Risky Business: Evaluating the Dynamic Risk Assessment for Offender Re-entry for

Use with New Zealand Youth

By

Jonathan Muirhead

A thesis submitted to Victoria University of Wellington in fulfilment of the

requirements for the degree of Master of Science in Forensic Psychology

Victoria University of Wellington

2016

Abstract

An important assumption that decisions based on criminal risk assessments rely on is that our assessments of someone's likelihood of reoffending are accurate. It is well known that young people share many risk factors for criminal conduct with adults, but there is also research to suggest that some factors may be more important at different ages. This research examined how well an adult dynamic risk assessment tool, The Dynamic Risk Assessment for Offender Re-entry (DRAOR), was able to predict any new criminal conviction as well as any new violent conviction in a sample of New Zealand youth (17-19 years) serving community supervision sentences. It was found that DRAOR scores were moderately strong predictors of future criminal conduct for youth, with better results being found for any reconvictions compared to violent reconvictions. The more recent an assessment was, the more accurate it was too. It was also found that those who did not go on to be reconvicted showed greater improvements on the risk scale throughout the course of their sentence than those who were reconvicted. These findings support the continued use of the DRAOR for youth in New Zealand who are serving community supervision sentences.

Acknowledgements

Firstly, I would like to thank my supervisors, Dr. Clare-Ann Fortune and Professor Devon Polaschek, without whom I would have been completely lost. Your encouragement, feedback, and support were invaluable in helping me complete this thesis. You managed to pull me back onto my feet after a rather large hurdle a few months in and helped me find a new topic that I thoroughly enjoyed researching. I feel as though I have gained a great deal of knowledge from both of you, as well as everyone else in the criminal justice lab. To the lab, I want to thank you all for the suggestions and feedback, as well as showing me that I was not alone in my research struggles. Thank you as well to Professor Paul Jose for your statistical advice; I am not sure if I would have ever got my head around multilevel modelling without your help. I would also like to thank the New Zealand Department of Corrections for supplying me with the dataset that this thesis was based on.

To my office mates Oliver Kitto and Merryck Anderson, you two deserve a huge thank you for keeping me sane. Be it the tea breaks, Stuff quizzes, or random Youtube videos, without you guys I would have had a far less entertaining year. Thanks as well to everyone else in the Forensic and Clinical programmes: not only did you listen to my research woes, but you also gave some great advice on how to overcome many of the obstacles that come with doing research. To my family, thank you for supporting and encouraging me throughout my studies, especially my sister Susannah who was always happy to get a beer with me when I needed a break.

Lastly, I would like to thank all of my friends for patiently listening to me talk and/or moan about my research, as well as distracting me when I needed it. I would not have finished without you guys.

Table of Contents

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

Imagine you are a probation officer given the job of assessing an individual named Ben for his risk of reconviction, while he serves a community supervision sentence. Although Ben continues to consort with the friends who originally got him into trouble, and is still struggling with anger and impulsivity issues, he has stopped abusing drugs and alcohol, and is very responsive to your advice as his probation officer. How likely do you think Ben is to reoffend? Do you think an assessment of Ben's risk of recidivism would differ if he were 17 compared to 40 years old?

Risk assessment is an important area of forensic psychology, and a lot of work has been done identifying factors that influence someone's likelihood of offending. The majority of this work in correctional settings has focused on adult males, with the assumption that these findings will translate to other populations of people who offend (e.g., youth and women). However, despite a large overlap of factors relating to offending for different groups, there are also a number of differences that are often overlooked (e.g., mental health is a risk factor for youth offending, but not for adult offending; Borum, 2003). There is also the fact that some factors are more influential at different points in one's life (e.g., peers have been found to be more of a risk factor during adolescence than adulthood; Hoge, Vincent, & Guy, 2012) which could have an impact on a risk assessment's accuracy.

Despite the fact that a number of risk assessment tools have been created to address the differences between populations, and studies have validated many risk assessment tools for different populations, there is still work to be done. One area suffering from a lack of research is that of older youth (17-19 years old), leaving uncertainty as to whether they should be assessed as children or adults. This research

aims to validate an adult dynamic risk assessment tool on a sample of 17 to 19 year old New Zealand (NZ) youth who are serving community supervision sentences, in order to shed light on how well a tool designed for use with adults works with an older youth population.

Risk and Protective Factors

There has been a lot of research on risk factors in a number of different fields (e.g., medical, criminal) with the basic assumption being that the more risk factors an individual has, the higher the likelihood of the outcome of interest. A risk factor has been defined as "a measurable *characterisation* of each *subject* that precedes the *outcome* of interest" (Kraemer et al., 1997, p. 338). In the forensic field, the outcome of interest is typically to do with criminal activity. Risk factors for criminal conduct can be categorised as either *static* or *dynamic*.

Static risk factors. *Static risk factors* are named due to their relatively unchangeable nature. Static risk factors measure historical or enduring aspects of an individual's life that are related to offending, but cannot be easily changed through interventions (Douglas & Kropp, 2002). Although historical aspects of a person's life are generally fairly accurate predictors of future behaviour (i.e., past behaviour is a good indicator of future behaviour), evaluating only static risk factors makes it difficult for someone's risk to go down; no matter how much rehabilitation a person receives, their criminal history will not be erased (Douglas & Skeem, 2005). Douglas and Skeem (2005) argue that static or historical risk factors only assess *risk status*, or the risk category a person fits into (e.g., low-risk, medium-risk, high-risk), without allowing for much change in risk over time. Static variables struggle to assess *risk state*, or the ongoing fluctuations that occur surrounding one's risk of reoffending

(e.g., a person considered to be high risk is not at a high risk of reoffending every minute of every day; Douglas & Skeem, 2005).

Dynamic risk factors. *Dynamic risk factors*, on the other hand, include variables that can change through intervention programmes, such as substance abuse, and thus are able to be used to monitor a person's risk state. When changed, dynamic risk factors are expected to alter the likelihood that someone will offend (Andrews & Bonta, 2010; Kraemer et al., 1997). In saying that, many researchers disagree that dynamic risk factors should be considered as simply causal mechanisms that lead to crime (Heffernan & Ward, 2015; Ward & Beech, 2005, 2015); however, that debate is beyond the scope of this project.

Dynamic factors can be further separated into two categories: *Stable* and *acute*. Stable dynamic risk factors are skill deficits, attitudes, and behaviours that, while changeable, will likely take a long period of time to make any meaningful change (e.g., criminal attitudes). Acute dynamic risk factors on the other hand can be personal or environmental, are often temporary (sometimes only lasting a few hours e.g., intoxication), and are considered good indicators of imminent offending. Acute dynamic risk factors are theorised to be useful in monitoring the likelihood of imminent offending, and thus aid the prediction of *when* someone is likely to reoffend (Risk state; Douglas & Skeem, 2005). Risk state is important for the monitoring of people on parole, as it allows for intervention when an individual needs it most (Douglas & Skeem, 2005).

Protective factors. As well as using risk factors to help predict future offending, recent research has begun looking at *protective* factors as well. Research into resilience and desistance from criminal conduct has helped to identify factors that may reduce someone's likelihood of engaging in criminal conduct. Resilience, in this

sense, is related to one's ability to abstain from criminal activities, despite being potentially predisposed to that lifestyle through an abundance of risk factors (e.g., a procriminal family; Hoge, Andrews, & Leschied, 1996). Desistance, on the other hand, refers to the reduction and eventual cessation of criminal conduct after having already engaged in offending (de Ruiter & Nicholls, 2011; Serin & Lloyd, 2009).

Protective factors can be conceptualised in a number of ways. They can be seen as the polar opposite of risk factors (e.g., high impulsivity vs. low impulsivity), as factors that directly reduce the likelihood of offending without the opposite being a risk factor (e.g., religiosity is often seen as a protective factor, but atheism is not considered a risk factor; de Vogel, de Vries Robbé, de Ruiter, & Bouman, 2011), or as factors that reduce the impact of particular risk factors (e.g., a strong emotional bond with a caregiver can help protect against the potential negative effects of growing up in a deprived neighbourhood; Lösel & Farrington, 2012). Regardless of the type of protective factor, it is agreed that they can be both internal and external resources that act to improve resilience and promote desistance from criminal behaviour. According to Lösel and Farrington (2012) protective factors that act by directly influencing potential offending are referred to as *direct protective factors*, and those that shield against the impact of risk factors are *buffering protective factors*. A major argument for the inclusion of protective factors in modern risk assessment tools is that they allow for consideration of a person's strengths as well as their weaknesses (Serin, Chadwick, & Lloyd, 2016; Thornton, 2013). This consideration can greatly improve rapport and help the clinician or probation officer when setting goals with the individual. Highlighting protective factors has also been theorised to help reduce the chance of overestimating an individual's level of risk, as assessments are not solely focusing on negative aspects of the person (de Ruiter & Nicholls, 2011).

Summary. The identification and assessment of risk and protective factors has been enormously valuable for the risk classification and management of people who have offended. Although there is still a lot that we do not know about the exact function of risk and protective factors (Ward & Beech, 2015), it is undeniable that they are relevant to criminal justice. The correlation between risk factors and recidivism has led to a large field of research that focuses on estimating the likelihood of someone's continued criminality, based on the volume of risk factors present.

Risk Assessment

Criminal risk assessments have important implications for a range of areas including parole decisions, treatment considerations, and the monitoring of people who have been released from custody. Risk assessments use expertise and empirically derived tools to predict a number of outcomes, one of which is the likelihood that someone will reoffend upon release. Offence prediction is important not only from a public safety point of view (preventing harm to future potential victims), but also when making decisions around treatment (giving more intense treatment to those who pose the greatest risk to society), and considering the rights of the person who has offended (Andrews & Bonta, 2010).

Risk assessment has enjoyed a great deal of research in the past few decades. The implementation of risk assessments has changed from unstructured clinical judgement, which was very inaccurate, to using very comprehensive tools that take well known risk and protective factors for criminal activity into account to give a reasonable level of predictive accuracy for reoffending (Andrews & Bonta, 2010). The different forms of risk assessment can be classified into four distinct generations (Andrews, Bonta, & Wormith, 2006).

First generation. *First generation* risk assessment relied exclusively on the clinical judgement and experience of professionals. Unfortunately, this method was unreliable and subject to the clinicians' biases (Hsu, Caputi, & Byrne, 2009). Results have not been favourable for first generation risk assessments, with one meta-analysis finding an effect size of $d = 0.11$ for unstructured assessments, compared to $d = 0.97$ for more modern assessments (Hanson & Morton-Bourgon, 2009). Clinicians tended to overestimate the risk of offending for an individual and did not attend to empirically validated risk factors or to the base rates of particular types of crime (Andrews & Bonta, 2010).

Second generation. *Second generation* risk assessments involve the statistical consideration of common factors among offending populations. The factors considered in these second generation tools are typically static. Although tools measuring these factors have shown good correlations with reoffending, they have not been derived from any relevant theory of offending, and thus do not include many important areas of concern, such as antisocial attitudes (Andrews & Bonta, 2010). Despite this, second generation tools have been found to outperform unstructured clinical judgement (e.g., Bengtson & Långström, 2007; Grove, Zald, Lebow, Snitz, & Nelson, 2000). One meta-analysis found that second generation tools outperformed first generation tools in 33-47% of the studies examined, whereas only 6-16% of the studies showed a significant difference in the other direction (Grove et al., 2000). Second generation tools are still used today due to their high reliability, speed and ease of use, and the very little expertise or training required for most tools, as administration often only entails ticking off a checklist of easily rateable items and adding up the score (e.g., STATIC-99; Hanson & Thornton, 2000). Although these static tools tend to be predictive of reoffending, they do not help in terms of

identifying targets for treatment, nor are they very responsive to changes a person

may make in rehabilitation programmes, as they mainly consider variables that cannot

change through intervention (Miller, 2006).

Third generation. The next advancement in risk assessment tools was the

inclusion of dynamic risk factors, and in some cases protective factors, to actuarial

risk scales. This marked the beginning of the *third generation* of risk assessment

(Andrews & Bonta, 2010). Dynamic risk factors are very useful when it comes to risk

assessment tools, not only for their correlation with offending, but also due to their

ability to highlight problem areas for people who have offended so that treatment can

be tailored to their specific needs (Andrews & Bonta, 2010) and track changes in risk

in order to intervene when necessary or step back if appropriate (see Change below).

The dynamic risk factors that are included in third generation risk assessment tools

(e.g., Level of Service Inventory – Revised; Andrews & Bonta, 1995) are

theoretically relevant to people who have offended, and often derived from

comprehensive theories of criminal conduct (Andrews & Bonta, 2010). The inclusion

of dynamic risk factors in third generation risk assessments is often considered to

slightly increase the predictive validity above relying solely on static items (e.g.,

studies have shown the LSI-R to have an AUC of .71; Andrews et al., 2006; Hanson

& Harris, 2000); however, not all researchers agree (Caudy, Durso, & Taxman, 2013).

Despite the contention around whether or not dynamic risk factors add incremental

validity to static tools, it is undeniable that dynamic items give the added benefit of

being able to guide treatment decisions and attend to any changes in risk made by the

individual being assessed. This added benefit of guiding treatment interventions

makes third generation tools useful, not only for predicting offending, but also for the

more valuable goal of reducing offending.

Fourth generation. *Fourth generation* risk assessments also take theoretically derived, dynamic risk factors into account; however, there is an emphasis on individual case management, as well as room for clinicians to use their *structured professional judgement* when assessing a client. It was found that although previous generation risk tools were being administered regularly, they were not always being used to inform treatment targets (Bonta, Rugge, Scott, Bourgon, & Yessine, 2008). Fourth generation assessments specifically guide treatments towards identified risk targets, thus giving the opportunity for better treatment outcomes (e.g., Level of Service/Case Management Inventory; Andrews & Bonta, 2010). The fourth generation of risk assessments also typically encourages administrators to use structured professional judgement when completing the assessment. Structured professional judgement is based on the empirically and theoretically based items in the risk assessment tool; however, the administrator makes the final call as to which risk band an individual fits in (e.g., low-, medium-, or high-risk). This professional freedom allows for the consideration of other salient factors that may improve the prediction of an individual's risk of reoffending, but may not be included in the risk assessment tool.

Summary. Criminal risk assessment has come a long way since the early attempts to predict reoffending. There are now countless tools that have been designed for and validated on a wide range of populations, with some of the best tools showing a high degree of accuracy for predicting future criminality. With that being said, there is still work to be done in terms of improving the accuracy of our predictions. This is especially true for deciding *when* a particular person may need increased monitoring. The use of dynamic risk factors in more recent tools has improved our ability to forecast behaviour; however, there has only been a handful of

studies looking at how these changeable factors vary over time, and when they do change, what that means in terms of behaviour.

Change

The ability to see the changes in risk an individual makes throughout rehabilitation programmes, community sentences, or parole, can help in a number of ways. Firstly, this can help produce a more accurate picture of current risk, as it allows for a person's risk level to improve if they have made positive changes, or deteriorate if they have made negative changes - something historical items are not particularly good at (Douglas & Skeem, 2005). The inclusion of dynamic risk factors in modern risk assessment tools also allows assessors to track the changes a person might make through the course of treatment. Amongst other uses, the amount of change an individual makes can have predictive validity in terms of recidivism outcomes (i.e., the more positive change an individual makes, the less likely they are to reoffend; Howard & Dixon, 2013). Unfortunately, a lot of the studies that have looked at the predictive validity of dynamic risk tools have done so using only one time point. Using only one time point means change on these dynamic tools cannot be measured, and essentially turns the dynamic tools into static tools for the purpose of the study (Olver, Wong, Nicholaichuk, & Gordon, 2007). In order to see if changes in dynamic risk factors do in fact predict recidivism outcomes, more than one point in time must be considered, so changes on the factors can be measured and then analysed in relation to recidivism outcomes.

In a validation study of the Violence Risk Scale-Sexual Offender Version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003) it was found that changes on the dynamic variables measured at two time points were negatively correlated to risk of reoffending for high-risk individuals (i.e., positive changes were related to

reduced reoffending; Olver et al., 2007). Olver and colleagues (2007) used file based VRS-SO scores of 321 individuals who had offended sexually and had been followed up for an average post-release period of 10 years. Measurements were available both pre- and post-treatment, allowing for the analysis of change scores. It was found that after controlling for static risk and follow-up time, positive dynamic change scores significantly predicted reductions in the likelihood of recidivism for people of high risk, but not low-risk of reoffending sexually (Olver et al., 2007). In another study, Lewis, Olver, and Wong (2013) evaluated change scores from pre- to post-treatment for people at a high risk of violent reoffending who had psychopathic personality traits, using the Violence Risk Scale (VRS; Wong & Gordon, 2006). It was found that post-treatment change scores were predictive of reoffending in this study after controlling for pre-treatment risk levels. Together, these two studies suggest, at least for people considered to be at a high risk of reoffending, the amount of change on dynamic risk factors that occurs during treatment programmes can be related to recidivism outcomes.

Other studies using a range of other risk assessment tools have also found evidence to suggest change scores can be predictive of recidivism outcomes (De Vries Robbé, de Vogel, Douglas, & Nijman, 2015). A retrospective analysis was conducted on pre- and post- treatment scores for a dynamic risk assessment tool (Historical, Clinical, Risk Management-20) as well as a tool assessing protective factors (Structured Assessment of Protective Factors for Violence Risk) for a sample of high-risk forensic psychiatric patients in the Netherlands. It was found that although recidivists' and non-recidivists' scores on the two scales did not differ significantly at pre-treatment, those who ultimately reoffended showed significantly fewer improvements at post-treatment. The total change scores (risk change scores minus

protective change scores) were predictive of group membership (recidivist vs. non-recidivist) at both 1-year follow-up and 11-year follow-up after controlling for both baseline risk scores and time at risk (De Vries Robbé et al., 2015).

An evaluation of the Offender Assessment System (OASys; Home Office, 2006) also found promising results (Howard & Dixon, 2013). The authors looked at a large sample ($N = 196,493$) of people who had offended and were serving a community supervision sentence or post-custodial supervision sentence of more than four months, who were assessed multiple times with the OASys between October 2004 and March 2008. It was found that the dynamic risk factors assessed were subject to change, and those changes added incremental predictive validity to initial risk scores for the prediction of future offending, after controlling for time at risk. Added to the above studies, these results are promising, as changes made on these dynamic factors have been associated with reductions in offending behaviour for people who are at a high risk of reoffending (De Vries Robbé et al., 2015; Lewis et al., 2013; Olver et al., 2007) and individuals on community sentences (Howard & Dixon, 2013).

However, not all studies have found that change scores are predictive of future offending (Hanson, Harris, Scott, & Helmus, 2007). Hanson et al. (2007) looked at multiple risk assessments of 997 individuals on community supervision in Canada and the United States, who had committed a sexual offence. Although the dynamic risk factors (both stable and acute) were predictive of reoffending in this group, there was little evidence to say the amount of change that occurred on the risk assessment tools across two time points was predictive. It was also found that over a 6-month period, there was very little change seen on the acute dynamic factors that were assessed, suggesting they were not as 'dynamic' as originally thought. Another notable

outcome from this study was that the most recent acute scores were not as good as an average over the last 6 months at predicting recidivism, a finding that goes against current theory (Douglas & Skeem, 2005). The authors proposed that perhaps the acute scores were more of an indicator of underlying dispositional traits, rather than simply situational variables (Hanson et al., 2007).

There has also been some work done evaluating change on the Dynamic Risk Assessment for Offender Re-Entry (DRAOR, see below for a description; Serin, 2007). Two recent projects have explored the DRAOR's predictive validity for older youth (17-19 years old; Ferguson, 2015) and females (Scanlan, 2015) in NZ and have included analyses of change scores between two time points. Area under the receiver operating characteristics curve (AUC) analyses found that 73% of the time, a randomly selected youth who was reconvicted of a new offence had a lower total change score (i.e., less improvement) than a randomly selected youth who was not reconvicted. Similar results were found with adults in this study, with a 65% likelihood that a randomly selected reconvicted adult would have a lower total change score than a randomly selected adult who was not reconvicted (Ferguson, 2015). DRAOR total score changes were also found to be predictive of criminal conviction for women, with AUCs suggesting that reconvicted women would have a 63% likelihood of a lower change score than non-reconvicted women (Scanlan, 2015). Another study was conducted using a large NZ sample of 3498 adult parolees (Hanby, 2013). Similar results to those above were found in terms of the predictive validity of the DRAOR. Hanby (2013) also assessed change in risk scores over time and found that there was a difference in the rate at which recidivists' acute dynamic risk changed, compared to non-recidivists. This finding highlights the importance of

examining dynamic risk assessment tools at multiple time points, in order to better understand how to identify those who may be more likely to reoffend.

Summary. Overall, the amount of change an individual makes on dynamic risk factors may be related to recidivism outcomes. In saying that, there is still very limited research and much of the current research has a number of limitations. The main limitation of current research looking at change on dynamic factors is that studies very rarely, with the exclusion of Hanby (2013), take multiple assessments (more than two) over time or the rate of change into account. As it currently stands, we need further research that takes multiple time points into account before we can understand the utility of dynamic risk change.

Youth Development

The other area that this research will look into is how criminal risk assessments relate to youth. Youth is a term that is used a lot in developmental literature, but the definition is somewhat unclear. Youth is generally considered to incorporate adolescence and young adulthood in its broadest sense (10-24 years; Arain et al., 2013); however, there is no set age range with different researchers using different age brackets. Whatever the definition used, youth can be considered as a time of great change, socially, developmentally, and legally.

Youth who commit criminal offences are of serious concern due in part to age being a very good predictor of reoffending, with youth offending more often and more quickly than older people (Indig, Frewen, & Moore, 2014). In fact, adolescence is a time where engaging in antisocial behaviour is considered normative. In her ground-breaking paper, Terrie Moffitt (1993) found there were broadly two types of antisocial adolescents: the small group who continuously engaged in antisocial behaviour from early childhood until well into their adult years (Life-course-

persistent; LCP), and those who only committed antisocial behaviour for a short

period of their teenage years and early adulthood (Adolescent-limited; AL). The peak

age for engagement in antisocial behaviour is around 17 years old, a period where

abstinence from antisocial behaviour becomes statistically uncommon (Moffitt, 1993;

Moffitt, Caspi, Harrington, & Milne, 2002). After 17, the AL group's involvement in

antisocial behaviour drops considerably into the mid-twenties, leaving only the LCP

group to continue offending.

There are many reasons for this large spike in offending during the late

teenage years. Firstly, there are a number of social aspects that may lead the AL group

to join their LCP peers in committing antisocial acts. The maturity gap was a term

that Moffitt (1993) used to describe the role vacuum that adolescents find themselves

in. Despite being physically mature enough to be biologically considered an adult,

socially they are not considered adults in Western societies and thus are not treated as

such. Adolescents who fit into the LCP group tend to be more autonomous than their

peers and often have access to adult privileges like alcohol and low levels of

supervision. Access to commodities such as alcohol can be valuable for adolescence,

which leads the AL group to befriend and mimic the LCP group, thus resulting in

increased antisocial behaviour during adolescence.

Breaking the rules is often done to impress peers, making the commonly

regarded risk factor of peer influence (Andrews & Bonta, 2010) even more influential

for this age group than for adults (Steinberg, Cauffman, Woolard, Graham, & Banich,

2009). As adolescents grow older, they typically become more independent of their

parents, and are influenced more easily by the social interactions they have with their

peers (Hoge et al., 2012). In line with peer approval for antisocial behaviour, many

youth tend to engage in more criminal activity within group situations where it is

possible to reap the social rewards for their actions, on top of any tangible ones (Scott & Steinberg, 2008). This is in contrast to the majority of the LCP group, as well as many adults who offend, who will often not need any peer influence to commit an offence (Moffitt, 1993; Moffitt et al., 2002).

Teenagers also tend to be more concerned with immediate rather than future consequences. They are said to be lacking 'future orientation' (Scott & Steinberg, 2008). Although teenagers may be aware of the negative long-term consequences of their actions, they are more focused on the immediate gratification, either social or material, that comes with their antisocial behaviour. It is only during early adulthood that individuals begin to become more future oriented (Scott & Steinberg, 2008). Not only do adolescents not weigh future consequences as heavily as immediate ones, but they also tend to be less risk averse than adults, making them even more likely to engage in risky behaviour (Scott & Steinberg, 2008). Gambling studies have also shown that adolescents tend to discount potential risks of a situation while promoting the potential rewards compared to adults (Steinberg, 2007). So it is not that adolescents are incapable of identifying the risks of engaging in certain behaviours, rather they tend to evaluate the rewards risk taking provides more profitably (Scott & Steinberg, 2008).

The neurological maturation of adolescents' and young adults' brains has been well researched and a number of the findings are able to shed more light on why adolescents may be more predisposed to offend (Arain et al., 2013). It has been argued that individuals aged 18-24 have many similar psychosocial capacities to those aged 15-17 due to the prefrontal cortex not having fully matured (Farrington & Loeber, 2002). Although intellectual maturity may have been reached, psychosocial maturity and self-regulation are still under developed until mid-twenties, thus

potentially impeding one's ability to abstain from criminal conduct (Prior et al., 2011). The prefrontal lobes, the last area of the brain to develop, play an integral role in the analytical cost-benefit decisions required when weighing up the pros and cons of risky behaviour (Arain et al., 2013). Although an older youth who offends may have the cognitive capacity to understand the consequences of their actions (usually comparable to adults by about 15 years old), their immature prefrontal lobes may restrict their ability to use that information to guide their actions away from particular activities (Steinberg et al., 2009). The prefrontal lobes are also used in relation to impulse control, thus youth may struggle with restraining from potentially rewarding activities, regardless of the cost (Arain et al., 2013).

In a legal sense, youth tend to represent the bridging stage between children and adults. This is a period where they may meet a number of legal age thresholds that, depending on the country, could include the age of consent, the driving age, the drinking age, or the way that they are held responsible for offending. In NZ, the Children, Young Persons and Their Families Act of 1989 defines youth as aged 14 years old to 16 years old. From their 17th birthday, young people are legally considered adults and are processed by the adult justice system of the New Zealand Department of Corrections. In spite of people aged 17 and upwards being legally defined as adults by the courts, those under 20 years of age are still considered youth by the New Zealand Department of Corrections, leaving those aged 17 to 19 in the middle of two contradicting definitions (Department of Corrections, n.d.)

The differing definitions of youth used in NZ are in part due to the ongoing and variable development of youth; one youth may be at a completely different developmental stage to another youth of the same age (Steinberg et al., 2009). For legal purposes, a major reason that different rules are used for different age groups

comes down to levels of maturity (Prior et al., 2011). Unfortunately, maturity is not a straightforward concept. Also, based on what we know about developmental science, it is not possible to give a cut-off age that signifies a certain level of maturity. No matter how one defines it, one 18-year-old may have an "adult level" of maturity and reasoning while another may not.

Defining a person's maturity based on their number of revolutions around the sun seems arbitrary at best. Although it is clear that maturity does have a linear relationship with age, there is no one set date that an adolescent shifts from the maturity of a child to that of an adult; the shift is likely to take a number of years, with different aspects of maturity developing at different speeds for different people (Prior et al., 2011). Some researchers have even argued that young adults (those aged 18-22) are more similar to adolescents than they are to older adults in their reasoning for committing a crime (Modecki, 2008).

Summary. Youth can vary from adults in a number of different ways that relate to offending. Even older youth are still developing cognitively up until their mid-twenties. These variations may have an impact on how we assess their risk of reoffending. Currently, the majority of risk assessment research is focused on adults, with some research on children, but very little research looking at those who fall into the older youth/ young adult bracket.

Youth Risk Assessment

An issue that arises when the above differences between youth and adults are considered is the practice of risk assessment. Accurate risk assessments are very important, especially for a group as malleable as youth, due to risk assessment's large influence over punishment and rehabilitation. The majority of risk assessment tools have been designed and validated for the assessment of adults (Singh, Grann, &

Fazel, 2011). This can be problematic due to the fact that some of the risk factors for youth offending differ from those for adult offending (Borum, 2000). Although there is considerable overlap with many risk factors between the two populations, such as drug and alcohol use or antisocial peers (Andrews & Bonta, 2010), there are also many characteristics considered to be far more influential for youth (e.g., mental health, parental discipline; Borum, 2000). Due to differences between youth and adults in terms of risk factors for future offending, it stands to reason that risk assessment tools designed for adults may vary in terms of accuracy when used for assessing risk with youth.

Although there have been some adaptations of adult tools for youth (e.g., Youth Level of Service/Case Management Inventory), there has been very little consideration of those who fall into the older youth/ young adult category. When it comes to validating risk tools, it is generally accepted that those aged 18 and above should be classed as adults, neglecting any differences that may be present for younger adults. As mentioned above, the period between childhood and adulthood is an important time developmentally (Prior et al., 2011), so it is essential to understand how well risk tools work for this age group.

There is evidence to suggest risk assessment tools designed for a specific population perform better on said group compared to those designed for more general populations (e.g., Dernevik, Beck, Grann, Hogue, & McGuire, 2009; Singh et al., 2011). In a comprehensive meta-analysis looking at the predictive validity of 9 commonly-used risk assessment tools over 68 samples, it was found that the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2002) had the highest level of predictive accuracy (Singh et al., 2011). This superior level of predictive accuracy was thought to be due in part to all of the studies

assessing the SAVRY using participants under 25 years old (i.e., the intended population). It was also found that the predictive validity of the risk assessment tools designed for more general populations tended to increase as age increased. One explanation for this finding was that the general tools were typically designed and validated on released prisoners, who would have been in their late-twenties and thirties and would typically have extensive criminal histories (Singh et al., 2011).

There has also been some contradictory evidence that suggests youth and adult risk tools are equivalent when it comes to predicting youth offending (Ralston & Epperson, 2013). These results highlight the fact that there is considerable overlap in the items included in the risk assessment tools for youth and adults. However, the results could have been influenced by the fact that the authors were unable to utilise a number of dynamic variables included in the tools that were used, due to a retrospective design (e.g., scales 3 and 4 of the Juvenile–Sex Offender Assessment Protocol–II; Ralston & Epperson, 2013). Had the authors been able to include all of the dynamic variables in their assessments, they might have ended up with different results.

Summary. Given the developmental distinctions that have already been discussed between youth and adults, it is important to evaluate whether our current risk assessment tools are sensitive to these differences. This is especially true for those who often fall into the "adult" category of a given risk tool, but are objectively different in a number of developmental areas to the average adult that most risk assessment tools are developed on (e.g., men in their late 20s and 30s; Singh et al., 2011). This current study aims to evaluate the predictive validity of a risk assessment tool that was developed on an adult population and is currently used to assess youth (17-19 years old) and adults (20+ years old) by the New Zealand Department of

Corrections, the Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007).

Introduction to Current Study

The DRAOR (Serin, 2007) is a dynamic risk assessment tool used to predict the likelihood that an individual will reoffend after release from custody. The DRAOR comprises 19 items across three subscales: Acute dynamic risk factors, Stable dynamic risk factors, and Protective factors. The DRAOR is a relatively new risk assessment tool, and thus has not been as heavily validated as many other tools, especially when it comes to its use with youth. Although one study (Ferguson, 2015) has looked at the DRAOR's validity with youth in NZ, the sample size was small and due to matching criteria the final sample used was not representative of typical NZ youth who have offended (the sample obtained had more extensive criminal histories than was typical for the whole dataset).

Currently, in NZ, the DRAOR is being used by the Department of Corrections to assess people in a number of situations. This study will look into the DRAOR's use with youth who were aged 17-19 and were serving a community supervision sentence of between 6 and 18 months. Community supervision sentences in NZ require the individual to regularly meet their probation officer, who then uses information from these meetings to regularly update their DRAOR scores. Community supervision sentences are of particular importance when considering those under 20 years old, as this is the most common form of sentence for this group, with almost 5000 youth starting community sentences in 2013 (Department of Corrections, 2013).

Validation of risk assessment tools across different populations is an important task, as it gives us confidence that we are getting an accurate assessment of what we are trying to measure. The DRAOR is being used for regular assessments of youth on

community supervision, but there has been little work done to evaluate the DROAR's use with youth. This study will partly be an extension of the findings from Ferguson (2015) but using a larger sample. The sample will be drawn from the same dataset as the previous study; however, this study aims to use as many youth as possible from that dataset, in an attempt to bolster the results. This study will be rerunning a number of the analyses that were used in Ferguson (2015) for the prediction of any criminal reconvictions. We will also include analyses for the prediction of violent reconvictions, as well as looking into how change in risk scores may be of use to the youth forensic field. As the same dataset will be used for this study, we have opted not to match the youth with adults in order to retain as many of the youth in our sample as possible – the matching criteria was a major component for the small sample size in Ferguson (2015).

The aims of this study will be to evaluate: 1) Whether the initial DRAOR assessment predicts reconvictions in a youth community supervision sample; 2) Whether more up-to-date DRAOR scores are better predictors of reconvictions than initial scores; 3) Whether DRAOR scores for youth change over time and if they do, whether that change is predictive of reconvictions; and 4) Whether the amount of change and the rate of change made on the DRAOR during a community sentence differs for youth who are reconvicted compared to youth who are not.

Chapter 2

Method

This chapter will outline the procedures used throughout the study, including the measures and other information used, the statistical analyses used, and the data clean up and exclusion criteria that resulted in our final sample of youth serving community sentences in NZ.

Data

The archival dataset used for this study was provided by the New Zealand Department of Corrections. The dataset contained information on a sample of male and female youth (<20 years old) who served a community supervision sentence of 6-18 months between 1 January 2011 and 31 December 2013. The initial dataset provided by Corrections contained information about 547 youth who had been administered the DRAOR during their community sentence. This is the same dataset as was used in the Ferguson (2015) study.

Measures

Dynamic Risk Assessment for Offender Re-entry (DRAOR). The Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007) is a risk assessment tool designed for use with people serving community supervision sentences or parole (Serin, 2015). The DRAOR has been fully implemented in NZ since April 2010 and research is being done with the DRAOR in Australia, NZ, Canada, and a few US states. Probation officers regularly meet with people who have offended and who are living in the community, as part of the person's parole or supervision sentence, and probation officers complete a DRAOR assessment on each occasion. The assessments are conducted by way of an interview with the person of interest, as well as taking third-party information, such as police records or information from family members,

into account. These regular assessments are intended to allow probation officers to monitor a person's risk of reoffending over time, to ascertain not only if the person is likely to reoffend, but also when (Serin, 2015). Although it is recommended that the DRAOR be readministered regularly, in practice this is not always the case, with some of the people in the current study going more than 100 days without any reassessments.

The DRAOR contains 19 theoretically-derived dynamic risk and protective factors that are distributed across three subscales: Stable Dynamic Risk, Acute Dynamic Risk, and Protective (see Table 1 below). Each item is scored on a 3-point scale from 0 to 2. For the stable and acute risk subscales, a score of 0 indicates that the particular item is not considered problematic for the individual. A score of 2 indicates that the item presents a significant problem for them. A score of 1 is used for items that may be slightly present, or if there is uncertainty as to whether the item is present (e.g., contradictory information from different sources). For the protective subscale, a score of 0 indicates that the item is not present for the individual, whereas a 2 indicates that the item is an asset. A score of 1 is again used for uncertainty or slight assets. Although in practice these subscales are generally used to guide clinical judgement of an individual's level of risk, for research purposes the scores are often combined into a *total score*. The total score is the sum of the acute and stable risk scores minus the protective score. This allows for the total score to fall between a minimum of -12 (scoring 0 for all risk factors and subtracting 2 for all protective factors) and a maximum of 26 (scoring 2 for all risk factors and 0 for all protective factors).

The DRAOR is still relatively new and less researched than many risk prediction tools (e.g., Level of Service/ Case Management Inventory). However, there

have been a number of recent studies that have found the DRAOR to be predictive and reliable for a number of NZ populations including women, youth, and adults (AUC range: .62 - .74; Ferguson, 2015; Hanby, 2013; Scanlan, 2015; Tamatea & Wilson, 2009; Yesberg & Polaschek, 2014).

One study has examined the factor structure of the DRAOR using principal component analysis (Yesberg & Polaschek, 2014). It was found that, instead of the current 3-factor structure, the results suggested the items better fitted a 4-factor model, with the main difference being the acute scale splitting into an internal acute scale and an external acute scale. This new structure is yet to be implemented in practice, so this study will aim to evaluate the original 3-factor structure that is currently in use (see Table 1).

Table 1

DRAOR Items by Subscale

Acute	Stable	Protective
Substance abuse	Peer associations	Responsive to advice
Anger/hostility	Attitudes towards authority	Prosocial identity
Opportunity/access to victims	Impulse control	High expectations
Negative mood	Problem solving	Costs/benefits
Employment	Sense of entitlement	Social support
Interpersonal relationships	Attachment with others	Social control
Living situation		

Risk of re-Conviction X Risk of re-Imprisonment (RoC*RoI). The RoC*RoI is a static risk assessment tool that uses a computer algorithm to evaluate an individual on 16 items in order to produce an estimation of their likelihood of reconviction (Bakker, Riley, & O'Malley, 1999). The algorithm provides a score between 0 and 1 that indicates the likelihood of reoffending resulting in imprisonment

within 5 years. A score of 0.20 would indicate a 20% chance of reimprisonment, whereas a score of 0.6 would indicate a 60% likelihood of reimprisonment within 5 years. People's scores are also given a corresponding risk band of low (0.00-0.29), medium (0.30-0.69), or high (0.70-1.00). The RoC*RoI has been shown to perform well in predicting reoffending (AUC = .76; Bakker et al., 1999).

Data Preparation and Exclusion Criteria

There was a lot of information available for each person in the dataset, some of which required cleaning up before it would be appropriate for analysis. The variables that were included for each youth were: Sentence commencement date, Sentence end date, Index offence, Gender, DOB, Age at the beginning of the sentence, Ethnicity, Static risk score and dates of administration of the risk assessment, Dynamic risk scores and dates of administration of the risk assessment, Number of previous convictions (any or violent), Age at first conviction, Number of previous imprisonments, Date of first breach offence since the beginning of the sentence, Date of first criminal offence since the beginning of the sentence, Date of first violent offence since the beginning of the sentence, whether or not they were reimprisoned, and the reimprisonment date.

Age. Originally, the age of all of the participants in the data file had been rounded to the nearest year (e.g., 19.7 was rounded to 20 years old). The sample's dates of birth had also been changed to the first day of the month each person was born, to anonymise the data. Because age was an important factor for this study, it was crucial to be as close to each person's true age as possible. Due to age's importance, we followed the same procedure as Ferguson (2015) to reduce errors in the reported ages. Firstly, all of the birthdates were changed to the 15th of the month, rather than the first. This change allows for a maximum of 16 days error from the true

birthdate (e.g., if someone was born on the 24th of a given month, the data would be 23 days out if their birthdate was still reported as the 1st versus only 9 days out if we use the 15th). An unrounded age was then calculated by subtracting this new date of birth from the sentence commencement date. All people above the age of 19.99 at the beginning of their sentence were excluded from the study as they are not considered youth by The New Zealand Department of Corrections (Department of Corrections, n.d.).

Number of DRAOR administrations. There was variability with the number of DRAOR administrations for each participant in the original dataset (range: 1-43). As the DRAOR administration should be fairly regular in frequency, in order to track changes made, all individuals who had fewer than three DRAOR administrations in the first 100 days of their sentence commencing were excluded.

DRAOR scores. The dataset included the scores on all DRAOR items for each administration. From this information, each individual's scores on each of the three subscales (acute, stable, and protective) and the total score were computed. The scores that were used for the analyses included the initial scores for each subscale and total score, the proximal scores for each subscale and total score, the change score for each subscale and total score, and the average monthly scores for each subscale and the total score (all explained below). The *initial score* was taken from the DRAOR administration closest to two weeks from the beginning of a person's sentence, as this interval allows time for the probation officer to get to know the individual a little better, thus making for a more accurate score. This initial score is different to the protocol used by Ferguson (2015) and Hanby (2013), who both opted to use a person's third score as the initial score. The reasoning for using the score closest to two weeks instead was due to the large variability in the time until the third DRAOR

administration across the sample. Some people had had three DRAOR administrations within one week, whereas others took closer to 100 days until they had their third assessment. The *proximal score* was defined as the score from the last DRAOR administration prior to the type of offence of interest (e.g., first violent offence or first general offence leading to reconviction). For those who were not reconvicted, the proximal score was from the last DRAOR administration taken before data extraction. The *change score* was calculated as the difference between the initial and the proximal scores for each person. The *average monthly scores* were used for multilevel modelling (see below). Averaging all of an individual's DRAOR administrations for each month since the beginning of their sentence created these average scores. If someone only had one administration for a given month, that administration was used as his or her average. If someone did not have any administrations, no data was entered for that month.

Final sample. The final sample for the majority of the analyses was 398 out of the original 547 people. Demographic information on both the original and the final sample can be seen below.

Demographic Information

Sample characteristics. The final sample was statistically equivalent to the full sample in all areas except for time at risk ($M_{diff} = 49.02$ days longer for the initial sample) and the mean number of DRAOR administrations, which rose from $M = 12.03$ to $M = 14.21$. The sentence length for both groups was also very close to a significant decrease from the initial sample to the final sample ($M_{diff} = 9.05$ days). The demographic information for both groups can be seen in Table 2, but for now a summary of the main characteristics of the final group will be given.

The age of the group ranged from 17.04 years to 19.99 years at the start of their sentence, with a mean age of 18.54 years ($SD = 0.62$). The final sample was mainly Māori (48.2%, $n = 192$) and European (37.4%, $SD = 149$). The average sentence length was 262.06 days (range = 180-554, $SD = 68.84$), with each person having an average of 8.27 previous convictions ($SD = 6.49$), 1.56 of those being for a violent offence ($SD = 0.89$). Most of the people in the sample were sentenced for a non-violent index offence (65.8%, $n = 262$), while about a quarter were serving a sentence for a sexual or violent index offence (25.9%, $n = 103$). Many people were assessed with the RoC*RoI before the beginning of their sentences, with the mean score being .33 ($SD = .18$), indicating the group was considered medium risk and to have a 33% likelihood of imprisonment within 5 years, on average.

Table 2

Summary of Sample Demographics Before and After Exclusion Criteria

	Full sample M (SD) N = 547	Final sample M (SD) N = 398	t-test statistic
Sentence length (days)	271.11 (71.27)	262.06 (68.84)	$t(943) = 1.96, p = .051,$ $d = 0.13, M_{diff} = 9.05,$ 95% CI [-0.04, 18.14]
Age at sentencing (years)	18.54 (0.63)	18.54 (0.62)	$t(943) = 0.60, p = .55,$ $d = 0.04, M_{diff} = 0.26,$ 95% CI [-0.60, 1.12]
Number of previous convictions (any)	8.53 (6.73)	8.27 (6.49)	$t(943) = 0.00, p = .99,$ $d = 0.00, M_{diff} = 0.00,$ 95% CI [-0.08, 0.08]
Number of previous violent convictions	0.73 (0.98)	0.71 (0.97)	$t(943) = 0.31, p = .76,$ $d = 0.02, M_{diff} = 0.02,$ 95% CI [-0.15, 0.11]
Number of previous imprisonments	0.45 (1.12)	0.39 (1.07)	$t(943) = 0.81, p = .42,$ $d = 0.05, M_{diff} = 0.06,$ 95% CI [-0.08, 0.20]
Age at first conviction (years)	17.09 (1.02)	17.06 (1.05)	$t(943) = 0.45, p = .66,$ $d = 0.03, M_{diff} = -0.03,$ 95% CI [-0.10, 0.16]

Initial RoC*RoI score	.34 (.19)	.33 (.18)	$t(724) = 0.59, p = .56$, $d = 0.04, M_{diff} = 0.01$, 95% CI [-0.04, 0.02]
Number of DRAOR administrations	12.53* (6.80)	14.21 (6.44)	$t(943) = 3.83, p < .01$, $d = 0.25, M_{diff} = -1.68$, 95% CI [0.82, 2.54]
Time at risk (days)	1021.82 (176.13)	972.80* (167.11)	$t(943) = 4.32, p < .01$, $d = 0.28, M_{diff} = 49.02$, 95% CI [26.71, 71.34]

	Full sample $N = 547$ (%)	Final sample $N = 398$ (%)	Chi-square statistic
Gender			
Female	92 (16.8)	72 (18.1)	$\chi^2(1, N = 945) =$
Male	455 (83.2)	326 (81.9)	$0.26, p = .61$, $\phi = 0.02$
Ethnicity			
Māori	265 (48.4)	192 (48.2)	
European	206 (37.7)	159 (37.4)	$\chi^2(4, N = 945) =$
Pasifika	47 (8.6)	35 (8.8)	$0.16, p = .99$,
Asian	5 (0.9)	3 (0.8)	$\phi = 0.01$
Other/Unknown	24 (4.4)	19 (4.8)	
Index offence			
Non-violent	357 (65.3)	262 (65.8)	$\chi^2(3, N = 945) =$
Violent/sexual	148 (27.1)	103 (25.9)	$0.34, p = .95$,
Justice/administrative	39 (7.1)	30 (7.5)	$\phi = 0.02$
Unknown	3 (0.5)	3 (0.8)	
Any recidivism			
Yes	400 (73.1)	290 (72.8)	$\chi^2(1, N = 945) =$
No	147 (16.9)	108 (17.2)	$0.01, p = .93$, $\phi = 0.00$
Violent recidivism			
Yes	109 (19.9)	77 (19.3)	$\chi^2(1, N = 945) =$
No	438 (80.1)	321 (80.7)	$0.05, p = .83$, $\phi = 0.01$
Reimprisonment			
Yes	117 (21.4)	75 (18.8)	$\chi^2(1, N = 945) =$
No	430 (78.6)	323 (81.2)	$0.85, p = .35$, $\phi = 0.03$

*$p < .05$

Offence-related Variables

    Recidivism. This study was interested in two types of recidivism outcomes: a

new conviction of any type excluding breaches of supervision (*any reconviction*), and

any new violent conviction excluding breaches of supervision (*violent reconviction*). Breach convictions were not examined in this study, as they often occur for very minor sentence infractions rather than criminal offences (e.g., not reporting to a probation officer), and a previous study found the DRAOR to be a poor predictor of breach offences for youth (Ferguson, 2015). The DRAOR was also not designed to predict breach convictions (Serin, 2007).

Time at risk. The time that an individual was at risk of reconviction varied from 403 days to 1193 days ($M = 972.80$, $SD = 167.11$). Each person's time at risk was calculated by subtracting the date at the start of his or her sentence from the data extraction date (13th June 2014).

Survival days. Each person's survival days also differed. For those who were reconvicted, the number of survival days was calculated by subtracting the date at the start of their sentence from the date of their offence (any or violent). For those who were not reconvicted (any or violent), the number of survival days was the same as their time at risk.

Analyses

All analyses were conducted using IBM SPSS Statistics version 22, except for the multilevel analyses, which were completed in HLM version 7.01. The statistical procedures used in the current study are outlined below.

Kaplan-Meier survival analysis. The Kaplan-Meier survival analysis is a descriptive procedure for time-to-event (time-to-reconviction in this case) variables (Garson, 2012). The analysis used survival days and reconviction outcomes to show the rate of reconviction for a given group. The Kaplan-Meier survival analysis is sophisticated in that it is able to incorporate the number of days a person is at risk of reconviction, in order to account for varying days of follow-up (e.g., one person who

began their sentence one year earlier than another person has an extra year of being at risk of reconviction before the data extraction date). By controlling for varying times at risk, we are able to calculate the cumulative survival of the group (i.e., the proportion of people who have not been reconvicted at any given time).

Cox regression. Cox regression is a form of non-parametric analysis that models one or more predictor variables' (covariates') relationship with an outcome or event (reconviction for the purposes of this study). As with Kaplan-Meier survival curves, Cox regression is also used for time-to-event analyses, but is able to incorporate how much a particular covariate is contributing to the likelihood of an event occurring. Cox regression was chosen over other forms of regression as it allows us to take time-dependent variables into account (Garson, 2013). This was important for the current study, as we were interested in the predictive power of the DRAOR while taking variation in survival days into account.

The effect size for Cox regression is given as a hazard ratio, which can be used to assess how much of a contributor a given covariate is to the likelihood of a particular hazard (reconviction in this case; Garson, 2013). A hazard ratio in the context of a Cox regression is the increase or decrease in the odds of an event occurring, given a one-unit increase in the predictor variable. For example, a hazard ratio of 1.15 indicates a 15% increase in the likelihood of the particular hazard for every one-unit increase of the covariate, whereas a hazard ratio of 0.85 would indicate a decrease in likelihood of 15% for that particular hazard, given a 1-unit increase in the covariate. Similarly, a hazard ratio of 1.00 suggests that the covariate is not influencing the probability of a hazard occurring.

In a multivariate Cox regression, each covariate is entered into a regression equation together. This allows us to see whether the unique contributions of each

predictor add any predictive power to the regression model while controlling for the contributions of the other covariates.

Area Under the receiver operating Curve (AUC). AUC analyses test the accuracy of a regression model using the X*Beta scores generated from a Cox regression analysis. The AUC gives the probability that a random subject from one group (e.g., recidivists) has a higher score on a particular covariate or multiple covariates (e.g., proximal DRAOR acute score) than a random subject from another group (e.g., non-recidivists). An AUC of 0.50 indicates that the covariate is discriminating at no better than chance level, whereas an AUC of 0.80 indicates that the covariate can correctly identify the group membership 80% of the time. AUCs are the preferred measure of predictive accuracy in forensic psychology, as scores for the population in question do not need to be continuous or normally distributed (criterion rarely fulfilled in this field; Rice & Harris, 2005).

Multilevel modelling. Multilevel modelling allows for the examination of changes over time within the sample (e.g., changes in dynamic risk or protective factors), whilst also taking between subject factors into account (e.g., reconviction status). By using multilevel modelling, we were able to examine the average rates of change that people made in risk over time (level-1), as well as examining how particular variables (e.g., reconviction outcomes) may have influenced these rates of change (level-2). This allows us to see how those who were reconvicted may differ from those were not reconvicted, in terms of their risk trajectories over time.

Chapter 3

Results

The chapter starts with a brief look at the rates of reconviction for any reconvictions as well as for violent reconvictions. The remainder of Chapter 3 will then focus on the results from all of the analyses for any new reconvictions excluding breaches, with Chapter 4 focusing on the results from the analyses concerning violent reconvictions. After the survival analyses, Cox regression results are displayed for how well initial DRAOR scores predict any reconviction, followed by how well proximal DRAOR scores predict any reconviction, and then a comparison between the two assessments. This chapter finishes with a look at how those who are reconvicted (any offence) differ in terms of how their DRAOR scores change over time compared to those who are not reconvicted.

Rates of reconviction

Figures 1 and 2 show the survival graphs for any reconviction and for violent reconviction in this youth sample respectively. For both graphs, the x-axis shows how many days it took until reconviction or data extraction, and the y-axis represents the proportion of people who at a given time are yet to be reconvicted.
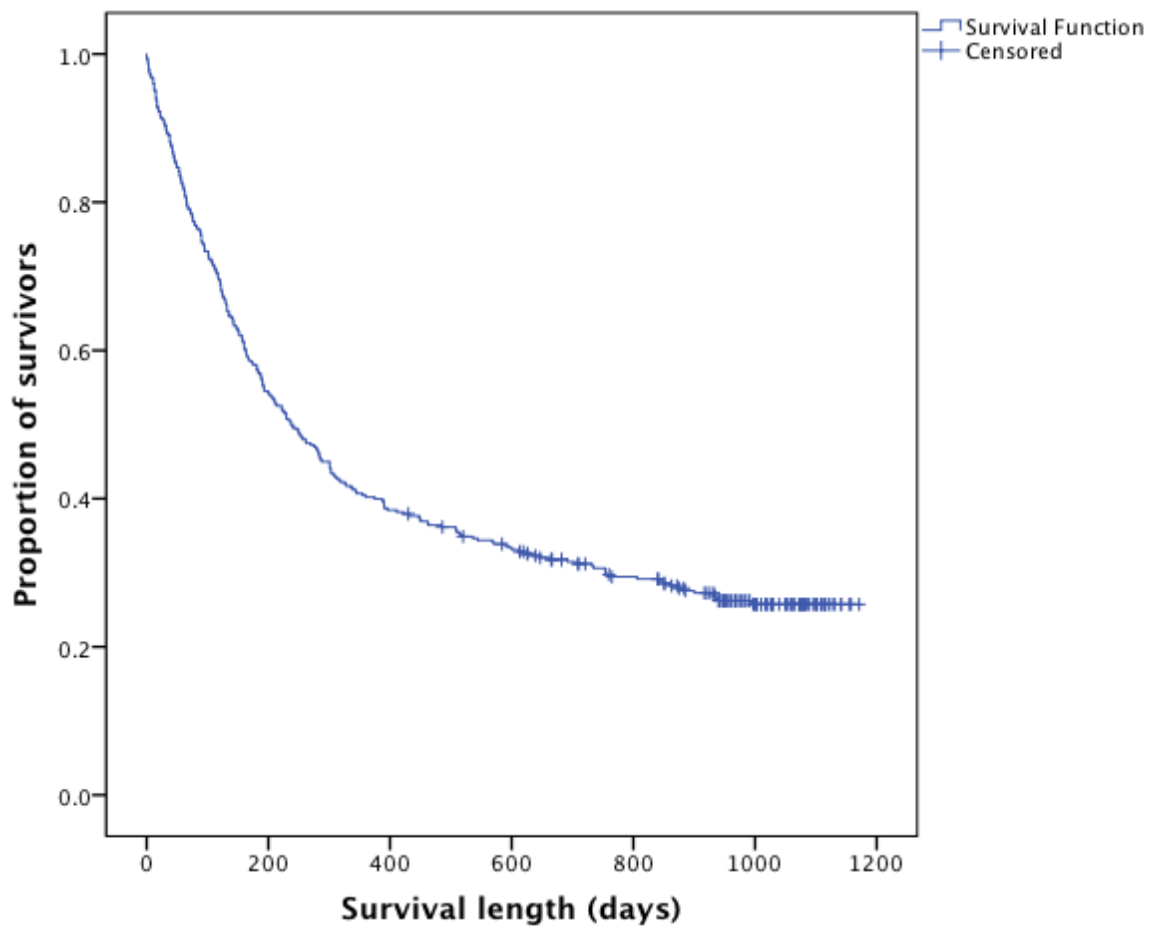
Figure 1. Survival curve for any reconviction.

Figure 1 shows the survival curve for any criminal reconviction. The median survival time before any reconviction was 237.00 days (*SE* = 23.94), 95% CI [190.08, 283.92], or in other words, it took 237 days for half of the sample to be reconvicted. An examination of the slope shows a large failure rate for any reconviction at the early stage of the individuals' sentences, followed by a tapering off after approximately one year.
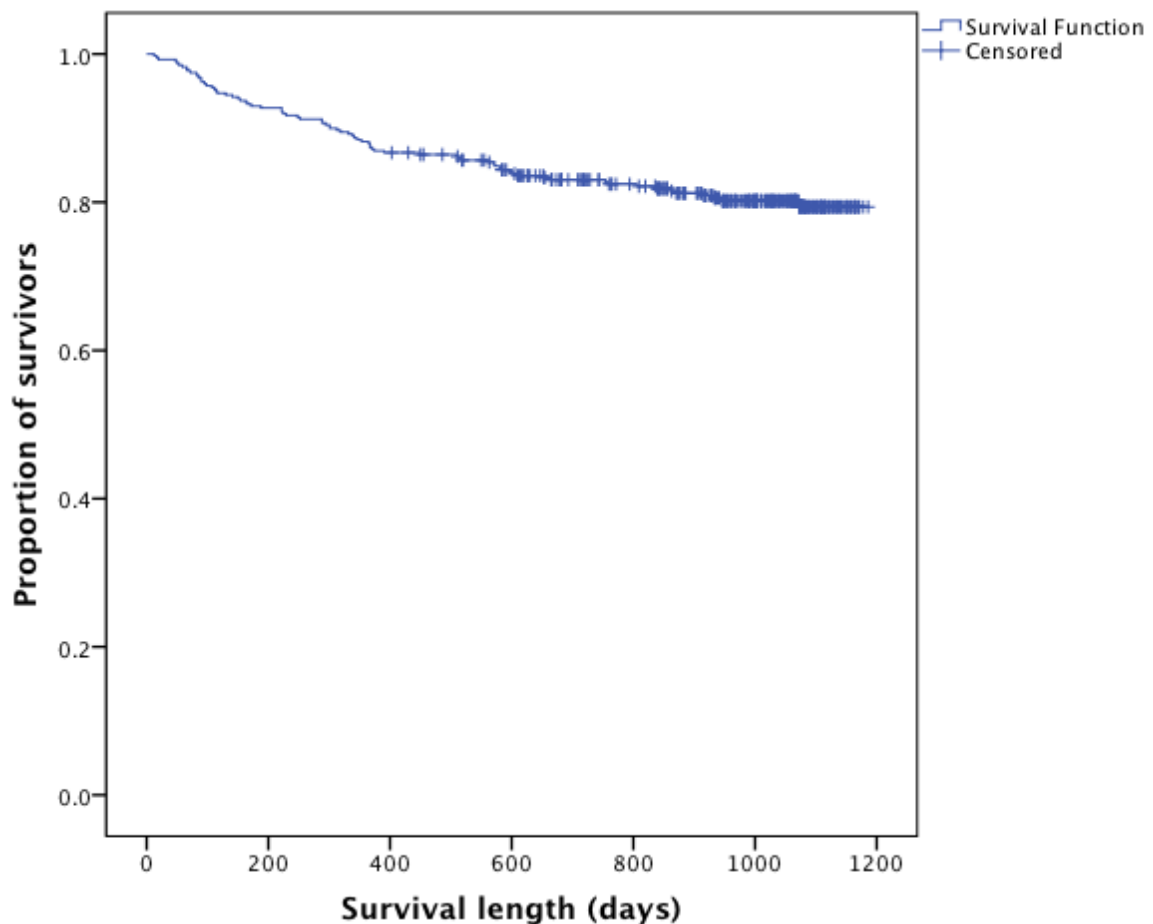
Figure 2. Survival curve for violent reconviction.

Figure 2 shows the survival rate for violent reconvictions. Due to the proportion of failures not exceeding 50% of the sample by the end of the study, it is not possible to report the median (Garson, 2012). An examination of the slope suggests that the rate of violent reconviction did not follow the same course as for any reconvictions, with a fairly steady rate of failure throughout the study instead.

Reporting Any Reconviction Versus Violent Reconviction

Answers to each research question will be given separately for both any reconviction and for violent reconviction. This remainder of this chapter will focus on any reconviction and the following chapter will focus on violent reconviction.

Results: Any Reconviction

How Well Do Initial DRAOR Scores Predict Any Reconviction in a Youth Sample?

The first aim of this study was to replicate the findings of Ferguson (2015), but with a larger sample of youth. Firstly, analyses were conducted to compare the sample used by Ferguson (2015) with the sample used in the current analysis. The means for both groups' initial DRAOR scores are displayed in Table 3. Although the mean initial scores were slightly higher for all subscales and the total score in the current study, they did not reach statistical significance due to the large standard deviations. This suggests that the groups are equivalent in terms of DRAOR scores, and thus we should expect similar results to those found in Ferguson (2015). With that being said, comparisons were also conducted on a number of demographic variables, which showed a significantly lower number of previous convictions (any), a lower number of previous violent convictions, and lower initial RoC*RoI scores for the current sample. These comparisons can be seen in Appendix A. The demographic comparisons suggest that the current sample may have had a marginally lower risk of reoffending than the sample of Ferguson (2015).

Table 3

t-test Comparisons for Initial DRAOR Subscale Scores and Total Score: Any Reconviction

Subscale	Ferguson (2015) M (SD), $N = 100$	Current study M (SD), $N = 369$	t-test statistic
Acute subscale	5.06 (2.63)	5.23 (2.36)	$t(467) = 0.62$, $p = .53$, $d = 0.07$, $M_{diff} = -0.17$, 95% CI [-0.71, 0.37]
Stable subscale	5.82 (2.22)	6.01 (2.10)	$t(467) = 0.79$, $p = .43$, $d = 0.09$, $M_{diff} = -0.19$, 95% CI [-0.66, 0.28]
Protective subscale	6.19 (2.38)	6.25 (1.99)	$t(467) = 0.26$, $p = .80$, $d = 0.03$, $M_{diff} = -0.06$, 95% CI [-0.52, 0.40]
Total score	4.69 (5.81)	4.99 (5.15)	$t(467) = 0.50$, $p = .62$, $d = 0.05$, $M_{diff} = -0.30$, 95% CI [-1.47, 0.87]

In order to assess for convergent validity of the DRAOR subscales and total score, a Pearson bivariate correlation analysis was performed. The results for this analysis are reported in Table 4. All of the initial DRAOR subscales and the total score had medium or large correlations according to the criteria given by Cohen (1992). The largest correlations were found between each subscale and the total score, which was expected, as the total score is a construct of the three subscales. The protective subscale's correlations were all in the negative direction because, as previously mentioned, the protective subscale is negatively scored (i.e., a higher protective score is indicative of a lower likelihood of reconviction).

Table 4

Correlations Between Initial DRAOR Subscales and Total Score

	Acute	Stable	Protective	Total
Acute	1			
Stable	.44**	1		
Protective	-.40**	-.53**	1	
Total	.79**	.82**	-.79**	1

** $p < .01$

One issue that can arise when two or more predictors being used in a multivariate regression model are highly correlated is that of multicollinearity. Multicollinearity can cause problems with the precision of estimates for each particular predictor in a model. In other words, multicollinearity can be an issue when trying to establish how much variance each individual predictor in a multivariate regression model accounts for (O'Brien, 2007).

A widely used method to test for multicollinearity is by calculating the *Variance Inflation Factor (VIF)*. The VIF is an indication of the amount the variance of the regression model is increased above what it would be if the independent variable in question were not correlated with any other predictor variables (O'Brien, 2007). Different statisticians have recommended different critical values for when the VIF becomes a problem, with some going as low as 4.00, but the standard cut-off before multicollinearity is likely to be an issue is a VIF of 10.00 or above for any predictor variable (O'Brien, 2007). VIF values were calculated for all of the DRAOR subscales (the total score will not be used in multivariate regression models as it is a composite of the other three subscales) and it was found that VIFs were all below 2.00. This suggests that multicollinearity is not likely to pose a problem in the current research.

Univariate predictive validity. As DRAOR administrations begin very early on in an individual's sentence, it is important to know how well these initial assessments perform when it comes to predicting criminal conduct (any offence). Ideally, those who received higher initial DRAOR risk scores (and lower protective scores) would be more likely to be reconvicted of a new offence, while those with lower risk scores and higher protective scores would have a reduced likelihood of reconviction. In order to test how well the initial DRAOR scores predict any reconviction, univariate Cox regressions were performed separately on all initial DRAOR subscale scores and the total score. The DRAOR subscale (or total score) was entered as a covariate in the first block of the Cox regression, with the time variable being the number of survival days until any reconviction (or data extraction for those who were not reconvicted). The status variable was the presence or absence of a reconviction.

Table 5 provides the results for the univariate Cox regression models for the predictive validity of the initial scores for all subscales and the total score. It was found that all initial scores on the subscales and the initial total score were predictive of reconviction (any offence); acute, $\chi^2(1, N = 398) = 10.96$, $p < .01$; stable, $\chi^2(1, N = 398) = 16.21$, $p < .01$; protective, $\chi^2(1, N = 398) = 26.95$, $p < .01$; total score, $\chi^2(1, N = 398) = 26.41$, $p < .01$.

The hazard ratios for each Cox regression model are also reported in Table 5. As explained previously (see method), the hazard ratio can be interpreted as the strength of a predictor variable for predicting an outcome (any reconviction). The hazard ratios for all four models were in the expected directions, with the acute, stable, and total scores showing hazard ratios of greater than 1.00 (increased scores indicate an increased chance of any reconviction), and the protective subscale

showing a hazard ratio of below 1 (indicating a reduced likelihood of any

reconviction for higher scores). The strongest hazard ratio for a risk subscale was for

the stable subscale (hazard ratio = 1.13, 95% CI [1.06, 1.19]), which means that for

every one-point increase on the initial stable subscale the likelihood of any

reconviction is increased by 13%. The strongest hazard ratio across all of the four

models was for the protective subscale, which had a hazard ratio of 0.85 (95% CI

[0.79, 0.90]). This is in the other direction to the risk scales as it indicates that for

every one-point increase on the initial protective subscale, the chance of any

reconviction is reduced by 15%.

Table 5

Univariate Cox Regression Models for Initial DRAOR Scores Predicting Any
Criminal Offending

Model for initial scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.09 (.03)	11.12**	1.09 [1.04, 1.15]	.59** [.53, .66]
Stable subscale	.12 (.03)	16.26**	1.13 [1.06, 1.19]	.60** [.54, .66]
Protective subscale	-.17 (.03)	26.85**	0.85 [0.79, 0.90]	.61** [.55, .67]
Total score	.06 (.01)	26.40**	1.06 [1.04, 1.09]	.63** [.57, .69]

**p < .01

Another important aspect to consider when evaluating the predictive power of

a risk assessment tool is the area under the receiver operating characteristic curve

(AUC). As mentioned earlier, the AUC shows the model's ability to distinguish group

membership (recidivists vs. non-recidivists in this case) based on the predictor

variable being used (subscale score or total score). A score of .50 indicates that the

model is no better than chance at distinguishing group membership. As can be seen in

Table 5, all of the AUCs for the initial DRAOR scores predicting any reconviction

were significantly greater than .50. The strongest AUC came from the total score,

which had an AUC of .63 (95% CI [.57, .69]). This means that if one random recidivist and one random non-recidivist were compared, the recidivist would have a higher total score 63% of the time.

Summary. The results from the four univariate Cox regression models for the initial subscales and the initial total score, suggest that the initial DRAOR assessment is predictive of any reconviction for youth serving community supervision sentences. The most useful score in terms of predictive accuracy was the total score, which showed that individuals with higher total scores were more likely to be reconvicted. These results replicate those found with the youth sample in Ferguson (2015), but with our larger sample size we were able to find significant AUCs whereas the AUCs in Ferguson (2015) did not quite reach significance. With this being said, the AUCs were still quite low in terms of practical use, so it seems that the initial DRAOR scores are not particularly strong indicators of any future reconviction (see Rice & Harris, 2005 for a discussion on the interpretation of AUC power).

Multivariate predictive validity. Although the univariate Cox regression models were all found to be predictive, they do not tell us if any of the predictors are tapping into any unique variance that the other scales are not. There is likely to be shared variance between the three subscales, due to their medium-strength correlations. In order to assess whether any predictors were explaining significantly more, or unique, variance than the other subscales, all three initial DRAOR subscale scores were entered simultaneously into the first block of a Cox regression. As before, the time variable was the number of survival days and the status variable was the presence or absence of any reconviction. The results from this multivariate Cox regression are presented in Table 6.

Table 6

Multivariate Cox Regression for Initial DRAORs Predicting Any Reconviction

Model for initial scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.03 (.03)	0.87	1.03 [0.97, 1.09]	
Stable subscale	.04 (.04)	1.23	1.04 [0.97, 1.12]	.63** [.56, .69]
Protective subscale	-.13 (.04)	11.34**	0.88 [0.81, 0.95]	

**$p < 0.01$

Overall, the multivariate Cox regression model was significant, χ^2(3, *N* = 398) = 30.17, *p* < .01, which is unsurprising given the significance of all of the univariate models. The model also found the initial protective subscale to be a unique predictor of any reconviction, with a hazard ratio of 0.88 (95% CI [0.81, 0.95]). This suggests that initial scores on the protective subscale are capturing something different to the two risk subscales. The AUC of .63 reported for this model suggests that a randomly selected recidivist is 63% more likely to have higher risk scores and a lower protective score than a randomly selected non-recidivist from this sample.

Summary. It was found that although the initial scores on the stable and acute subscales were similar in terms of their ability to predict any reconviction, the initial scores on the protective subscale contain a significant amount of unique variance in the prediction of any reconviction. The results from Ferguson (2015) were trending in this direction; however, the initial protective subscale did not reach significance for any reconviction in that study.

How Well Do Proximal DRAOR Scores Predict Any Reconvictions in a Youth Sample?

Again, as we were interested in replicating the Ferguson (2015) study, *t*-tests were conducted to examine the mean scores for each proximal subscale and the total

score for each group. Recall from the method that the proximal scores were taken

from the DRAOR assessment immediately prior to the first reoffence leading to a

reconviction or prior to the end of a youth's sentence if they were not reconvicted.

The group means are displayed in Table 7. In the instance of the proximal stable

subscale score, the current study had a higher mean, $t(467) = 2.13$, $p = 0.03$, $d = 0.25$,

$M_{diff} = $ -0.63, 95% CI [-1.21, -0.05], but the other comparisons were non-significant.

Table 7

t-test Comparisons for Proximal DRAOR Subscale Scores and Total Score: Any Reconviction

Score	Ferguson (2015) M (SD), $n = 100$	Current study M (SD), $n = 369$	t-test statistic
Acute subscale	4.11 (2.47)	4.36 (2.45)	$t(467) = 0.90$, $p = .37$, $d = 0.10$, $M_{diff} = $ -0.25, 95% CI [-0.79, 0.30]
Stable subscale	4.82 (2.58)	5.45 (2.36)*	$t(467) = 2.32$, $p = .02$, $d = 0.25$, $M_{diff} = $ -0.63, 95% CI [-1.16, -0.10]
Protective subscale	6.80 (2.86)	6.81 (2.52)	$t(467) = 0.03$, $p = .97$, $d = 0.00$, $M_{diff} = $ -0.01, 95% CI [-0.59, 0.57]
Total score	2.13 (6.86)	3.00 (6.44)	$t(467) = 1.18$, $p = .24$, $d = 0.13$, $M_{diff} = $ -0.87, 95% CI [-2.32, 0.58]

*$p < .05$

As with the initial scores, we assessed for convergent validity of the proximal

scores by performing a Pearson bivariate correlation (see Table 8). All of the

proximal subscales and the total score were correlated above Cohen's (1992) criteria

of $r < .50$ for being a large effect size. Again, the correlations of the subscales with

the total score were the strongest as a result of the proximal total score being a

composite of the other proximal subscales.

Table 8

Correlations Between Proximal DRAOR Subscales and Total Score

	Acute subscale	Stable subscale	Protective subscale	Total score
Acute subscale	1			
Stable subscale	.52**	1		
Protective subscale	-.54**	-.66**	1	
Total score	.81**	.87**	-.87**	1

**$p < .01$

As the correlations were stronger with the proximal scores than they were with the initial scores, the subscales were assessed for multicollinearity again. The VIF scores were all below 2.00, so again multicollinearity is unlikely to be an issue when interpreting the multivariate models.

Univariate predictive validity. As it is recommended that the DRAOR be regularly readministered, it is important to evaluate how well up-to-date DRAOR scores perform when predicting any reconvictions. Ideally, those assessed as having high risk scores and low protective scores should be the most likely to reoffend imminently, compared to those with lower risk scores and higher protective scores. In order to assess how well an up-to-date DRAOR assessment predicts reconvictions, proximal DRAOR subscales and the total score were entered into the first block of separate univariate Cox regressions. The time variable was the number of survival days and the status variable was whether any reconviction occurred. Table 9 provides the results from the Cox regression analyses. All of the models for the proximal subscale scores and the proximal total score were significantly better than chance at predicting any reconviction; acute, $\chi^2(1, N = 398) = 57.07$, $p < .01$; stable, $\chi^2(1, N = 398) = 49.72$, $p < .01$; protective, $\chi^2(1, N = 398) = 47.15$, $p < .01$; total score, $\chi^2(1, N = 398) = 72.50$, $p < .01$.

Table 9

Univariate Cox Regression Models for Proximal DRAOR Scores Predicting Any Reconviction

Model for proximal scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.20 (.03)	61.33**	1.22 [1.16, 1.28]	.69** [.63, .74]
Stable subscale	.17 (.02)	50.65**	1.18 [1.13, 1.24]	.70** [.64, .76]
Protective subscale	-.17 (.03)	47.34**	0.84 [0.80, 0.88]	.68** [.62, .74]
Total score	.08 (.01)	74.83**	1.09 [1.07, 1.11]	.73** [.67, .78]

**$p < .01$

The highest hazard ratio came from the proximal acute subscale, HR = 1.22, 95% CI [1.16, 1.28]. This is indicative of a 22% increase in the likelihood of any reconviction for every 1-unit increase in the proximal score for the acute subscale. With that being said, the proximal stable subscale was not considered to have a statistically different hazard ratio, due to the confidence intervals overlapping. The AUCs for all of the subscale and the total score models were operating with a moderate to high level of accuracy for predicting any reconviction.

Summary. The results from the four univariate Cox regression models indicate that the proximal DRAOR subscale scores and proximal total score are able to predict any reconvictions in a sample of community supervision sentenced youth. These results are similar to those found by Ferguson (2015).

Multivariate predictive validity. As with the initial DRAOR scores, we were interested in whether any of the proximal DRAOR scores were explaining any unique variance over and above the others. All three of the proximal DRAOR subscales were entered into the first block of a Cox regression simultaneously, with survival days set as the time variable and the presence of any reconviction as the status variable. It was found that overall the regression model was a significant predictor of any criminal

offending, $\chi^2(3, N = 398) = 74.43$, $p < .01$, and was able to show a high level of

accuracy with an AUC of .72 (95% CI [.67, .78]).

Table 10 shows the output from the multivariate regression model. As can be

seen, the proximal scores for both the acute subscale and the stable subscale were

significantly accounting for unique variance in this model; however, the protective

subscale was not. This suggests that most of the predictive power of the proximal

DRAOR scores is coming from the two risk scales as opposed to the protective scale.

This is in stark contrast to our findings with the initial scores, where the protective

subscale was the strongest predictor of any reconviction in the multivariate regression

model.

Table 10

Multivariate Cox Regression for Proximal DRAOR Subscales Predicting Any Reconviction

Model for proximal scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.13 (.03)	16.42**	1.13 [1.07, 1.20]	
Stable subscale	.07 (.03)	4.76*	1.08 [1.01, 1.15]	.72** [.67, .78]
Protective subscale	-.06 (.04)	3.00	0.94 [0.88, 1.01]	

**$p < .01$

Summary. This model found that the strongest predictors of imminent youth

reconviction (any offence) were the acute subscale, followed by the stable subscale.

This is different to the findings of Ferguson (2015), who did not find the acute

subscale to be significantly explaining any unique variance. Our results are more in

line with current theory though, as the acute subscale is theoretically the more

dynamic of the risk scales, and thus one would expect that an up-to-date score on the

acute subscale would greatly aid in the identification relevant factors for any

imminent offending.

Do Proximal DRAOR Scores Outperform Initial Scores When Predicting Any Reconvictions?

Although it is assumed that the most up-to-date DRAOR assessment will be the most informative, it is important to test this empirically. Results from the separate Cox regression outputs indicate that the proximal scores are the stronger predictors, especially in the case of the acute subscale where the confidence intervals for the initial and proximal hazard ratios do not overlap. However, for the other subscales and for the total score, the confidence intervals for the initial and proximal scores do overlap, therefore we cannot be certain that the two scores have different levels of predictive power. In order to assess whether there is in fact a difference between the two scores, we must compare the scores directly using a Cox regression.

We first examined how highly the initial and proximal scores were correlated with each other. As expected, all of the initial scores were highly correlated with the corresponding proximal scores (e.g., the initial stable score was highly correlated with the proximal stable score, $r = .66$) as shown in bold in Table 11.

Table 11

Correlations Between Initial and Proximal DRAOR Scores: Any Reconviction

	Initial acute	Initial stable	Initial protective	Initial total score
Proximal acute	**.56****	.31**	-.33**	.51**
Proximal stable	.37**	**.66****	-.41**	.60**
Proximal protective	-.40**	-.45**	**.69****	-.63**
Proximal total score	.52**	.56**	-.57**	**.69****

** $p < .01$

Incremental validity for initial and proximal DRAOR scores predicting any offending. In order to assess whether the proximal scores were adding

incremental validity on top of the initial scores, four multivariate Cox regressions were conducted. For each subscale and for the total score, the initial and proximal scores were entered together into the first block of a Cox regression. As with the previous Cox regressions, the time variable was the number of survival days and the status variable was the presence of any reconviction.

It was found that all four of the Cox regression models containing both the initial and the proximal scores were predictive of recidivism (any offence): acute, $\chi^2(2, N = 398) = 57.10, p < .01$; stable, $\chi^2(2, N = 398) = 50.10, p < .01$; protective, $\chi^2(2, N = 398) = 48.25, p < .01$; and total score, $\chi^2(2, N = 398) = 72.76, p < .01$. This was to be expected, as all of the univariate models were predictive as well.

Table 12 displays the regression outputs for each of the separate models. As can be seen, the proximal assessments are accounting for significantly more variance than the initial assessments for all four models. For the risk scores the proximal hazard ratios are significantly higher than the initial risk scores' hazard ratios; however for the protective subscale, there is overlap between the confidence intervals of the initial and proximal hazard ratios. This implies that the proximal risk scores are better predictors of any criminal recidivism when compared to initial scores, but we are unable to say the same for the protective subscale. The strongest hazard ratio from these models came from the proximal acute score, 1.22, which indicates a 22% increase in offending for every 1-unit increase on the proximal acute scale.

Table 12

Multivariate Regression Models Containing Initial and Proximal DRAOR Scores for Predicting Any Reconviction

Multivariate model for initial and proximal scores	β (SE)	Wald	Hazard ratio [95% CI]
Acute score			
Initial	-.01 (.03)	0.03	1.00 [0.94, 1.06]
Proximal	.20 (.03)	48.74**	1.22 [1.15, 1.29]
Stable score			
Initial	-.03 (.04)	0.38	0.98 [0.90, 1.06]
Proximal	.18 (.03)	32.20**	1.20 [1.13, 1.28]
Protective score			
Initial	-.04 (.04)	1.10	0.96 [0.89, 1.04]
Proximal	-.15 (.03)	21.02**	0.86 [0.81, 0.92]
Total score			
Initial	-.01 (.02)	0.26	0.99 [0.96, 1.02]
Proximal	.09 (.01)	45.00**	1.09 [1.07, 1.12]

** $p < .01$

Summary. The above results indicate that the proximal scores are able to explain more variance than the initial scores; however, for the protective subscale the confidence intervals of the hazard ratios overlap slightly, meaning that we cannot say that the proximal assessment is a significantly better predictor than the initial assessment for predicting any reconviction in this sample. These results are similar to Ferguson (2015); however, the proximal score on the acute subscale had the highest hazard ratio in this study, whereas the stable subscale was the strongest in Ferguson (2015).

Do DRAOR Scores Change From Initial to Proximal Assessment and Is That

Change Indicative of Any Reconviction Outcomes?

As discussed earlier, the DRAOR is a dynamic risk assessment tool. Therefore, it stands to reason that the DRAOR should be able to pick up changes made by youth in terms of their risk of recidivism. Firstly, we were interested in whether there was any change overall from initial scores to proximal scores for our sample. In order to examine this, a paired samples t-test was conducted for initial and proximal DRAOR scores. Table 13 shows the outcomes of the t-test, with significantly lower proximal scores for the acute subscale, $t(368) = 7.36$, $p < .01$, $d = 0.77$, $M_{diff} = 0.86$, 95% CI [0.63, 1.10]; stable subscale, $t(368) = 5.37$, $p < .01$, $d = 0.56$, $M_{diff} = 0.56$, 95% CI [0.36, 0.77]; and total score, $t(368) = 8.09$, $p < .01$, $d = 0.84$, $M_{diff} = 1.99$, 95% CI [1.51, 2.47]; and the proximal score being significantly higher for the protective subscale, $t(368) = 5.87$, $p < .01$, $d = 0.61$, $M_{diff} = -0.56$, 95% CI [-0.75, -0.37]. These changes were all in the expected direction (towards improvement).

Table 13

Mean Change Made from Initial to Proximal DRAOR Score

	Mean initial score (*SD*)	Mean proximal score (*SD*)	Mean change score (*SD*)	Range of change scores	No change (%)
Acute	5.23 (2.36)**	4.36 (2.45)	0.86 (2.26)	[-7, 7]	28.2
Stable	6.01 (2.10)**	5.45 (2.63)	0.56 (2.01)	[-6, 8]	49.3
Protective	6.25 (2.00)	6.81 (2.52)**	-0.56 (1.85)	[-6, 5]	54.5
Total score	4.99 (5.15)**	3.00 (6.44)	1.99 (4.73)	[-14, 18]	17.6

**$p < .01$

In order to assess DRAOR change between initial and proximal assessments further, a change score was created (see method). The change score represents the size and direction of change a youth's DRAOR scores made from the initial assessment to

the proximal assessment. For the risk subscales and the total score, a positive change

score indicates a change in the direction of reduced risk (a lower score at the proximal

assessment), whereas a negative change score represents a change in the direction of

greater risk. For the protective subscale, the score is reversed, with a positive change

score indicating a lower protective score at the proximal assessment, and a higher

protective score representing a change towards a greater protective score at the

proximal assessment. Table 13 shows the mean change scores for each group, as well

as the range of change, and the percentage of participants whose score did not change

from initial to proximal assessment. As expected, the subscale that had the highest

number of participants showing change was the acute subscale (only 28.2% did not

show change). This is promising, as theoretically the acute subscale is the most

dynamic. The stable and protective subscales were comparable when it came to the

number of participants who did not change (49.3% and 54.5% respectively),

suggesting that they are not as changeable as the acute subscale.

Do Change Scores Predict Outcomes for Any Reconvictions?

In order to better understand the relationship between change scores and

reconvictions, we first conducted a Pearson bivariate correlation between change

scores and initial DRAOR scores (see Table 14). It was found that all of the initial

DRAOR scores had a small to medium positive correlation with their corresponding

change score, that is, the amount of change for a given score was related to the initial

score. The strongest correlation was found for the acute subscale ($r = .44$). Positive

correlations for the risk subscales and the total score indicate that the higher the risk

was at the initial assessment, the more change in the direction of reduced risk was

made. For the protective subscale, a positive correlation indicates that the fewer

protective factors someone started with, the greater the change in the direction of

increased protective factors was found. This makes intuitive sense, as the amount of

change for a given scale is limited by one's starting point. Someone who has an initial

score of 0 on a risk scale cannot decrease any more, whereas someone with a score of

10 has a lot of room for improvement.

Table 14

Correlations Between Initial DRAOR Subscales and Total Score and DRAOR Change
Scores: Any Reconviction

	Initial acute	Initial stable	Initial protective	Initial total
Acute change	**.44****	.12*	-.06	.27**
Stable change	-.03	**.18****	-.02	.07
Protective change	.11*	.04	**.14****	.02
Total change	.15**	.12*	-.09	**.15****

**p < .01, *p < .05

The next step was to see how well the change scores performed in Cox

regression models, while controlling for the initial scores. Four separate multivariate

Cox regressions were conducted, with the initial score for a subscale in the first block,

and the corresponding change score in the second block. This allowed us to see

whether change scores were able to predict reconvictions (any offence) after

controlling for the sample's baseline dynamic risk (their initial DRAOR assessment).

Table 15

Multivariate Regression Models for Change Scores Controlling for Initial DRAOR Scores Predicting Any Reconviction

Model	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Initial acute score	.19 (.03)	42.28**	1.21 [1.14, 1.29]	.69** [.63, .75]
Acute change score	-.20 (.03)	48.74**	0.82 [0.78, 0.87]	
Initial stable score	.16 (.03)	26.68**	1.17 [1.10, 1.24]	.71** [.65, .77]
Stable change score	-.18 (.03)	32.15**	0.83 [0.78, 0.89]	
Initial protective score	-.19 (.03)	35.81**	0.82 [0.77, 0.88]	.68** [.62, .74]
Protective change score	.15 (.03)	21.02**	1.16 [1.09, 1.24]	
Initial total score	.08 (.01)	43.42**	1.08 [1.06, 1.11]	.73** [.67, .78]
Total score change score	-.09 (.01)	45.00**	0.92 [0.89, 0.94]	

**$p < .01$

It was found that the Cox regression models comprising of initial DRAOR scores and the corresponding change score were all significant predictors of any reconvictions: acute, $\chi^2(1, N = 398) = 46.14$, $p < .01$; stable, $\chi^2(1, N = 398) = 33.89$, $p < .01$, protective, $\chi^2(1, N = 398) = 21.31$, $p < .01$; and total, $\chi^2(1, N = 398) = 46.34$, $p < .01$. As stated above, for the risk subscales and the total score, a positive change score indicates a change in the direction of reduced risk (lower risk score). For the protective subscale a positive change score represents a shift in the direction of fewer protective factors (lower score). When interpreting the hazard ratios in Table 15 for the above models, a hazard ratio of less than 1.00 for the change score on the risk subscales indicates a reduction in the likelihood of recidivism the more a youth's risk score decreases from initial to proximal assessment. As an example, for the acute subscale, the hazard ratio for the change score was 0.82. This suggests that for every

additional point on the acute change score (more reduction in risk from initial to proximal assessment), a youth is 18% less likely to be reconvicted. Thus, a reduction in risk is associated with a reduction in the likelihood of reconviction. For the protective subscale, a hazard ratio of greater than 1.00 indicates that a higher change score increases a youth's chance of reconviction (recall that the protective change score is reversed). The protective change score had a hazard ratio of 1.16, which means that for every additional point in the direction of fewer protective factors from initial to proximal assessment, a youth's risk of recidivism goes up by 16%.

The AUCs for all of the subscale models showed a moderate level of accuracy (range: .68 - .71) for distinguishing between recidivists and non-recidivists, while the total score showed a strong level of accuracy ($AUC = .73$). For the total score, this is interpreted as there being a 73% likelihood that a randomly selected recidivist from this sample would have a lower total change score than a randomly selected non-recidivist. That is, a recidivist is likely to have made less change in the direction of a reduction in risk as assessed by the DRAOR total score.

Summary. The results from the above Cox regressions suggest that, after controlling for youth's baseline DRAOR subscale scores, the amount and direction of change a youth makes from their initial DRAOR assessment to their proximal DRAOR assessment is predictive of reconviction (any offence). These results replicate the findings of Ferguson (2015), giving us more confidence in this assertion.

Do Recidivists (Any Reconviction) Have a Different Rate of Change to Non-recidivists?

After finding that change scores on the DRAOR from the initial assessment to the proximal assessment were able to predict reconvictions (any offence), we wanted to know if there was a difference between recidivists and non-recidivists in the rate at

which their DRAOR scores change. In order to assess this, we first separated our sample into two groups, recidivists ($n = 261$) and non-recidivists ($n = 108$). An independent-samples t-test was conducted on all of the DRAOR change scores in order to see if there was a statistical difference in the amount of change made between the two groups. It was found that the recidivists had lower reductions in risk and lower increases in protection compared to non-recidivists: acute, $t(189.87) = 3.09$, $p < .01$, $d = 0.44$, $M_{diff} = -0.81$, 95% CI [-1.32, -0.29]; stable, $t(367) = 4.53$, $p < .01$, $d = 0.47$, $M_{diff} = -1.01$, 95% CI [-1.45, -0.57]; protective, $t(367) = 3.02$, $p < .01$, $d = 0.31$, $M_{diff} = 0.63$, 95% CI [0.22, 1.04]; and total score, $t(367) = 4.66$, $p < .01$, $d = 0.48$, $M_{diff} = -2.45$, 95% CI [-3.49, -1.42]. The mean initial assessment and proximal assessment scores for both groups are displayed in Figure 3.
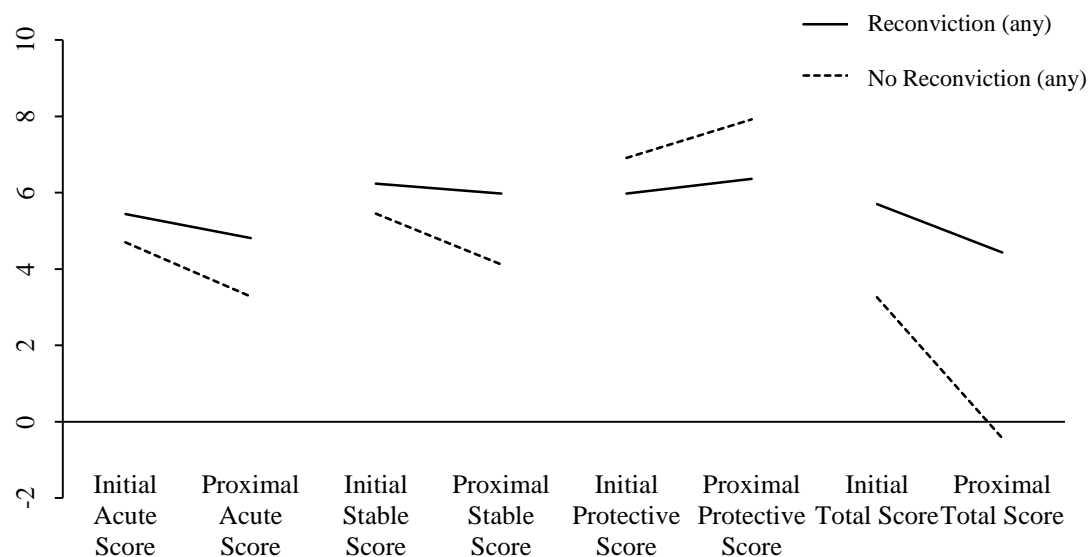
Figure 3. Mean scores at initial and proximal DRAOR assessment for recidivists and non-recidivists: Any reconviction.

As can be seen from the t-test results above and from Figure 3, there is a marked difference in score change between those who were reconvicted (any offence) and those who were not. However, this result could be a function of recidivists having a shorter time for change (i.e., the proximal DRAOR assessment for participants who

were reconvicted before the end of their sentence was not necessarily their last

DRAOR assessment, whereas the proximal score for non-recidivists was always the

last DRAOR assessment of their sentence). Firstly, it was important for us to conduct

an independent-samples t-test to compare the time between initial and proximal

DRAOR assessments for both groups. As expected, the non-recidivists had a longer

period of time between their initial assessment and their proximal assessment ($M =$

207.14, $SD = 66.89$) than the recidivists ($M = 122.67$, $SD = 83.71$). This difference

was statistically significant, $t(367) = 9.33$, $p < .01$, $d = 0.97$, $M_{diff} = 84.47$, 95% CI

[66.66, 102.28].

      Due to the large discrepancy in time between initial and proximal assessments

for the two groups, simply comparing the amount of change made by recidivists and

non-recidivists is not very informative. The next step was to create a model for the

change that incorporated time.

      Multilevel modelling. Multilevel modelling was used to determine the rates

of change for both recidivists and non-recidivists over the first eight months of their

sentences for all three subscales as well as the total score. The multilevel modelling

did not go beyond eight months due to a large percentage of the sample having shorter

periods of data collection (i.e., there would be too much missing data if the model was

extended beyond eight months). The scores used for the models were each youth's

average scores for each month since the beginning of their sentence. Where there was

no data available for a given month, the cell was left blank and later removed during

the analyses.

      To begin with, unconditional means models were created for the average

monthly scores of all subscales and the total score. The composite models are as

follows:

$$ACUTE_{ti} = \beta_{00} + r_{0i} + e_{ti}$$

$$STABLE_{ti} = \beta_{00} + r_{0i} + e_{ti}$$

$$PROTECT_{ti} = \beta_{00} + r_{0i} + e_{ti}$$

$$TOTAL_{ti} = \beta_{00} + r_{0i} + e_{ti}$$

Using the acute subscale monthly scores as an example, $ACUTE_{ti}$ refers to the average acute subscale score for month t for participant i. β_{00} refers to the mean monthly acute subscale score, while r_{0i} and e_{ti} refer to errors at the group and individual level. All of these unconditional models found a significant amount of variance across scores, justifying the use of more complex models (see Appendix B).

The next step was to add the time variable *MONTH* into the equations. The resulting models were as follows:

$$ACUTE_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

$$STABLE_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

$$PROTECT_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

$$TOTAL_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

These models were used to test whether there was any within-subjects change in DRAOR scores over time (by month). It was found that this was the case for all models, with a significant amount of variance still to be explained, justifying the addition of a level-2 variable into the models (see Appendix B). The next step was to see whether there was a difference in the rate of change for DRAOR scores over time between those who were reconvicted (any offence) and those who were not. For this, we added a dichotomous variable, *OFFENCE* (0 = non-recidivist, 1 = recidivist), into level-2 of the models. The resulting composite models were as follows:

$$ACUTE_{ti} = \beta_{00} + \beta_{01}*OFFENCE_{i} + \beta_{10}*MONTH_{ti} +$$

$$\beta_{11}*OFFENCE_{i}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

$$STABLE_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} +$$

$$\beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

$$PROTECT_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} +$$

$$\beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

$$TOTAL_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} +$$

$$\beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$$

The results from these models will be reported individually alongside corresponding graphs. The corresponding statistical outputs can be found in Appendix B.

Acute subscale. On average, those who were reconvicted (any offence) began the first month of their sentences with an acute score of 5.50 ($SE = 0.29$), 0.68 higher than their non-reconvicted counterparts ($M = 4.83$, $SE = 0.25$). Those who were not reconvicted had an average decrease in acute scores of 0.20 ($SE = 0.04$) per month, whereas those who were reconvicted (any offence) decreased by 0.12 ($SE = 0.04$) per month on average. The results from the acute subscale model indicate that although some variance over time was explained by group membership (those who were reconvicted versus those who were not), there was not enough of a difference in slopes to reach statistical significance ($p = .08$). The trajectories can be seen in Figure 4.
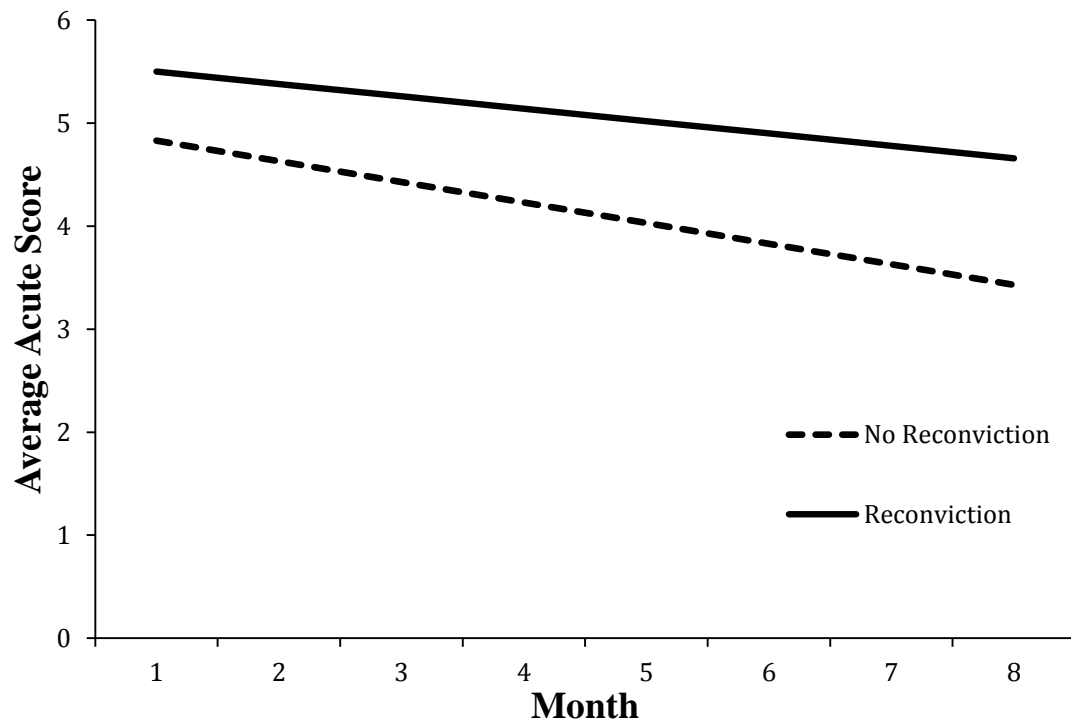
Figure 4. Change in average acute subscale scores over time by reconviction status.

Stable subscale. For the stable subscale, those who ended up with a reconviction (any offence) had an average score of 6.42 ($SE = 0.26$), 0.77 higher than those who were not reconvicted ($M = 5.65$, $SE = 0.22$) for the first month of their sentences. Those who were not reconvicted decreased their average monthly scores by 0.20 ($SE = 0.03$) per month, whereas those who were not reconvicted only decreased by 0.03 ($SE = 0.04$) per month. This difference was statistically significant ($p < 0.01$). A visual representation of these results can be found in Figure 5.
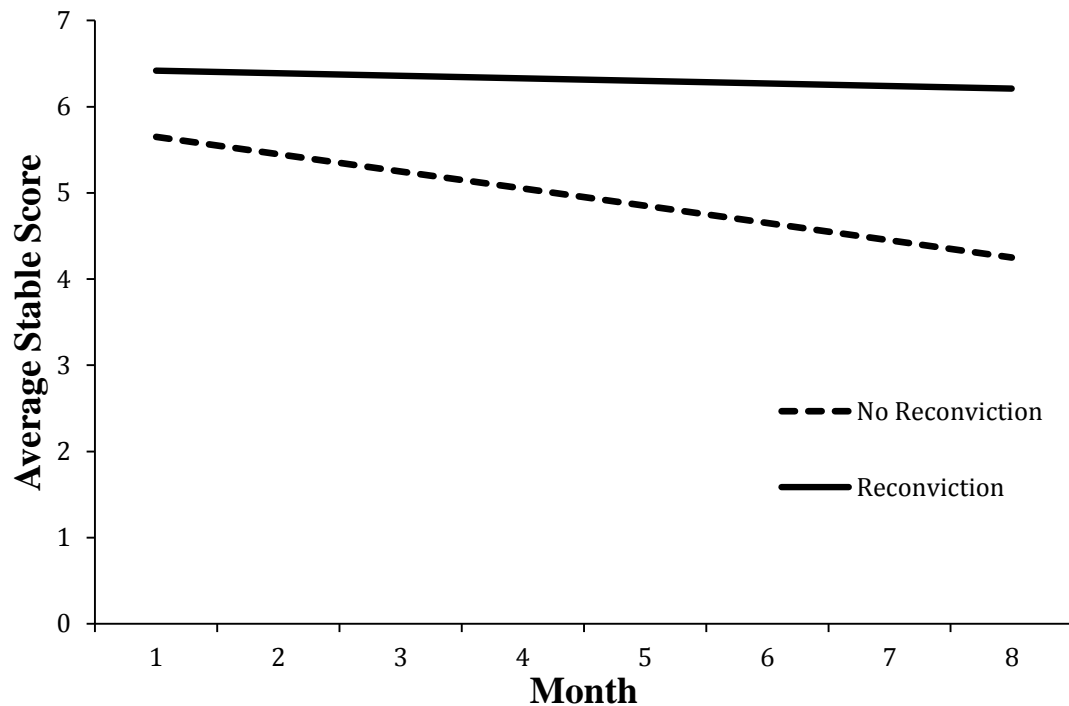
Figure 5. Change in average stable subscale scores over time by reconviction status.

Protective subscale. For the protective subscale, the pattern was in the opposite direction because the scale is reverse scored (a higher protective score is meant to indicate a lower chance of reconviction). Those who were reconvicted (any offence) had an average score for the first month of 5.90 ($SE = 0.23$), 0.76 lower than those who did not have any further convictions ($M = 6.65$, $SE = 0.21$). Those who were reconvicted also had a slower rate of change, increasing their protective subscale score by 0.03 ($SE = 0.04$) per month on average, compared to those who did not get any further convictions, who increased their protective subscale score by 0.16 ($SE = 0.03$) per month on average. This difference was found to be significant ($p < .01$). The results are displayed graphically in Figure 6.
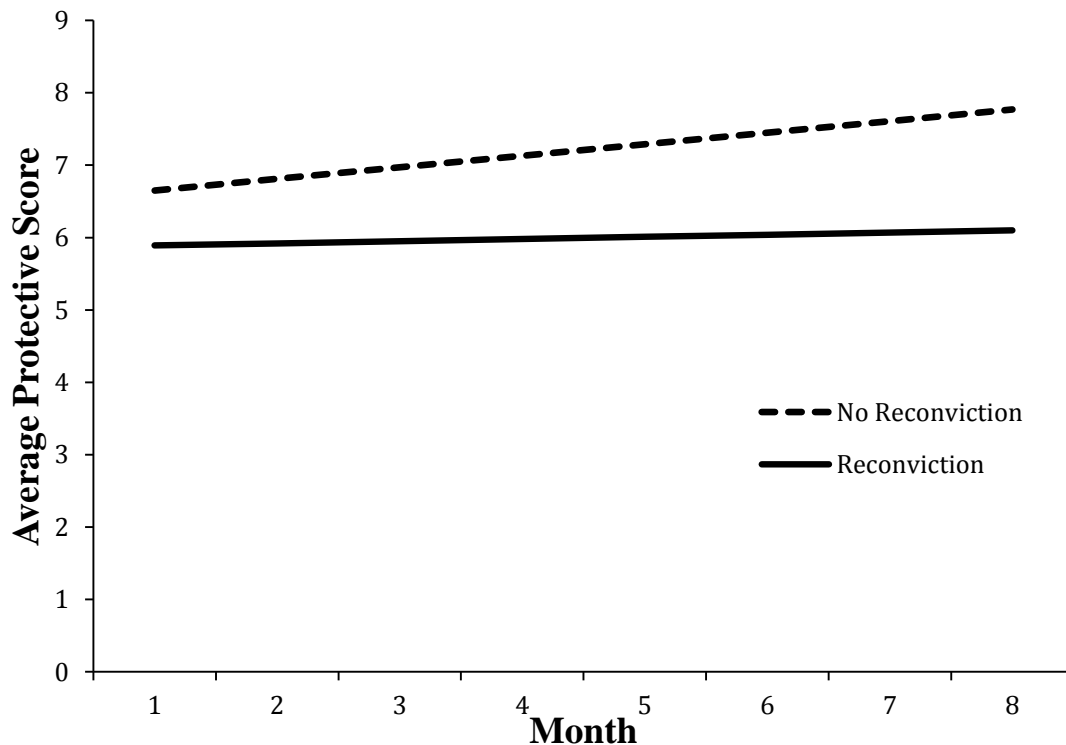
Figure 6. Change in average protective subscale scores over time by reconviction status.

DRAOR total score. As was to be expected, the pattern of change between the two groups was more pronounced for the total score, given that the total score is a composite of the three subscales. Those who committed an offence (any) that resulted in a reconviction had an average total score over the first month of 6.01($SE = 0.62$). This was 2.26 points higher than the average total score for the first month of youth who were not reconvicted ($M = 3.76$, $SE = 0.53$). The rate of change was also different between the two groups, with the total scores of those who were reconvicted decreasing by an average of 0.19 ($SE = 0.09$) points per month, compared to those who were not reconvicted, whose total scores decreased by 0.55 ($SE = 0.07$) per month ($p < .01$) on average. These results are depicted graphically in Figure 7.
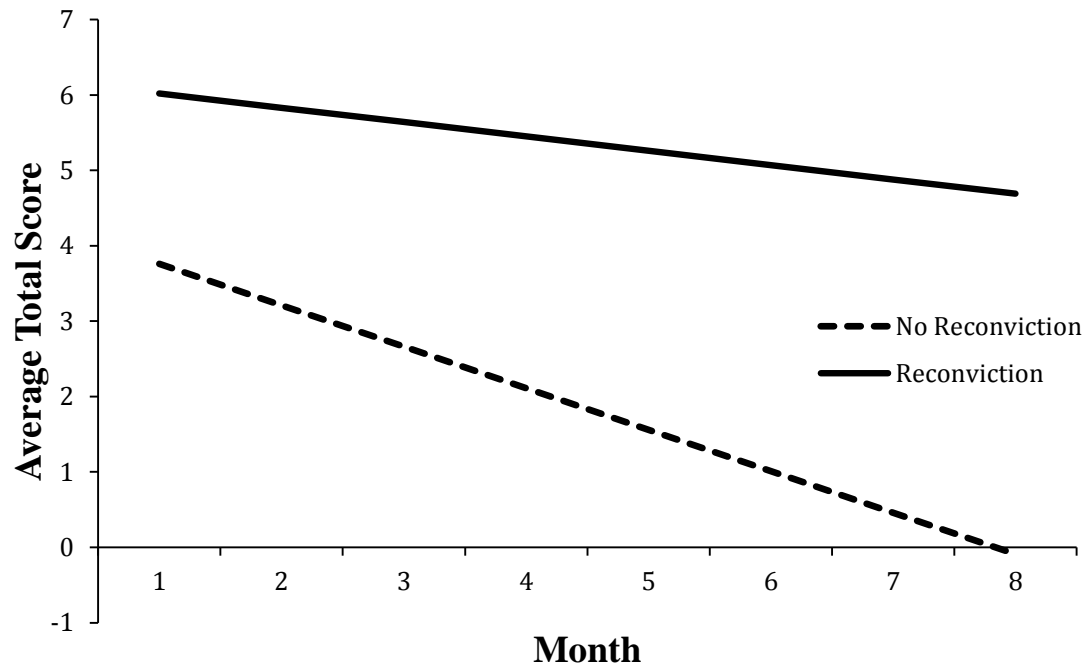
Figure 7. Change in average DRAOR total scores over time by reconviction status.

Summary. For the DRAOR total score, the stable subscale, and the protective subscale, there was a significant difference in the rate of change between youth who were reconvicted (any offence) and those who were not. For the stable subscale and total score, those who were reconvicted were not reducing their risk scores as quickly as those who were not reconvicted. For the protective subscale the youth who were reconvicted were not increasing their scores as quickly. For the acute subscale, although a trend was found in the expected direction, the difference in rates of change did not reach statistical significance; therefore, we cannot say confidently that the two groups differed in terms of how rapidly they improve on acute risk. A more comprehensive table of the results from the multilevel analyses can be found in Appendix B.

Chapter 4

Results: Violent Reconviction

How Well Do Initial DRAOR Scores Predict Violent Reconvictions in a Youth Sample?

As well as being interested in the DRAOR's performance for predicting any form of criminal recidivism, we were also curious as to how well the DRAOR would perform in the more specialised task of predicting violent recidivism. In order to assess this, we followed the same procedure as the analysis for any reconviction; however, the survival days were calculated up to a youth's first violent offence leading to a reconviction, and the status variable was the presence of a violent reconviction, rather than just any reconviction. The sample size for this analysis was also slightly larger ($N = 397$) than the sample for any reconvictions ($N = 369$), as fewer participants had been reconvicted for violent offences prior to their second DRAOR assessment. The mean initial DRAOR scores for this sample are reported in Table 16. These scores were statistically equivalent to the sample used in the analyses for any reconviction. No comparison was made to the sample used in Ferguson (2015), as there were no analyses run on violent reconvictions in the previous study.

Table 16

Mean Initial DRAOR Scores: Violent Reconviction

Score	*M* (*SD*)
Acute subscale	5.27 (2.37)
Stable subscale	6.10 (2.13)
Protective subscale	6.17 (2.05)
Total score	5.20 (5.26)

As was the case for the section above, we first wanted to assess the initial DRAOR subscales and total score for convergent validity. This was done through performing a Pearson bivariate correlation on all of the subscales and the total score.

It was found that there were moderate to large correlations between the subscales and the total score (see Table 17). The strongest subscale correlation was between the protective subscale and the stable subscale ($r = -.54$), indicating that those who had higher stable risk scores at their initial assessment also had lower protective scores. The total score was highly correlated with all subscales, which was expected, as the total score is a construct of the three subscales. Despite the high level of relatedness, the VIFs were all below 2.00, indicating multicollinearity was not likely to be an issue.

Table 17

Correlations Between Initial DRAOR Subscale Scores and Total Score: Violent Reconviction

	Acute	Stable	Protective	Total score
Acute subscale	1			
Stable subscale	.44**	1		
Protective subscale	-.42**	-.54**	1	
Total score	.79**	.82**	-.80**	1

** $p < .01$

Univariate predictive validity. We were interested in whether youth who scored higher on the initial DRAOR assessment were more likely to reoffend violently resulting in a reconviction than those who do not score as high. In order to assess this, the three initial DRAOR subscales and the initial total score were entered into the first block of four separate Cox regressions. The time variable was the number of survival days until a violent offence resulting in a reconviction (or data extraction for those who did not reoffend violently) and the status variable was the presence or absence of a violent reconviction.

Table 18

Univariate Cox Regression Models for Initial DRAOR Subscale and Total Scores
Predicting Violent Reconviction

Model for initial scores	β (SE)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.14 (.05)	8.89**	1.15 [1.05, 1.27]	.61** [.54, .68]
Stable subscale	.13 (.05)	5.48*	1.14 [1.02, 1.26]	.59* [.52, .66]
Protective subscale	-.19 (.05)	12.16**	0.83 [0.75, 0.92]	.61** [.54, .68]
Total score	.08 (.02)	13.40**	1.08 [1.04, 1.13]	.63** [.57, .70]

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

It was found that the regression models were significant for all subscales; acute, $\chi^2(1, N = 397) = 8.88$, $p < .01$; stable, $\chi^2(1, N = 397) = 5.47$, $p < .01$; protective, $\chi^2(1, N = 397) = 11.67$, $p < .01$; as well as the total score, $\chi^2(1, N = 397) = 13.25$, $p < .01$. Table 18 shows that the largest hazard ratio came from the protective subscale ($HR = 0.83$) indicating that a one-point increase on the initial acute subscale decreases a youth's risk of violent reconviction by 17%. The most accurate model was for the total score, which had an AUC of .63, indicating that a randomly selected youth who received a violent reconviction would have a 63% likelihood of having a higher initial DRAOR total score than a randomly selected youth who did not receive a violent reconviction.

Summary. From the initial univariate Cox regression models, it seems that all initial DRAOR scores are able to explain a significant amount of variance for violent reconvictions in a sample of community-sentenced youth. The amount of variance explained is comparable to that for any reconvictions.

Multivariate predictive validity. The next step in determining the validity of the initial DRAOR scores for predicting violent reconvictions was to look at the subscales together in a multivariate Cox regression. All three subscales were entered

into the first block of a Cox regression, with the time variable being the number of

survival days until a violent offence that led to a reconviction, and the status variable

being the presence or absence of a violent reconviction. The overall model was able to

significantly predict violent reconvictions, $\chi^2(3, N = 397) = 14.75, p < .01$. The model

also had a moderate level of accuracy ($AUC = .64$), being able to distinguish violent

recidivists from those who did not receive a violent reconviction 64% of the time (see

Table 19). As well as the overall model being significant, it was found that the

protective subscale was explaining a significant amount of variance not accounted for

by the other two subscales, showing a hazard ratio of 0.87 (95% CI [0.76, 0.99]). This

is in line with the results for any reconviction, again suggesting that the initial

protective subscale is capturing something distinct.

Table 19

Multivariate Cox Regression Model with Initial DRAOR Subscales Predicting Violent Reconviction

Model for initial scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.09 (.06)	2.80	1.10 [0.98, 1.22]	
Stable subscale	.00 (.07)	0.00	1.00 [0.88, 1.15]	.64** [.57, .70]
Protective subscale	-.14 (.07)	4.67*	0.87 [0.76, 0.99]	

*p < .05

Summary. Results suggest that the multivariate regression model is a

significant predictor of violent reconviction in a sample of youth serving community

supervision sentences. It was also found that the initial protective subscale was

explaining significant variance for violent reconvictions that was not explained by the

other two initial subscales.

How Well Do Proximal DRAOR Scores Predict Violent Reconvictions in a Youth

Sample?

As for any offending, we were interested in seeing how well a more up-to-date

DRAOR assessment would predict violent offending. The mean proximal scores for

violent offending are displayed on Table 20.

Table 20

Mean Proximal DRAOR Scores for Violent Reconviction

Score	M (SD)
Acute subscale	4.20 (2.36)
Stable subscale	5.57 (2.70)
Protective subscale	6.64 (2.67)
Total score	3.13 (6.56)

The proximal DRAOR scores for violent reconvictions were put into a

Pearson bivariate correlation to assess for convergent validity. The results of this

analysis are displayed in Table 21. As can be seen, there were large correlations

between all subscales and the total score, suggesting that all of the variables are

measuring a similar thing. Despite the high correlations, all VIFs were below 2.0,

indicating multicollinearity would not pose a problem.

Table 21

Correlations Between Proximal DRAOR Scores and Total Score for Violent Reconviction

	Acute subscale	Stable subscale	Protective subscale	Total score
Acute subscale	1			
Stable subscale	.51**	1		
Protective subscale	-.53**	-.69**	1	
Total score	.78**	.88**	-.88**	1

$**p < .01$

Univariate predictive validity. We first wanted to assess how well each

proximal score for the subscales and the total score predicted violent reconvictions

separate from the other factors. For this, each subscale and the total score were

entered into the first block of four different Cox regressions. The time variable was

entered as the survival days until a violent offence leading to a reconviction (or data

extraction for those who were not reconvicted for violence) and the status variable

was the presence or absence of a violent offence that led to a reconviction. The

outcomes of these Cox regressions can be seen in Table 22. Each model was found to

be significantly predicting violent reconvictions: acute, $\chi^2(1, N = 397) = 24.85, p <$

.01; stable, $\chi^2(1, N = 397) = 7.71, p < .01$; protective, $\chi^2(1, N = 397) = 10.72, p < .01$;

total score, $\chi^2(1, N = 397) = 18.69, p < .01$. The subscale that showed the highest

level of predictive power was the acute subscale, with a hazard ratio of 1.26. This

indicates that for every additional point on a youth's proximal acute subscale, they are

26% more likely to be reconvicted (violent offence). The AUCs for all of the

subscales were all comparable with highly overlapping confidence intervals,

suggesting that neither any one subscale nor the total score was any better than any

other subscale at distinguishing youth who were reconvicted for violence from those

who were not.

Table 22

Univariate Cox Regression Models for Proximal DRAOR Scores Predicting Violent Reconviction

Model for proximal scores	β (SE)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.23 (.04)	26.59**	1.26 [1.15, 1.37]	.68** [.61, .75]
Stable subscale	.12 (.04)	7.68**	1.13 [1.04, 1.23]	.61** [.54, .68]
Protective subscale	-.14 (.04)	10.73**	0.87 [0.80, 0.95]	.62** [.55, .68]
Total score	.08 (.02)	18.89**	1.08 [1.04, 1.12]	.66** [.59, .73]

**$p < .01$

Summary. The univariate Cox regressions show that the proximal DRAOR

subscales and total score were able to predict violent reconvictions in a sample of

youth serving community supervision sentences. Although the proximal acute subscale had the highest hazard ratio, it should be noted that there was overlap with the confidence intervals of the other subscales, though not for the total score.

Multivariate predictive validity. After establishing that all of the proximal subscales were able to predict violent reconvictions by themselves, it was important to see how well they performed together. The three subscales were entered into the first block of a Cox regression, with the time variable set as survival days until a violent offence resulting in a reconviction, and the status variable as the presence or absence of a violent reconviction. The overall model was significant, which was expected due to the fact that all of the covariates were predictive by themselves, $\chi^2(3, N = 397) = 25.51$, $p < .01$. It was also found that the acute subscale was predicting violent reconvictions above and beyond the other two subscales. When taking the variance explained by the other two subscales into account, a one-unit increase on the acute subscale corresponded with a 23% increase in the likelihood of reconviction (*HR*: 1.23; see Table 23). This result suggests that the acute subscale was explaining more variance than the other two subscales, which is in line with a theoretical assumption of the DRAOR, that the acute subscale is the most important when predicting imminent offending (Serin, 2015).

Table 23

Multivariate Cox Regression Model for Proximal DRAOR Scores Predicting Violent Reconviction

Model for proximal scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Acute subscale	.21 (.05)	14.53**	1.23 [1.11, 1.37]	
Stable subscale	-.01 (.06)	0.05	0.99 [0.87, 1.11]	.68** [.61, .75]
Protective subscale	-.05 (.06)	0.60	0.96 [0.85, 1.07]	

**$p < .01$

Summary. Overall, the proximal DRAOR scores are supported as being good predictors of violent reconvictions. The proximal acute subscale was explaining the most variance for violent reconvictions, even when the variance explained by the other proximal subscales was taken into account.

Do Proximal DRAOR Scores Outperform Initial DRAOR Scores for Predicting Violent Reconvictions?

As was the case for predicting any reconvictions, we wanted to see if the more up-to-date proximal scores would be better predictors of violent reconvictions than their initial counterparts. Firstly, a Pearson bivariate correlation was performed to assess for convergent validity between the initial and proximal scores. Each proximal score had a moderate or strong correlation with its initial partner, with the weakest correlation coming from the acute subscale (see Table 24 in bold). This suggests that a youth's scores are likely to be similar from initial assessment to proximal (e.g., if an individual scores highly on any subscale at the initial assessment, it is likely that they will score highly on that same subscale at the proximal assessment).

Table 24

Correlations Between Initial and Proximal DRAOR Scores: Violent Reconviction

	Initial acute	Initial stable	Initial protective	Initial total score
Proximal acute	**.44****	.21**	-.23**	.37**
Proximal stable	.35**	**.58****	-.38**	.55**
Proximal protective	-.36**	-.40**	**.58****	-.55**
Proximal total score	.45**	.48**	-.48**	**.58****

**p < .01

Incremental validity for initial and proximal DRAOR scores predicting violent reconviction. In order to assess whether the proximal DRAOR scores are better predictors of violent reconvictions than their initial counterparts, four

multivariate Cox regressions were used. For each subscale and for the total scores, both the initial and the proximal scores were entered into the first block of four separate Cox regressions. The time variable was survival days until a violent offence that led to a reconviction (or data extraction for those who did not receive a violent reconviction), and the status variable was the presence or absence of a violent reconviction.

All four models containing both the initial and the proximal scores for a particular subscale or for the total score were found to be significant: acute, $\chi^2(2, N = 397) = 26.06$, $p < .01$; stable, $\chi^2(2, N = 397) = 8.52$, $p < .05$; protective, $\chi^2(2, N = 397) = 14.46$, $p < .01$; total score, $\chi^2(2, N = 397) = 20.92$, $p < .01$. This was to be expected, as all of the univariate models for both initial and proximal scores were found to be significant as well.

In terms of whether the proximal scores are better at explaining variance than the initial scores, the picture is mixed. For the stable and protective subscales, the proximal scores were not performing better than the initial scores, as can be seen by the nonsignificant Wald scores and the confidence intervals of the hazard ratios including 1.00 (see Table 25). In terms of the acute subscale and the total score, both of the proximal scores received significant Wald scores and the confidence intervals of their hazard ratios did not include 1.00. It should be noted, however, that for both the acute subscale and the total score, the confidence intervals for the hazard ratios of the initial and proximal scores overlap each other. This means that we cannot rule out the fact that the initial and proximal scores may be explaining an equal amount of variance.

Table 25

Multivariate Cox Regression Model with Initial and Proximal DRAOR Scores Predicting Violent Reconviction

Multivariate model for initial and proximal scores	β (SE)	Wald	Hazard ratio [95% CI]
Acute score			
Initial	.06 (.05)	1.21	1.06 [0.96, 1.17]
Proximal	.21 (.05)	18.41**	1.23 [1.12, 1.35]
Stable score			
Initial	.06 (.07)	0.81	1.06 [0.93, 1.21]
Proximal	.09 (.05)	3.02	1.10 [0.99, 1.21]
Protective score			
Initial	-.13 (.07)	3.83	0.88 [0.77, 1.00]
Proximal	-.09 (.05)	2.79	0.92 [0.83, 1.02]
Total score			
Initial	.04 (.03)	2.24	1.04 [0.99, 1.10]
Proximal	.06 (.02)	7.69**	1.06 [1.02, 1.10]

**$p < .01$

Summary. Each of the individual models was found to be a significant predictor of violent recidivism. However, only two of the proximal scores were found to be potentially performing at a greater level than the initial scores – the acute subscale and the total score. Unfortunately, we cannot be certain of those two scales' better performance due to overlapping confidence intervals of the hazard ratios for the initial and the proximal scores.

Do DRAOR Scores Change From Initial to Proximal Assessment and Are Those Changes Indicative of Violent Reconviction Outcomes?

As for any reconviction we wanted to evaluate the dynamic nature of the DRAOR, this time using the violent reconviction initial and proximal scores. Firstly, we examined the amount of change made between initial and proximal assessments

for each subscale as well as the total score. In order to assess the amount of change

made, paired-samples *t*-tests were conducted. It was found that there was a significant

change in the direction of reduced risk (or enhanced protective factors) for all three

subscales: acute, $t(396) = 8.52$, $p < .01$, $d = 0.86$, $M_{diff} = $ -1.07, 95% CI [0.74, 1.40];

stable, $t(396) = 4.66$, $p < .01$, $d = 0.47$, $M_{diff} = $ -0.53, 95% CI [0.19, 0.87]; protective,

$t(396) = 4.16$, $p < .01$, $d = 0.42$, $M_{diff} = $ 0.47, 95% CI [-0.80, -0.14]; and for the total

score, $t(396) = 7.48$, $p < .01$, $d = 0.75$, $M_{diff} = $ -2.07, 95% CI [1.24, 2.90].

Table 26

Mean Change Made from Initial to Proximal DRAOR Assessments: Violent
Reconviction

	Mean initial score (*SD*)	Mean proximal score (*SD*)	Mean change score (*SD*)	Range of change scores	No change (%)
Acute	5.27 (2.37)**	4.20 (2.36)	1.07 (2.51)	[-8, 8]	19.6
Stable	6.10 (2.13)**	5.57 (2.70)	0.53 (2.26)	[-7, 8]]	45.1
Protective	6.17 (2.05)	6.64 (2.67)**	-0.47 (2.23)	[-8, 8]	42.6
Total score	5.20 (5.26)**	3.13 (6.56)	2.07 (5.51)	[-21, 22]	13.1

**$p < .01$

Table 26 provides the mean change score (an inversion of the mean change

from initial to proximal, where a positive score indicates a reduction on that subscale

from initial to proximal), the range of change scores for the sample, and the

percentage of the sample who did not change from initial to proximal assessment for a

given subscale (or total score). As for the analyses concerning any reconviction, the

subscale with the most participants changing was the acute subscale (only 19.6% of

participants did not change). Both the stable and protective subscales had close to half

of the sample showing no change from initial to proximal assessments. This

reinforces the idea that the acute subscale is the more dynamic of the DRAOR

subscales.

Does Change Predict Violent Reconviction Outcomes in a Youth Sample?

Again, it was important to investigate the relationship between initial DRAOR scores and the amount of change a youth made by their proximal assessment for violent reconvictions. First, the initial scores for each subscale and for the total score were entered into a Pearson bivariate correlation with the change scores. As can be seen in Table 27, the correlations were small for the stable subscale ($r = .25$), the protective subscale ($r = .23$), and the total score ($r = .26$). The correlation for the acute subscale was large ($r = .54$). These positive correlations indicate that those whose initial risk scores were high showed more change in the direction of a reduction in risk. For the protective subscale the positive correlation indicates that the higher the initial score the more change in the direction away from improved protective factors was shown (as the protective subscale is reverse scored). Again, this is intuitive as the starting point for a youth dictates how much he or she can move in a certain direction. These results are similar to those found when looking at the change for any reconviction, but they are slightly stronger in the case of violent reconviction.

Table 27

Correlations Between Initial DRAOR Subscales and Total Score and DRAOR Change Scores: Violent Reconviction

	Initial acute	Initial stable	Initial protective	Initial total
Acute change	**.54****	.22**	-.17**	.40**
Stable change	-.01	**.25****	-.05	.12*
Protective change	.05	-.02	**.23****	-.07
Total change	.22**	.21**	-.19**	**.26****

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

After conducting the correlations, the next step was to assess the predictive power of the DRAOR change scores. As shown by the correlations above, a youth's initial score limits the change score, therefore we opted to control for initial scores in our next analysis. Four separate multivariate Cox regressions were performed with the

initial score entered into the first block, and the corresponding change score entered into the second block. The time variable was survival days until a violent offence that led to a reconviction (or data extraction for those who did not receive a reconviction for violence), and the status variable was the presence or absence of a violent reconviction. Unlike for any reconvictions, not all of the models were significant for violent reconvictions. Neither the stable subscale model, $\chi^2(2, N = 397) = 3.05, p = .08$, nor the protective subscale model, $\chi^2(2, N = 397) = 2.79, p = .10$, reached significance for predicting violent reconvictions. However, the acute subscale model, $\chi^2(2, N = 397) = 17.29, p < .01$, and the total score model, $\chi^2(2, N = 397) = 7.67, p < .01$, were found to be significantly predicting violent recidivism.

The strongest hazard ratio for a change score came from the acute change score ($HR = 0.81$; see Table 28), indicating that for every one-point shift in the direction of reduced risk, youth were 19% less likely to receive a violent reconviction. Looking at the accuracy of the models, it seems that all four were capable of distinguishing between youth who were reconvicted for violence and those who were not with a medium effect size (all AUCs fall between .62 and .68; see Table 28). For the total score, as an example, the model suggests that a randomly selected youth who did not receive a violent reconviction would have a 67% likelihood of having a larger change score than a youth who did receive a violent reconviction.

Table 28

Multivariate Cox Regression Models Containing Change Scores Controlling for Initial DRAOR Scores Predicting Violent Reconviction

Model for change scores	β (*SE*)	Wald	Hazard ratio [95% CI]	AUC [95% CI]
Initial acute score	.26 (.06)	22.76**	1.30 [1.17, 1.45]	.68**
Acute change score	-.21 (.05)	18.41**	0.81 [0.74, 0.89]	[.62, .75]
Initial stable score	.15 (.06)	7.27**	1.16 [1.04, 1.30]	.62**
Stable change score	-.09 (.05)	3.02	0.91 [0.82, 1.01]	[.55, .68]
Initial protective score	-.21 (.06)	14.21**	0.81 [0.72, 0.90]	.63**
Protective change score	.09 (.05)	2.79	1.09 [0.99, 1.21]	[.57, .70]
Initial total score	.10 (.02)	18.14**	1.10 [1.05, 1.15]	.67**
Total score change score	-.06 (.02)	7.69**	0.94 [0.91, 0.98]	[.60, .73]

**$p < .01$

Summary. The above Cox regressions suggest that, for youth serving a community supervision sentence, the amount of change made on the acute subscale and on the total score, between initial and proximal assessments, was predictive of violent reconvictions. For the stable and protective subscales, this was not the case. This pattern was in contrast to the prediction of any reconvictions, where all four models were found to be significant.

Do Violent Recidivists Have a Different Rate of Change to Those Who Are Not Reconvicted for Violence?

After finding that two of the change scores assessed were predictive of violent reconvictions, the next step was to see if violent recidivists showed a different rate of change to those who were not reconvicted for violence. We first split the sample into two groups, those who were reconvicted for a violent offence before data extraction ($n = 76$), and those who were not ($n = 321$). An independent-samples *t*-test was run to

examine the difference between the two groups on the four change scores. It was

found that the only significant difference came from the acute change score, $t(395) =$

1.97, $p < .05$, $d = 0.20$, $M_{diff} = 0.63$, 95% CI [0.00, 1.25]. No significant differences

were found for the stable subscale, $t(395) = 1.14$, $p = .26$, $d = 0.11$, $M_{diff} = 0.33$, 95%

CI [-0.24, 0.90]; the protective subscale, $t(395) = 0.77$, $p = .44$, $d = 0.08$, $M_{diff} = -0.22$,

95% CI [-0.78, 0.34]; or the total score, $t(395) = 1.67$, $p = .10$, $d = 0.17$, $M_{diff} = 1.17$,

95% CI [-0.20, 2.55]. The mean changes for each group from initial to proximal

assessment are displayed visually in Figure 8.
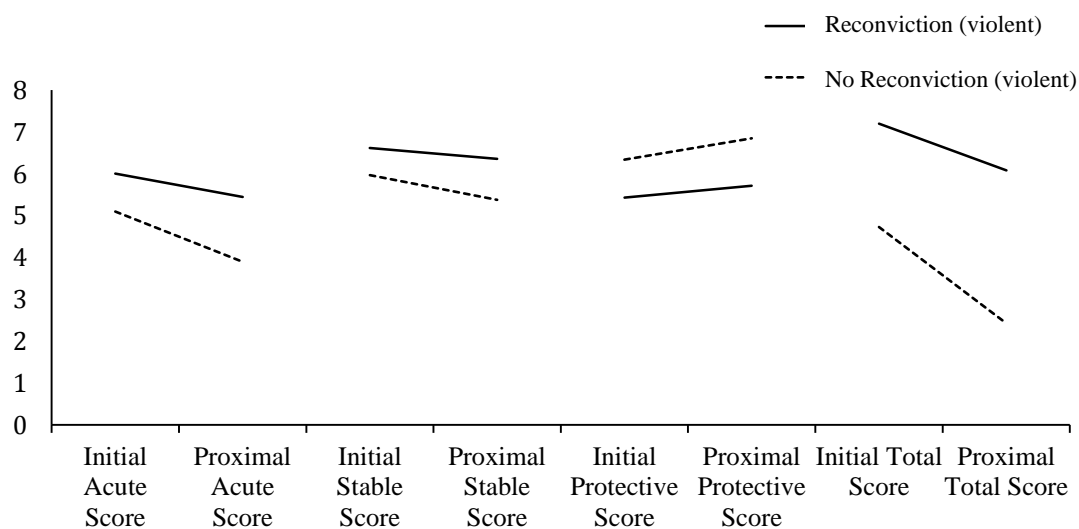
Figure 8. Mean scores at initial and proximal DRAOR assessment for recidivists and

non-recidivists: Violent reconviction.

As can be seen in Figure 8 and from the *t*-test results, there is a less obvious

difference between the two groups than there was for any reconvictions. Again, an

independent-samples *t*-test was run to examine the difference in the time between

initial and proximal assessments for the two groups, as the significant result for the

acute subscale may have arisen due to less time for those who were reconvicted for

violence to change. It was found that those who were not reconvicted for a violent

offence had a longer period of time from their initial DRAOR assessment to their

proximal ($M = 210.97$, $SD = 74.82$) than those who were reconvicted for violent

offending ($M = 154.38$, $SD = 10.66$). This difference was significant, $t(99.26) = 4.95$,

$p < .01$, $d = 0.99$, $M_{diff} = 56.59$, 95% CI [33.88, 79.30]. As the multilevel modelling

analyses that were done for any reconvictions were exploratory in nature, and the

DRAOR was not designed for the prediction of violent reconvictions, we opted not to

conduct further analyses with multilevel modelling for violent reconvictions.

Summary. For the stable risk subscale, the protective subscale, and the total

score, violent reoffence status did not significantly impact the rate of change that

youth made from initial assessment to proximal assessment. There was a significant

difference between those who were reconvicted for violent offending and those who

were not in acute subscale score change, but this should be considered with caution

due to the large difference in time from initial to proximal assessments between the

groups. It is unsurprising that these results for violent reconvictions are not as

impressive as those found for any reconvictions, because the DRAOR is not

specifically designed to assess risk of violent reconviction.

Chapter 5

Discussion

This research evaluated how well a dynamic risk assessment tool (the DRAOR) was able to predict future reconvictions, for both any reconvictions and violent reconvictions, in a sample of NZ youth serving community supervision sentences. The study also explored how changes in scores over time influenced these predictions and how those who were reconvicted differed from those who were not in terms of their rates of change. Although both predicted reconviction, it was found that more up-to-date assessments (proximal) more accurately predicted recidivism than the earlier assessments (initial). It was also found that those who were reconvicted (any offence) have lower rates of change on their DRAOR scores over time than those who were not reconvicted.

How Well Do Initial DRAOR Assessments Predict Reconvictions?

The first research question concerned whether a DRAOR assessment, taken near the beginning of a youth's sentence (initial assessment), would predict future offending for our sample of NZ youth. It was found that all three DRAOR subscales and the total score were able to predict reconvictions (for both any reconvictions and violent reconvictions) but that the protective subscale was the best predictor for both violent and any reconvictions. This means that youth who had higher risk scores and lower protective scores near the beginning of their sentence were more likely to be reconvicted (any reconviction or violent reconviction) than those who had lower risk scores and higher protective scores; however, the level of accuracy was not very high.

These results are somewhat different to Ferguson (2015), who only looked at the DRAOR's prediction of any criminal offending for NZ youth, and found only the DRAOR total score to be a significant predictor of future reconvictions, with the three

subscales failing to reach significance. In saying that, Ferguson (2015) did find trends in the same direction as the current study's results, so the discrepancies between these two studies could be an artefact of a small sample size in the previous study.

Ferguson (2015) found that the stable subscale was the most informative of the initial subscale scores when predicting any criminal offence for youth. In contrast, the current study found that the protective subscale was the strongest initial predictor of the three subscales, giving weight to the argument for the inclusion of protective factors in dynamic risk assessment tools (Serin et al., 2016). This was supported by our multivariate regression analysis, which found the protective subscale explained significantly more of the variance than the other two subscales when contributions from all subscales were accounted for. This result is in line with the Central Eight risk factors of Andrews and Bonta (2010), as a number of the items on the protective subscale are practically opposite to what are considered to be the most important risk factors for predicting offending (e.g., antisocial attitudes, antisocial associates, family/ marital situation, leisure/recreation).

Both Ferguson (2015) and Scanlan (2015) found that the initial acute subscale, not the protective, was the strongest predictor of reconvictions when assessing adult males. This could suggest differences in how the protective subscale of the DRAOR performs with youth. Although more research will need to be done, it could be that the protective factors assessed by the DRAOR have more of an impact for youth than for adults. Sampson and Laub (2005) theorised that protective factors can act like turning points for adolescents, giving them a reason to steer away from crime. In contrast, for adults, many of these opportunities for turning points have already passed and thus protective factors may not have as much of an impact. This finding also fits with the adolescent-limited theory of offending (Moffitt, 1993), as youth who

are theorised to be less likely to continue offending are proposed to have more protective factors (e.g., social support; Moffitt et al., 2002).

How Well Do Proximal DRAOR Assessments Predict Reconvictions?

The second research question addressed whether a more up-to-date DRAOR assessment (proximal assessment) would be more indicative of future behaviour than the initial assessments. The results indicate that proximal DRAOR assessments were able to predict both violent and any reconvictions in a sample of NZ youth.

The proximal DRAOR assessments were greater predictors of reconvictions for any new criminal offence than the initial assessments. This was true for both risk subscales and the total score, but there were slightly overlapping confidence intervals when analysing the contributions of the initial and proximal protective subscale scores. So, although the regression analysis found a significant difference between the two assessments for the protective subscale, we cannot say with confidence that more up-to-date assessments of protective factors are better than initial assessments for predicting future criminal offences in a NZ youth sample. The strongest predictor of recidivism from the proximal assessments was the acute subscale, meaning that the acute subscale was able to give the most information relevant to any reconvictions compared to the other subscales of the DRAOR. This fits with the purpose of the acute subscale, as it was developed to be the most rapidly changeable of the subscales (Serin, 2015), so it is promising that the more up-to-date an acute assessment is, the better a predictor it is. This also fits with the current literature around youth specific offending, as the acute subscale contains a number of items that are theorised to be the most informative for recidivism with youth (e.g., substance abuse, employment, interpersonal relationships; Cuervo & Villanueva, 2015).

The results when examining proximal DRAOR assessments for predicting violent reconvictions were not as definitive as for any reconvictions. Only the acute subscale and the DRAOR total score showed improvement from initial to proximal assessments in terms of predictive power. The confidence intervals for the hazard ratios between the two assessments overlapped as well, so our confidence in these results is not strong. The fact that the DRAOR is not as good at predicting violent offending as it is for any offending is not surprising given the DRAOR was designed as a general risk assessment tool, rather than a violent risk assessment tool (Serin, 2015).

The current study's results were consistent with Ferguson (2015) in that the proximal measurements were stronger predictors than the initial assessments of future reconvictions. The reason that the initial scores were not as good at predicting reconvictions could have arisen, in part, from the fact people serving community sentences are given the opportunity to undergo rehabilitation for their offending (Department of Corrections, n.d). Any benefits from said rehabilitation would not be captured in the initial assessment but would be seen in the proximal assessment, thus making for a more accurate measurement. This can be supported by the fact the youth in this study had lower risk and greater protective scores at the proximal assessments compared to the initial assessments.

The above findings mean that we can now be more confident in using the DRAOR to assess risk of any reconvictions for NZ youth (17-19 years) serving community sentences. This is reassuring given that the DRAOR has been used for youth assessments for the last few years. There is still more work to be done in order to better understand how well different subscales of the DRAOR perform with this population, but knowing that the DRAOR has a moderate to high ability to predict

any reconvictions in youth, so long as the assessments are kept up-to-date, is promising. In terms of the prediction of violent reconvictions with youth, the DRAOR did not perform as well, with only a low to moderate level of predictive validity. This suggests that if the main concern for a youth is that they will reoffend violently, it may be beneficial to use a more specific violence risk assessment tool (e.g., SAVRY; Borum et al., 2002).

The finding that different subscales of the DRAOR perform with varying levels of accuracy depending on how up-to-date an assessment was (e.g., initial versus proximal) will help in terms of interpreting a youth's scores. If the assessment that is available is quite out of date, then it may be useful to pay more attention to the protective subscale score, whereas if the assessment is quite recent (more proximal) then the acute subscale should be given more weight. This finding applied to both the prediction of any reconvictions and violent reconvictions. With that being said, efforts should still be made to keep DRAOR assessments up to date, since more recent assessments are more informative.

Do DRAOR Scores Change Over Time and Are Changes Predictive of Reconvictions?

As the DRAOR is said to be a *dynamic* risk assessment tool, it was important to evaluate how people's scores can change. We first created change scores for each individual by looking at the total amount of change made on each subscale and the total score from their initial assessment to their proximal assessment (note that different proximal assessments were used for violent reconvictions compared to any reconvictions due to different offence dates). Firstly, it was found that when looking at the assessments for both violent reconvictions and any reconvictions, the youths'

scores on the DRAOR changed significantly between initial and proximal assessments, in the direction of reduced risk (or increased protective factors).

When looking at any reconvictions, the amount of change made between youths' initial assessment and proximal assessment was found to be predictive of reconvictions after controlling for initial scores for all subscales and for the DRAOR total score. This result was in line with the findings from Ferguson (2015), and another DRAOR study looking at an adult population of males and females (Scanlan, 2015).

For violent reconvictions, the results were not as promising. Although it was found that changes made on the acute subscale and the DRAOR total score were predictive of violent reconvictions after controlling for initial DRAOR scores, the results for the stable and the protective subscales did not indicate that changes made on these subscales predict violent reconvictions for youth. The less promising result for the prediction of violent offending with the DRAOR may be due again to the fact that the DRAOR was not designed to look specifically at violent offending (Serin, 2015), despite a similarity of risk factors between the two types of offending.

The amount of research looking at how change on dynamic risk assessment tools is related to recidivism is currently limited, with even less looking into youth change, but the majority of studies suggest that there is utility in looking at how people's risk scores change. This study's findings, as well as much of the extant literature (e.g., Howard & Dixon, 2013; Olver et al., 2007), suggest that the more someone's risk moves in the direction of reduced risk and increased protective factors, the less likely they are to be reconvicted. This is intuitive, as a lower risk score is intended to signify a lower likelihood of offending. With more research, especially in the area of youth risk assessment, we may be able to improve our

predictions of future offending through a better understanding of what changes on dynamic tools may mean.

Do Those Who Were Reconvicted Differ From Those Who Were Not Reconvicted in Terms of How Their DRAOR Scores Change Over Time?

After finding the amount of change a youth made on the DRAOR from their initial to their proximal assessment was predictive of reconvictions, we were next interested in how those who were reconvicted differed from those who were not. When looking at any reconvictions, it was found there was a significant difference in the amount of change made between those who were reconvicted (any offence) compared to those who were not reconvicted, for all subscales as well as the total score. For violent reconvictions this was not the case, with only the acute subscale showing a significant difference for the two groups.

An issue with a lot of studies that have looked at how change on dynamic risk assessment tools is related to reoffending is that usually only two time points are taken into account and often time is not adequately controlled for (e.g., Hanson et al., 2007; Howard & Dixon, 2013; Olver et al., 2007). This can be especially problematic, as was the case for this study, if the length of time between assessments varies between individuals. In order to overcome the issue of inconsistent times at risk, this study used multilevel modelling to control for time while looking at changes in dynamic risk each month.

For the stable subscale, the protective subscale, and the total score, it was found that those who were ultimately reconvicted (any offence) were showing a slower rate of change, on average, compared to those who were not reconvicted. This was not the case for the acute subscale though, as the difference was not strong enough to meet statistical significance. The non-significant finding for the acute

subscale was surprising as theoretically it should be the most dynamic of the subscales, therefore one would expect change on this scale to be the most informative. However, the fact that the acute subscale is so dynamic may have been why we did not find a significant difference between the two groups, as there may have been too many fluctuations in scores to fit the data to a linear model. Future studies may want to consider looking into curvilinear models in order to better understand how changes on the acute subscale relate to reconvictions.

Interestingly, Hanby (2013) found different results when looking at adults' changes in DRAOR scores over time. Rather than a non-significant finding for the acute subscale, the acute subscale was found to be the only subscale to show a difference in rates of change. More research will need to be conducted to clarify whether this difference was due to the current research using youth rather than adults, but there has been speculation that changes made by youth resulting in desistance from crime could be very distinct from adult change (Serin & Lloyd, 2009). One other possibility is that some of the items on the acute subscale might fluctuate a lot with youth regardless of whether they continued to offend (e.g., anger, negative mood; Larson, Csikszentmihalyi, & Graef, 1980). It has also been speculated that internal factors such as negative mood may be very difficult for probation officers to rate, thus change on these items may be hard to score (Jones, Brown, & Zamble, 2010). There is also the fact that our sample were at an age where engagement in antisocial activity is statistically normative (Adolescent-limited; Moffitt, 1993), compared to an adult offending sample where a large group would fit the Life-course-persistent mould (Moffitt, 1993). It could be that different rates of change on the more dynamic items are more informative for those who are likely to engage in offending for a long period

of time, compared to those who are less likely to continue offending into their adult years.

The findings on the rates of change for DRAOR scores over time could provide valuable information to probation officers who are monitoring youths' progress. If one of the youth on their caseload was not showing improvements in their DRAOR scores as quickly as others, then it could suggest they were more likely to be reconvicted (any offence). This could provide an opportunity for the probation officer to intervene early and give the youth the support they need before any further offending is committed. This would not only help the youth to avoid trouble, but also reduce the potential harm a reoffence could cause to society. Not only do the findings on change in risk have clinical relevance, but also from a research perspective, they have filled a gap in our understanding about what changes in dynamic risk might mean.

Limitations

Although this study highlighted a number of important aspects as to how the DRAOR performs when assessing youth in NZ, there are some limitations that should be considered. Firstly, the data used for this project were not originally collected for the purpose of research; instead, probation officers collected the data while scoring the DRAOR as part of their normal practice. A notable issue with the collection method was that the timing and frequency of DRAOR assessments for each participant were not uniform. Some participants may have had three assessments in one week, whereas others could have had a period of three months without an assessment. This posed a problem when it came to analysing the proximal assessments, as some youths' proximal assessments were a lot closer to the reoffence or end of sentence than others'. In order to fix this issue, future researchers may want

to consider having a uniform frequency of DRAOR assessments (e.g., once a fortnight) for all individuals throughout the duration of the study. Also, a lot of information that would have been helpful to assess, such as the interrater reliability of the probation officers, was not available.

This research did not include a comparison group of adults. Validating a risk assessment tool for different populations is extremely important; however, results are made more informative when a comparison to the intended population is used. Ferguson (2015) compared the predictive validity of the DRAOR between youth and adult males, but the matching criteria used resulted in a small sample of NZ youth from the available dataset. Because of the difficulty in getting a large representative sample of NZ youth to match an adult sample from our dataset, we opted to explore only the DRAOR's predictive validity with a larger NZ youth sample, without an adult comparison group. It should also be noted that there have already been a number of studies that have looked into the validity of the DRAOR with an adult population (e.g., Hanby, 2013; Scanlan, 2015; Yesberg & Polaschek, 2014), so the main purpose of this study was to replicate the findings for youth from Ferguson (2015).

Future Directions

In addition to the recommendations made above to overcome some of the limitations of this study, there are a number of directions that future researchers may want to explore. One area that would be very interesting to explore is how the DRAOR performs for youth at the item level. Although we were able to examine the DRAOR's performance at the factor level, we are still uncertain as to how each item performs. It is possible that some items are far more informative than others for youth, and thus we may be able to refine the DRAOR if that is the case. Previous research has also found that the DRAOR better fits a four-factor model when used to

assess high-risk NZ adults (Yesberg & Polaschek, 2014), as opposed to the currently used three-factor model. Future studies may want to conduct a factor analysis on the DRAOR with a youth sample to determine how well the current three-factor model works with youth, and whether a different factor structure may be beneficial for the prediction of youth offending.

Another area that would be interesting to look into further is to do with rates of change. The procedure used in this research created a number of linear models to depict the changes youth make in their DRAOR scores over time. Future studies may wish to look into curvilinear models to see if there is variation in the rates of change over time (e.g., if some youth improve quickly followed by a period of no change, compared to someone who may have a more erratic path of ups and downs). It could also be informative to look into models that take time until a reconviction or until the end of the sentence into account, as opposed to this study, which looked into time from the beginning of the sentence. By looking at how DRAOR scores change leading up a youth's proximal assessment, we may be able to establish if there are any changes made right before a reconviction occurs (e.g., a spike in their DRAOR score).

Other researchers may also want to look into how well the DRAOR performs with different youth populations, such as those from different countries. Seeing as the DRAOR is beginning to be used in both the US and Canada, it is important that this research is replicated in those respective countries, as the cultural differences may influence how well the DRAOR performs with youth around the world.

Conclusion

The results from this research, as well as those from Ferguson (2015), support the use of the DRAOR with youth in NZ on community supervision sentences. However, the DRAOR should continue to be validated not only for youth, but for

other populations as well. Since the DRAOR is used across NZ, as well as in a number of other countries, it is crucial that we increase our understanding of how the DRAOR performs. With a better understanding we can not only improve our confidence in its use, but also potentially refine the tool for different populations if it is found that particular items are better indicators of future behaviour for certain groups.

The assessment of youths' risk of reconviction is an important area that is often overlooked, with most risk assessment research only looking at very young people or adults, neglecting those who fall in between. It is especially important to understand and monitor older youths' levels of risk, given the high rates of crime for those in late adolescence (Moffitt, 1993). This research has managed to validate the DRAOR's use with older youth (17-19 years) serving community supervision sentences in NZ, which will allow for confident use of the DRAOR with this population in the future.

Cast your mind back to Ben who was introduced at the beginning of this thesis. Regardless of whether Ben was 17 or 40 years old we can still be confident in using the DRAOR to predict his risk of reconviction. This is important due to the potentially severe consequences of an inaccurate risk assessment. Not only will we be able to provide the correct amount of monitoring for Ben, but with our new understanding of rates of change for the DRAOR, we may also be able to intervene if and when necessary, and reduce the likelihood of further offences.

References

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory - Revised*. Toronto: Multi-Health Systems.

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. New Providence, NJ: LexisNexis/ Anderson Publishing.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency, 52*, 7-27. doi: 10.1177/0011128705281756

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., . . . Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment, 9*, 449-461. doi: 10.2147/NDT.S39776

Bakker, L. W., Riley, D., & O'Malley, J. (1999). *ROC, risk of reconviction: Statistical models predicting four types of re-offending*. Wellington, NZ: Department of Corrections Psychological Service.

Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse: Journal of Research and Treatment, 19*, 135-153. doi: http://dx.doi.org/10.1177/107906320701900205

Bonta, J., Rugge, T., Scott, T.-L., Bourgon, G., & Yessine, A. K. (2008). Exploring the black box of community supervision. *Journal of Offender Rehabilitation, 47*, 248-270. doi: http://dx.doi.org/10.1080/10509670802134085

Borum, R. (2000). Assessing violence risk among youth. *Journal of Clinical Psychology, 56*, 1263-1288.

Borum, R. (2003). Managing at-risk juvenile offenders in the community: Putting

evidence-based principles into practice. *Journal of Contemporary Criminal

Justice, 19*, 114-137. doi: http://dx.doi.org/10.1177/1043986202239745

Borum, R., Bartel, P., & Forth, A. (2002). *Manual for the structured assessment of

violence risk in youth (SAVRY)*. Tampa, FL: University of South Florida.

Caudy, M. S., Durso, J. M., & Taxman, F. S. (2013). How well do dynamic needs

predict recidivism? Implications for risk assessment and risk reduction.

Journal of Criminal Justice, 41, 458-466. doi:

http://dx.doi.org/10.1016/j.jcrimjus.2013.08.004

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155.

Cuervo, K., & Villanueva, L. (2015). Analysis of risk and protective factors for

recidivism in Spanish youth offenders. *International Journal of Offender

Therapy and Comparative Criminology, 59*, 1149-1165. doi:

10.1177/0306624X14557917

de Ruiter, C., & Nicholls, T. L. (2011). Protective factors in forensic mental health: A

new frontier. *The International Journal of Forensic Mental Health, 10*, 160-

170. doi: http://dx.doi.org/10.1080/14999013.2011.600602

de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. A. (2011).

Assessing protective factors in forensic psychiatric practice: Introducing the

SAPROF. *The International Journal of Forensic Mental Health, 10*, 171-177.

doi: http://dx.doi.org/10.1080/14999013.2011.600230

De Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2015).

Changes in dynamic risk and protective factors for violence during inpatient

forensic psychiatric treatment: Predicting reductions in postdischarge

community recidivism. *Law and Human Behavior, 39*, 53-61. doi:

http://dx.doi.org/10.1037/lhb0000089

Department of Corrections. (2013). *Trends in the offender population (2013 calendar

year).* Wellington, New Zealand: Retrieved from

http://www.corrections.govt.nz/__data/assets/pdf_file/0012/738939/Trends_in

_the_Offender_Population_2013_updated.pdf.

Department of Corrections. (n.d). Supervision. from

http://www.corrections.govt.nz/working_with_offenders/community_sentence

s/sentences_and_orders/supervision.html

Department of Corrections. (n.d.). Young offenders. from

http://www.corrections.govt.nz/resources/best-use-of-psychological-service-

treatment-resources/young_offenders.html

Dernevik, M., Beck, A., Grann, M., Hogue, T., & McGuire, J. (2009). The use of

psychiatric and psychological evidence in the assessment of terrorist

offenders. *The Journal of Forensic Psychiatry & Psychology, 20*, 508-515.

Douglas, K. S., & Kropp, R. P. (2002). A prevention-based paradigm for violence risk

assessment clinical and research applications. *Criminal Justice and Behavior,

29*, 617-658.

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific

about being dynamic. *Psychology, Public Policy, and Law, 11*, 347-383. doi:

http://dx.doi.org/10.1037/1076-8971.11.3.347

Farrington, D. P., & Loeber, R. (2002). Serious and violent juvenile offenders. In M.

K. Rosenheim, F. E. Zimring, D. S. Tanenhaus & B. Dohrn (Eds.), *A century

of juvenile justice*. Chicago, IL: The University of Chicago Press.

Ferguson, A. (2015). *An age-old issue: Evaluating the applicability of adult criminal risk assessment tools for use with youth offenders*. Unpublished master's thesis. Victoria University of Wellington, Wellington, NZ.

Garson, G. D. (2012). *Life Tables and Kaplan-Meier Analysis.* Asheboro, NC: Statistical Associates Publishers.

Garson, G. D. (2013). *Cox Regression*. Asheboro, NC: Statistical Assosciates Publishers.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30. doi: http://dx.doi.org/10.1037/1040-3590.12.1.19

Hanby, L. (2013). *A longitudinal study of dynamic risk, protective factors and criminal recidivism: Change over time and the impact of assessment timing.* Unpublished doctoral thesis. Carleton University, Ottawa, Canada.

Hanson, R. K., Harris, A. J., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Vol. 5): Public Safety Canada Ottawa.

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene?: Dynamic predictors of sexual assault recidivism. *Criminal Justice and Behavior, 27*, 6-35. doi: http://dx.doi.org/10.1177/0093854800027001002

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1-21. doi: 10.1037/a0014421

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119-136. doi: 10.1023/A:1005482921333

Heffernan, R., & Ward, T. (2015). The conceptualization of dynamic risk factors in child sex offenders: An agency model. *Aggression and Violent Behavior, 24*, 250-260. doi: http://dx.doi.org/10.1016/j.avb.2015.07.001

Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1996). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry, 37*, 419-424.

Hoge, R. D., Vincent, G. M., & Guy, L. S. (2012). Predction and risk/needs assessment. In R. Loeber & D. P. Farrington (Eds.), *From juvenile delinquency to adult crime: Criminal careers, justice policy, and prevention*. New York, NY: Oxford University Press.

Home Office. (2006). *Offender Assessment System OASys user manual (Revised July 2006)*. London, UK: Home Office.

Howard, P. D., & Dixon, L. (2013). Identifying change in the likelihood of violent recidivism: Causal dynamic risk factors in the OASys violence predictor. *Law and Human Behavior, 37*, 163-174. doi: http://dx.doi.org/10.1037/lhb0000012

Hsu, C.-I., Caputi, P., & Byrne, M. K. (2009). The Level of Service Inventory—Revised (LSI-R): A useful risk assessment measure for Australian offenders? *Criminal Justice and Behavior, 36*, 728-740. doi: http://dx.doi.org/10.1177/0093854809335409

Indig, D., Frewen, A., & Moore, E. (2014). Predictors and correlates of re-incarceration among Australian young people in custody. *Australian & New Zealand Journal of Criminology*, 1-17. doi: 10.1177/0004865814550534

Jones, N. J., Brown, S. L., & Zamble, E. (2010). Predicting criminal recidivism in adult male offenders: Researcher versus parole officer assessment of dynamic

risk. *Criminal Justice and Behavior, 37*, 860-882. doi:

10.1177/0093854810368924

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer,

D. J. (1997). Coming to terms with the terms of risk. *Archives of General*

Psychiatry, 54, 337-343. doi:

http://dx.doi.org/10.1001/archpsyc.1997.01830160065009

Larson, R., Csikszentmihalyi, M., & Graef, R. (1980). Mood variability and the

psychosocial adjustment of adolescents. *Journal of Youth and Adolescence, 9*,

469-490.

Lewis, K., Olver, M. E., & Wong, S. C. P. (2013). The Violence Risk Scale:

Predictive validity and linking changes in risk with violent recidivism in a

sample of high-risk offenders with psychopathic traits. *Assessment, 20*, 150-

164. doi: http://dx.doi.org/10.1177/1073191112441242

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective

factors in the development of youth violence. *American Journal of Preventive*

Medicine, 43(2, S1), S8-S23. doi:

http://dx.doi.org/10.1016/j.amepre.2012.04.029

Miller, H. A. (2006). A dynamic assessment of offender risk, needs, and strengths in a

sample of pre-release general offenders. *Behavioral Sciences & the Law, 24*,

767-782. doi: http://dx.doi.org/10.1002/bsl.728

Modecki, K. L. (2008). Addressing gaps in the maturity of judgment literature: Age

differences and delinquency. *Law and Human Behavior, 32*, 78-91. doi:

http://dx.doi.org/10.1007/s10979-007-9087-7

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial

behavior: A developmental taxonomy. *Psychological Review, 100*, 674-701.

doi: http://dx.doi.org/10.1037/0033-295X.100.4.674

Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-

course-persistent and adolescence-limited antisocial pathways: Follow-up at

age 26 years. *Development and Psychopathology, 14*, 179-207. doi:

http://dx.doi.org/10.1017/S0954579402001104

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation

factors. *Quality & Quantity: International Journal of Methodology, 41*, 673-

690. doi: http://dx.doi.org/10.1007/s11135-006-9018-6

Olver, M. E., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity

and reliability of the Violence Risk Scale-Sexual Offender version: Assessing

sex offender risk and evaluating therapeutic change. *Psychological

Assessment, 19*, 318-329. doi: http://dx.doi.org/10.1037/1040-3590.19.3.318

Prior, D., Farrow, K., Hughes, N., Kelly, G., Manders, G., White, S., & Wilkinson, B.

(2011). *Maturity, young adults and criminal justice: A literature review*.

Birmingham, UK: University of Birmingham.

Ralston, C. A., & Epperson, D. L. (2013). Predictive validity of adult risk assessment

tools with juveniles who offended sexually. *Psychological Assessment, 25*,

905-916. doi: http://dx.doi.org/10.1037/a0032683

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies:

ROC area, Cohen's *d*, and *r. Law and Human Behavior, 29*, 615-620. doi:

10.1007/s10979-005-6832-7

Sampson, R. J., & Laub, J. H. (2005). A life-course view of the development of crime. *Annals of the American Academy of Political and Social Science, 602*, 12-45. doi: http://dx.doi.org/10.1177/0002716205280075

Scanlan, J. (2015). *Assessing women's risk of recidivism: An investigation into the predictive validity of the Dynamic Risk Assessment for Offender Re-entry (DRAOR) with matched samples of community-sentenced women and men.* Unpublished master's thesis. Victoria University of Wellington, Wellington, NZ.

Scott, E. S., & Steinberg, L. (2008). Adolescent development and the regulation of youth crime. *The Future of Children, 18*(2), 15-33.

Serin, R. C. (2007). *The dynamic risk assessment for offender re-entry (DRAOR).* Unpublished user manual.

Serin, R. C. (2015). *The dynamic risk assessment for offender re-entry (DRAOR).* Unpublished user manual.

Serin, R. C., Chadwick, N., & Lloyd, C. D. (2016). Dynamic risk and protective factors. *Psychology, Crime & Law, 22*, 151-170. doi: 10.1080/1068316X.2015.1112013

Serin, R. C., & Lloyd, C. D. (2009). Examining the process of offender change: The transition to crime desistance. *Psychology, Crime & Law, 15*, 347-364. doi: http://dx.doi.org/10.1080/10683160802261078

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*, 499-513. doi: 10.1111/j.1745-9133.2009.00551.x

Steinberg, L. (2007). Risk taking in adolescence: New perspectives from brain and

behavioral science. *Current directions in psychological science, 16*(2), 55-59.

Steinberg, L., Cauffman, E., Woolard, J., Graham, S., & Banich, M. (2009). Are

adolescents less mature than adults?: Minors' access to abortion, the juvenile

death penalty, and the alleged APA "flip-flop". *American Psychologist, 64*,

583-594. doi: http://dx.doi.org/10.1037/a0014763

Tamatea, A., & Wilson, N. (2009). *Dynamic risk assessment for offender re-entry*

(DRAOR): A pilot study. New Zealand: Department of Corrections.

Unpublished government report.

Thornton, D. (2013). Implications of our developing understanding of risk and

protective factors in the treatment of adult male sexual offenders.

International Journal of Behavioral Consultation and Therapy, 8(3), 62-65.

doi: http://dx.doi.org/10.1037/h0100985

Ward, T., & Beech, A. R. (2005). Looking to the future: Risk and explanation. In T.

Ward, D. L. L. Polaschek & A. R. Beech (Eds.), *Theories of sexual offending*.

Chichester, UK: Wiley Blackwell.

Ward, T., & Beech, A. R. (2015). Dynamic risk factors: A theoretical dead-end?

Psychology, Crime & Law, 21, 100-113. doi:

http://dx.doi.org/10.1080/1068316X.2014.917854

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk

scale: A treatment friendly risk assessment tool. *Psychology, Public Policy,*

and Law, 12, 279-309. doi: 10.1037/1076-8971.12.3.279

Wong, S. C. P., Olver, M. E., Nicholaichuk, T., & Gordon, A. (2003). *The violence*

risk scale: Sexual offender version (VRS-SO). Saskatoon, Saskatchewan,

Canada: Regional Psychiatric Centre and University of Saskatchewan.

Yesberg, J. A., & Polaschek, D. L. L. (2014). Assessing dynamic risk and protective

    factors in the community: Examining the validity of the Dynamic Risk

    Assessment for Offender Re-entry. *Psychology, Crime, & Law, 21*, 88-99. doi:

    10.1080/1068316X.2014.935775

Appendices

Appendix A

Comparisons Between the Sample Used in Ferguson (2015) and the Current Study

	Ferguson (2015) M (SD), N = 122	Current study M (SD), N = 398	t-test statistic
Sentence length (in days)	270.25 (67.09)	262.06 (68.84)	$t(518) = 1.16$, $p = .24$, $d = 0.12$, $M_{diff} = 8.19$, 95% CI [-5.72, 22.10]
Number of previous convictions (any)	9.87 (7.56)	8.27* (6.49)	$t(518) = 2.29$, $p = .03$, $d = 0.23$, $M_{diff} = 1.60$, 95% CI [0.23, 2.97]
Number of previous violent convictions	0.93 (1.10)	0.71* (0.97)	$t(518) = 2.12$, $p = .03$, $d = 0.21$, $M_{diff} = 8.19$, 95% CI [0.02, 0.42]
Number of previous imprisonments	0.39 (1.09)	0.39 (1.07)	$t(518) = 0.00$, $p = .99$, $d = 0.00$, $M_{diff} = 0.00$, 95% CI [-0.22, 0.22]
Initial RoC*RoI score	.37 (.16)	.33* (.18)	$t(518) = 2.20$, $p = .03$, $d = 0.23$, $M_{diff} = -0.04$, 95% CI [0.004, 0.08]

	Ferguson (2015) (%)	Current study (%)	Chi-square statistic
Ethnicity			
Māori	52 (42.6)	192 (48.2)	$\chi^2(3, N = 520) = 1.46$,
European	49 (40.2)	149 (37.2)	$p = .69$,
Pasifika	12 (9.8)	35 (8.8)	$\phi = 0.05$
Other	9 (7.4)	22 (5.6)	
Index offence			
Non-violent	76 (62.3)	262 (65.8)	$\chi^2(3, N = 520) = 1.52$,
Violent/sexual	36 (29.5)	103 (25.9)	$p = .68$,
Justice/admin	8 (6.6)	30 (7.5)	$\phi = 0.05$
Unknown	2 (1.6)	3 (0.8)	

*$p < .05$

Appendix B

Multilevel Modelling Outputs for Change in DRAOR Scores Over Time

Table B1

Multilevel Models for Change for DRAOR Acute Subscale

Model	Level-1 Model	Level-2 Model	Composite Model
A	$ACUTE_{ti} = \pi_{0i} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$	$ACUTE_{ti} = \beta_{00} + r_{0i} + e_{ti}$
B	$ACUTE_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$	$ACUTE_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$
C	$ACUTE_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + \beta_{10}*(OFFENCE) + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}*(OFFENCE) + r_{1i}$	$ACUTE_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} +$ $\beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$

These models predict the acute risk scores of NZ youth serving community sentences as a function of the number of months since the beginning of their sentences (level-1) and their reconviction status (level-2).

Table B2

Results from Multilevel Models for the Acute Subscale

		Model A	Model B	Model C
Fixed Effects		Estimate (Standard Error)		
Intercept	Initial	4.75*** (0.10)	5.31*** (0.13)	4.83*** (0.24)
	Reconviction			0.67* (0.28)
Rate of	Initial		-0.15*** (0.02)	-0.20*** (0.04)
change	Reconviction			0.08 (0.04)
(month)				
Variance Components		Estimate (Standard Deviation)		
Level-1	Within-person	1.68*** (1.30)	0.95*** (0.98)	0.96*** (0.98)
Level-2	In initial status	3.76*** (1.94)	5.80*** (2.41)	5.73*** (2.39)
	In rate of			
	change		0.14*** (0.37)	0.14*** (0.37)

Model A is the unconditional means model and Model B is the unconditional growth model. Model C includes reconviction status as a predictor of both initial status and rate of change. These models predict the DRAOR acute subscale scores for NZ youth serving community supervision sentences as a function of the number of months since the beginning of their sentences (level-1) as well as their reconviction status (level-2). *p < .05, ***p < .001

Table B3

Multilevel Models for Change for DRAOR Stable Subscale

Model	Level-1 Model	Level-2 Model	Composite Model
A	$STABLE_{ti} = \pi_{0i} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$	$STABLE_{ti} = \beta_{00} + r_{0i} + e_{ti}$
B	$STABLE_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$	$STABLE_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$
C	$STABLE_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + \beta_{10}*(OFFENCE) + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}*(OFFENCE) + r_{1i}$	$STABLE_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} + \beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$

These models predict the stable risk scores of NZ youth serving community sentences as a function of the number of months since the beginning of their sentences (level-1) and their reconviction status (level-2).

Table B4

Results from Multilevel Models for the Stable Subscale

		Model A	Model B	Model C
Fixed Effects		Estimate (Standard Error)		
Intercept	Initial	5.91*** (0.11)	6.21*** (0.12)	5.65*** (0.22)
	Reconviction			0.77*** (0.26)
Rate of change (month)	Initial		-0.08*** (0.02)	-0.20*** (0.03)
	Reconviction			0.17*** (0.04)
Variance Components		Estimate (Standard Deviation)		
Level-1	Within-person	1.00*** (1.01)	0.50*** (0.71)	0.50*** (0.71)
Level-2	In initial status	4.81*** (2.19)	4.83*** (2.20)	4.73*** (2.18)
	In rate of change		0.10*** (0.32)	0.10*** (0.31)

Model A is the unconditional means model and Model B is the unconditional growth model. Model C includes reconviction status as a predictor of both initial status and rate of change. These models predict the DRAOR stable subscale scores for NZ youth serving community supervision sentences as a function of the number of months since the beginning of their sentences (level-1) as well as their reconviction status (level-2). ***$p < .001$

Table B5

Multilevel Models for Change for DRAOR Protective Subscale

Model	Level-1 Model	Level-2 Model	Composite Model
A	$PROTECT_{ti} = \pi_{0i} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$	$PROTECT_{ti} = \beta_{00} + r_{0i} + e_{ti}$
B	$PROTECT_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$	$PROTECT_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$
C	$PROTECT_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + \beta_{10}*(OFFENCE) + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}*(OFFENCE) + r_{1i}$	$PROTECT_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} + \beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$

These models predict the protective scores of NZ youth serving community sentences as a function of the number of months since the beginning of their sentences (level-1) and their reconviction status (level-2).

Table B6

Results from Multilevel Models for the Protective Subscale

		Model A	Model B	Model C
Fixed Effects		Estimate (Standard Error)		
Intercept	Initial	6.38*** (0.11)	6.21*** (0.12)	6.65*** (0.19)
	Reconviction			-0.76*** (0.23)
Rate of change (month)	Initial		-0.08*** (0.02)	0.16*** (0.03)
	Reconviction			-0.13*** (0.04)
Variance Components		Estimate (Standard Deviation)		
Level-1	Within-person	0.94*** (0.94)	0.46*** (0.68)	0.46*** (0.68)
Level-2	In initial status	4.43*** (2.11)	4.27*** (2.07)	4.17*** (2.04)
	In rate of change		0.10*** (0.31)	0.10*** (0.31)

Model A is the unconditional means model and Model B is the unconditional growth model. Model C includes reconviction status as a predictor of both initial status and rate of change. These models predict the DRAOR protective subscale scores for NZ youth serving community supervision sentences as a function of the number of months since the beginning of their sentences (level-1) as well as their reconviction status (level-2).

***$p < .001$

Table B7

Multilevel Models for Change for DRAOR Total Scores

Model	Level-1 Model	Level-2 Model	Composite Model
A	$TOTAL_{ti} = \pi_{0i} + e_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$	$TOTAL_{ti} = \beta_{00} + r_{0i} + e_{ti}$
B	$TOTAL_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$	$TOTAL_{ti} = \beta_{00} + \beta_{10}*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$
C	$TOTAL_{ti} = \pi_{0i} + \pi_{1i}*(MONTH) + e_i$	$\pi_{0i} = \beta_{00} + \beta_{10}*(OFFENCE) + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}*(OFFENCE) + r_{1i}$	$TOTAL_{ti} = \beta_{00} + \beta_{01}*OFFENCE_i + \beta_{10}*MONTH_{ti} +$ $\beta_{11}*OFFENCE_i*MONTH_{ti} + r_{0i} + r_{1i}*MONTH_{ti} + e_{ti}$

These models predict the DRAOR total scores of NZ youth serving community sentences as a function of the number of months since the beginning of their sentences (level-1) and their reconviction status (level-2).

Table B8

Results from Multilevel Models for the DRAOR Total Score

		Model A	Model B	Model C
Fixed Effects		Estimate (Standard Error)		
Intercept	Initial	4.28*** (0.27)	5.38*** (0.28)	3.76*** (0.53)
	Reconviction			2.26*** (0.62)
Rate of change (month)	Initial		-0.29*** (0.04)	-0.55*** (0.07)
	Reconviction			0.36*** (0.09)
Variance Components		Estimate (Standard Deviation)		
Level-1	Within-person	6.08*** (2.47)	2.92*** (1.71)	2.93*** (1.71)
Level-2	In initial status	28.08*** (5.30)	28.78*** (5.37)	27.88*** (5.28)
	In rate of change		0.61*** (0.58)	0.76*** (0.76)

Model A is the unconditional means model and Model B is the unconditional growth model. Model C includes reconviction status as a predictor of both initial status and rate of change. These models predict the DRAOR total scores for NZ youth serving community supervision sentences as a function of the number of months since the beginning of their sentences (level-1) as well as their reconviction status (level-2). ***$p < .001$