Frequency and severity of offending by young people in New Zealand: Descriptive analysis and development of a predictive model.

A thesis

Submitted in partial fulfilment

Of the requirements for the degree

Of

Masters of Arts in Psychology

At the

University of Canterbury

By

Anna Galletly

University of Canterbury

2006

Acknowledgements

Special recognition must go to Randy Grace, who was my primary supervisor. Thank you Randy for all your support and input into this thesis. You taught me so much about statistics and creating statistical models and helped to motivate me to put in the extra effort.

I would also like to thank the Christchurch Police Department, especially the people who work in records. They made me very welcome while I was collecting data and helped me with any questions that I had. I would also like to thank Child, Youth and Family for their contribution in providing the data that was the basis for this thesis.

Finally, thanks must go to my partner, Josh, who has been incredibly supportive and helped me through some stressful times.

Table of Contents

List of Tables

List of Figures

Abstract

Youth offending is an increasingly major problem in many countries and cultures. Several theories imply that a subset of young people display delinquent behaviour at a young age and go on to have an extensive and serious criminal career. Recently, there has been interest in the literature in identifying these young people early on and carrying out interventions in order to deter them from a criminal career. Many studies have examined the development and usefulness of actuarial measures of risk of future violence or recidivism in adult offenders. However, the same attention has not been paid to the youth offender population.

The present study gathered data from the population (N = 4307) of all young persons in New Zealand whose antisocial behaviour resulted in a Youth Justice intake from the Department of Child, Youth, and Family (CYF) in 2002. Information was obtained about this population from the CYF database, CYRAS, and from the Police National Intelligence Application database for a stratified random sample (N = 500). Three models were developed using Hierarchical Cox regression to predict recidivism, and they each used a different definition of recidivism. The performance of the models was assessed using ROC analysis and they were found to predict recidivism with a moderately good level of accuracy. A validation sample (N = 500), different from the sample on which the models were developed, was used to further assess the performance of the models by showing that they were able to generalize to a new data set and continue to perform at an adequate level. An actuarial model, like the one developed in the present study, could be used to help make decisions about which young people within the Youth

Justice System require intervention in order to reduce the likelihood of subsequent reoffending.

Introduction

Youth offending has recently received increasing attention in the literature. This is partly due to the general public expressing grave concerns about youth offending in recent times, in response to the highly publicised serious crimes committed by adolescence in New Zealand and around the world. In the following literature review statistics describing the nature of youth crime are reported to show why it is an area of concern. The following section describes studies that have examined risk factors for the development of youth offending. These studies have been carried out in an attempt to better the understanding of how young people ultimately become involved in the Youth Justice system. Theories have attempted to explain the developmental trajectory of youth offenders and these are described also. The literature review then moves to a discussion of the prediction of recidivism. With the acceptance that certain risk factors are well linked with the development of youth offending there have been attempts to use these risk factors to assess the recidivism risk level of an individual. Debate has occurred over the years concerning the use of actuarial versus clinical judgement in risk assessment, however, studies have generally found that actuarial methods outperform clinical judgement. Following these findings many actuarial measures have been developed to predict adult recidivism but the equivalent studies examining children and adolescence are somewhat lacking. The literature review describes the few instruments that have been developed for use with young people and also how well they perform at predicting recidivism. Methods for evaluating the predictive accuracy of instruments are discussed and in particular Receiver-Operator Characteristic (ROC) analysis is reviewed.

Statistics

Youth crime is a major concern for people in New Zealand and in other countries. In general, rates of crime worldwide have increased dramatically since World War II, especially during the 1950's and 1970's. It is likely that juvenile crime has also risen as a proportion of overall crime. In 1995, England and Wales reported that 26% of people cautioned or convicted for an offence were aged between 10 and 17 years. The ratio of juveniles to adults who were prosecuted in Norway in 1996 was 1.23 whereas in Bangladesh it was only 0.03 (Pease & Tseloni, 1996). This reflects the substantial variation in youth crime statistics between countries. One reason for such variation is differing definitions of criminal activity throughout the world and the ways in which youth crime is dealt with and recorded. For example, the age of criminal responsibility varies around the world from 7 years in the USA, Ireland, Singapore and Switzerland, to 18 years in Belgium, Peru, Syria and Romania (Rutter, Giller, & Hagell, 1998).

However, it is important to examine self-report data in addition to official statistics, because the latter likely underestimate the actual rate of youth offending. For example, only a proportion of actual crimes that are committed are documented, and not all people committing criminal acts are apprehended. One study examined longitudinal research that asked children under the age of 13 about their antisocial behaviour. For children aged 7-12 years, approximately 20% said that they had committed one or more offences (Espiritu, Huizinga, Crawford, & Loeber, 2001). By contrast, Snyder (2001) examined FBI crime statistics and found that only about 1% of youth aged 7-12 years had been involved with the juvenile justice system in 1997. This figure is quite different from the self-report data (Espiritu et al., 2001). Overall, it is estimated that about one quarter to one third of total offences in England, Wales and the USA are committed by young people under the age of 18 years (Rutter et al., 1998).

There has been relatively little research on youth crime in New Zealand. One important source of information is the Ministry of Justice Webpage. It is pointed out on this website that there needs to be a more scientific approach to collecting and analysing data on youth offending in New Zealand (Becroft, 2003). In 2002, there were 44,533 resolved offences for youth under the age of 17 in New Zealand. In that year the apprehension rate for young people aged 10-13 was 44.7 per 1,000 in that population . The proportion of crime attributed to youth under the age of 17 was 21.9% in 2002. These figures have remained relatively stable over the last 5 years. Only a small proportion of these crimes were classified as violent, with over 50% being dishonesty offences (Becroft, 2003). Maxwell, Robertson and Anderson (2002) reported that there were about 150 serious crimes per year committed by youth aged between 10 and 13 years. Of young persons under the age of 17 that come into contact with Youth Aid workers, approximately 4% are under the age of 10 years, 30% are aged 10-13, and the rest are aged between 14-17 years (Grace & McLean, 2003). Overall, these data suggest that there is a substantial amount of crime committed by young persons in New Zealand, similar to other developed Western countries.

Risk Factors

Much recent research has examined risk factors and their role in the development of antisocial behaviour in young people. Risk factors can be divided into those which are individual factors, family variables, school variables, peer variables and community and cultural variables (Kashani, Jones, Bumby, & Thomas, 1999). In general, the evidence suggests that there is no one factor that is important, rather there is a cumulative effect of multiple factors that lead to antisocial behaviour. Risk factors can be divided into those that are static and those that are dynamic. Static risk factors are those that are stable and cannot be changed such as age, gender, and criminal history. Dynamic risk factors can be changed which can decrease the likelihood of recidivism and thus can represent targets for intervention. Dynamic risk factors are also referred to as criminogenic needs because it is believed that they contribute to the development of criminal behaviour (Andrews & Bonta, 1994; Hoge, 2002). These include factors such as alcohol and drug use, and association with antisocial peers. Responsivity factors also need to be considered when selecting a treatment intervention that will be most effective for the youth in question. These factors are characteristics of the young person or are other variables that will effect how well the young person responds to treatment (Hoge, 2002). Responsivity factors include variables such as motivation, cognitive style and mood.

Gender is a variable that has been approached in different ways in terms of describing it as a risk factor for antisocial behaviour. Current risk assessment measures tend to ignore gender or include it as a risk factor for males, however, there is some research to suggest that different risk factors may exist for males and females (Mazerolle, 1998). There is a debate around whether the established criminal theories are relevant for describing female criminality. Some argue that females experience less exposure to risk factors for offending while others argue there are differences in the risk factors that predict offending across the genders (Funk, 1999). One argument is that domestic abuse is an especially significant risk factor for female offending (Chesney-Lind, 1997) while another is that social bonds to significant others is a more important risk factor for the development of offending in females.

Funk (1999) carried out a study to examine the differences between males and females in risk factors for juvenile offending. Data from 388 males and 112 females that were referred to the local Department of Juvenile Justice in 1993 was collected. This included information on prior offending and offending after 1993 along with other possible risk variables. However, several potential predictor variables had to be disregarded due to the somewhat limited nature of the court file information. Funk (1999) found that risk assessment measures that were developed from a sample of males and females accounted for less of the variance associated with female recidivism, moreover, a risk assessment measure developed solely on females performed over twice as well when predicting female recidivism than a combined measure did. He also found that combined measures typically do not include risk factors relevant to females such as child abuse and running away from home. From these findings he concluded that it would be beneficial to develop separate risk assessment instruments for males and females.

The following table lists studies that have examined the development of antisocial behaviour in young people and also the risk factors they have found to be significant.

Table 1.

Risk Factors Found to be Significantly Associated with the Development of Antisocial Behaviour.

Authors	**Risk Factors**
Benda, Flynn Corwyn, and Toombs (2001)	• prior incarcerations • age of first criminal offence • gang membership • age of initiation of alcohol or drug use • scores on the psychopathic deviate subscale of the Minnesota Multiphasic Personality Inventory (MMPIpd) • Scores on the denial and asocial subscales of the Jesness Inventory • scores on the Carlson Psychological Inventory
Ge, Donnellan and Wenk (2001)	• adverse family environment • low cognitive ability • early use of alcohol • early age at first arrest • large number of arrests before the age of 17 years
White, Moffit, Caspi, Bartusch, Needles, and Southamer-Leber (1994)	• Impulsivity
Broidy, Nagin, Tremblay, Bates, Brame, and Dodge (2003)	• Early physically-aggressive behaviour • Childhood non-aggressive conduct problems and oppositional behaviour
Gretton, Hare, and Catchpole (2004)	• psychopathy
Douglas and Skeem (2005)	• impulsiveness • negative affect • psychosis • antisocial attitudes • substance use and related problems • problems with interpersonal relationships • problems with treatment adherence

Haines and Case (2005)	• lack of parental supervision • poor emotional attachment • parental offending • inconsistent and severe discipline • parental relationship problems • parental rejection
Roosa, Deng, Ryu, Burrell, Tein, and Jones (2005)	• high risk neighbourhoods

Protective factors as well as risk factors have been examined in relation to youth offending. In one study, archival data was analysed from 76 youth offenders (Carr & Vandiver, 2001). Aspects of personal characteristics, family situations, social situations, and academic ability were found to be protective factors and discriminated between recidivists and nonrecidivists. For example nonrecidivists had more positive attitudes towards authority in general and in particular towards police and school rules. The familial protective factors included having more structure, support and guidance within the family and also fewer siblings. In terms of risk factors, specific personal characteristics and family factors differentiated between recidivists and nonrecidivists. However, the total number of stressors, risk factors and prior offences did not differentiate between recidivists and nonrecidivists.

Douglas and Skeem (2005) distinguished between risk status and risk state. Risk status refers to static risk factors which are identified by comparing people at high risk for violent behaviour with those at low risk. Risk state takes into account changes in level of risk over time. Risk state can be understood as a person's probability of being violent at a certain time taking into account biological, psychological and social factors in their life. Therefore risk state is made up of static and dynamic risk factors. More recent actuarial instruments such as the Historical/Clinical Risk Management-20 (HCR-20) and the Level of Service Inventory-Revised (LSI-R), have tried to take into account risk state as well as risk status.

Overall, studies suggest that the most important risk factors are, gender, an early onset of antisocial behaviour, problems with attention, low intelligence, and low Socio-economic Status (SES) (Grace & McLean, 2003; Hill, 2002) A history of convictions or antisocial behaviour is one risk factor that stands out in the literature as very significant in predicting future risk (Sheldrick, 1999). This review shows that there are several risk factors that have strong links to the development of offending behaviour and they are well supported in the literature. This preponderance of risk factors implies that it should be possible to develop an actuarial instrument to predict antisocial behaviour. As Catchpole and Gretton (2003) point out, although there is a substantial amount of research relating to developing instruments to predict adult recidivism, the equivalent research for children and adolescents is much less prominent.

Developmental Trajectories

Among young people who engage in antisocial behaviour there is a large degree of variation in the type, frequency, and duration of criminal activity. Moffitt (1993) has proposed that the population of youth offenders can be split into two main groups. The first group is called “life-course persistent” offenders. She states that this group is born with a difficult temperament and more often into low SES families who do not have the resources to be able to cope with a difficult child. The discipline style in these families is also more likely to be harsh and inconsistent. These factors influence the child’s development and eventually lead to restricted choices of roles in society. Antisocial behaviour is evident in this group from a very early age and it continues through into adulthood. The expression of the antisocial tendencies changes as new social opportunities arise. The prognosis is bleak for this group because they fail to learn conventional pro-social alternatives to antisocial behaviour. This group commits a wide variety of offences and are more likely than adolescent-limited offenders to be lone offenders and to commit more victim-oriented offences. Moffitt (1993) proposes that life-course persistent offending is a form of psychopathology.

The second group is called “adolescent-limited offenders”. Individuals in this group typically have no history of antisocial behaviour in childhood and engage in antisocial behaviour as an adult to a much lesser degree than the life-course persistent offenders. Their antisocial behaviour is not consistent across situations and it may be under the control of reinforcement and punishment contingencies. Moffitt (1993) proposes that this group begins to engage in antisocial behaviour because of a gap between physical maturity and social maturity. They are between childhood and adulthood, and are ready, developmentally speaking, to be acknowledged as adults. The small numbers of life-course persistent individuals who are already engaging in “adult” activities provide a model that the adolescent-limited group try to imitate. Formal and informal sanctions are thought to maintain this behaviour in the adolescent-limited group as they strive for independence from their parents. This behaviour can be seen as developmentally normal.

If Moffitt’s theory is valid, it becomes important to be able to identify the individuals who are in the life-course persistent group as early in their criminal careers as possible. Doing so would enable intervention and hopefully prevention of more serious crime as the individual moves into adulthood. Presumably, identifying this group could be achieved by examining the history of antisocial behaviour and the other risk factors mentioned in the discussion above. However, there is a problem with this logic because not all children who engage in antisocial behaviour become criminals as adults. For an instrument based on childhood factors, there is likely to be a high false-positive rate. Therefore, a balance needs to be drawn when developing a risk assessment tool so that it does not have a false-positive rate that is too high but still identifies those that actually are at risk of chronic antisocial behaviour.

Chung, Hill, Hawkins, Gilchrist and Nagin (2002) examined the offending trajectories in a population of children from a low socio-economic status background. They identified five offending patterns: nonoffenders, late onsetters, desisters, escalators and chronic offenders. Nonoffenders do not report any offending at any point in time while late onsetters have no offence history prior to age 13 years but from this point on they increase slowly to reach low levels of offending. Desisters have a history of low level offending at age 13 but then generally stop offending by 18 years of age. Escalators lave a low level of offending at age 13 years and have increased to a high level of offending at age 18 years. Chronic offenders have a consistently high level of offending across their adolescent years. A follow-up study found evidence for the same patterns in a population from the Seattle Social Development Project (Chung, Hawkins, Gilchrist, Hill, & Nagin, 2002). Factors that predicted membership in one trajectory or another were identified at age 10-12. For example, young people who drank alcohol were more likely to be late onsetters than nonoffenders. Those at 10-12 years of age who belonged to families with good management strategies and lower levels of conflict, and who lived in neighbourhoods characterised by low numbers of antisocial youth were more likely to be desisters rather than escalators.

Taken together these studies show that there are identifiable patterns associated with the development of youth offending. There is still debate in the literature as to which proposed pattern best describes the youth offending developmental process. However, these patterns can help drive the development of measures to assess recidivism risk. They show which areas are important to focus on when developing a risk assessment tool.

Watt, Howells, and Delfabbro (2004) discuss different theories, namely criminal propensity theories and social control theories, that have been used to explain the development of a criminal career in a given individual. According to criminal propensity theories, there are differences between individuals in factors that lead them to delinquency and crime. However, different criminal propensity theories emphasise different individual factors as the most important in the development of antisocial behaviour. Some say low self-control is at the centre of delinquency development (Gottfredson & Hirschi, 1990), while others argue that psychopathy is the most important variable (Hare, 1996). They all agree however, that these individual differences emerge early in life. In contrast, social control theory implies that social factors are more important in the development of delinquency than individual factors. These social factors include family dynamics and involvement in the community. When an individual has weakened links to these social elements they become more likely to perform antisocial acts. It is probable that both of these theories describe aspects of the development of delinquency in young people (Paternoster, Dean, Piquero, Mazerolle, & Brame, 1997). Social learning theory differs from the other two already discussed because it includes both social and individual factors. Social learning theory argues that delinquent behaviour is driven by the perceived rewards and costs associated with the behaviour. Individual and social factors determine how the individual perceives the rewards from antisocial behaviour. For example the young person may have positive attitudes towards delinquency or they may be involved in a peer group where antisocial behaviour will gain them status.

Watt et al. (2004) examined how well the literature on juvenile recidivism supported each of these three theories. They found some evidence in favour of each theory; however, social learning theory had the least consistent support. They argued that it is important when undertaking risk assessment to take into account these theories that explain how the predictive factors are linked to the problem behaviour. The variables associated with each of the theories will indicate different outcomes, for example, criminal propensity variables are proposed to be stable individual factors that are unlikely to change. If one wanted to ascertain who was most likely to reoffend then criminal propensity variables would probably be the best factors to assess. However, social control and social learning theory variables are proposed to be dynamic, and therefore able to change. If the purpose of the assessment was to decide upon a treatment approach then these variables should be the focus of the assessment. Watt et al. (2004) proposed that future research on youth offenders should be driven by current criminological and psychological theory as this has not always been the case in the past. Specifically, it is important to keep these theories in mind when developing instruments to predict recidivism in youth offenders.

Why we Want to Predict Recidivism

It is important that we develop a way to identify those young offenders who are at high risk of going on to have a serious criminal career. In this way, resources can be directed at this group of young offenders in an effort to divert them from a path of criminal behaviour. It has also been reported that targeting interventions at those who are at a low risk of going on to reoffend can encourage them to become more involved with criminal behaviour (Krysik & LeCroy, 2002). This may be because they are exposed to higher-risk young offenders in the treatment program who influence them in a negative way and make them aware of new ways to engage in antisocial behaviour. An alternative explanation is that the attention directed towards their antisocial behaviour during the treatment may unfortunately lead them to rebel in response to that attention. Another important reason for being able to identify and thus intervene at an early stage with youth offenders is the high social and economic cost to society that is incurred if they continue along their criminal career path.

Actuarial Versus Clinical Risk Assessment

Whether actuarial assessment or clinical judgement is superior for making diagnostic decisions has been debated within several different professional fields including medicine, psychiatry, and psychology. Over the years actuarial measures have developed, their predictive validity has become more impressive and for approximately the last 50 years they have generally been found to be superior to clinical judgement in the literature (Craig, Browne, Stringer, & Beech, 2004; Dawes, Faust, & Meehl, 1989; Mills, 2005). The superiority of actuarial methods has been found across many different situations where prediction is required. Dawes et al. (1989), state that at the time of their publication there were almost 100 studies in the social science field comparing actuarial assessment with clinical judgment and in almost all of them the actuarial method was found to be superior to clinical judgement. Studies have shown that even when the professional is given access to additional information to aid in the decision process the actuarial method still outperforms them.

Dawes et al. (1989) discuss some of the reasons for the superiority of the actuarial method. Firstly, an actuarial measure will always provide the same prediction or outcome with the same data. However, given the same data, different professionals may come to different conclusions or the same professional may come to different conclusions at different points in time. Secondly, actuarial measures take into account only valid variables but professionals find it difficult to discriminate between valid and invalid variables. Professionals are also prone to making other judgement errors like using the more recent outcome information to inform decisions which may not necessarily be the most valid, or they may decide prematurely on their prediction of outcome and ignore data that disagrees with their decision. However, Dawes et al., (1989) point out that actuarial measures need to be carefully developed and their reliability and validity satisfactorily tested for them to be useful. Also, when they are applied to new settings or populations their predictive power needs to be reassessed to show they are useful within this new situation.

Baumann, Law, Sheets, Reid, and Graham (2005) discussed actuarial versus clinical risk assessment in relation to child welfare decisions. They pointed out that the large amount of literature on actuarial versus clinical methods of risk assessment generally favoured actuarial methods. However, there are some practical limitations associated with actuarial methods. These include the problems with the level of complexity of the instrument, often the more complicated an instrument is the more impractical it becomes (Gardner, Lidz, Mulvey, & Shaw, 1996).

More recently Gardner, Lidz, Mulvey, and Shaw (1996) have shown that clinician ratings were better than chance when predicting recidivism but actuarial measures outperformed clinicians. Fuller and Cowan (1999) examined the predictive validity of the consensus of a multi-disciplinary team and found that they performed as well as some actuarial instruments over comparable time-scales.

Overall these findings suggest that actuarial methods will perform better than clinical judgement when decisions are required about the recidivism risk level of a young person. This validates the need for the development of actuarial measures for use with young people that have high predictive properties and sound psychometric properties. However, under certain circumstances it is important to consider clinician predictions. At times the actuarial instrument may miss an important factor that can be identified by the clinician.

Dolan and Doyle (2000) discuss the structured clinical judgement approach which combines empirical knowledge and clinician expertise. There are guidelines provided to ensure that the clinical judgement is as accurate as possible. These include having high levels of agreement between assessors by ensuring their training is of a high standard, and making sure the prediction is for a defined type of behaviour. Within this model the actuarial risk score is only adjusted if there is adequate reason to do so. The Historical/Clinical/Risk Management-20 (HCR-20) is an instrument that is based on this model.

Mills (2005) pointed out that actuarial instruments were developed to avoid the processing errors inherent in clinical judgement decisions. He proposed that actuarial measures that assess change in risk of recidivism, in other words that include dynamic risk factors, will represent an important area of future development for violence risk prediction. He also anticipated that the statistical procedure of Cox regression would be useful for future research when predicting violence risk through survival or hazard functions. Cox regression allows for survival curves to be calculated for each score of a risk measure and is able to accommodate censored variables which are the cases for which the terminal event does not occur during the observational period. Therefore cases for which only partial information is known can still be considered within the analysis. Using Cox regression a hazard rate is calculated which is defined as the probability per unit time that the terminal event will occur. Predictor variables that are associated with variance in the hazard rate are identified and coefficients are estimated. Inferential techniques can be used to determine whether coefficients for particular variables are statistically significant.

Tiffin and Kaplan (2004) stated that when assessing risk in children and adolescents important factors to consider include personality, socialisation, substance use and past antisocial behaviour. They advise that actuarial instruments are best used together with a thorough assessment and applied with caution to younger children.

Another advantage to actuarial risk assessment measures is that the same factors are considered across individuals. This makes the process of decision-making more standardized and decreases the influence of individual biases. Actuarial measures also facilitate the use of operational definitions of decisions about risk. They also make it easier to evaluate the effectiveness of a structured decision-making process (Hoge, 2002).

A problem with actuarial risk prediction instruments is the definition of recidivism which is employed for use with the particular measure. This can vary across measures and make it difficult to compare the accuracy of prediction across measures. The base rate will vary according to what measure of recidivism is employed.

Sheldrick (1999) points out that professionals should refrain from making long-term predictions about risk when an assessment is based on dynamic risk factors because these can change over time. Assessments should be carried out periodically in order to create an accurate picture of what is going on for the person at the time. When assessments are made based on static risk factors the outcome will always be the same because these risk factors do not change. It is advisable to develop risk measures that take into account both static and dynamic risk factors so both aspects are considered within a risk score. Overall the findings reported above suggest that actuarial measures that are developed using Cox regression are the best method available at the present time to predict recidivism.

Problems with Predicting Behaviour

Craig et al. (2004) discuss the methods that have been used over the years to predict criminal and antisocial behaviour. It is difficult to accurately predict behaviour because there are often many factors involved that influence whether the behaviour will occur or not. The most basic method of prediction is the use of a 2 x 2 contingency table. This results in 4 outcomes; true negative where the prediction is low risk and the outcome is non-recidivism, false negative where the prediction is low risk and the outcome is reoffending, false positive where the prediction is high risk and the outcome is non-recidivism, and true positive where the prediction is high risk and the outcome is reoffending. The aim of any actuarial risk prediction measure is to minimise the number of false positives and false negatives and increase the number of true negative and true positives. The problem with this method is that it is centred around base rates which are known to be unstable and unreliable. Base rates are known to differ between subgroups of offenders and across age. They are also known to increase over time and vary between studies depending on the definition of offending employed in the study.

As the base rate differs from 0.50 it becomes more difficult statistically to develop a predictive model because the correlations are attenuated as the base rate goes down. We are most accurate at predicting outcome when the base rate of the given behaviour is at 0.50. The problem is that with rare behaviours even the best measures will identify a large number of false positives.

Assessing Predictive Accuracy Using ROC Analysis

Therefore, the methods like percent correctly classified, sensitivity and specificity, and false-positive and false-negative rates, used in the past to measure the predictive accuracy of various instruments are flawed in that it is difficult to make comparisons of the accuracy of different measures using these methods when the base rates are different (Rice & Harris, 1995). Other measures that have been suggested as alternatives include correlation coefficients and odds ratios, but these are also affected by the base rate or selection ratios.

The use of receiver-operating characteristic (ROC) curves has been offered as another way to assess the predictive accuracy of instruments that measure risk (Kroner, 2005; Rice & Harris, 1995). ROC curves are derived from signal-detection theory and this method is now used in several different fields of expertise.

A ROC graph plots sensitivity (or true-positive rate) versus one minus the specificity (or the false-alarm rate) at different cut-off (i.e., criterion) scores. The area under the curve (AUC) varies from 0.5 (indicating chance performance) to 1.0 (indicating perfect discriminative accuracy), and may be interpreted as the probability that a randomly selected participant who reoffended will obtain a higher score on the measure under scrutiny than a randomly selected participant who did not reoffend. The greater the AUC, the greater the difference between the false-alarm and true-positive rates. A higher AUC indicates that the instrument in question is performing well. By looking at a ROC curve it is easy to see how changing the cut-off score on the risk measure will alter the true-positive and false-alarm rates. The ROC analysis is favoured because it is less affected by base rates and selection ratios than other forms of analyses. ROCs are also useful for evaluating decision rules or selection ratios once costs are known. Measures that are derived using different base rates are easily comparable using ROCs and therefore decisions can be made about which would be the best risk assessment instrument to use given the circumstances. The other traditional measures of predictive accuracy can be derived from ROCs and therefore make data easily interpretable.

Rice and Harris (1995) point out some cautions when using ROCs. It is important to have more than one data point or sensitivity-specificity pair on the ROC curve in order to make the ROC interpretable at other points. It is also important to know the standard error of the AUC. In conclusion, ROC curves are the best way to measure predictive accuracy because they are independent of the base rate.

Risk Instruments for Young Offenders

Although the literature on risk instruments for adults is far more substantial, there are still several instruments that have been, or are in the process of being developed for use with children and adolescence. The instruments with the most literature associated with them are discussed below.

Bloom, Webster, Hucker, and De Freitas (2005) review Canadian research on the development of actuarial instruments for predicting violent recidivism. Canada has led the way in terms of research on actuarial risk assessment. This line of research began with a paper describing the development of the Violence Risk Appraisal Guide (VRAG) (Harris, Rice, & Quinsey, 1993). This instrument was based on a group of 618 male mentally-disordered offenders. The authors followed this group for 7 years and used multiple regression and logistic regression to develop a predictive model for violent recidivism. The VRAG includes variables related to 12 factors including childhood history, adult criminal history, demographics and psychiatric diagnosis. It also includes the PCL-R score (described below) which was found to have the heaviest weighting. A cross-validation study found that the VRAG performed as well as it did with the original sample on which it was developed (Rice & Harris, 1995). Originally Harris et al. (1993) obtained prediction-outcome correlations of approximately 0.50 and more recently they found the VRAG obtained an AUC of .73 when using a conservative definition of violent recidivism (Rice & Harris, 1995).

The Hare Psychopathy Checklist – Revised (PCL-R) is an instrument used to measure psychopathy in adults (Hare, 1991). A large amount of literature shows that this instrument has good psychometric properties and people gaining high scores on the PCL reoffend more often and more violently compared to those who do not obtain high scores (Salekin, Rogers, & Sewell, 1996). This instrument has greatly influenced the development of other actuarial instruments such as the HCR-20. A youth version of the PCL-R has also been developed (PCL-YV) which can be used with 12 to 18 year olds. This has 20 items that assess personality and behavioural factors which are considered to be a part of psychopathy and yields an overall score, as well as two subscores called callous/deceitful and conduct disorder. The first refers to personality while the second refers to lifestyle. Hoge (2002) points out that some items require professional judgement in order to score them and the instrument should therefore only be used by those with the appropriate experience. The adult and youth versions are very similar. Some changes were made however, to accommodate its use with adolescence, such as including items that reflect the higher level of involvement of peers, family and school, and also giving weight to the individual factors that are lasting and exist across settings. Information is obtained from multiple sources and each item is coded for intensity, frequency and duration.

Studies show that the PCL:YV can predict violence in an adolescent sample (Catchpole & Gretton, 2003) and that those obtaining high scores on the PCL:YV are more likely to have experienced a maladaptive family environment, display an absence of attachment, and have a history of performing antisocial and violent acts (Gretton et al., 2004). Despite the development and use of the PCL:YV there has been some debate about the validity of psychopathy within the adolescent population (Seagrave & Grisso, 2002). There is some concern about whether psychopathy as a construct remains stable through into adulthood and whether one should be identified as ‘psychopathic’ as a young person because of the negative consequences of receiving this label. It may be that while one can find features in a young person that reflect psychopathic traits, these may be part of a changing developmental process and will not persist into the adult life of the person in question. Therefore, some researchers have cautioned that there is a risk of a high false positive rate when psychopathy is examined in young people. Seagrave and Grisso (2002) explored the developmental characteristics of children and adolescents that may be mistakenly labelled as psychopathic traits and also go on to give guidelines for the development of instruments that measure psychopathy to help ensure that these measures are performing to a high standard. However, it has also been argued that few adolescents obtain a score on the PCL:YV that would warrant the label of psychopathic and this would indicate a young person with serious antisocial problems, therefore, the PCL:YV is only identifying those with severe problems (Gretton et al., 2004).

To investigate psychopathy in adolescence Gretton et al. (2004) obtained extensive file information and followed up a group of 157, 12 to 18 year olds for a period of 10 years after an initial conviction. A cut-off score is yet to be established for the PCL:YV, however, for the adult version scores of 30 and above are considered high risk and this criterion was used in this study along with a score of 18 to 29 as medium risk and below 18 as low risk. They found that the PCL:YV was particularly good at predicting violent reoffending, not only the actual event but also how long it took for the young person to reoffend. The PCL:YV did not predict non-violent reoffending but this could be in part due to the fact that 95% of the sample reoffended non-violently. It also did not predict sexual reoffending but this had a very low base rate in the population under study. However, the PCL:YV did predict the latency of non-violent offending.

The authors also tested whether the PCL:YV provided meaningful information above factors such as age at first arrest and a diagnosis of Conduct Disorder and found that high scores on the PCL:YV was a potent risk factor in its own right. The authors point out that this sample consisted of young offenders who had been referred for a psychological assessment and were therefore at a high risk for recidivism, therefore the PCL:YV may perform even better in a more varied sample. The authors also point out that much more research is needed on the psychometric properties of the PCL:YV before it is accepted and used within a clinical setting.

Corrado, Vincent, Hart, and Cohen (2004) carried out a study to assess the predictive validity of the PCL:YV, to determine whether a shorter version was as predictive of recidivism as the whole measure, and to investigate which clusters of psychopathic features were most important in predicting youth recidivism. Their sample included 161 boys aged 12 to 18 years who were currently in custody of some kind and were followed up to assess recidivism. They found that those who had a large number of psychopathic traits reoffended much more quickly after release than those with a low number of psychopathic traits. The predictive efficiency of the whole measure as measured by the AUC was approximately 0.65 to 0.68 which is lower than that generally found with adults. Corrado et al. suggested that their results implied that other variables were important in the development of criminal activity in young people. They also found that the shorter version was as good at predicting recidivism as the longer version particularly for predicting violent recidivism. A major conclusion drawn from this study is that it appears that the predictive ability of the PCL:YV is largely due to the behavioural features recorded by the instrument and not the interpersonal and affective personality constructs which are a major part of the psychopathy phenomena.

Therefore, as the research stands at present, there is substantial support for the association of psychopathy with reoffending in adults, but divided opinion on its association with reoffending in young people. Studies examining the PCL:YV have found some positive results in terms of its ability to predict recidivism but there are some cautions to keep in mind when using the PCL:YV or the construct of psychopathy with young offenders.

The Youth Offender Level of Service Inventory (YO-LSI) has 76 items which gathers information on risk factors that are organised into seven areas: Criminal History, Substance Abuse, Education or Employment, Family, Peer Relations, Accommodation, and Psychological Variables. The total score achieved can be used to indicate the level of intervention required or scores on specific areas can be used to indicate specific areas requiring intervention (Ilacqua, Coulson, Lombardo, & Nutbrown, 1999).

Ilacqua et al. (1999), carried out a study of 82 male and 82 female youth offenders to assess the validity of the YO-LSI. They defined recidivism as a conviction or charge with at least one offence in the following year. Their results showed that higher scores on the YO-LSI were associated with a higher level of recidivism. They also examined the use of the YO-LSI with female young offenders and found it to be just as effective as it was with male young offenders.

The Structured Assessment for Violence Risk in Youth (SAVRY) is a 20 item questionnaire (Borum, Bartel, & Forth, 2002). It includes items that measure historical, social-contextual, and individual-clinical risk factors. The construct of psychopathy is measured within the SAVRY and it also has items that measure protective factors. The structure of the SAVRY is modelled upon the HCR-20. When using the SAVRY to predict recidivism risk the items can be added to give a total risk score, and also clinical judgement can be incorporated into the process to give a rating of high, moderate, or low risk.

The Youth Level of Service/Case Management Inventory (YLS/CMI) grew out of the YO-LSI and has 42 items (Hoge & Andrews, 2002). The idea behind both of these instruments is that when making predictions about recidivism, risk factors and need characteristics must be taken into account and individual characteristics and circumstances are assumed to be the cause of youth offending (Andrews & Bonta, 1994). Items fall into one of eight categories: prior and current offences and dispositions, family circumstances and parenting, education and employment, peer relations, substance use, leisure and recreation, personality and behaviour, and attitudes and orientation. Items are noted as present or absent and then summed to give a total score which represents the risk level of the individual. The instrument also leads to direct recommendations about interventions appropriate for the individual in question. An advantage of this measure is that it is based on current theory about the development of delinquency in youth (Hoge, 2002). This instrument is used in Ontario, Canada with young offenders aged 12 to 15 years and is also known as the Ministry Risk/Need Assessment Form (MRNAF).

Generally the YLS/CMI has been found to have good psychometric properties (Watt et al., 2004). Jung and Rawana (1999) point out that the YLS/CMI was developed and normed on a population made up of mostly male non-Native American juvenile offenders therefore its ability to predict recidivism in other populations like females or Native Americans is unknown. Therefore Jung and Rawana (1999) carried out a study to assess the validity of the YLS/CMI in general and in relation to Native Americans and female young offenders in particular. A total of 263 young offenders took part in the study including 173 males, 90 females, 134 Native Americans and 129 non-Native Americans. The results showed that scores on the YLS/CMI differentiated between those who reoffended and those who did not. Also each of the eight categories listed above on their own distinguished between recidivists and nonrecidivists. The most significant factors were education and employment, negative peer relationships and antisocial attitudes. The Native American sample obtained higher scores on the YLS/CMI. They were more likely to have negative peer associations, more alcohol or drug use, and less participation in prosocial activities. However, total scores on the YLS/CMI were able to predict reoffending for both groups. No differences in total scores were found between male and female youth offenders. Overall the YLS/CMI predicted recidivism for females and Native Americans just as well as it did for the population it was developed for (i.e., non-Native male youth offenders).

Catchpole and Gretton (2003) carried out a study assessing the predictive validity of the SAVRY, YLS/CMI and PCL:YV using ROC analyses. They examined file information from 74 youth offenders and used this information to code the various instruments. Catchpole and Gretton (2003) found that for general reoffending ROC AUCs for the three instruments ranged from .74 to .78. For violent reoffending ROC AUCs of .73 were obtained for all three measures. Also the YLS/CMI was inclined to predict few participants as low risk and more participants as high risk in comparison to the SAVRY. The authors explain this in terms of the differing purposes of the two instruments. They also point out that when assessing risk in adolescence it is important to keep in mind that this is a time of developmental change. Therefore it is necessary to maintain up-to-date assessments to take these changes into account.

Risler, Sutphen, and Shields (2000) discuss the validity of the First Offender Risk Assessment Index (FORAI) which they originally developed at the University of Georgia in conjunction with the Athens/Clark Count juvenile court. After carrying out a literature review seven variables were selected and included in the FORAI to predict recidivism: age at first court referral, seriousness of referring offence, parental supervision, school functioning, peer group association, alcohol and drug use, and family's history of criminal involvement (Risler et al., 2000). After testing and refining, the FORAI contained 41 items in two forms for both the youth and the parent. The items fall into one of the seven categories mentioned above and responses on each of the two forms are compared with a score that takes into account the amount of agreement between the parent and child. Risler et al. (2000) carried out a validation study of the FORAI on 181 first time offenders and their parents. During a 4 year follow-up period recidivism was measured as any new offence or charge that came before the court. The results showed that the FORAI was a valid measure of recidivism as it correctly predicted the future offending of 71.8% of the sample. It also had a low rate of false negatives but a comparatively high rate of false positives. Of the seven constructs four significantly predicted reoffending (family's history of criminal involvement, school functioning, age at first court referral, and seriousness of the referring offence). The remaining three were not significant in predicting reoffending and Risler et al. (2000) discussed several possible reasons for this result. The authors concluded that when compared to other assessment measures, the FORAI was one of the most accurate but still requires further validation and testing to enhance its predictive ability.

Thompson and Putnin (2003) discuss two instruments that are currently used in Australia to assess the risk level of youth offenders. The Secure Care Psychosocial Screening (SECAPS) assessment is used to assess criminogenic needs and risk (Putnins, 1999). It also includes items that assess self-harm risk and what is labelled responsivity factors like numeracy, literacy and intelligence. Responses are entered into the SECAPS database and a report is generated automatically. The SECAPS is brief, easy to use and its use ensures that important information is obtained for each youth offender in the setting in which it is employed. Reassessment occurs over time which means the young person's level of improvement can be monitored. During research to establish the predictability power of the SECAPS it was found that the most significant factors for predicting recidivism were age, age at first offence, number of prior offences, recent alcohol and or inhalant use and ADHD-related signs. SECAPS risk scores were found to correlate significantly (r = .34) with recidivism at 6 months after release from care. Normative data are available for the SECAPS and development of the measure continues.

Another instrument used in Australia is the Australian Adaptation of the Youth Level of Service/Case Management Inventory (YLS/CMI-AA; (Hoge & Andrews, 1995). It has 47 items that fall into the categories of prior and current offences, family and living circumstances, education/employment, peer relations, substance abuse, leisure/recreation, personality/behaviour, attitudes/beliefs, and major strengths. Some of the changes to the instrument to adapt it for Australian use include wording of some items, and including new items that have been found to be relevant risk factors within the Australian youth offender population. The YLS/CMI-AA assesses dynamic risk factors and therefore has an intervention focus. Thompson and Putnin (2003) report on some psychometric data that has recently been collected on the YLS/CMI-AA. They report a point biserial correlation between recidivism and total risk score of .28 ($p < .001$, 2-tailed).

Leistico and Salekin (2003) examined the reliability and validity of the Risk, Sophistication-Maturity, Treatment Amenability-instrument (RST-i), designed for use with young offenders. This measure has 45 items and assesses aspects of personality, antisocial behaviours, attitudes towards legal difficulties and environmental factors. The items fall into three subscales which are risk, sophistication-maturity and treatment amenability. They found that scores on the RST-i significantly predicted whether a participant was likely to be transferred to the adult court.

Krysik and LeCroy (2002) assessed the psychometric properties of the risk prediction instrument employed by the Arizona Supreme Court. This measure is comprised of nine items which include number of counts involved in offence, weapons or drug involvement, assaultive behaviour, nature of the status offence, prior property offences, whether the juvenile is detained or not, age at first complaint, most serious offence class, and time between offence and referral. Scores from this instrument range from -1 to 9 with scores of 4 and above indicating high risk. They defined recidivism as any delinquent complaint happening up to 1 year after the initial referral. The authors found that the recidivism rate for the high risk group was 50% and the recidivism rate for the low risk group was 26%. Therefore they concluded that this instrument was not performing to an adequate level in terms of identifying those young offenders who were at a high risk of developing into long term recidivists.

Krysik and LeCroy (2002) cited problems with the development of the instrument as contributing to its poor performance. These problems included only 3 months of data being used as a basis for its development, not using separate samples for development and validation of the instrument, and the reliability of scoring the instrument was not considered during its development. Krysik and LeCroy (2002) go on to explain how they modified the instrument to improve its predictive ability. They calculated bivariate relationships between predictor variables and recidivism. The predictors that were significant at the .05 level were kept for use. These predictors were entered into a stepwise logistic regression equation to ascertain which were the most significant factors. The five variables thus obtained were drug use in the past year, truancy, problems within family relationships, assaultive behaviour, and type of offence. This new measure which included these five factors performed at a much higher level compared with the old instrument. Another benefit of the new instrument was that it elicited a score that was much easier to understand, that is the percentage likelihood of recidivism rather than the score ranging from -1 to 9.

Ashford and LeCroy (1990) examined the predictive ability of three actuarial instruments, the Conta Costa Risk Assessment Instrument, the Orange County Risk Assessment Instrument, and the Arizona Juvenile Risk Assessment Form on a sample of young offenders from Arizona. They found that the Contra Costa Risk Assessment Instrument could not differentiate between those who reoffended and those who did not. The other two instruments could differentiate between the two groups but still did not perform to a satisfactory level, only predicting reoffending 18-22% better than chance. They also found that the delinquent history of the young person was the best predictor of future recidivism. The authors warned of the possible problems associated with using actuarial instruments developed in one geographic area with young people living in another geographic area.

The Child and Adolescent Functional Assessment Scale (CAFAS) was developed by Hodges (1994). This measure assesses behavioural and emotional functioning. It provides a total score and five subscores measuring role performance, cognition, behaviour toward others and self, moods or emotion, and substance abuse. It also measures how well the basic needs of the young person are being met and the calibre of parenting the young person is receiving. Severity of impairment is recorded across all of these subscales and overall level of impairment is also calculated. Behavioural markers are used to guide the rater with assigning severity scores and therefore it is quite easy to use (Hoge, 2002). The instrument has been found to have good psychometric properties and the total score has been found to significantly predict recidivism across an 18 month follow-up interval.

The preceding section describes different risk assessment instruments developed for use with children and adolescence. Overall, there have been positive results associated with the performance of these instruments. However, more research is needed that further assesses their psychometric properties and this research is only the beginning of establishing a substantial body of work associated with predicting recidivism in young offenders.

Statistical Models

It is important to examine the methods employed to develop the risk assessment instruments discussed above in order to ascertain the best statistical processes to use. Statistical modelling has been employed to predict recidivism or future violent behaviour among many different populations. The most popular statistical method employed thus far has been logistic regression, however, classification and regression tree modelling has also received some attention as a method that captures more of the clinical thinking processes within the decision process (Thomas et al., 2005). A classification tree splits factors up in binary form in a way to predict an outcome. The tree begins with a single root node at the top and each branch represents a single binary decision which leads onto another branch and so on until the decision is reached at the bottom of the tree. Thomas et al. (2005) carried out a study comparing logistic regression to classification trees to see which statistical method was best at predicting aggressive behaviour in a sample of patients with psychosis. They found that the full logistic regression model was best at predicting future violence in terms of both sensitivity and specificity. The classification tree model performed well in terms of sensitivity but its specificity was lower. Therefore, it can be concluded that logistic regression is the best method to employ when developing a risk assessment instrument.

It is evident from the section above that several actuarial risk prediction instruments have been developed for use with youth offenders, however, these are predominantly psychometric instruments that require some level of subjective judgement as opposed to true actuarial models or statistical models that do not.

Purpose of the Present Study

The present study proposes to develop a statistical model to predict recidivism based on a sample of youth offenders in New Zealand. The increasing use of computer databases by government agencies for archiving large amounts of information and availability of software for analysis means that it is now easier to carry out studies which analyse this data. Combining multiple databases could result in more accurate statistical models. New Zealand represents an ideal location in which to carry out this type of research because there is only one database used by the police, and one database used by Child, Youth and Family. Combining information from both of these databases would allow us to obtain information about a young person over a relatively long period of time.

The model will be developed from archival information obtained from the Child Youth and Family (CYF) database, CYRAS, and the police database, National Intelligence Application (NIA). We will obtain CYRAS records for all young persons who had a Youth Justice intake in 2002 (N = 4307). We will then identify a random sample of 500 cases stratified by region and collect NIA data for these cases. It was necessary to stratify the sample because NIA data cannot be automatically downloaded, therefore it is necessary to search for information on each individual separately. A second random stratified sample of 500 will be identified and NIA data will also be collected for these cases. This second sample will be used to validate the model by testing its generalizability.

Method

Participants

The sample consisted of all young persons that had a Youth Justice (YJ) Intake in the year 2002 (N = 4307). The number of males in the sample was 3440 and the number of females was 867. The ethnic make-up of the sample consisted of 1396 (32.4%) Pakeha individuals, 1660 (38.5%) Maori individuals, 389 (9%) Pacific Island individuals and 90 (2.1%) individuals identified as belonging to another ethnic group. For 772 (17.9%) individuals ethnicity was not recorded.

Procedure

Statistical Analysis

The data were modelled in two stages. The first stage consisted of creating a model (called **M1**) using the entire CYF sample. Recidivism was defined as the occurrence of at least one additional YJ Intake after the criterion date (i.e., date of the young persons first YJ intake in 2002) and prior to the 31st December 2004 (YJPost). The second stage consisted of using the output from **M1** (i.e., risk score generated by **M1** for individual cases) and the historical data from NIA to create two other models for predicting recidivism in the follow-up period, one for general recidivism (**M2**) and one for serious recidivism (**M3**). The outcome variables from NIA included the prosecutions for any offence after the criterion date for M2 and prosecutions for any serious offence for M3.

Different measures of reoffending are used because there is no single measure that adequately describes offending; therefore separate models were developed for general recidivism and violent/serious recidivism. Given that young people in New Zealand are processed through different administrative systems (CYF, Police, Courts) when they offend it is likely that a composite measure will provide the most comprehensive measure of reoffending. To test this idea a cross-tabulation analysis was carried out on the two measures of reoffending from the CYF and NIA databases, YJPost and AnyProsecution. The cross-tabulation was carried out on the combined Developmental and Validation samples (N = 1000).

Table 2.

Cross-tabulation Results for the Developmental and Validation Samples Combined on Two Measures of Reoffending

		AnyProsecution		Total
		0	1	
YJPost?	0	258	212	470
	1	126	404	530
Total		384	616	1000

In Table 2 it can be seen that 53% of the sample had an additional YJ intake and 62% had an additional prosecution during the follow-up period. There was a positive and significant ($p < .001$) correlation between YJPost and AnyProsecution of $r = 0.32$, however, Table 2 shows that 23.8% of those who obtained a second YJ intake did not receive a new prosecution, and 34.4% of those who obtained a second prosecution did not receive a new YJ intake. A total of 74.2% received either a second YJ intake or prosecution. This analysis demonstrates that a combined measure of reoffending YJPostOrAnyProsecution is a better measure of general reoffending than either measure on its own. Therefore this composite measure was used in M2.

Survival analysis, specifically, Cox regression, was the statistical method selected to develop the models, because it naturally accommodates variable follow-up times and ‘censored’ variables in which the ‘terminal’ event may or may not occur during the observational period. Cox regression will be used to estimate the hazard rate or probability per unit time that the terminal event will occur. Predictor variables associated with variance in the hazard rate will be identified and coefficients estimated. Inferential techniques will be used to determine which coefficients for particular variables are statistically significant and should therefore be included in the model. For each case the model will predict a standardized hazard rate or ‘Xbeta’ score which can then be interpreted as a risk score with positive Xbeta scores indicating higher than average risk and negative Xbeta scores indicating lower than average risk. An Xbeta score of 0 represents the average risk level of the sample in question. After the model has been developed it will be tested on the second validation sample of 500 youth offenders using ROC analysis. The development and validation samples will also be compared in terms of whether they differ significantly across predictor variables.

CYF data set

The individuals in the sample were identified in the CYRAS database and information was obtained for each person including details of intakes, family group conferences (FGC), social worker findings, court appearances and demographics. This information provided an entire CYF history for each person up until 31 December 2004. All of the data was incorporated within a Microsoft Access database and potential predictor and outcome variables were generated. The criterion date was defined as the date of the first YJ Intake in 2002. The follow-up period consisted of the time from the first YJ Intake in 2002 until the individuals 18th birthday or 31 December 2004, whichever came first. The final sample size was 4295 after 12 cases were omitted because their birthdates listed in CYRAS showed that they were over 18 years of age or less than 10 years of age at the time of the criterion date. Below is a list of the predictor variables derived from the CYRAS database and the label given to each variable during analysis. The following is an explanation of what the predictor variables mean. A youth justice intake occurs when a young person aged 10-16 years commits a criminal act. These young people are dealt with by the youth justice system which is a special section of the law that aims to deal with the offending without the young person receiving a criminal conviction. Therefore, the aim of the youth justice system is to keep the young person out of the criminal justice system and out of the courts. The youth justice system is based on a restorative justice philosophy. For cases that are more serious and any case going through the courts a Family Group Conference is held where the young person, their family, the Police Youth Aid and Youth Advocate and the victim all meet to decide how the young person will be held accountable. A Care and Protection intake occurs when a child is believed to be at risk of harm. The harm may be in the form of physical or sexual abuse, violence or conflict between caregivers, or emotional or physical neglect. The outcome of a youth justice intake and Family Group Conference may be a placement of the young person in a residential care facility. The placement generally occurs when it is believed that the child is a risk to themselves or others if they remain in the community. The result of a Care and Protection intake may be the placement of the young person in a foster home or other care facility.

Table 3.

Predictor Variables for CYF Data

Predictor Variable	Variable Designation
Gender (Male/Female)	MF; 1 = Male; 0 = Female
Ethnicity (Pakeha/Maori/Pacific/Other)	MP; 1 = Maori/Pacific 0 = Not Maori/Pacific
Age at First CYF Intake	Age1stIntake
Age at Criterion	AgeatCriterion
Rate of CYF Intakes (number of intakes divided by difference between criterion and first intake dates)	RateofCYFIntakes
Age at First YJ Intake	AgeatFirstYJIntake
Number of Prior Care & Protection (C&P) Intakes	NumPriorC&P
Number of Prior YJ Intakes	NumPriorYJ
Number of Prior Intakes	NumPriorIntakes
Number of Prior Intakes – Section 15	NumPriorIntakesSec15
Number of Prior Intakes – Urgent	NumPriorIntakesUrgent
Number of Prior Placements	NumPlacements
Number of Prior Findings	NumPriorFindings
Number of Prior Findings – Behavioural/Relationship Difficulties	NumPriorFindingsBRDifficulties
Number of Prior Findings – Emotional Abuse	NumPriorFindingsEmotAbuse
Number of Prior Findings – Neglected By	NumPriorFindingsNeglected
Number of Prior Findings – Not Found	NumPriorFindingsNotFound
Number of Prior Findings – Physical Abuse	NumPriorFindingsPhysAbuse
Number of Prior Findings – Self-Harm/Suicidal	NumPriorFindingsSelfHarm
Number of Prior Findings – Sexual Abuse	NumPriorFindingsSexualAbuse
Number of Prior YJ-FGCs	NumPriorYJ-FGCs
Number of Prior YJ-FGCs – No Agreement	NumPriorYJ-FGCsNoAgree
Number of Prior Court Outcomes	NumPriorCourtOutcomes
Number of Prior Court Outcomes – Supervision	NumPriorOutcomes-Supervision
Number of Prior Court Outcomes - Custody	NumPriorOutcomes-Custody
Number of Prior Court Outcomes – Other YJ	NumPriorOutcomes-OtherYJ
Number of Prior Court Dates	NumPriorCourtDates

Recidivism was defined as the occurrence of at least one additional YJ intake during the follow-up period. Other variables were considered but it was decided that YJ intake was the best measure. Court outcomes could have been used but they were problematic because they could have been procedurally related to the criterion YJ intake.

The outcome variables for the CYF sample were defined as any additional YJ intake after the criterion date (YJPost) and the follow-up time or time between the criterion date and the individuals 18^{th} birthday or the 31^{st} of December 2004 for those who did not have another YJ intake (TimeAtLarge). For those that did it was the time between the criterion date and the second YJ intake. These outcome variables were used in the development of M1.

NIA data set.

The Police NIA database was used to obtain information about records and prosecutions for a random sample of 500 individuals from the CYF sample, stratified by geographic region. These initial 500 cases were termed the developmental sample because they were used in generating statistical models for predicting recidivism. Information was also obtained for a second stratified random sample of 500 individuals from the CYF sample and this group was used as a validation sample for the model. In the CYF sample 842/4307 (19.55%) individuals were from the Southern Region, 740 (17.18%) were from the Auckland Region, 611 (14.18%) were from the East-West Region, 875 (20.32%) were from the Midlands Region, 510 (11.84%) were from the Wellington-upper South Region, 649 (15.07%) were from the Northern region and 80 (1.86%) were listed as “Co-ordinators SDN”. The latter category was left out of the stratified sample. The number of cases per region for both the developmental and the validation stratified samples were as follows: Southern, 100; Auckland, 90; East-West, 70; Midlands, 105; Wellington-Upper South, 60; and Northern, 75.

Two primary sources of information were extracted from the NIA database to be used in the present study, **prosecutions** (or convictions) and **records.** The prosecutions information included the court hearing date, disposition, offence date, offence details and offence code for each prosecution that the young person had obtained. The records information included such things as Charge Sheets, Intelligence Notings, Occurrences, Youth Aid Case Folders, and other less common events like DNA recordings, Drug Seizures, Telephone Changes and Address Changes. An intelligence noting could include things like a noting that the young person was seen at an address where a crime was committed or a noting of the young persons involvement with other known criminals. An occurrence generally describes an interaction that the young person has with the police. For example, the young person may have been found in possession of drugs, or they may have been found to be driving an unregistered or stolen car. The date, time, location and scene station associated with each of these events was also recorded. A prosecution corresponds to a single offence and if it is a recent offence then it is likely to have a remand disposition which, after the court hearing, will be changed to convicted, case proved, or it will be deleted. Traffic prosecutions were not included in the analysis because preliminary analyses showed that they were not significantly related to prosecutions for general and violent offences, which was the primary purpose of the study. Table 4 lists the predictor variables and outcome variables that were derived from the NIA data.

Table 4

Predictor and Outcome Variables Derived From NIA Data

Predictor Variables	**Variable Designation**
Number of Prior Prosecutions	NumPriorProsecutions
Number of Prior Records	NumPriorRecords
Number of Prior Charge Sheets	NumPriorChargeSheets
Number of Prior Occurrences	NumPriorOccurrence
Number of Prior Intelligence Notings	NumPriorIntelligence
Number of Prior Youth Aid	NumPriorYouthAid
Outcome Variables	
Number of Prosecutions Follow-Up	NumProsecutionsFollowUp
Prosecution During Follow-Up	Prosecution?
Prosecution for Serious Offence During Follow-Up	ProsecutionSerious?
Number of Records Follow-Up	NumRecordsFollow-Up
Maximum Offence Severity	MaxSeverity
Total Severity of Offences During Follow-Up	TotalSeverityIndex
Time At Large	TimeAtLarge

A 5 point scale was used to measure the severity of an offence (Maxwell et al., 2002). The scale classifies violent offences like homicide, sexual assault, grievous and serious assaults, and robbery as the most serious (5), burglaries, car conversions, minor assaults and drug offences (excluding cannabis) as medium seriousness (3), and vagrancies, administrative offences and cannabis possession as minimum seriousness (1).

TimeAtLarge was calculated separately for each survival analysis depending on whether YJPost, AnyProsecutionOrYJPost or SeriousProsecution was used as the outcome variable. The TimeAtLarge was equal to the number of days between 31 December 2004 and the criterion date if the individual had not reoffended. If the individual had reoffended then the TimeAtLarge was equal to the number of days between the offence date and the criterion date.

Results

Preliminary analysis of CYF 2002 sample data

Some of the predictor variables that were positively skewed were recoded so that extreme values did not have undue influence. Frequency tables and histograms showing the distribution of each variable were examined and cut-off points were determined so there were a reasonable number of cases in each category. A list of the variables that were recoded and their new values can be found in Appendix A. Table 4 displays descriptive statistics for all the predictor and outcome variables for the entire CYF 2002 sample.

Table 5

Descriptive Statistics for the Predictor and Outcome Variables for the CYF 2002 Sample

	N	Minimum	Maximum	Mean	Std. Deviation
AgeAtFirstIntake	4295	.16	17.99	12.59	4.05
AgeAtCriterion	4295	10.16	17.99	15.76	.98
AgeAtFirstYJ	4295	7.38	17.99	15.34	1.11
NumPriorCP	4295	0	4	1.31	1.532
NumPriorYJ	4295	0	4	.63	1.161
NumPriorIntakes	4295	0	7	2.21	2.445
NumPriorIntakesSec15	4295	0	12	.70	1.165
NumPriorIntakesUrgent	4295	0	8	.71	1.204
NumPriorIntakes>Age10	4295	0	25	1.91	2.628
NumPriorPlacements	4295	0	4	.71	1.327
NumFindings	4295	0	4	1.00	1.387
NumFindingsBRDifficulty	4295	0	5	.42	.804
NumFindingsEmotAbuse	4295	0	3	.05	.245
NumFindingsNeglected	4295	0	5	.18	.557
NumFindingsNotFound	4295	0	6	.23	.581
NumFindingsPhysicalAbuse	4295	0	7	.16	.490
NumFindingsSelfHarm	4295	0	2	.01	.111
NumFindingsSexualAbuse	4295	0	5	.10	.374
NumPriorYJ-FGC	4295	0	11	.52	1.080
NumPriorYJ-FGC-NoAgree	4295	0	4	.04	.235
NumPriorCourtOutcomes	4295	0	14	.40	1.160
NumPSupervisionOutcomes	4295	0	6	.07	.380
NumPCourtDates	4295	0	12	.34	.959
NumPCustodyOutcomes	4295	0	5	.13	.477
NumCustSuperOutcomes	4295	0	9	.20	.698
NumPOtherYJOutcomes	4295	0	10	.09	.467
NumFGCCustSuper	4295	0	15	.72	1.532
YJPost?	4295	0	1	.52	.500
TAL-YJ	4295	1	1277	388.73	313.617
Valid N (listwise)	4295				

The model was developed in conjunction with the evaluation of an intensive treatment programme based on multi-systemic therapy (MST) for high risk youth offenders being piloted in Auckland and Christchurch between 2003 and 2006 (the

Reducing Youth Offending Programme [RYOP]), therefore the cases were subdivided for analysis into three groups, Christchurch (1), Auckland (2) and Elsewhere (3). This was done in order to determine whether it was necessary to take into account regional differences within the model. The subdivision also made sense in terms of having adequate numbers in each group and representing the largest city in both the North and South Islands. This comparison of the two cities and the rest of the country was considered to be a valid representation of the differences across New Zealand. For each case the region and area specified at the time of the criterion date was used to represent it. The numbers of cases in each region were as follows; 381 in Christchurch, 740 in Auckland, and 3186 in Elsewhere. The following tables show the averages for all predictor variables across regions and therefore show any differences between the three regions.

Table 6.

Averages of Predictor Variables across Regions (= p < .05, ** = p < .01, *** = p <.001).*

Predictor Variable	**Christchurch**	**Auckland**	**Elsewhere**	**All Groups**
MaleFemale	0.81	0.82	0.79	0.80
MaoriPacific***	0.29	0.64	0.46	0.48
AgeAtFirstIntake***	11.43	13.22	12.58	12.59
AgeAtCriterion	15.78	15.77	15.74	15.75
AgeAtFirstYJ*	15.20	15.35	15.35	15.34
NumPriorCP***	1.67	1.08	1.32	1.31
NumPriorYJ***	0.87	0.68	0.59	0.63
NumPriorIntakes***	2.87	1.96	2.19	2.21
NumPriorIntakesSec15***	0.89	0.58	0.71	0.70
NumPriorIntakesUrgent*	0.80	0.61	0.73	0.71
NumPriorIntakes>Age10***	2.58	1.79	1.86	1.91
NumPriorPlacements***	0.97	0.76	0.67	0.71
NumFindings***	1.24	0.84	1.01	1.00
NumFindingsBRDifficulties***	0.56	0.34	0.42	0.42
NumFindingsEmotAbuse*	0.05	0.03	0.05	0.05
NumFindingsNeglected	0.19	0.15	0.19	0.18
NumFindingsNotFound***	0.33	0.18	0.23	0.23
NumFindingsPhysicalAbuse	0.17	0.14	0.17	0.16
NumFindingsSelfHarm	0.00	0.01	0.01	0.01
NumFindingsSexualAbuse	0.12	0.10	0.10	0.10
NumPriorYJ-FGC***	0.77	0.54	0.49	0.52
NumPriorCourtOutcomes***	0.65	0.33	0.38	0.40
NumPSupervisionOutcomes***	0.19	0.06	0.05	0.07
NumPCourtDates***	0.52	0.29	0.33	0.34
NumPCustodyOutcomes**	0.20	0.12	0.13	0.13
NumPOtherYJOutcomes**	0.04	0.05	0.10	0.09
NumPriorYJ-FGCNoAgree	0.07	0.03	0.04	0.04
YJPost?***	0.60	0.58	0.50	0.52

It can be seen from Table 6 that there were differences between the regions. The proportion of Maori/Pacific young people was greater in Auckland (64%) than in Christchurch (29%) or Elsewhere (46%). There was also a larger number of CYF Intakes (including IntakesSec15, IntakesUrgent, Care and Protection, and Intakes>Age10), Placements, Findings, FGCs, and Court Outcomes in Christchurch compared with Auckland and Elsewhere. Christchurch also had a lower Age at First Intake. These data indicate that young people in Christchurch have earlier and more frequent CYF contact compared with the rest of the country. This may be in part because of the lower population density in Christchurch and the Southern Region and therefore more time and resources to assign to these young people. Reoffending, as measured by a second YJ Intake during the follow-up period, was highest in Christchurch (60%) followed by Auckland (58%) and lowest in the rest of the country (50%).

Overall these differences across regions in both predictor and outcome variables suggest that one model will not adequately describe the whole sample. Therefore the data was analyzed separately for each region in the survival analysis and a hybrid model was developed which allowed for differences in predictor and outcome variables across regions. This way all of the information from the CYF variables, including regional differences could be included in the model.

The Kaplan-Meier method was used to calculate survival plots for each region as shown in the graph below.

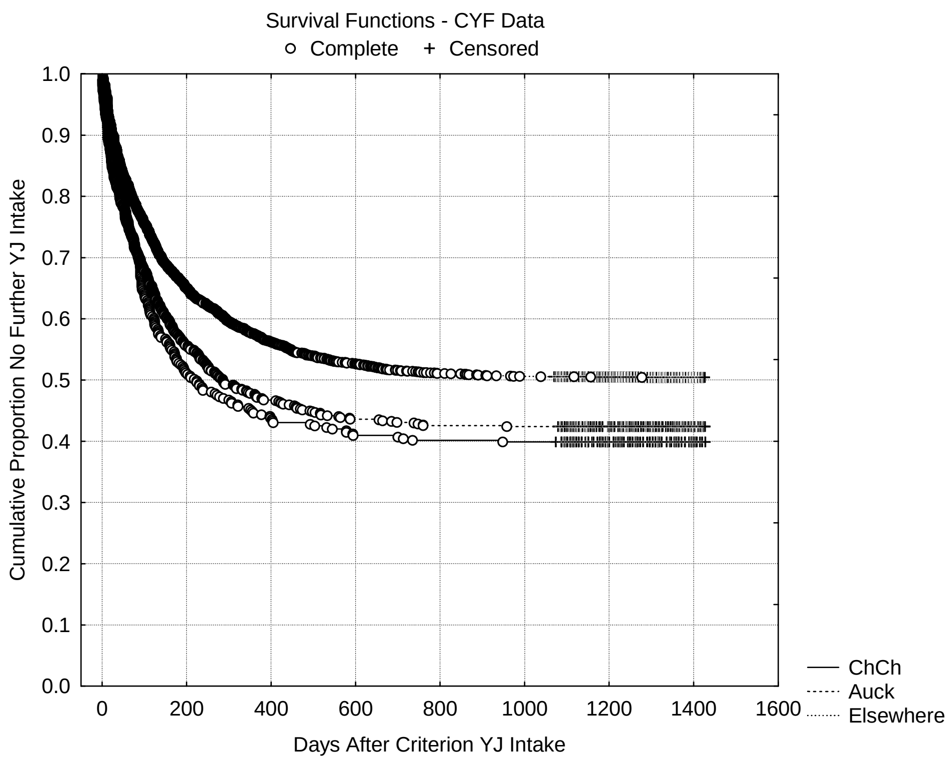

Figure 1. **Survival functions (cumulative proportion with no second YJ Intake) for the CYF 2002 data. Data are shown separately for Christchurch, Auckland and Elsewhere.**

Figure 1 shows the proportion of cases that have not reoffended as measured by a second YJ Intake, as a function of days following the criterion date. The figure shows that the rate of second YJ Intake was highest in Christchurch followed by Auckland and lowest in the rest of the country which is consistent with the data from Table 6. The difference between Christchurch and Auckland was not significant (Gehan's Wilcoxon = -.5, $p > .10$), however the difference between Christchurch and the rest of the country was statistically significant (Gehan's Wilcoxon = -4.6, $p < .001$) as was the difference between Auckland and the rest of the country (Gehan's Wilcoxon = -5.1, $p < .001$). These differences across regions evident from Figure 1 lend further support to the idea that the model needs to be designed so risk is assessed separately for each region.

Model 1: Predicting Subsequent YJ Intake for the CYF 2002 Sample

Different combinations of predictor variables were examined using exploratory survival analyses. Forward and backward stepwise regression and best-subsets regression were used to select the predictor variables that performed the best for each of the regions of Auckland, Christchurch and the rest of the country. When developing any statistical model there is a trade-off between how well the model explains the data and the generalizability of the model. Generally, the better the model explains the data the worse it performs on other data sets. It is important to balance these two factors and ensure that the model does not become so complicated that it does not generalize to other samples (Pitt, Myung, & Zhang, 2002). The decision about how complex to make the model is subjective but it needs to be simple enough that it generalizes. Therefore the selection of predictor variables for the present model, after many analyses were conducted, was made with this point in mind. It was also important that the predictor variables were as consistent across regions as they could be while still explaining the data well.

The overall performance of the model for each region was statistically significant as can be seen from table 7.

Table 7.

Performance of the Model for Each Region

Region	Degrees of Freedom	χ^2 $p < .001$
Christchurch	4	97.90
Auckland	4	51.52
Elsewhere	6	364.64

The following tables show which predictor variables were included in the model for each region.

Table 8.

Predictor Variables for Auckland Cases. B = coefficient, SE = standard error, Wald = Wald statistic, DF = degrees of freedom, Sig = obtained significance level, Exp(B) = exponential function of coefficient

	B	SE	Wald	df	Sig.	Exp(B)
MaoriPacific	.615	.115	28.700	1	.000	1.849
AgeAtFirstIntake	-.045	.014	10.099	1	.001	.956
AgeAtFirstYJ	-.242	.049	24.492	1	.000	.785
Male_Female	.295	.136	4.709	1	.030	1.343

Table 9.

Predictor Variables for Christchurch Cases. B = coefficient, SE = standard error, Wald = Wald statistic, DF = degrees of freedom, Sig = obtained significance level, Exp(B) = exponential function of coefficient

	B	SE	Wald	df	Sig.	Exp(B)
MaoriPacific	.254	.144	3.118	1	.077	1.289
AgeAtFirstIntake	-.045	.016	7.914	1	.005	.956
AgeAtFirstYJ	-.248	.064	14.874	1	.000	.780
NumFindings SexualAbuse	-.364	.166	4.811	1	.028	.695

Table 10.

Predictor Variables for Cases from Elsewhere. B = coefficient, SE = standard error, Wald = Wald statistic, DF = degrees of freedom, Sig = obtained significance level, Exp(B) = exponential function of coefficient

	B	SE	Wald	df	Sig.	Exp(B)
MaoriPacific	.241	.052	21.568	1	.000	1.273
AgeAtFirstIntake	-.022	.010	4.954	1	.026	.979
AgeAtFirstYJ	-.164	.023	49.916	1	.000	.849
Male_Female	.513	.070	54.367	1	.000	1.671
NumPriorCP	.105	.025	17.513	1	.000	1.110
NumPCourtDates	.081	.024	11.629	1	.001	1.084

The term Exp(B) can be interpreted as an odds ratio for a 1 unit positive change in the predictor variable. For all regions young people were more likely to reoffend if they were of Maori or Pacific Island ethnicity and younger at the time of their first CYF and YJ intake. Maori and Pacific Island ethnicity were coded as a single category because it was found that separate coding did not provide a significant improvement in the models performance Ethnicity was still included in the model for Christchurch even though it approached significance ($p < .08$), in order to be consistent with the other regions. Because there were fewer young people of Maori or Pacific Island ethnicity in Christchurch it was expected that statistical power would lower for this region. Males were more likely to offend everywhere except in Christchurch and for Christchurch only, the number of prior findings of sexual abuse was negatively related to reoffending. The number of prior CP intakes and number of prior court dates was positively related to reoffending in the rest of the country but not in Auckland or Christchurch.

The appropriate coefficient was multiplied by the centred predictor variable (i.e., predictor variable – average predictor variable) and summed across variables for each case in order to create an XBeta score (Cleves, Gould, & Guitierrez, 2002). Figure 2 shows the distribution of Xbeta scores for M1 for the entire CYF sample.

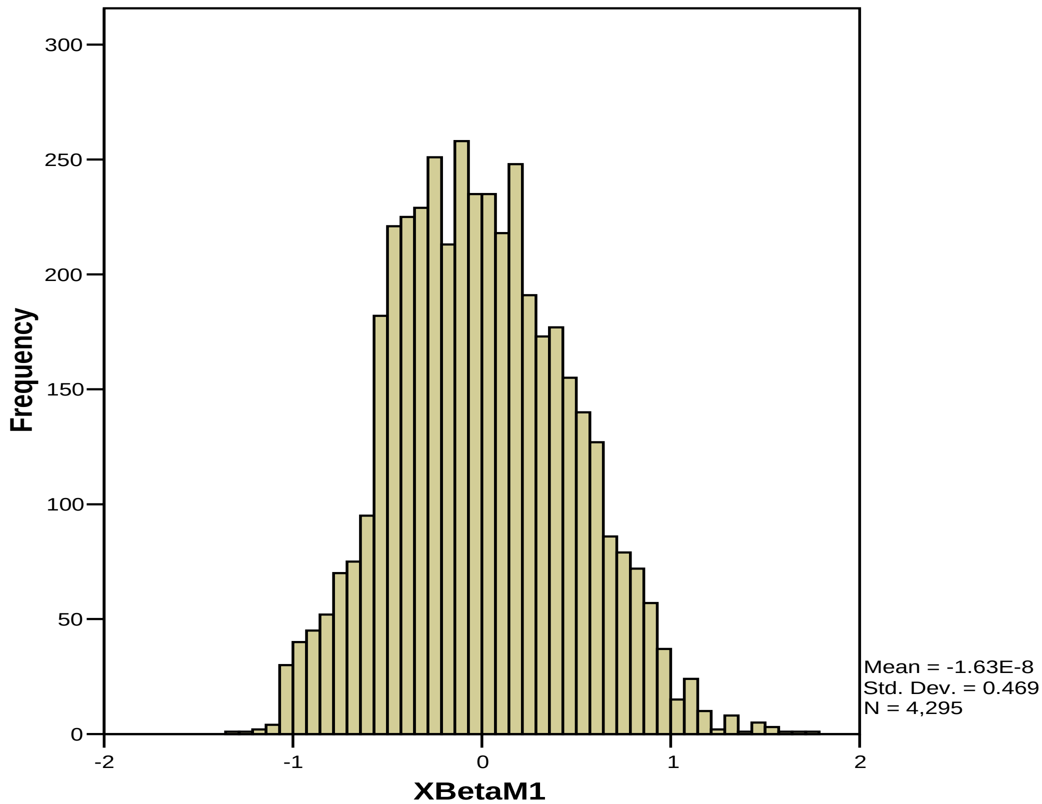

***Figure 2*. Distribution of XBetaM1 scores for the CYF sample.**

The XBetaM1 score provides a risk measure in that the higher the XBetaM1 score the more likely it is that the young person will receive a second YJ Intake. The overall predictive accuracy of the model was assessed by creating a ROC curve using YJPost as the criterion variable and XBetaM1 as the test variable. Figure 3 shows the ROC curve.

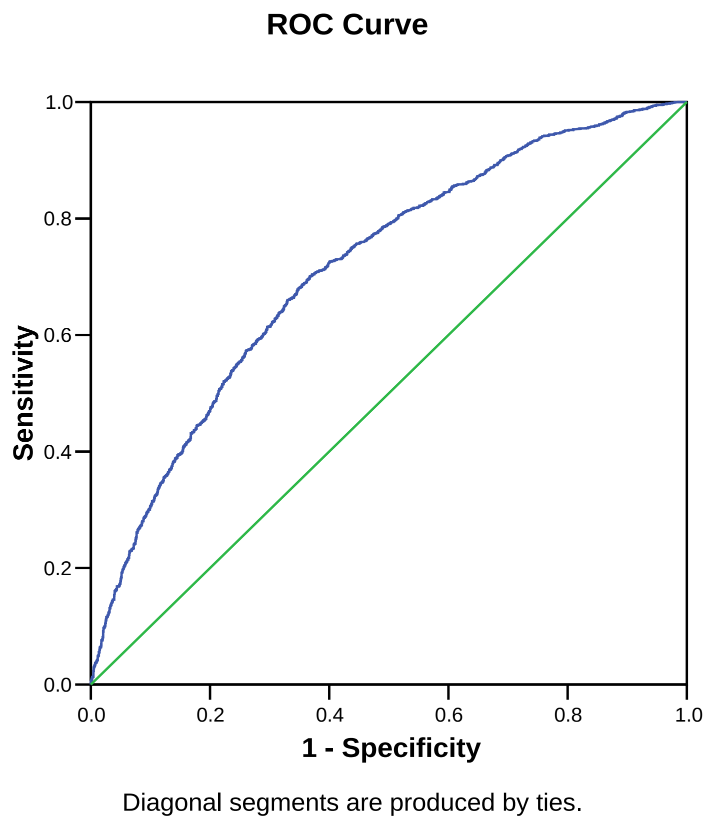

Figure 3. **ROC curve for M1 predicting subsequent YJ Intake for the CYF 2002 sample.**

The AUC value for M1 was 0.713 indicating a moderately good level of predictive performance.

Model 2: Predicting Offending (AnyProsecutionOrYJPost) for a Stratified Sample of CYF 2002 Cases.

Two random stratified samples of 500 cases were obtained from the NIA database. The first was collected in January 2005 and was termed the developmental sample because it was used in the initial specification of the models. The second sample of 500 was collected in January 2006 and was termed the validation sample because it was used to test the predictive accuracy of the models. Table 11 displays descriptive statistics for the developmental sample in terms of predictive and outcome variables.

Table 11.

Descriptive Statistics from NIA database for a stratified random sample of CYF 2002 cases.

	N	Minimum	Maximum	Mean	Std. Deviation
NumPriorProsecutions	500	0	90	1.43	6.275
NumPriorRecords	500	0	170	10.00	15.277
NumPriorChargeSheets	500	0	18	1.23	2.201
NumPriorIntelligence	500	0	86	2.94	6.928
NumPriorOccurrence	500	0	33	3.00	3.934
NumPriorYouthAid	500	0	16	1.31	2.465
NumProsecFollowUp	500	0	61	3.88	7.430
MaxSeverity	500	0	5	2.28	2.125
TotalSeverity	500	0	183	13.35	21.670
ConvictOrYJPost?	500	0	1	.70	.459
Valid N (listwise)	500				

Some NIA variables were recoded in the same fashion and for the same reasons as discussed above in relation to the CYF variables. The new values for the recoded NIA variables can be found in Appendix A.

Figure 4 displays a survival plot for the NIA developmental sample using the Kaplan-Meier method and any prosecution or YJ Intake during the follow-up period as the measure of reoffending.

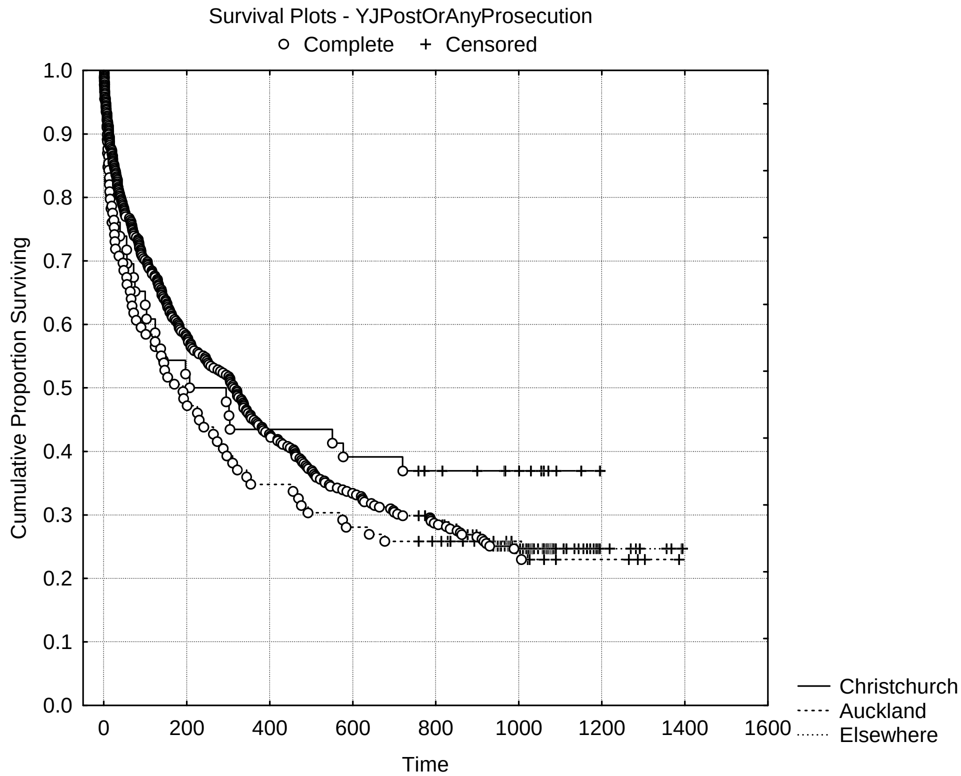

Figure 4. **Survival functions (cumulative proportion of cases with no prosecution or additional YJ intake) for the NIA developmental sample shown separately for Christchurch, Auckland and the rest of the country.**

The time at large (either time to conviction or follow-up time) averaged across groups was 429 days. The difference in survival rate was not statistically significant between Christchurch and Auckland (Gehan's Wilcoxon = .74, $p > .10$) or Christchurch and the rest of the country (Gehan's Wilcoxon = -.46, $p > .10$) but the difference was statistically different between Auckland and the rest of the country (Gehan's Wilcoxon = 2.0, $p < .05$).

Cox regression was used to analyse the NIA developmental sample and create the model (M2). Different combinations of predictor variables were considered in a series of exploratory analyses, and ultimately the variables XBetaM1, NumPriorIntelligence, and NumPriorOccurrence were found to be the combination with the best performance.

Hierarchical Cox regression analyses showed that XBetaM1, NumPriorIntelligence and NumPrior Occurrence made independent contributions to risk prediction. The model with these variables was significant χ^2 (df = 3) = 93.49, $p < .001$. Table 12 shows the predictor variables for M2 and their associated coefficients and level of statistical significance.

Table 12.

Predictor Variables for M2 Based on NIA Developmental Sample With YJPostOrAnyProsecution as the Outcome Measure. B = coefficient, SE = standard error, Wald = Wald statistic, df = degrees of freedom, Sig = obtained significance level, Exp(B) = exponential function of coefficient

	B	SE	Wald	df	Sig.	Exp(B)
XBetaM1	.704	.124	32.326	1	.000	2.022
NumPriorIntelligence	.129	.033	14.896	1	.000	1.137
NumPriorOccurrence	.063	.030	4.450	1	.035	1.065

Figure 5 shows the distribution of XBeta scores generated with the M2 model (XBetaM2) which is similar to that of XBetaM1 scores and is approximately normal with a slight positive skew.

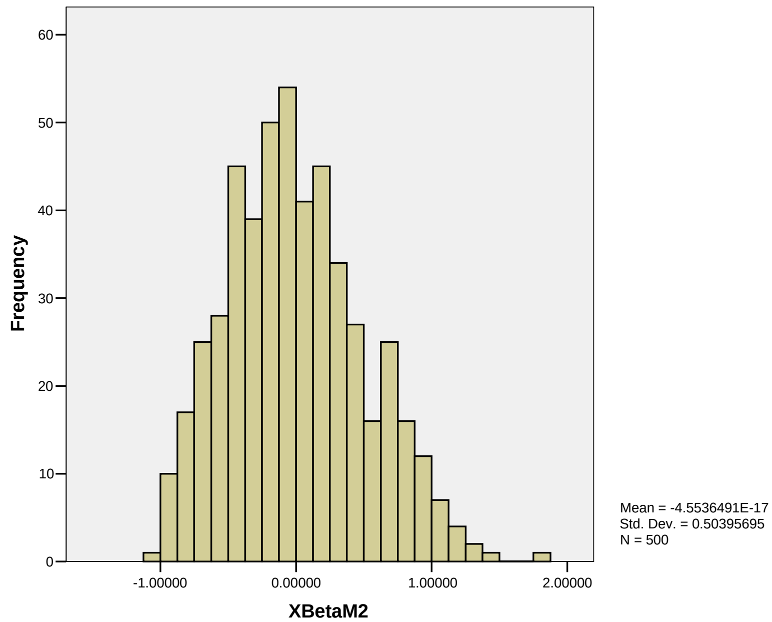

Figure 5. **Distribution of XBetaM2 scores for the NIA developmental sample.**

ROC analysis was used to assess the overall predictive accuracy of M2. The AUC was .73 which corresponds to a reasonably good level of predictive accuracy. M2 combined information from CYF and the police database to provide a measure of risk. These two different sources of information provided independent contributions to the performance of the model as was shown in the Hierarchical Cox regression analysis described above. In other words, using both sources of information is better than using either one on its own.

Model 3: Predicting Serious Offending (ProsecutionSerious?) for a Stratified Random Sample of CYF 2002 Cases

The analysis for developing M3 was the same as that used for M2 except that any serious prosecution during the follow-up period was used as the outcome variable instead of any prosecution or YJ intake. A serious prosecution was defined as a prosecution receiving a rating of 5 on the scale described above (Maxwell et al., 2002). Figure 6 displays the survival plots for the 3 regions. The difference in survival rate was not significantly different between Christchurch and Auckland (Gehan's Wilcoxon = .16, $p > .10$) or Christchurch and the rest of the country (Gehan's Wilcoxon = .98, $p > .10$) or Auckland and the rest of the country (Gehan's Wilcoxon = 1.1, $p > .10$) indicating similar rates of offending across the country.

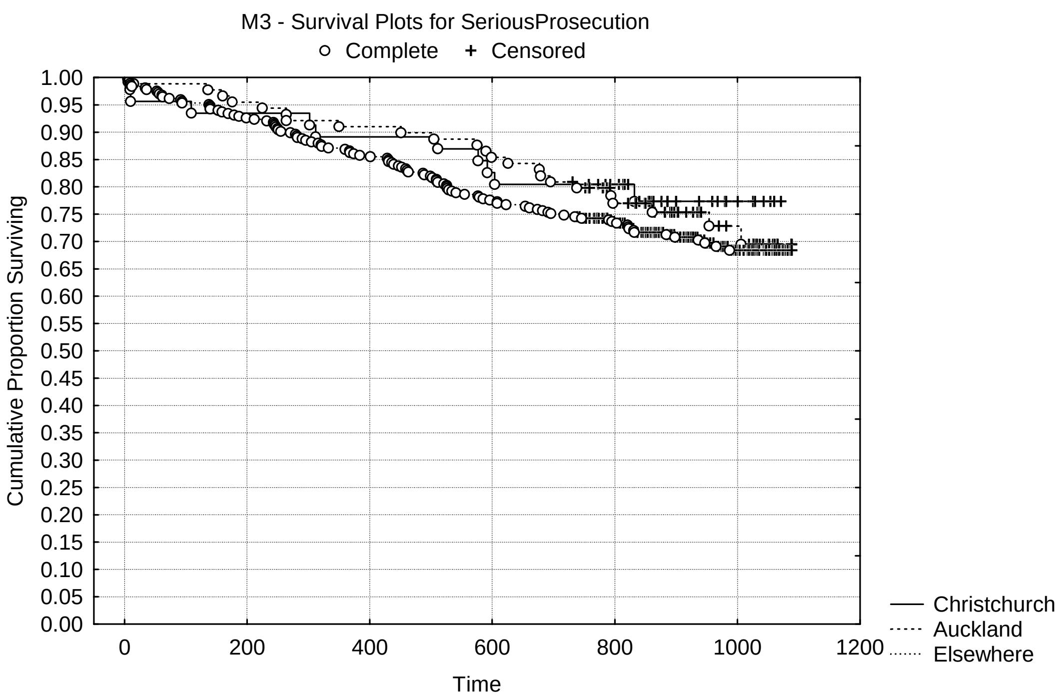

***Figure 6.* Survival functions (cumulative proportion of cases with no prosecution for a serious offence in the follow-up period) for the NIA developmental sample, shown separately for Christchurch, Auckland and the rest of the country.**

Cox regression was used to develop M3 with SeriousProsecution? as the outcome variable. As for the development of M2, different combinations of predictor variables were considered and tested for predictive accuracy. The resulting model with the best predictive accuracy included the predictor variables XBetaM1, NumPriorIntelligence, and NumPriorOccurrence – the same predictor variables that were used in M2. The overall performance of the model was statistically significant χ^2 (df = 3) = 70.55, $p < .001$. Table 13 shows the predictor variables for M3 and their associated coefficients and level of statistical significance.

Table 13.

Predictor Variables for M3 Based on the NIA Developmental Sample With SeriousProsecution? as the Outcome Measure. B = coefficient, SE = standard error, Wald = Wald statistic, df = degrees of freedom, Sig = obtained significance level, Exp(B) = exponential function of coefficient

	B	SE	Wald	df	Sig.	Exp(B)
XBetaM1	.823	.200	16.931	1	.000	2.277
NumPriorIntelligence	.172	.051	11.412	1	.001	1.187
NumPriorOccurrence	.110	.045	5.999	1	.014	1.116

Figure 7 displays the distribution of XBetaM3 scores which is similar to that of XBetaM2 and M1. They are all approximately normally distributed which indicates a reasonable degree of natural variation in risk within the population.

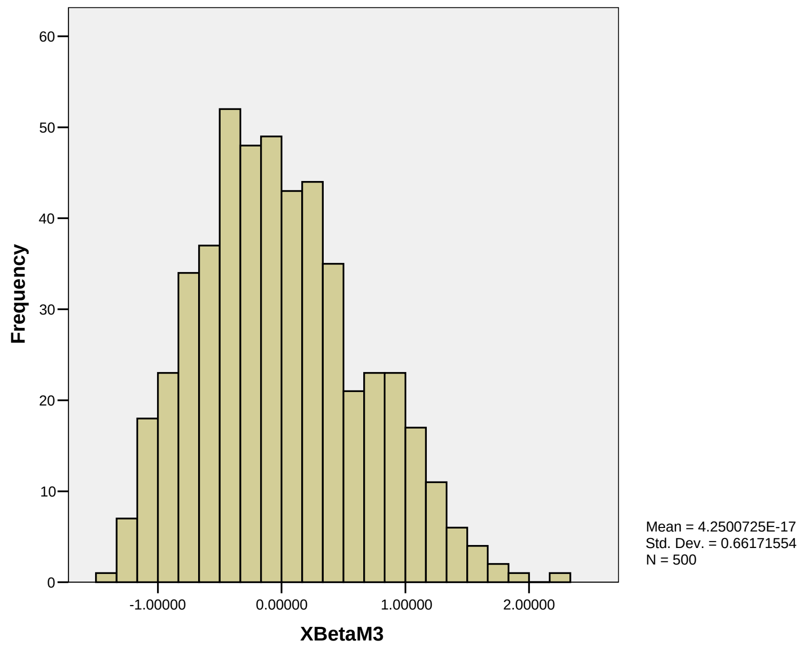

Figure 7. **Distribution of XBetaM3 scores for the NIA developmental sample.**

ROC analysis was used to measure the overall predictive accuracy of M3. An AUC value of 0.723 was achieved which indicates a moderately good level of predictive accuracy.

Comparison of the Developmental and Validation Samples.

Before testing the predictive accuracy of M2 and M3 on the validation sample it was important to compare the Developmental and Validation samples to see whether there were differences in predictive and outcome variables across the two samples. In order to assess this t tests and Chi squares (for continuous and categorical variables, respectively) were calculated and displayed in Tables 14, 15 and 16.

Table 14.

Results of t *Tests Comparing the Developmental and Validation Samples on Predictor Variables.*

	Dev Mean	Valid Mean	*t*	*df*	*P*
AgeAtFirstIntake	12.61	12.75	-0.58	998	0.56
AgeAtFirstYJ	15.25	15.28	-0.46	998	0.64
NumPriorCP	1.25	1.27	-0.23	998	0.82
NumFindingsSexAbuse	0.11	0.10	0.50	998	0.62
NumPCourtDates	0.33	0.37	-0.53	998	0.59
NumPriorIntelligence	1.69	1.83	-1.22	998	0.22
NumPriorOccurrence	2.28	2.67	-3.01	998	0.00

Table 15.

Results of Chi Square Tests Comparing Dichotomous Predictor and Outcome Variables for the Developmental and Validation Samples.

	Dev %	Valid %	χ^2	*df*	*p*
%Male	80%	79%	.06	1	0.88
%MaoriPacific	50%	46%	2.12	1	0.15
YJPost	54%	52%	0.26	1	0.66
YJPostOrAnyProsecution	74%	75%	0.19	1	0.72
SeriousProsecution	28%	31%	1.22	1	0.30

Table 16.

Results for t *Tests Comparing the Developmental and Validation Samples on XBeta M1, M2 and M3 Values.*

	Dev Mean	Valid Mean	*t*	*df*	*p*
XBetaM1	0.02	-0.01	0.90	998	0.37
XBetaM2	0.00	0.04	-1.03	998	0.30
XBetaM3	0.00	0.02	-0.72	998	0.47

As can be seen from Tables 14 and 15 there were no significant differences between the Developmental and Validation samples on any of the predictor variables or outcome variables apart from NumPriorOccurrence. This difference is possibly a sampling artefact given the number of comparisons, and overall the values are very similar across the two groups. Therefore it was concluded that the Developmental and Validation samples were highly similar which was to be expected given that they were both random samples from the same population. The XBeta scores were computed for the Validation sample using the coefficients estimated for M2 and M3 from the Developmental sample. The equation XBetaM2 = .704*(XBetaM1 - .019) + .129*(NumPriorIntelligence - 1.694) + .063*(NumPriorOccurrence - 2.284) was used to develop the XBetaM2 scores, and the equation XBetaM3 = .823*(XBetaM1 - .019) + .172*(NumPriorIntelligence - 1.694) + .11*(NumPriorOccurrence - 2.284)was used to calculate the XBetaM3 scores. The coefficients are multiplied by the centred predictor variables (i.e., predictor variable – the average value of the predictor variable in the Developmental sample). As can be seen from Table 16 the XBeta scores were all close to zero and there were no significant differences between the Developmental and Validation samples.

Survival plots were calculated comparing the two samples on each of the three measures of reoffending and are shown in Figures 8, 9 and 10 below.

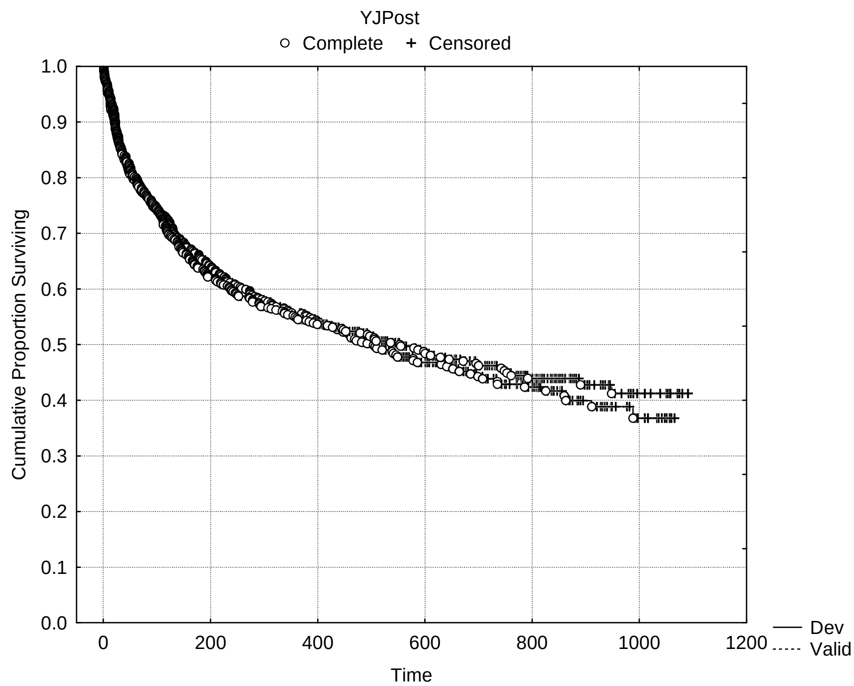

Figure 8. **Survival functions (cumulative proportion of cases with no additional YJ Intake in the follow-up period) for the NIA developmental and validation samples.**

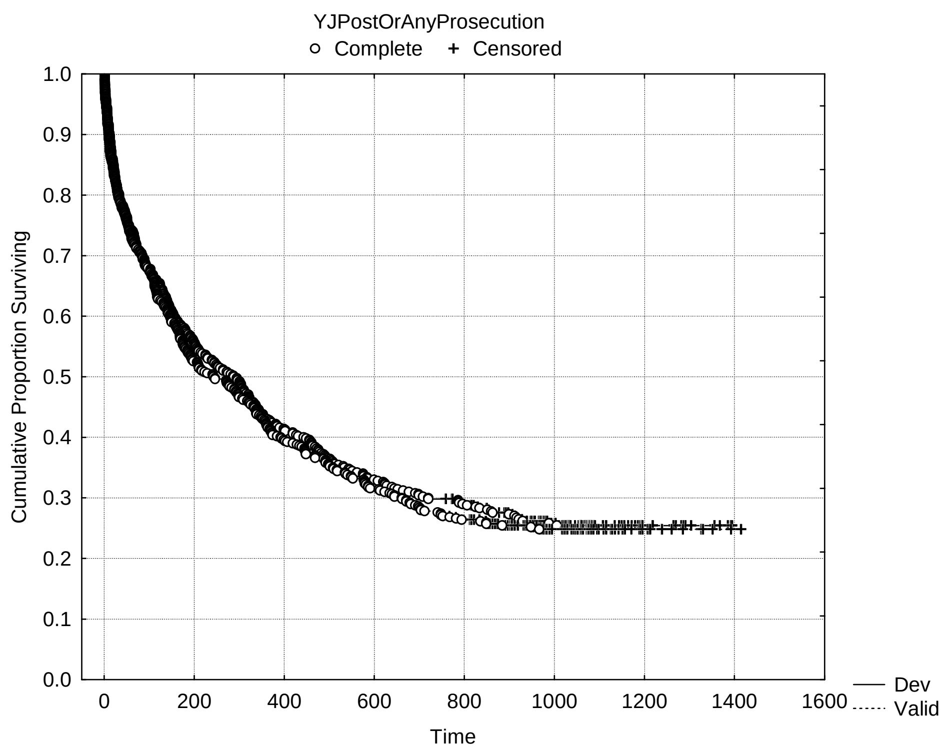

Figure 9. **Survival functions (cumulative proportion of cases with no additional YJ Intake or prosecution in the follow-up period) for the NIA developmental and validation samples.**

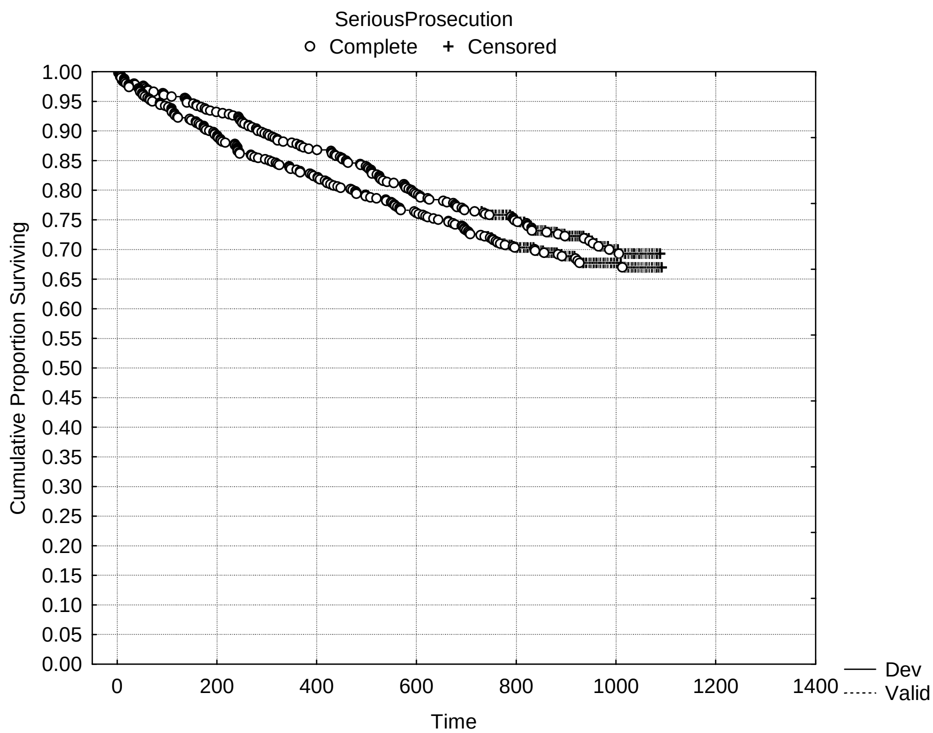

Figure 10. **Survival functions (cumulative proportion of cases with no prosecution for a serious offence in the follow-up period) for the NIA developmental and validation samples.**

As can be seen from Figure 8 the difference in survival rates when the outcome measure was YJPost between the developmental and validation sample was not significantly different (Gehan's Wilcoxon = -.07, p >.10). The difference was also not significantly different when the outcome measure was AnyProsecutionOrYJPost (Gehan's Wilcoxon = .44, p > .01) as can be seen from Figure 9. The Validation sample offended at a slightly higher rate when the outcome measure was SeriousProsecution as can be seen from Figure 10, but this difference was not statistically significant (Gehan's Wilcoxon = 1.61, p > .10). In conclusion the two samples had very similar rates of reoffending across all three outcome measures of recidivism.

Cross-Validation Analysis.

ROC analysis was used to test the predictive validity of using XBetaM2 with AnyProsecutionOrYJPost as the measure of reoffending for the validation sample. The same was done using XBetaM3 with SeriousProsecution as the measure of reoffending. This was carried out in order to see how well the model could generalize to explain other data sets. The AUC was 0.74 for XBetaM2 and 0.68 for XBetaM3 for the validation sample. Both of these values are very similar to those obtained with the developmental sample which were 0.73 for XBetaM2 and 0.72 for XBetaM3. Therefore it was concluded that M2 and M3 were valid as measures of relative risk in the validation sample.

Further analysis was carried out to provide more evidence of the generalizability of the model to predict recidivism for new data. This analysis consisted of comparing predicted and obtained survival curves for the lower and upper risk quintiles of the validation sample for M2 and M3. XBetaM2 and M3 scores for cases below the 20th percentile were designated to the lower risk quintile and scores for cases above the 80th percentile were designated to the upper risk quintile. Average XBeta scores were calculated for each quintile and are presented in Table 17 below.

Table 17.

Average XBetaM2 and M3 Scores for Quintiles.

	20th Percentile Score	80th Percentile Score	Average XBeta Score in Quintile	
			Lower	Upper
M2	-0.47	0.47	-0.63	0.85
M3	-0.62	0.66	-0.82	1.14

Average XBeta scores were used to calculate predicted survival curves for both quintiles and the Kaplan-Meier method was used to generate obtained survival curves for both quintiles. Figures 11 and 12 show the predicted and obtained survival curves for the lower and upper quintiles for M2.

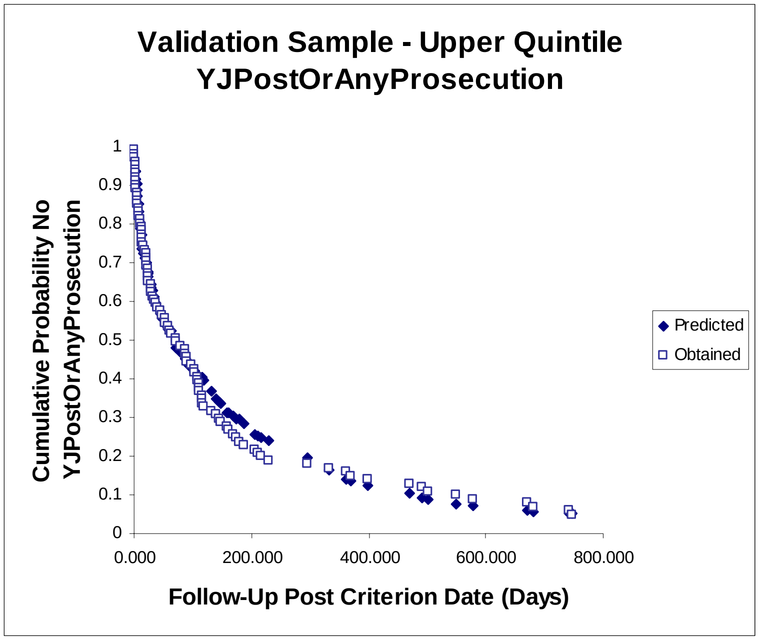

Figure 11. **Predicted and obtained survival curves for the upper quintile for M2.**

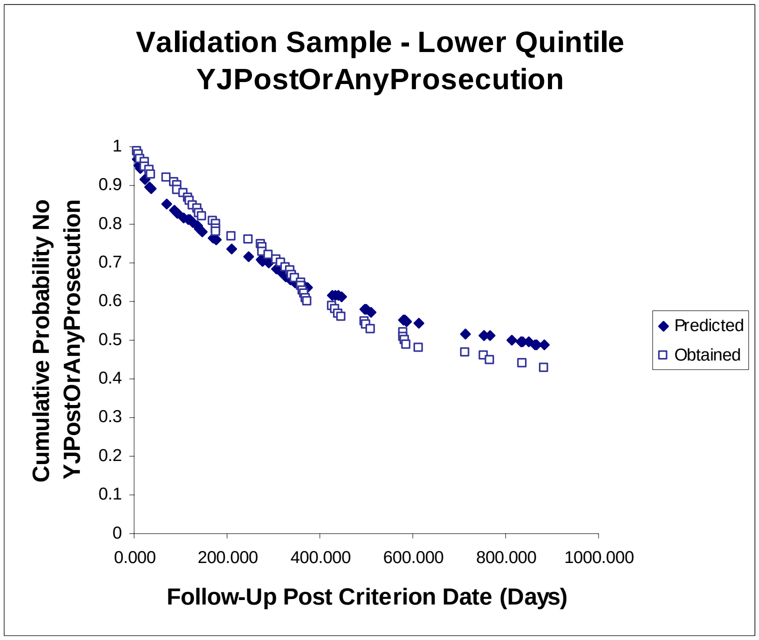

Figure 12. **Predicted and obtained survival curves for the lower quintile for M2.**

The figures above show that for the lower and upper risk quintiles the predicted and obtained survival curves are very similar for M2. It can also be seen from the figure that there was a large difference between the upper and lower quintiles in terms of the rate of reoffending with approximately 25% of the upper quintile having not reoffended after 200 days compared to approximately 75% of the lower quintile.

The same analysis was carried out for M3 and is shown in Figure 13 and 14 below.

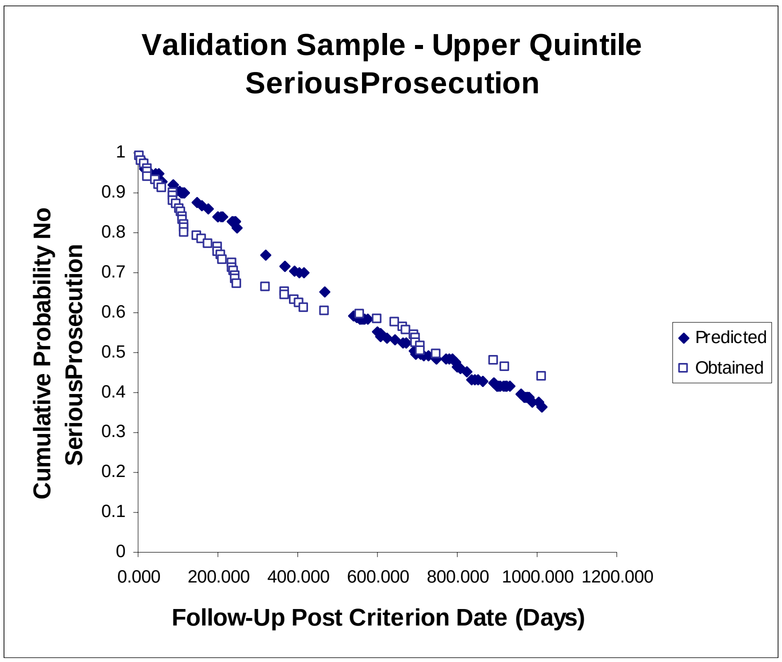

Figure 13. **Predicted and obtained survival curves for the upper quintile for M3.**

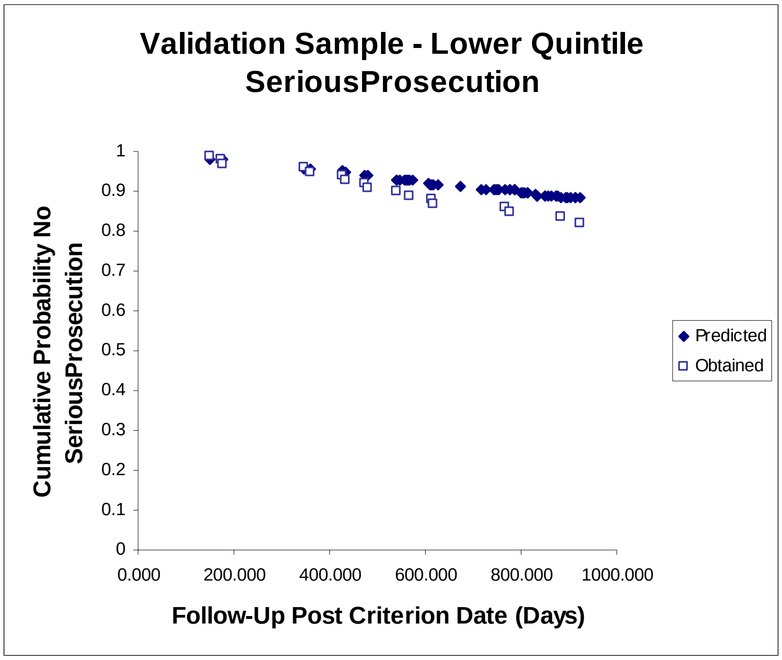

Figure 14. **Predicted and obtained survival curves for the lower quintile for M3.**

The prediction curves for M3 are very similar to the obtained survival curves, however, they do underestimate the rate of reoffending in the lower quintile slightly. This is probably due to the low base rate of serious offending in this group. As there was with M2, there was a large difference in the rate of reoffending between the upper and lower quintiles for M3.

The analysis confirmed that M2 and M3 performed well when calculating predictions of survival curves for groups with different risk levels. Overall the cross-validation analysis provided strong evidence for the generalizability of M2 and M3 to new data.

Discussion

The aim of the present study was to develop a model that predicted recidivism based on a sample of young people from New Zealand who had contact with CYF in 2002. The first model (M1) used information obtained from the CYF CYRAS database. The population consisted of all those young people who had a Youth Justice intake in 2002. Recidivism was defined as the occurrence of a subsequent YJ intake prior to December 2004. Demographic and historical variables such as age at first YJ intake, ethnicity, gender, and number of prior placements were included in the model. M1 also accounted for regional differences in the predictor variables.

The second (M2) and third (M3) models combined information from CYF in the form of inclusion of M1 within the models, and information from the police NIA database. Information like police contacts and arrests or prosecutions was obtained from NIA for a random stratified sample of 500 young people from the CYF 2002 sample. The predictor variables included in both of the models were M1, number of intelligence notings, and number of police contacts (occurrences). A composite measure, the occurrence of a subsequent YJ intake or prosecution for any offence, was used as the definition of recidivism for M2. For M3 the occurrence of a subsequent prosecution for a serious offence was used as the definition of recidivism. ROC analysis was used to assess the performance of the models. An additional random stratified sample of 500 young people from the CYF 2002 sample was used as a validation sample. The performance of M2 and M3 was assessed with this sample using ROC analysis. To further illustrate the validity of the model the upper and lower quintiles of the validation sample were computed and the obtained and predicted survival curves for each quintile were compared.

Summary of Findings

The descriptive statistics of the CYF 2002 sample have provided some important findings. The analysis has shown that we can predict, to a degree, who will reoffend in a population of first time offenders. Therefore, there is a pattern to reoffending and this pattern can be assessed. There are certain variables that are associated with recidivism and those young people that possess or score highly on these variables are more likely to reoffend than those that do not possess or score highly on the variables. As can be seen from the survival plots above, the rate of reoffending for these young people is relatively high across all three measures of reoffending (subsequent YJ Intakes, general prosecutions and YJ Intakes, and prosecutions for serious offences). Approximately 50% of the sample went on to have a secondary YJ Intake within about one year. However, this figure varied across the regions with the highest rate of reoffending in Christchurch, followed by Auckland and the lowest rate of reoffending in the rest of the country. Approximately 65% go on to have a subsequent prosecution in the next 2.5 years and approximately 30% go on to have a subsequent prosecution for a serious offence. The differences across regions possibly reflect the different ways youth offenders are dealt with in different areas of New Zealand. In Christchurch young people have earlier and more frequent CYF contact compared to young people in the other two regions. This may be due to the lower population density in Christchurch and therefore, the higher level of resources able to be allocated to these young people.

The overall performance of M1 was statistically significant for each of the three regions. The model had a high level of predictive validity. The performance of both M2 and M3 was also statistically significant. The predictor variables that were found to be significant and included in the models varied across regions. Being of Maori or Pacific Island ethnicity, younger at your first intake into CYF and younger at your first Youth Justice contact was significantly associated with recidivism in each region. However, males were more likely than females to reoffend in Auckland and the rest of the country but not in Christchurch. In Christchurch the number of findings recorded in the CYRAS database of sexual abuse was associated with increased risk of recidivism, but findings of sexual abuse was not a significant factor for the other two regions. The number of prior court dates and care and protection intakes was significantly associated with recidivism in the rest of the country but not in Auckland or Christchurch. The Xbeta scores generated from M1 were included in M2 and M3 and this meant that the variation across regions was incorporated within these two models. For both M2 and M3 having a higher number of intelligence notings and police contacts (occurrences) from the NIA database increased the likelihood of recidivism, as did obtaining a high XbetaM1 score.

The predictor variables that were found to be significant and therefore included in the models were consistent with previous research on risk factors for the development of offending in young people. For example, a strong finding to come out of prior research is that past offending significantly predicts future offending (Benda et al., 2001; Sheldrick, 1999). This finding was replicated in the present study with the number of prior court dates and police contacts increasing the likelihood of recidivism. Past research has also found that the younger a person is when a number of events occur including using alcohol and drugs, having their first contact with Youth Justice, and displaying physically aggressive behaviour, the more likely they are to subsequently reoffend (Benda et al., 2001; Broidy et al., 2003; Ge et al., 2001). This was replicated in the present study with the finding that the younger a person was at age of first Youth Justice contact and CYF intake the more likely they were to reoffend.

All three models performed well as measured by ROC curves. The test for M2 and M3 was how well they performed with the validation sample. This sample consisted of a completely different group of young people compared with the group that the models were developed on. The developmental and validation sample were compared to ascertain any differences between them. The only statistically significant difference was the number of prior police contacts (occurrences) with the validation sample having a higher average number of police contacts. Generally the two samples were very similar across all predictor and outcome variables. Xbeta scores were calculated for the validation sample. The survival curves across the three measures of reoffending were also very similar for the developmental and validation samples.

The ROC analysis showed that M2 and M3 performed well when predicting recidivism for the validation sample. The AUC values were very similar for the developmental and validation samples. They were also very similar to those obtained in studies validating other risk assessment measures (Catchpole & Gretton, 2003; Corrado et al., 2004). By calculating the upper and lower quintiles for the validation sample comparisons could be carried out to assess how well the models performed for groups with different risk levels. Predicted and obtained survival curves were calculated for the two quintiles. The survival curves were very similar for both M2 and M3. This demonstrated that M2 and M3 performed well when predicting recidivism for a new data set. These analyses used XbetaM2 and XbetaM3 scores calculated on the basis of parameters estimated from the developmental sample, therefore they were a priori predictions and provide strong evidence of the models generalizability. From this analyses it can be concluded that the relationships between the variables included in M2 and M3, and reoffending are genuine and not due to chance in the developmental sample.

Methodological Considerations

It is important to keep in mind that the models were developed on a population of young people aged between 14 and 18. Therefore, it must not be assumed that they would have the same predictive validity if used with a population of people of a different age group. Also the study was carried out on a youth offender population in New Zealand and may not generalize well to other countries and cultures. Further study would need to be carried out to assess the models predictive validity in other settings before it could be used officially as a measure of risk of recidivism.

The present study examined risk factors that were readily available from the CYF and police NIA databases. This unfortunately limited the choice of risk factors that could be examined. There are other risk factors that have been identified in the literature that could also be associated with recidivism and their inclusion within the models could improve the level of predictive performance. Some of these risk factors include psychopathy and a diagnosis of Attention Deficit Hyperactivity Disorder or Conduct Disorder. Findings about these risk factors are discussed in the introduction.

The present study also only examined risk factors that were rooted in the immediate past of the young person and other important variables that contributed to the development of delinquency may be evident by delving further back into the young persons childhood. For example family dynamics has been found to be linked with the development of offending (Haines & Case, 2005) and this was not touched on in the present study.

It is important to keep in mind that the statistical model developed in the present study is not a risk measure that can be used with all young people as a risk screening instrument. It was developed on a population of young people who had already had one YJ Intake, therefore the model is a measure of risk of future recidivism rather than a measure of risk of offending in general.

The M3 model that assessed recidivism measured as a subsequent prosecution for a serious offence contained the same predictor variables as the M2 model that measured general recidivism. It is likely that there are risk factors specifically associated with violent recidivism that were not considered to be included in M3. Some possibilities are serious traffic offences like Drunk In Charge, or Reckless Driving. Future research could investigate alternative risk factors to develop a model that performs at a higher level when predicting violent recidivism.

It was mentioned in the method section that it was important to create a model that performed at a high level but was also generalizable to new data sets other than the population it was developed on. It is possible that in trying to keep the models as simple as possible to increase the level of generalizability, risk factors that may be important were left out. In other words, it is possible that the models in the present study were too simple. Future research could investigate the inclusion of other risk factors to improve the performance of the models.

The models were developed on data obtained from the CYF and police NIA databases therefore the models are only as good as the information provided in the databases. Accuracy and authenticity of the information provided in the databases in taken for granted. It is possible that some information was inaccurate and this is important to keep in mind when carrying out any type of archival research. A more accurate method of data collection would be to interview the young offenders as part of the study itself and therefore ask them for more specific information that is related to the study. However, this method is very time consuming and would limit the number of young people involved in the study. By carrying out archival research it is possible to study a very large population and this is the optimal situation for developing statistical models.

Implications of Study and Ideas for Future Research

The literature that examines the use of actuarial models for predicting risk is predominantly focussed on adult offender populations. The present study has demonstrated the feasibility of actuarial models within the youth offender population. The developmental models discussed in the introduction, including Moffit's (1993) theory about adolescent-limited and life-course persistent offenders, imply that young people follow this developmental trajectory from a young age. A risk measure, like the one developed in the present study, can be used to identify the life-course persistent offenders at a young age and therefore divert them from this criminal path. This approach of early intervention is an alternative to the current model of punishment employed in most westernized countries and involves changing the justice systems attitudes toward youth offenders. The development of reliable actuarial models that measure risk level in young offenders makes this alternative more feasible and hopefully more appealing to government policy makers and decision makers within the justice system.

The present study also demonstrated what can be achieved when information is pooled from different organisations. The combination of data from the police NIA database and the CYF database made the creation of the models possible. The positive results that have come out of the present study advocate for more of this type of research to take place and also for more formal collaboration between organisations to be instigated. When information is shared between organisations it can only improve the level of service that they offer which has positive implications for all involved, including the young offenders themselves.

There are several ways in which the model developed in the present study could be used in a practical setting. The Xbeta scores would need to be split up into categories to show the level of risk associated with the score. For example, four groups could be used, namely low, medium, high and severe risk. Most actuarial measures provide a risk level score which falls into a category. Cut-off scores could be established to define what scores fall within each risk level group. An organisation like CYF could enter information obtained from young people that are referred to them into the model and therefore generate a risk level score that would help them ascertain how likely it was that the young person would go on to reoffend in the future. This could help them decide what, if any, intervention to implement and how intensive the intervention should be. The model could also be used by the Youth Justice system to help make decisions about what consequences are appropriate for an offence committed by a young person with those young people achieving high scores on the risk measure receiving more intensive intervention.

One drawback of the statistical models developed in the present study is that no dynamic risk factors are considered within them, the models are solely made up of static factors. As was discussed in the introduction, assessing dynamic risk factors provides a guide about where to intervene with the young person because dynamic risk factors are changeable. Future research could further develop the model and use the same analyses to select dynamic risk factors that add predictive validity to the models. With the addition of dynamic risk factors the models would not only tell the user how likely it is that the young person in question would reoffend but would also help them ascertain how to intervene with that person. Some examples of dynamic risk factors that could be assessed include association with antisocial peers, alcohol and drug use, impulsiveness, negative mood, and features of interpersonal relationships. All of these dynamic risk factors have been identified in the literature as promising links to recidivism (Douglas & Skeem, 2005).

Another interesting area that could be examined in future studies are the differences in criminal behaviour and the risk factors that contribute to the development of criminal behaviour between males and females. Past research has found that there may be constructs that are risk factors for females and not males and vice versa. For example, child abuse and running away from home have been found to be significantly associated with juvenile offending for females but not for males (Funk, 1999). It is important that we are aware of these differences, if they exist, within a New Zealand youth offender population. As is outlined in a recent report put out by the Ministry of Justice, there is growing concern amongst such agencies as the police that young females in New Zealand are committing increasingly violent, and more serious criminal acts

(Carruthers, 2002). The number of young females apprehended for a crime has risen from 96 per 10,000 in the population in 1994 to 108 per 10,000 in the population in 2000. The rate of apprehensions for males is very similar across 1994 to 2000. However, it is difficult to comment on the significance of these numbers given that there are a comparatively small number of females committing crimes.

Another area that is important to research within a New Zealand population is the difference in risk factors and criminal patterns within the Maori and Pacific Island population compared to the remaining population of youth offenders. The present study found that being of Maori or Pacific Island ethnicity increased the probability of recidivism and ethnicity was incorporated within M1. The Ministry of Justice reported that approximately half of all those young people who are apprehended by the police, involved in Family Group Conferences, and prosecuted before the courts are of Maori descent (Carruthers, 2002). The percentage varies across regions and this was reflected within the present study with being of Maori or Pacific Island descent being more highly associated with recidivism in Auckland than in the rest of the country. This trend is also evident from examinations of the adult prison population. The proportion of Maori under the age of 17 in 1996 was 24%, therefore it is evident that they are significantly overrepresented within the youth offender population. This problem is set to increase in the future with an increase in the numbers of young people in New Zealand. Specifically, the number of Maori people is set to rise to approximately 27% of the whole population by 2016. This is an increase of 3% from 1996.

The Ministry of Justice also reports that although young Pacific Island people are not overrepresented in the general youth offender population in New Zealand they are overrepresented in terms of the number of serious or violent offences they commit (Carruthers, 2002). They are also more likely to commit a serious or violent offence as their first offence and then may not go on to reoffend. The proportion of people of Pacific Island descent under the age of 17 in New Zealand is also set to rise by 3% to 13% by 2016.

These figures call for more research that examines the question of why so many young people of Maori ethnicity are among the youth offending population. Research specifically looking at young Maori people and detecting important risk factors for this population is important for the future. Actuarial models that measure risk could be developed separately for this population and therefore capture more specifically what is occurring for young Maori people that makes them more prone to engage in delinquency. The same could occur for the population of young female offenders in order to understand what is specifically occurring for them that leads them into criminal activity.

Conclusion

The present study demonstrated the development of a model that predicted recidivism in a group of young offenders. The model was validated using a separate sample and was found to generalize well to a new data set. This study represents an important step towards understanding the nature of the youth offending population in New Zealand and continues the research that has been carried out on youth offenders in other countries. The models demonstrated that it is predictable to an extent who will go on to reoffend and this information can be used in a number of ways, including aiding decisions about what interventions to put in place and what consequences should occur for the young person in question following a criminal act.

References

Andrews, D., & Bonta, J. (1994). *The Psychology of Criminal Conduct*. Cincinnati, OH: Anderson Publishing Company.

Ashford, J. B., & LeCroy, C. W. (1990). Juvenile recidivism: A comparison of three prediction instruments. *Adolescence, 25*(98), 441-450.

Baumann, D. J., Law, J. R., Sheets, J., Reid, G., & Graham, J. C. (2005). Evaluating the effectiveness of actuarial risk assessment models. *Children and Youth Services Review, 27*(5), 465-490.

Becroft, A. (2003, August 2003). *Youth Offending: Putting the Headlines in Context*. Retrieved January, 2006, from http://www.justice.govt.nz/youth/media/rates0803.html

Benda, B. B., Flynn Corwyn, R., & Toombs, N. J. (2001). Recidivism among adolescent serious offenders: Prediction of entry into the correctional system for adults. *Criminal Justice and Behavior, 28*(5), 588-613.

Bloom, H., Webster, C., Hucker, S., & De Freitas, K. (2005). The Canadian Contribution to Violence Risk Assessment: History and Implications for Current Psychiatric Practice. *Canadian Journal of Psychiatry, 50*(1), 3-11.

Borum, R., Bartel, P., & Forth, A. (2002). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY). Consultation version.* Tampa: Florida Mental Health Institute, University of South Florida.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., et al. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology. Special Issue: Violent children, 39*(2), 222-245.

Carr, M. B., & Vandiver, T. A. (2001). Risk and protective factors among youth offenders. *Adolescence, 36*(143), 409-426.

Carruthers, D. (2002). *Youth Offending Strategy*. Retrieved January, 2006, from http://www.justice.govt.nz/pubs/reports/2002/youth-offending-strategy/youth-strategy/index-strategy.html

Catchpole, R. E. H., & Gretton, H. M. (2003). The Predictive Validity of Risk Assessment With Violent Young Offenders: A 1-Year Examination of Criminal Outcome. *Criminal Justice and Behavior, 30*(6), 688-708.

Chesney-Lind, M. (1997). *The female offender.* Thousand Oaks, CA: Sage.

Chung, I.-J., Hawkins, J. D., Gilchrist, L. D., Hill, K. G., & Nagin, D. S. (2002).

Identifying and predicting offending trajectories among poor children. *Social Service Review, 76*(4), 663-685.

Chung, I.-J., Hill, K. G., Hawkins, J. D., Gilchrist, L. D., & Nagin, D. S. (2002). Childhood predictors of offense trajectories. *Journal of Research in Crime and Delinquency, 39*(1), 60-90.

Cleves, M., A., Gould, W., W., & Guitierrez, R., G. (2002). *An Introduction to Survivial Analysis Using Stata*. Texas: Stata Corporation.

Corrado, R. R., Vincent, G. M., Hart, S. D., & Cohen, I. M. (2004). Predictive validity of the Psychopathy Checklist: Youth Version for general and violent recidivism. *Behavioral Sciences & the Law, 22*(1), 5-22.

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2004). Limitations in actuarial risk assessment of sexual offenders: A methodological note. *British Journal of Forensic Practice, 6*(1), 16-32.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668-1674.

Dolan, M., & Doyle, M. (2000). Violence risk prediction: Clinical and actuarial measures and the role of the Psychopathy Checklist. *British Journal of Psychiatry, 177*, 303-311.

Douglas, K. S., & Skeem, J. L. (2005). Violence Risk Assessment: Getting Specific About Being Dynamic. *Psychology, Public Policy, and Law, 11*(3), 347-383.

Espiritu, R. C., Huizinga, D., Crawford, A. M., & Loeber, R. (2001). Epidemiology of self-reported delinquency. In R. Loeber & D. P. Farrington (Eds.), *Child delinquents: Development, intervention and service needs.* (pp. 47-66). Thousand Oaks, CA: Sage.

Fuller, J., & Cowan, J. (1999). Risk assessment in a multi-disciplinary forensic setting: Clinical judgement revisited. *Journal of Forensic Psychiatry, 10*(2), 276-289.

Funk, S. J. (1999). Risk assessment for juveniles on probation: A focus on gender. *Criminal Justice and Behavior, 26*(1), 44-68.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illnesses. *Law and Human Behavior, 20*(1), 35-48.

Ge, X., Donnellan, M. B., & Wenk, E. (2001). The development of persistent criminal offending in males. *Criminal Justice and Behavior, 26*(5), 731-755.

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

Grace, R. C., & McLean, A. P. (2003). *Child Offenders: What do we know?* Christchurch: University of Canterbury.

Gretton, H. M., Hare, R. D., & Catchpole, R. E. H. (2004). Psychopathy and Offending From Adolescence to Adulthood: A 10-Year Follow-Up. *Journal of Consulting and Clinical Psychology, 72*(4), 636-645.

Haines, K., & Case, S. (2005). Promoting Prevention: Targeting Family-Based Risk and Protective Factors for Drug Use and Youth Offending in Swansea. *British Journal of Social Work, 35*(2), 169-187.

Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised (PCL-R)*. Toronto, Ontario: Multi-Health Systems.

Hare, R. D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior, 23*(1), 25-54.

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.

Hill, J. (2002). Biological, psychological and social processes in the conduct disorders. *Journal of Child Psychology and Psychiatry, 43*(1), 133-164.

Hodges, K. (1994). *Manual for the Child and Adolescent Functional Assessment Scale*. Ypsilati: Eastern Michigan University, Department of Psychology.

Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior. Special Issue: Risk and assessment in contemporary corrections, 29*(4), 380-396.

Hoge, R. D., & Andrews, D. (1995). *Australian adaptation of youth level of service/case management inventory*. North Tonawanda, NY: Multi-Health Systems Inc.

Hoge, R. D., & Andrews, D. (2002). *The Youth Level of Service/Case Management Inventory*. Toronto, ON, Canada: Multi-Health Systems.

Ilacqua, G. E., Coulson, G. E., Lombardo, D., & Nutbrown, V. (1999). Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. *Psychological Reports, 84*(3, Pt 2), 1214-1218.

Jung, S., & Rawana, E. P. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior, 26*(1), 69-89.

Kashani, J. H., Jones, M. R., Bumby, K. M., & Thomas, L. A. (1999). Youth violence: Psychosocial risk factors, treatment, prevention, and recommendations. *Journal of Emotional and Behavioral Disorders, 7*(4), 200-210.

Kroner, D. G. (2005). Issues in Violent Risk Assessment: Lessons Learned and Future Directions. *Journal of Interpersonal Violence, 20*(2), 231-235.

Krysik, J., & LeCroy, C. W. (2002). The empirical validation of an instrument to predict risk of recidivism among juvenile offenders. *Research on Social Work Practice, 12*(1), 71-81.

Leistico, A.-M. R., & Salekin, R. T. (2003). Testing the reliability and validity of the Risk, Sophistication-Maturity, and Treatment Amenability Instrument (RST-i): An assessment tool for juvenile offenders. *International Journal of Forensic Mental Health, 2*(2), 101-117.

Maxwell, G., Robertson, J., & Anderson, T. (2002). *Police Youth Diversion*. Wellington: Victoria University of Wellington.

Mazerolle, P. (1998). Gender, general strain, and delinquency: An empirical examination. *Justice Quarterly, 15,* 65-91.

Mills, J. F. (2005). Advances in the Assessment and Prediction of Interpersonal Violence. *Journal of Interpersonal Violence, 20*(2), 236-241.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674-701.

Paternoster, R., Dean, C. W., Piquero, A., Mazerolle, P., & Brame, R. (1997). Generality, continuity, and change in offending. *Journal of Quantitative Criminology, 13*(3), 231-266.

Pease, K., & Tseloni, A. (1996). Juvenile-adult differences in criminal justice: Evidence from the United Nations Crime Survey. *The Howard Journal, 35,* 40-60.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*(3), 472-491.

Putnins, A. L. (1999). *Secure Care Psychosocial Screening: Manual for SECAPS v.4.* Adelaide: Department of Human Services.

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

Risler, E. A., Sutphen, R., & Shields, J. (2000). Preliminary validation of the juvenile First Offender Risk Assessment Index. *Research on Social Work Practice, 10*(1), 111-126.

Roosa, M. W., Deng, S., Ryu, E., Burrell, G. L., Tein, J.-Y., Jones, S., et al. (2005). Family and Child Characteristics Linking Neighborhood Context and Child Externalizing Behavior. *Journal of Marriage and Family*, *67*(2), 515-529.

Rutter, M., Giller, H., & Hagell, A. (1998). *Antisocial behavior by young people*. Cambridge: Cambrifge University Press.

Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist--Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, *3*(3), 203-215.

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

Sheldrick, C. (1999). Practitioner review: The assessment and management of risk in adolescents. *Journal of Child Psychology and Psychiatry*, *40*(4), 507-518.

Snyder, H. N. (2001). Epidemiology of official offending. In R. Loeber & D. P. Farrington (Eds.), *Child delinquents: Development, intervention, and service needs.* (pp. 25-46). Thousand Oaks, CA: Sage.

Thomas, S., Leese, M., Walsh, E., McCrone, P., Moran, P., Burns, T., et al. (2005). A comparison of statistical models in predicting violence in psychotic illness. *Comprehensive Psychiatry, 46*(4), 296-303.

Thompson, A. P., & Putnins, A. L. (2003). Risk-need assessment inventories for juvenile offenders in Australia. *Psychiatry, Psychology and Law, 10*(2), 324-333.

Tiffin, P., & Kaplan, C. (2004). Dangerous children: Assessment and management of risk. *Child and Adolescent Mental Health, 9*(2), 56-64.

Watt, B., Howells, K., & Delfabbro, P. (2004). Juvenile Recidivism: Criminal Propensity, Social Control and Social Learning Theories. *Psychiatry, Psychology and Law, 11*(1), 141-153.

White, J. L., Moffitt, T. E., Caspi, A., Bartusch, D. J., & et al. (1994). Measuring impulsivity and examining its relationship to delinquency. *Journal of Abnormal Psychology*, *103*(2), 192-205.

Appendix A

Variable Recoding Specifications

CYF Variables

NumPriorC&P
0 = 0
1 = 1
2 = 2
3 = 3
>3 = 4

NumPriorYJ
0 = 0
1 = 1
2 = 2
3 = 3
>3 = 4

NumPriorYJ-FGC
0 = 0
1 = 1
2 = 2
3 = 3
>3 = 4

NumPriorIntakes
0 = 0
1 = 1
2 = 2
3 = 3
4 = 4
5 = 5
6 = 6
>6 = 7

NumPriorCourtOutcomes
0 = 0
1 = 1
2 = 2
>2 = 3

NumPriorCourtDates
0 = 0
1 = 1
2 = 2
>2 = 3

NumPriorOtherYJOutcomes
0 = 0
1 = 1
>1 = 2

NumPriorFindings
0 = 0
1 = 1
2 = 2
3 = 3
>3 = 4

NumPlacements
0 = 0
1 = 1
2 = 2
3-6 = 3
>6 = 4

NIA Variables

NumPriorIntelligence
0 = 0
1 = 1
2 = 2
3-4 = 3
5-6 = 4
>6 = 5

NumPriorOccurrence
0 = 0
1 = 1
2 = 2
3 = 3
4 = 4
5-6 = 5
>6 = 6

NumPriorYouthAid
0 = 0
1 = 1
2 = 2
3 = 3
>3 = 4

NumPriorChargeSheets
0 = 0
1 = 1
2-3 = 2
>3 = 3

NumPriorConvictions
0 = 0
1 = 1
2 = 2
3-4 = 3
>4 = 4

NumConvictionsFollowUp
0 = 0
1 = 1
2 = 2
3-4 = 3
>4 = 4