what RAIs can and cannot do.

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S. et al. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34,* 341-382.

Achenbach, T. M. (1978). The child behavior profile: I. Boys aged 6-11. *Journal of consulting and Clinical Psychology, 46,* 478-488.

Achenbach, T. M. & Edelbrock, C. S. (1983). *Manual for the child behavior checklist and revised child behavior profile.* Burlington, Vt: University of Vermont, Department of Psychiatry.

Achenbach, T. M. & Edelbrock, C. S. (1987). *Manual for the child behavior checklist and revised child behavior profile,* Burlington, VT: University of Vermont, Department of Psychiatry.

Ageton, S. S. (1983). *Sexual assault among adolescents.* Lexington, MA: Lexington Books.

American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

American Psychiatric Association (1989). *Diagnostic and statistical manual of mental disorders.*

Andrews, D. A. (1996). Criminal recidivism is predictable and can be influenced: An update. *Forum on Corrections Research,* 8, 42-44.

Andrews, D. A., & Bonta, J. (1994). *The psychology of criminal conduct.* Cincinnati, OH:

Anderson.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). *Criminal Justice and Behavior, 17(1),* 19-52.

Ashford, J. B., & LeCroy, C. W. (1988). Predicting recidivism: An evaluation of the Wisconsin Juvenile Probation and Aftercare Risk Instrument. *Criminal Justice and Behavior, 15 (2),* 141-151.

Augimeri, L. K. (2005). Early Assessment Risk Lists for Boys and Girls. In T. Thomas, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice.* (pp. 324-338). New York, NY, US: Guilford Press.

Augimeri, L. K., Koegl, C. J., Webster, C. D., & Levene, K. S. (2001). *Early assessment risk list for boys: EARL-20B, Version 2.* Toronto: Earlscourt Child and Family Centre.

Austin, J. (1986). Evaluating how well your classification system is operating: A practical approach. *Crime & Delinquency, 32,* 302-322.

Austin, J., Holien, D., Chan, L., & Baird, C. (1989). *Reducing prison violence by more effective inmate management: An experimental field test of the Prisoner Management Classification (PMC).* San Francisco, CA: National Council on Crime and Delinquency.

Austin, J., Krisberg, B., DeComo, R., Del Rosario, D., Rudenstine, S., & Elms, W. (1994). *Juveniles taken into custody research program:* FY 1993 annual report (Draft). Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Baird, S. C. (1984). *Classification of juveniles in corrections: A model systems approach.* Madison, WI: National Council on Crime and Delinquency.

Baird, S. C. (1991). *Validating risk assessment instruments used in community corrections.* Madison, WI: National Council on Crime and Delinquency.

Barnoski, R. (1998). *Validation of the Washington State Juvenile Court Assessment: Interim report (Report No. 98-11-1201). Olympia: Washington State Institute for Public Policy*.

Barnoski, R. & Markussen, S. (2005). Washington State Juvenile Court Assessment. In T. Thomas, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice.* (pp. 324-338). New York, NY, US: Guilford Press.

Barnoski, R. & Matson, S. (1997). Evaluating early intervention in Washington State juvenile courts: A six-month progress report (Report No. 97-01-1202). Olympia: Washington State Institute for Public Policy.

Borum, R., Bartel, P., & Forth, A. (2005). Structured Assessment of Violence Risk in Youth. In T. Thomas, G. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice.* (pp. 324-338). New York, NY, US: Guilford Press.

Bartel, P., Borum, R., & Forth, A. (2000). *Structured Assessment for Violence Risk in Youth (SAVRY):* Consultation Edition.

Bazemore, G., Dicker, T. J., & Nyhan, R. (1994). Juvenile justice reform and the difference it makes: An exploratory study of the impact of policy change on detention worker attitudes. *Crime & Delinquency, 40*, 18-36.

Blumstein, A., Roth, J., and Cohen, J., eds. (1986). *Criminal careers and career criminals.* Vol. 1. Washington, DC: National Academy Press

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). C*lassification and regression trees.* Pacific Grove, CA: Wadsworth and Brooks/Cole.

Burgess, E.W. (1928). Factors determining success of failure on parole. In A.A. Bruce (Ed.), *The workings on the indeterminate sentence law and the parole system in Illinois.* Springfield: Illinois State Board of Parole.

Carroll, J. S. (1982). Evaluation, diagnosis, and prediction in parole decision making. *Law and Society Review 17(1),* 199-228.

Catchpole, R. & Gretton, H. (2003). The predictive validity of risk assessment with violent young offenders. A 1-year examination of criminal outcome. *Criminal Justice and Behavior*, *30,* 688-708.

Caulkins, J., Cohen, J., Gorr, W., & Wei, J. (1996). Predicting criminal recidivism: A comparison of neural network models with statistical methods. *Journal of Criminal Justice, 24* (3), 227-240.

Champion, D. J. (1998). *The Juvenile Justice System: Delinquency, Processing, and the Law*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Cleckley, H. (1976). *The mask of sanity.* St. Louis, MO: Mosby.

Clements, C. B. (1996). Offender classification: Two decades of progress. *Criminal Justice and Behavior,* 23 (1), 121-143.

Copas, J. B. (1985). Prediction equations, statistical analysis, and shrinkage. In D. P. Farrington & R. Tarling (Eds.), *Prediction in Criminology* (pp. 232-255). Albany: State University of New York Press.

Copas, J. B., & Loeber, R. (1990). Relative improvement over chance for 2 x 2 tables. *British Journal of Statistical and Mathematical Psychology, 43*, 293-307.

Corbett, R. P. Jr., & Harris, M. K. (1996). A review of research for practitioners. *Federal Probation, 60,* 64-70.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Coughlin, C. and Vuchinich, S. (1996). Family experience in preadolescence and the development of male delinquency. *Journal of Marriage and the Family, 58,* 491-501.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95-106.

Dawes, R., Faust, D., & Meehl, P. (1989). Clinical versus actuarial judgment, *Science, 243*, 1668-1674.

Dembo, R. & Brown, R. (1994). The Hillsborough County Juvenile Assessment Center, *Journal of Child and Adolescent Substance Abuse, 3*, 25-43.

Denno, D. W. (1990). *Biology and violence: From birth to adulthood.* New York: Cambridge University Press.

Dolan, M. C., & Rennie, C. E. (2005). Reliability and validity of the psychopathy checklist: Youth version in a UK sample of conduct disordered boys.

Duncan, O. D., & Ohlin, L. E. (1949). The efficiency of prediction in criminology. *American Journal of Sociology, 54,* 441-452.

Edelbrock, C., & Achenbach, T. (1984). The teacher version of the Child Behavior Profile: I. Boys aged six through eleven. *Journal of Consulting and Clinical Psychology, 52,* 207-217.

Elliott, D. S., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use.* Beverly Hills, CA: Sage.

Elliot, D. S., Huizinga, D., & Menard, S. (1989). *Multiple problem youth: Delinquency, substance use, and mental health problems.* New York, NY: Springer-Verlag.

Farrington, D. P. (1985). Predicting self-reported and official delinquency. In D. P. Farrington & R. Tarling (Eds.), *Prediction in criminology* (pp. 150-173). Albany: State University of New York Press.

Farrington, D. P. (1989). Early predictors of adolescent aggression and adult violence. *Violence and Victims, 4,* 79-100.

Farrington, D. P. (1997). Key issues in studying the biosocial bases of violence. In A. Raine & P. A. Brennan, D. P. Farrington, & S. A. Mednick (Eds.), *Biosocial bases of violence.* New York, NY: Plenum Press.

Farrington, D. P., & Loeber, R. (2000). Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health*, 10, 102-124.

Farrington, D. P., & Hawkins, D. (1991). Predicting participation, early onset and later persistence in officially recorded offending. *Criminal Behavior and Mental Health, 1,* 1-33.

Farrington, D. P. & Tarling, R. (1985). *Prediction in criminology.* Albany: State University of New York Press.

Farrington, D. P., Jolliffe, D., Loeber, R., Stouthamer-Loeber, M., & Kalb, L. (2001). The concentration of offenders in families, and family criminality in the prediction of boys' delinquency, *Journal of Adolescence, 24,* 579-96.

Ferguson, T. (1952). *The young delinquent in his social setting.* London: Oxford University Press.

Forth, A. E., Hart, S. D., & Hare, R. D. (1990). Assessment of psychopathy in male young offenders. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2,* 342-344.

Forth, A. E., Kosson, D. S., & Hare, R. D. (2003). *The Hare Psychopathy Checklist: Youth Version (PCL-YV).* Toronto, Ontario: Multi-Health Systems.

Frick, P. J., O'Brien, B. S., Wootton, J. M., & McBurnett, K. (1994). Psychopathy and conduct problems in children. *Journal of Abnormal Psychology, 103,* 700-707.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34,* 401-433.

Glaser, D. (1987). Classification for risk. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making* (pp. 249-292). Chicago: University of Chicago Press.

Glueck, S., & Glueck, E. (1950). *Unraveling juvenile delinquency.* Cambridge, MA: Harvard University Press.

Gottfredson, D. M. (1987). Prediction and classification in criminal justice decision making. In D.M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making* (pp. 1-51). Chicago: University of Chicago Press.

Gottfredson, S. D. & Gottfredson, D. M. (1980). Screening for Risk: A comparison of methods. *Criminal Justice and Behavior,* 8, 315-330.

Gottfredson, S. D. & Gottfredson, D. M. (1985). Screening for risk among parolees: Policy, practice, and method. In D. P. Farrington & R. Tarling (Eds.), *Prediction in criminology* (pp. 54-77). Albany: State University of New York Press.

Gottfredson, D. M., & Snyder, H. N. (2005). The mathematics of risk classification: Changing data into valid instruments for juvenile courts. National Center for Juvenile Justice Report. Department of Justice. Office of Juvenile Justice and Delinquency Prevention, Washington, DC.

Gretton, H., & Abramowitz, C. (2002, March). *SAVRY: Contribution of items and scales to clinical risk judgments and criminal outcomes.* Paper presented at the biennial conference of the American Psychology-Law Society, Austin, TX.

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

Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist-Revised.* Toronto: Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised* (2nd ed.). Toronto: Multi-Health Systems.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315-335.

Hart, S. D. and Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. M. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22-35). New York: Wiley.

Hawkins, J. D. & Catalano, R. F. Jr. (1992). *Communities that care.* San Francisco: Jossey-Bass.

Hawkins, J. D., Herrenkohl, T., Farrington, D. P., Brewer, D., Catalano, R. F., & Harachi, T. W. (1998). A review of predictors of youth violence. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 106-146). Thousand Oaks, CA: Sage.

Hechtman, L., & Weiss, G. (1986). Controlled prospective 15 year follow up of hyperactives as adults' non-medical drug and alcohol use and antisocial behavior. *Canadian Journal of Psychiatry, 31,* 557-567.

Herrenkohl, T. I., Maguin, E., Hill, K. G., Hawkins, J. D., Abbott, R. D., Catalano, R. F. (2000). Developmental risk factors for youth violence. *Journal of Adolescent Health, 26,* 176-186.

Howell, J. C. (2001). Risk/Needs assessment and screening devices. In R. Loeber & D. P. Farrington (Eds.), *Child Delinquents: Development, intervention, and service needs* (pp.395-404). Thousand Oaks, CA: Sage.

Huizinga, D., & Elliott, D. S. (1987). Juvenile offenders: Prevalence, offender incidence, and arrest rates by race. *Crime & Delinquency, 33 (2)*, 206-223.

Janda, L. H. (1998). *Psychological testing: Theory and applications.* Allyn & Bacon, Needham Heights, MA.

Kelley, B. T., Thornberry, T. P., and Smith, C. A. (1997). Juvenile Justice Bulletin. Department of Justice, Office of Juvenile Justice and Delinquency Prevention, Washington, DC.

Kendal, E., & Mednick, S. A. (1991). Perinatal complications predict violent offending. *Criminology, 29,* 519-529.

Kline, P. (2000). *The handbook of psychological testing.* 2nd ed. New York, NY: Routledge.

Kline, P. (1998). *The new psychometrics: Science, psychology, & measurement.* New York, NY: Routledge,

Le Blanc, M. (1998). Screening of serious and violent juvenile offenders: Identification, classification, and prediction. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 167-193). Thousand Oaks, CA: Sage.

Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 313-345). Thousand Oaks, CA: Sage.

Loeber, R. & Hay, D. F. (1994). Developmental approaches to aggression and conduct problems. In M. Rutter and D. F. Hay (Eds.), *Development through life: A handbook for clinicians* (pp. 488-516), Malden, MA: Blackwell Scientific.

Loeber, R. & Dishion, T. J. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin,* 94, 68-99.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). Antisocial behavior and mental health problems: Explanatory factors in childhood and adolescence. Mahwah, NJ: Lawrence Erlbaum Associates.

Loeber, R., Stouthamer-Loeber, M., Van Kammen, W. B., & Farrington, D. P. (1989). Development of a new measure of self-reported antisocial behavior for young children: Prevalence and reliability. In Klein, M. W., (Ed), *Cross-National Research in Self-Reported Crime and Delinquency*. Boston: Kluwer-Nijhoff.

Loeber, R., Farrington, D., Stouthamer-Loeber, M., Moffitt, T.E., & Caspi, A. (2001). The development of male offending: Key findings from the first decade of the Pittsburgh Youth Study. In R. Bull (Ed), *Children and the law: The essential readings.* (pp. 336-378), Malden, MA: Blackwell Scientific.

Loeber, R., Pardini, D., Homish, D. L., Wei, E. H., Crawford, A. M., Farrington, D. P., (2005). The prediction of violence and homicide in young men. *Journal of Consulting and Clinical Psychology, 73(6)*, 1074-1088.

Loeber, R., Wung, P., Keenan, K., Giroux, A., Stouthamer-Loeber, M., Van Kammen, W. B., et al. (1993). Developmental pathways in disruptive child behavior. *Development and Psychopathology,* 5, 101-132.

Lynam, D. R. (1996). Early identification of chronic offenders: Who is the fledgling psychopath? *Psychological Bulletin, 120(2)*, 209-234.

Lynam, D. R. (1997). Pursuing the psychopath: Capturing the fledgling psychopath in a nomological net. *Journal of Abnormal Psychology, 106*, 425-438.

Mannuzza, S., Klein, R. G., Konig, P. H., & Giampino, T. L. (1989). Hyperactive boys almost grown up: IV. Criminality and its relationship to psychiatric status. *Archives of General Psychiatry, 46(12)*, 1073-1079.

Maguin, E., & Loeber, R. (1996). Academic performance and delinquency. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol. 20, pp. 145-264). Chicago, IL: University Chicago Press.

McCord, W., McCord, J., & Zola, I. K. (1959). *Origins of crime*. New York, NY: Columbia University Press.

McManus, M., Alessi, N. E., Grapentine, W. L., & Brickman, A. S. (1984). Psychiatric disturbance in serious delinquents. *Journal of the American Academy of Child and Adolescent Psychiatry, 23*, 602-615.

Mednick, S. A., & Kendel, E. S. (1988). Congenital determinants of violence. *Bulletin of the American Academy of Psychiatry and the Law, 16(2)*, 101-109.

Megargee, E. I. (1984). A new classification system for criminal offenders, VI: Differences among the types of the adjective checklist. *Criminal Justice and Behavior, 11*, 349-376.

Mitchell, S., & Rosa, P. (1979). Boyhood behaviour problems as precursors of criminality: A fifteen-year follow-up study. *Journal of Child Psychology and Psychiatry, 22*, 19-33.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology, 13,* 355-375.

Moffitt, T. E., Mednick, S. A., & Gabrielli, W. F. (1989). Predicting careers of criminal violence: Descriptive data and predispositional factors. In D. A. Brizer & W. Crowner (Eds.), *Current approaches to the prediction of violence* (pp. 15-34). Washington, DC: American Psychiatric Association.

Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology, 14(1),* 179-207.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., et al. (2001). *Rethinking risk assessment: The MacArthur Study of Mental Disorder and Violence.* New York, NY: Oxford University Press Inc.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783-792.

Murrie, D. C., Cornell, D. G., Kaplan, S., McConville, D. & Levy-Elkon, A. (2004). Psychopathy scores and violence among juvenile offenders: A multi-measure study. *Behavioral Sciences and the Law, 22,* 49-67.

National Council on Crime and Delinquency, (1993). Annie E. Casey Foundation Juvenile Detention Alternatives Initiative: Detention utilization case level data and projections for Sacramento County. San Francisco: Author.

National Institute of Mental Health. Diagnostic Interview Schedule for Children (DISC), Version 2.3. New York: New York State Psychiatric Institute, Division of Child and Adolescent Psychiatry; 1992.

Peterson, P. L., Hawkins, J. D., Abbott, R. D., & Catalano, R. F. (1994). Disentangling the effects of parental drinking, family management, and parental alcohol norms on current drinking by Black and White adolescents. *Journal of Research on Adolescence, 4(2),* 203-227.

Quinsey, V. L., Harris, G. T., Rice, M. E., Cormier, C. A. (2006). Violent offenders: Appraising and managing risk (2nd ed.). Washington, DC: American Psychological Association.

Raine, A., Brennan, P., & Mednick, S. A. (1994). Birth complications combined with early maternal rejection at age 1 year predisposes to violent crime at age 18 years. *Archives of General Psychiatry, 51,* 984-988.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63(5),* 737-748.

Rumberger, R. W. (1995). Dropping out of middle school: A multilevel analysis of students and schools. *The Journal of American Education Research, 32,* 583-625.

Seagrave, D. and Grisso, T. (2002). Adolescent development and the measurement of juvenile psychopathy. *Law and Human Behavior, 26(2)*, 219-239.

Schneider, A.L., Ervin, L., & Snyder-Joy, Z. (1996). Further exploration of the flight from discretion: The role of risk/need instruments in probation supervision decisions. *Journal of Criminal Justice, 24,* 109-121.

Serin, R.C. and Amos, N.L. (1995). The role of psychopathy in assessment of dangerousness. *International Journal of Law and Psychiatry, 18 (2), 231-238.*

Shannon, L. (1988). *Career criminal continuity.* New York: Human Services Press.

Snyder, S. N. (1996). The juvenile court and delinquency cases. *The Future of Children, 6(3),* 53-63.

Silver, E., Smith, W. R., & Banks (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior, 27(6),* 733-764.

Snyder, S. N. (2000). Juvenile Arrests 1999. Juvenile Justice Bulletin. Department of Justice, Office of Juvenile Justice and Delinquency Prevention, Washington, DC.

Stattin, H., & Magnusson, D. (1989). The role of early aggressive behavior in the frequency, seriousness, and types of later crime. *Journal of Consulting and Clinical Psychology, 57*, 710-718.

Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., et al. (2000). Enhancing the clinical usefulness of actuarial violence risk assessment tools: A classification tree approach. *Law and Human Behavior*, *24,* 83-100.

Stouthamer-Loeber, M. (1991). Young children's verbal misrepresentations of reality. In K. J. Rotenberg (ed.), *Children's interpersonal trust: Sensitivity to lying, deception, and promise violations* (pp. 20-42). New York: Springer-Verlag.

Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D. P., & Wikström, P-O. H. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. *Journal of Consulting and Clinical Psychology, 70(1),* 111-123.

Straus, M. A. (1991). Discipline and deviance: Physical punishment of children and violence and other crime in adulthood. *Social Problems, 38,* 133-152.

Thornberry, T. P., Huizinga, D., & Loeber, R. (1995). The prevention of serious delinquency and violence: Implications from the program of research on the causes and correlates of delinquency. In J. C. Howell, B. Krisberg, J. D. Hawkins, & J. Wilson (Eds.), *Source book on serious, violent, and chronic juvenile offenders* (pp. 171-212). Thousand Oaks, CA: Sage.

Tolan, P. H., & Thomas, P. (1995). The implications of age of onset for delinquency risk:

II. Longitudinal data. *Journal of Abnormal Child Psychology, 23*, 157-181.

Verhulst, F. C., Koot, H. M. & Berden, G. F. M. J. (1990). Four-year follow-up of an epidemiological sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,*440-448.

Washington State Institute for Public Policy (1998). *Interim Validation of the Washington State Juvenile Court Administrators Risk Assessment.*

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence, Version 2.* Burnaby, BC, Canada: Mental Health, Law and Policy Institute, Simon Fraser University.

Wiebush, R. G., Baird, C., Krisberg, B., & Onek, D. (1995). Risk assessment and classification for serious, violent, and chronic juvenile offenders. In J. C. Howell, B. Krisberg, J. D. Hawkins, & J. Wilson (Eds.), *Source book on serious, violent, and chronic juvenile offenders* (pp. 171-212). Thousand Oaks, CA: Sage.

Wikström, P.-O., & Loeber, R. (2000). Do disadvantaged neighborhoods cause well-adjusted children to become adolescent delinquents? A study of male juvenile serious offending, individual risk and protective factors, and neighborhood context, *Criminology, 38,* 1109-1142.

Wilson, J. J., & Howell, J. C. (1993). *A comprehensive strategy for serious, violent, and chronic juvenile offenders.* Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.