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.

By

Jessica M. Scanlan

A thesis
submitted to Victoria University of Wellington in fulfillment of the requirements for the degree of Master of Science
in Forensic Psychology

Victoria University of Wellington, 2015

Abstract

Although men and women share risk factors for offending, some scholars claim these factors operate differentially by gender and that certain proposed women-specific risk factors are neglected in the existing gender-neutral risk assessment tools. The present research evaluated one such gender-neutral risk assessment tool used by New Zealand Department of Corrections: The Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007; Serin, Mailloux, & Wilson, 2012). The research was comparative and examined the predictive validity of the DRAOR for breaches of sentence and criminal reconvictions in matched samples of New Zealand women and men who had served community supervision sentences. Cox regression and AUC analyses showed the initial DRAOR had mixed predictive validity and the proximal DRAOR comparative predictive validity across gender. Additionally, the proximal DRAOR assessment consistently outperformed the initial DRAOR in the prediction of reconvictions for both women and men. Further, offenders made significant change on the DRAOR between two assessment points and overall the change made on the DRAOR was significantly related to reconvictions for women and men. For both samples, the RoC*RoI did not predict breach reconvictions; however, the proximal DRAOR TS provided incremental predictive validity above the RoC*RoI for criminal reconvictions. To conclude, the research supports the continued use of the DRAOR as a risk prediction tool with community-sentenced women and men and thus supports gender neutrality. Further, the research supports the dynamic nature of the DRAOR and highlighted the importance of updating dynamic risk assessments. Additionally, the research recommends that change made on a dynamic risk assessment tool over time be considered useful for predictive purposes for women and men alike.

Acknowledgements

First I would like to thank my supervisors, my primary supervisor Dr Clare-Ann Fortune and my secondary supervisor Professor Devon Polaschek. Thank you for the opportunity to conduct research in an area I am passionate about and for your constant guidance and encouragement over the course of the research. You both encouraged me to be the best I can be, which in addition to the skills I have learnt from you both, made this experience extremely valuable and rewarding.

Thank you to all the students, past and present, of the Victoria criminal justice lab. Our weekly meetings helped to keep my research on a forward trajectory. Getting to know each of you and seeing how passionate you all are about the research you are producing has been wonderful. A special thank you to Julia Yesberg for her mentorship. I am extremely grateful for your willingness to take the time to read pieces of my work and provide insightful feedback. I would also like to thank the New Zealand Department of Corrections for supplying the data that made this research possible.

I am very fortunate to have a wonderful family and terrific friends. To Anna Ferguson for your friendship and companionship. To my family, Mum, Dad, Daniel, Thomas, Meghan, Beth and Molly for your continued support throughout my tertiary studies. Thank you for taking an interest in my research and particularly to my Mum for not getting upset when it took me well over a week to return messages, I promise the delay will be considerably less from now on.

The biggest thank you of all goes to my partner Tim. You have given me unwavering support throughout my research endeavor. Your willingness to help and support me in so many different ways, including cooking fantastic meals, is appreciated beyond words. The belief you have in me helps me to do my best everyday. Finally, to my grandmother Mary, for such wonderful memories.

Table of Contents

Abstract ... **i**

Acknowledgements ... **ii**

Table of Contents ... **iii**

List of Tables ... **vii**

List of Figures ... **ix**

List of Appendices ... **x**

Chapter 1: Literature Review ... **1**

Introduction ... **1**

Criminal Risk Assessment ... **2**

The development of risk assessment ... 2

First-generation ... 2

Second-generation: Actuarial risk assessment tools ... 2

Third and fourth-generation tools ... 3

The predictive validity of dynamic versus static risk tools ... 4

Dynamic risk factors ... 4

Stable and acute dynamic risk factors ... 4

Resilience, desistence and protective factors ... 5

Incorporation of protective factors ... 6

The changeable nature of dynamic risk factors ... 7

Dynamic risk factors as causal factors ... 7

Summary ... **8**

Risk Assessment of Women Offenders ... **9**

The gender-neutral perspective ... 9

The Central eight risk factors ... 9

Empirical support for gender neutrality ... 10

The 'typical' woman offender ... 11

The gender-responsive perspective ... 11

Gender-responsive risk factors ... 12

Empirical support for the gender-responsive perspective**.** ... 15

The extant research on women's risk assessment**.** ... 17

Introduction to the Current Research ... **18**

Chapter 2: Method **19**
Data **19**
Sample Preparation **19**
Exclusion/inclusion criteria 19
Matching the Samples **20**
The success of PSM 21
Normality of Continuous Data **23**
Sample Characteristics **23**
Measures **23**
The Dynamic Risk Assessment for Offender Re-entry (DRAOR) 23
Existing research with the DRAOR 25
The Risk of Re-conviction X Risk of Re-imprisonment (RoC*RoI) 25
Procedure **26**
Calculation of DRAOR TS 26
Extraction of DRAOR scores 26
Recidivism 26
Time-at-risk 27
Survival days 27
Data Analysis **27**
Kaplan-Meier survival analyses 28
Cox regression 28
Area Under the receiver operating Curve (AUC) 28
Interpretation of Analyses **29**
Chapter 3: Results **31**
Part I Breaches 31
Base rates of breaches and survival analyses 31
Inclusion criteria 32
The predictive validity of the DRAOR 33
Initial DRAOR 33
ANOVA analysis of initial DRAOR scores 33
Univariate Cox regression models 34
Multivariate Cox regression models 38
Proximal DRAOR 39

ANOVA analysis of proximal DRAOR scores 39
Univariate Cox regression models 41
Multivariate Cox regression models 43
The incremental predictive validity of the proximal DRAOR 44
The dynamic nature of the DRAOR 47
Paired-samples *t*-test 48
Change made at the recidivist versus non-recidivist level 49
The predictive validity of DRAOR change scores 54
The predictive validity of the RoC*RoI 59
Part II Criminal Reoffending 60
Base rates of criminal reoffending and survival analyses 60
Inclusion criteria 61
The predictive validity of the DRAOR 61
Initial DRAOR 62
ANOVA analysis of initial DRAOR scores 62
Univariate Cox regression models 63
Multivariate Cox regression models 65
Proximal DRAOR 66
ANOVA analysis of proximal DRAOR scores 67
Univariate Cox regression models 68
Multivariate Cox regression models 69
The incremental predictive validity of the proximal DRAOR 70
The dynamic nature of the DRAOR 71
Paired samples *t*-test 72
Change at the recidivist versus non-recidivist level 73
The predictive validity of DRAOR change scores 79
The predictive validity of the RoC*RoI and the incremental predictive validity of the DRAOR 82
Chapter 4: Discussion 85
Summary of Findings 85
Theoretical and Practical Implications 86
The gender-neutral versus gender-responsive 86
The dynamic nature of the DRAOR 89

Static versus dynamic risk tools 91
Protective factors 93
Limitations and Future Directions 94
Conclusion 95
References 97
Appendices 108

List of Tables

Table 1 *Final Samples Categorical Matching Variables and Equivalance Test*.....................22
Table 2 *Final Samples Continuous Matching Variables and Equivalance Test* 22
Table 3 *The DRAOR* 24
Table 4 *Summary of Effect Sizes*...30
Table 5 *Initial DRAOR Scores and ANOVA Analysis for the Breach Outcome*......................33
Table 6 *Univariate Cox Regression Models of the Initial DRAOR Predicting Breaches* 35
Table 7 *Correlation Matrix of Initial DRAOR Scores for the Breach Outcome* 37
Table 8 *Multivariate Model of Initial DRAOR Subscale Scores Predicting Breaches*...........39
Table 9 *Correlation Matrix of Proximal DRAOR Scores for the Breach Outcome* 40
Table 10 *Proximal DRAOR Scores and ANOVA Analysis for the Breach Outcome*...............40
Table 11 *Univariate Models of the Proximal Proximal DRAOR Scores Predicting Breaches* 42
Table 12 *Multivariate Model of the Proximal DRAOR Subscales Predicting Breaches*.........44
Table 13 *Correlation Matrix of Initial and Corresponding Proximal DRAOR Scores for the Breach Outcome*...45
Table 14 *Incremental Predictive Validity of the Proximal DRAOR Scores Predicting Breaches*...46
Table 15 *Paired-Samples t-test of the Initial and Proximal DRAOR Scores for the Breach Outcome* 48
Table 16 *ANOVA of Mean DRAOR Change Scores for Women and Men who were Convicted of a Breach and those who were not*...53
Table 17 *Correlation Matrix of Initial DRAOR Scores and DRAOR Change Scores for the Breach Outcome*...55
Table 18 *Multivariate Model of DRAOR Change Scores Predicting Breaches Controlling for Initial DRAOR Scores*..58
Table 19 *Univariate Cox Regression Model of the RoC*RoI Predicting Breaches*................59
Table 20 *Correlation Matrix of Initial DRAOR Scores for the Criminal Outcome*.................62
Table 21 *Initial DRAOR Scores and ANOVA Analysis for the Criminal Outcome*.................62
Table 22 *Univariate Models of the Initial DRAOR Scores Predicting Criminal Reoffending* 63
Table 23 *Multivariate Models of the Initial DRAOR Subscales Predicting Criminal Reoffending* 65
Table 24 *Correlation Matrix of Proximal DRAOR Scores for the Criminal Outcome* 66
Table 25 *Proximal DRAOR Scores and ANOVA Analysis for the Criminal Outcome* 67

Table 26 *Univariate Models of the Proximal DRAOR Scores Predicting Criminal Reoffending* 68
Table 27 *Multivariate Model of the Proximal DRAOR Subscales Predicting Criminal Reoffending* 69
Table 28 *Incremental Predictive Validity of the Proximal DRAOR Predicting Criminal Reoffending* 71
Table 29 *Paired-Samples t-test of the Initial and Proximal DRAOR Scores for the Criminal Outcome* 72
Table 30 *ANOVA of the Mean Change Scores for Women and Men who were Convicted of a Criminal Reoffence and those who were not* 78
Table 31 *Correlation Matrix of Initial DRAOR Scores and DRAOR Change Scores for the Criminal Outcome* 79
Table 32 *Multivariate Model of DRAOR Change Scores Predicting Criminal Reoffending Controlling for Initial DRAOR Scores* 81
Table 33 *Correlation Matrix of RoC*RoI and Initial and Proximal DRAOR TS for the Criminal Outcome* 83
Table 34 *Univariate Model of the RoC*RoI Predicting Criminal Reoffending and Multivariate Model of the Incremental Predictive Validity of the Proximal DRAOR TS* 84

List of Figures

Figure 1. Survival curves of breaches for women and men 32

Figure 2. Initial and proximal DRAOR scores for women and men who were convicted of a breach and those who were not...51

Figure 3. Mean change scores for women and men who were convicted of a breach and those who were not...53

Figure 4. Survival curves of criminal reoffending for women and men 62

Figure 5. Initial and proximal DRAOR scores for women and men who were convicted of a criminal reoffence and those who were not 76

Figure 6. Mean change scores for those who were convicted of a criminal reoffence and those who were not 78

List of Appendices

Appendix A Demographic information for the female and male samples eligible for propensity score matching....108

Appendix B Propensity score matching....109

Appendix C Predictive validity of the proximal DRAOR and the non-significant initial DRAOR scores for women and men....110

Appendix D The RoC*RoI and breaches....112

Appendix E Multivariate model of the RoC*RoI and initial DRAOR TS predicting criminal reoffending....113

Chapter 1: Literature Review

Introduction

Imagine you are a probation officer. Your client, Jennifer, is a 34-year-old woman who has been sentenced to five months of community supervision following her fifth conviction for theft. As Jennifer's probation officer, part of your job is to monitor her risk of reoffending while in the community. You will have at your disposal a dynamic risk assessment tool designed to assess and manage Jennifer's risk over the coming months. However, it is likely that this tool was developed on a male offender population. Given Jennifer is a woman, will this tool be applicable to her? This research will address this question as one of the first empirical studies directly comparing the predictive validity of an existing gender-neutral risk tool with women and a matched sample of men.

Risk assessment is a critical, albeit challenging, task of the criminal justice system, where attempts are made to predict an offender's likelihood of future criminal behaviour (Hudson, Wales, Bakker, & Ward, 2002). Risk assessments govern numerous decisions made within the contemporary criminal justice context, including: parole and treatment eligibility, treatment targets, and the level of punitive sanctions imposed. Thus it is crucial that assessment tools used in practice are validated with diverse offender populations, including the growing population of women offenders (Department of Corrections, 2013).

The dominant perspective of risk assessment is referred to as the *gender-neutral* perspective. This perspective states that men and women share the same risk factors for criminal behaviour (Andrews & Bonta, 2010). On the other hand, feminist scholars and proponents of the *gender-responsive* perspective question the legitimacy of gender neutrality. They argue because women are not the normalised[1] population and because existing tools neglect factors pertinent to women's criminality they are not valid for use with women (Blanchette & Brown, 2006). The outcomes of risk assessments determine the balance between community safety and the criminal justice system's ethical obligation to do the best by offenders (Holtfreter & Cupp, 2007). Thus the concerns of feminist scholars warrant research attention, particularly as increasing numbers of women are being administered risk assessment tools. Before I review the controversy surrounding the assessment of women's risk, a broad introduction into the risk assessment literature is necessary.

[1] The normalised population is the population for which a risk assessment tool was developed.

Criminal Risk Assessment

The development of risk assessment. Over the past 40 years there have been significant developments in the area of offender assessment. The structured evaluation of empirical correlates of criminal behaviour is common in current risk assessment methods (Miller, 2006). However, the structured evaluation of such factors has been preceded by other assessment methods; this evolution of risk assessment is referred to as the 'generations' of risk assessment (Andrews & Bonta, 2010). Unless otherwise stated the empirical studies reviewed in the following section are based predominately on male offenders. The consequence of which will be discussed in the second half of this literature review.

First-generation. Unstructured professional judgment was the pioneering method of risk assessment. This approach was very subjective because the assessments were based on the clinician's 'gut-feeling'. Assessments were undermined by the susceptibility of clinicians and other trained professionals to the biases and heuristics of human nature. Additionally, the lack of scientific grounding meant the accuracy of assessments was rarely above chance level (Hsu, Caputi, & Byrne, 2009).

Second-generation: Actuarial risk assessment tools. The second-generation of risk assessment was hallmarked by the development of actuarial risk assessment tools. Actuarial tools stipulate strict sets of instructions for the evaluation and formulation of empirically derived risk factors into a numerical risk score. Risk factors are empirical correlates of criminal behaviour. A risk factor is an observable proxy that accounts for variance in the unobservable construct of future criminal behaviour; if all the variance in criminal behaviour was accounted for in a risk prediction, the prediction would be 100% accurate.

Actuarial risk assessment tools can be fully algorithmic (i.e., computer generated) or can involve trained professionals using guidelines to score the risk factors. In other words, actuarial risk assessment is a *process* of structured evaluation of specified risk factors that produces a numerical likelihood of recidivism (Yang, Wong, & Coid, 2010). Actuarial risk tools determine risk classifications (or risk status) by placing an offender with a given score within a group of representative offenders who share a similar score. Thus each actuarial risk assessment tool predicts the likelihood of future criminal behavior relative to the representative group.

Second-generation actuarial tools evaluate *static risk factors* only. Static risk factors are historical factors not amenable to targeted intervention. These unchangeable factors are commonly characteristics of an offender's history, for example, age at first offence; or characteristics of the offender, for example, age and gender. These risk factors have a strong and robust relationship with future criminal behaviour (Miller, 2006).

The accuracy of static actuarial tools is vastly superior to those of unguided clinical judgment and is particularly strong in the prediction of long-term recidivism (Andrews, Bonta, Wormith, 2006; Beech & Craig, 2012; Bengtson & Langström, 2007; Garcia-Mansilla, Rosenfeld, & Nicholls, 2009; Hanson & Morton-Bourgon, 2009). Meta-analytical reviews have shown the mean validity estimates of static risk tools to range between 0.24 and 0.46, compared to those of unstructured clinical judgment, which range between 0.03 and 0.14 (see Andrews et al., 2006).

The development of static actuarial tools to determine risk status moved the method of risk assessment forward dramatically. Despite their robust predictive relationship with criminal behaviour; second-generation tools are limited in contemporary criminal justice settings. Their reliance on static risk factors mean the tools are unable to monitor changes in risk across time, which is required for effective offender management. Additionally, they are unable to inform intervention or determine whether an offender has reduced their risk as a result of treatment.

Third and fourth-generation tools. Present-day correctional settings regularly use third and fourth-generation risk tools. These tools are predominately actuarial and distinguished from second-generation tools by the incorporation of *dynamic risk factors*[2] (see below). Unlike third-generation tools, fourth-generation tools include protective factors (see below) and are designed specifically for supervision of offenders from their intake through to their exit of the criminal justice system (Andrews & Bonta, 2010). Both generations of dynamic risk tools assess a broad range of factors and because these factors can be targeted through intervention, they are particularly useful for case management. Like, second-generation tools, dynamic risk tools are empirically derived; however, unlike the second generation they are theoretically grounded. The incorporation of dynamic risk factors means that third and fourth-generation tools provide information concerning the nature, imminence and severity of criminal behaviour (Sjöstedt & Grann, 2002). These tools are also referred to as *risk-need tools* because they serve multiple objectives: (1) identification of an offender's risk status, (2) identification of an offender's risk over time and (3) identification of treatment targets (Salisbury, Van Voorhis & Spiropoulos, 2009).

[2] An alternate risk assessment method is Structured Professional Judgment (SPJ). Unlike actuarial tools, the SPJ method does not rely on numerical scores. SPJ involves a set of guidelines that specify areas a trained professional (e.g., psychologist) should explore in the production of a risk assessment. Unlike actuarial tools the final risk classification is based solely on the professional's judgment.

The predictive validity of dynamic versus static risk tools. There is mixed empirical evidence concerning whether dynamic risk tools capture more variance in recidivism than static risk tools. For example, empirical evaluations of third and fourth-generation risk tools have shown dynamic risk factors account for unique variance in recidivism (Hanson, Harris, Scott & Helmus, 2007). In a sample of sexual offenders, Hanson and Harris (2000) showed that even after controlling for pre-existing differences in static predictors the dynamic risk factors continued to predict recidivism. In contrast, a recent study showed dynamic risk factors evaluated in the Level of Service Inventory-Revised[3] (LSI-R) were empirically predictive of recidivism; however, they did little to improve the predictive validity of the criminal history domain, the static domain of the tool. Specifically, the LSI-R total score, which is a composite score derived from the static and dynamic subdomains had predictive validity approximate to that of the criminal history subdomain itself (Caudy, Durso, & Taxman, 2013).

Dynamic risk factors. Despite mixed empirical evidence for the incremental predictive validity of dynamic risk tools, dynamic risk factors have uses beyond recidivism prediction. They are personal, situational and environmental factors that, as the name suggests, are amenable to change over time. Examples of dynamic risk factors for reoffending include: antisocial peers, impulsivity and antisocial attitudes. Additionally, dynamic risk factors recognise the unique and highly complex environments that exist for offenders (Yesberg & Polaschek, 2015). Given this, dynamic risk factors are potentially more psychologically meaningful than their static counterparts, particularly from a case management perspective (Mann, Hanson, & Thornton, 2010).

Stable and acute dynamic risk factors. Dynamic risk factors have been further refined into *stable* and *acute* dynamic risk factors. Stable dynamic risk factors are persistent or enduring characteristics of an offender, which can change gradually over months or years (e.g., antisocial attitudes). Stable dynamic risk factors have utility in long-term risk prediction (Hanson et al., 2007) and are considered the best target for interventions; positive change on these factors should foster enduring change in antisocial behaviour.

In contrast, acute dynamic risk factors can fluctuate rapidly over days, hours or even minutes (e.g., intoxication). Acute dynamic risk factors are particularly useful for the daily management of offenders (Hanson & Harris, 2002). They provide valuable information regarding the timing of criminal behaviour and are theorised to signal imminent recidivism

[3] The LSI-R (Andrews & Bonta, 2010) is a structured risk tool that includes 54 risk factors divided into 10 subdomains. The LSI-R includes one static domain; however, the tool is predominately dynamic.

(i.e., the prediction of short-term recidivism; Hanson et al., 2007). Acute factors are particularly important for case management; they provide information about the situations and contexts in which recidivism is more likely to occur and can detect fluctuations in risk in an offender's immediate environment (Campbell, French, & Gendreau, 2009).

Research with sexual offenders has shown stable dynamic risk factors—compared to static and acute dynamic risk factors—most strongly differentiate recidivists from non-recidivists (Hanson & Harris, 2000). The empirical utility of stable and acute dynamic risk factors remains uncertain. For example, Hanson and colleagues (2007) conducted an empirical evaluation of static, stable and acute dynamic risk factors. The study involved 156 parole and probation officers assessing 997 sexual offenders from 16 Canadian jurisdictions using the STATIC-99, STABLE-2007 and ACUTE-2007[4].. Unexpectedly, the six-month average of acute assessments was a stronger predictor of sexual and general recidivism than the most recent acute assessments. Contrary to their theoretical utility as indicators of imminent risk, the finding suggests acute factors were measuring a relatively enduring construct. Why is it that some offenders classified as high-risk do not go onto reoffend? Scholars are interested in this question and have looked to the developmental literature and the research on resilience for answers.

Resilience, desistence and protective factors. Resilience is the term used in developmental research to refer to a biopsychosocial adaptation to stress or troubling circumstances in one's life (Lösel & Farrington, 2012). Most resilience research comes from developmental studies of young people and has shown that a mix of community, family and personal factors are responsible for reducing the risk of an otherwise high-risk young person's involvement in the criminal justice system later in life.

Criminal desistance is defined as a process where "active offenders reduce and eventually terminate their criminal careers" (Skardhamar & Savolainen, 2014, p. 264). The desistance literature has identified a number of factors linked with criminal desistance. Sampson and Laub (1993) showed the formation of prosocial bonds through marriage promoted desistance in adult men. Age is also a factor that is linked with desistance. The age-crime curve, which examines the rate of crime over the life course, shows the prevalence of crime peaks for the period of adolescence and steadily decreases from the mid to late 20s (Moffitt, 1993).

[4] The STATIC-99, STABLE-2007 and ACUTE-2007 are all structured risk assessment tools designed to predict recidivism in sexual offenders. The STATIC-99 includes 10 static risk factors. While the STABLE-2007 and ACUTE-2007 are dynamic risk tools designed to tap the two types of dynamic risk factors (see Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012).

The factors that when present promote desistance, or in the developmental context promote resilience to high-risk environments, are generally referred to as *protective factors* in the risk assessment literature. Protective factors are strengths or resources, internal and external to the offender that may mitigate the likelihood of criminal behaviour. Relative to risk factors, protective factors are under-researched (Walker, Bowen, & Brown, 2013). There is a need for a theoretical model explaining how protective factors influence criminal behaviour; however, the development of such a model requires an understanding of the interaction between risk and protective factors (de Vogel, de Vries Robbé, de Ruiter, & Bouman, 2011; Walker et al., 2013). Different conceptualisations for the interaction between risk and protective factors have been offered (de Vries Robbe, 2014).

Lösel and Farrington (2012) propose two types of protective factors: direct and buffering. Direct protective factors refer to those factors that predict a low probability of future problem behaviour, independent of other factors. In contrast, buffering protective factors predict a low probability of a negative outcome in the presence of risk factors. According to Lösel and Farrington, direct protective factors have the larger effect on future problem behaviour, whereas the function of buffering protective factors is to weaken the impact of risk factors. Other scholars suggest protective factors are the opposite of risk factors (Spice, Viljoen, Lazman, Scalora, & Ullman, 2012) or independent of risk factors all together (Ullrich & Coid, 2011). Most recently, the label *strength* has been used to refer to factors that instinctively could reduce the effect of risk factors (Jones, Brown, Robinson, & Frey, 2015). The term *promotive factor* is assigned for factors that correlate negatively with recidivism regardless of risk-level, and a *protective effect* is reserved for factors that mitigate the likelihood of recidivism in higher-risk groups, but have less of an effect on lower-risk groups (Jones et al., 2015).

The incorporation of protective factors. Historically, criminal risk assessments have been deficit-oriented, exclusively focusing on the detection and evaluation of risk factors (Jones, et al., 2015). This deficit-oriented approach to risk assessment is criticised for ignoring protective factors, which runs the danger of over classifying risk (de Vogel et al., 2011). Although the conceptualisation of protective factors is unclear, they have been incorporated into structured risk assessment tools with the intention of providing a more balanced appraisal of risk and their incorporation has been justified empirically.

For example, in a study of protective factors for violence, five factors of the interpersonal domain (e.g., prosocial support, emotional support, time spent with friends and family, closeness to others and involvement in religious activities) were significantly associated with reduced violence, after excluding criminal family members and close friends from the relevant social networks (Ullrich & Coid, 2011). Other empirical studies have supported the value of including protective factors in structured risk assessments (de Vries Robbé, de Vogel, & de Spa, 2011; de Vries Robbé, de Vogel, & Douglas, 2013).

In addition to their predictive utility, the inclusion of protective factors has increasingly been acknowledged as necessary for strength-based rehabilitation (Ward & Brown, 2004). Like dynamic risk factors, some protective factors are able to change over time (de Vries Robbé et al., 2011) and thus present a valuable opportunity for their engagement and promotion in rehabilitative treatment programmes.

The changeable nature of dynamic risk factors. An offender's risk of recidivism fluctuates over time, and is referred to as an offender's risk state (Douglas & Skeem, 2005). The rate of change in risk can be rapid or gradual; regardless, it holds that an assessment of risk state becomes less informative as time passes. Given this, the predictive validity of dynamic risk factors is likely to improve with reassessment (Andrews et al., 2006).

A common method for evaluating dynamic factors has been to examine their predictive validity from a single time point; however, this single-wave design is unable to attest to the factors' changeability (Howard & Dixon, 2013) or whether an up to date assessment would be more predictively valid. In order to examine whether dynamic risk factors truly are dynamic, a minimum of two assessments is required.

Studies using two or more time points are relatively scarce in comparison to single-wave designs; however, those that exist show promising results for the practice of repeated assessment of dynamic risk factors. For example, Brown and colleagues (2009) assessed 136 federally sentenced men on a series of static and dynamic risk factors prior to release (i.e. time-invariant dynamic assessment) [5] and then reassessed the dynamic risk factors at two and three months post-release (i.e. time-dependent dynamic assessments). The combined static and time-invariant dynamic assessment provided incremental predictive validity above the purely static model, which attests to the unique predictive capabilities of dynamic risk factors above static factors. Importantly, the model that predicted recidivism most strongly included the static and time-dependent dynamic assessments.

Dynamic risk factors as causal factors. The *need* principle of the Risk, Need, Responsivity (RNR; Andrews, Bonta, & Hoge, 1990) framework of offender rehabilitation

[5] Time-invariant refers to a single dynamic assessment where change on the factors cannot be examined. A time-dependent dynamic assessment is the reassessment. The analysis of this later assessment relative to the time-invariant assessment enables the capacity of dynamic risk factors for change to be examined.

states that dynamic risk factors or criminogenic needs, when identified, should be the targets of effective treatment. Brown and colleagues established that dynamic risk factors can change and empirically support the predictive utility of reassessment; however, empirical research also needs to show the targeting of dynamic risk factors is a worthwhile exercise. According to Kraemer and colleagues (1997) a dynamic risk factor can be considered causal if[6]: (1) it correlates with recidivism, (2) it precedes the recidivist event in time, (3) is sensitive to change over time and (4) change on the factor is associated with changes in recidivism. Kraemer's fourth criterion is crucial because a key characteristic of dynamic risk factors is their ability to be targeted through intervention and, as a result, to lower an offender's likelihood of engaging in future criminal behaviour, as per the need principle (Lewis, Olver, & Wong, 2012).

In one study, 150 high-risk males who attended an institution based cognitive-behavioural treatment program for violence were assessed on the Violence Risk Scale (VR) pre and post-treatment[7]. Results showed that, after controlling for pre-treatment risk, the amount of change made by men predicted violent recidivism over a five-year follow-up period (Lewis et al., 2012). Additional support for dynamic risk factors as causal factors comes from research linking change made on the factors to criminal behaviour (see Brown et al., 2009; Olver, Lewis, & Wong, 2013).

Summary

Actuarial risk assessment tools are a necessary component of criminal justice systems. Dynamic actuarial tools serve three purposes (1) the classification of risk status, (2) the monitoring of risk state and (3) identification of treatment targets. Dynamic risk tools include changeable factors that have the capacity to monitor an offender's risk overtime and thus provide a viable method to effectively manage offenders. However, the risk assessment literature has almost exclusively been based on research of male offenders, which is of concern for the growing number of women offenders entering the criminal justice system. The following section reviews the literature on the risk assessment of women offenders.

[6] Causality is used here with reference to Kraemer and colleagues' criteria only. As previously stated risk factors are observable proxies for criminal behavior. However, the need principle of the RNR aligns dynamic risk factors as causal risk factors when referenced against Kraemer's criteria.

[7] The Violence Risk Scale (VRS; Wong & Gordon, 2000) is an actuarial risk tool designed to assist those who work with high-risk violent offenders, the tool is intended to integrate risk assessment and treatment. It includes six static and 20 dynamic risk factors.

Risk Assessment of Women Offenders

Most actuarial risk tools used in modern correctional services were originally developed on samples of male offenders and only later applied to women (Blanchette & Taylor, 2007). Thus, the principal assumption of contemporary risk assessment is that risk factors are the same for men and women, and that assessment tools developed on men are valid for use with women. It is common practice to evaluate women offenders' risk using male-derived risk assessment tools; however, the lack of research into the application of these tools to women has called to question the legitimacy of this practice (Zakaria, Allenby, Derkzen, & Jones, 2013).

The controversy around the assessment of women's risk is embedded within a larger debate regarding the causes of criminal behaviour; that is, whether they are general or gender-specific (Smith, Cullen, & Latessa, 2009). A review of the literature reveals two distinct schools of thought regarding the assessment of women's risk of reoffending (Nicholls, Ogloff, & Davis, 2004). The *gender-neutral* perspective considers gender to be a distal influence on criminal behaviour (Andrews & Bonta, 2010). The other side of the coin, the *gender-responsive* perspective, advocates that considering gender is central to producing accurate risk assessments. As the dominant school of thought, the gender-neutral perspective will be discussed first.

The gender-neutral perspective. Gender neutrality is theoretically grounded in the *Personal, Interpersonal, Community-Reinforcement* model of criminal behaviour (PIC-R; Andrews & Bonta, 2010). The PIC-R addresses criminal behaviour at the person level, recognising that the commission of criminal behaviour is the result of an interaction between the offender and their immediate environment. According to this perspective, an offender's gender is a distal influence, which resides in the broader socio-political context and as a result has minimal impact on criminality after the consideration of more immediate factors (Rettinger & Andrews, 2010). These immediate factors referenced by the PIC-R are the *Central eight* risk factors.

The Central eight risk factors. The most commonly used actuarial assessment tools have been developed in accordance with the PIC-R and evaluate the Central eight risk factors (Holtfreter & Cupp, 2007). The Central eight are primarily dynamic factors that have been identified as the strongest and most consistent predictors of criminal behaviour (Andrews & Bonta, 2010).

A hierarchy separates the Central eight into the *Big four* and *Modest four* based on each factor's empirical relationship with recidivism (Andrews & Bonta, 2010). The Big four are primarily dynamic factors internal and external to the offender that have the strongest empirical relationship with criminal recidivism. They are: 1) History of antisocial behaviour, (2) Antisocial personality pattern, (3) Antisocial cognitions, and (4) Antisocial associates (Andrews & Bonta, 2010). According to the gender-neutral perspective both static and dynamic risk factors are important for risk assessment; note, the predictive power of past criminal behaviour is the only static factor recognised within the Big four. The Modest four have an additional, albeit weaker, relationship with criminal recidivism. Three of the modest factors are the major settings for human exchange: family/marital, school/work, and leisure/recreation, with substance abuse being the fourth (Andrews & Bonta, 2010).

The gender-neutral perspective states the Central eight are responsible for the well-established gender differences (see section below) in criminal behaviour. Men commit more crime because they are exposed more often and for a longer duration to the correlates of criminal behaviour (Moffitt & Caspi, 2001). The perspective advocates that risk assessment is gender-neutral because once acquired, the Central eight operate equally across gender to increase the likelihood of future criminal behaviour (Andrews et al., 2012).

Empirical support for gender neutrality. A New Zealand longitudinal study of boys and girls born in Dunedin between 1972 and 1973 showed no striking gender differences in the risk factors for chronic antisocial behaviour across the life course (Moffitt & Caspi, 2001). Further, the Central eight risk factors have been empirically linked to recidivism in both men and women (Andrews et al., 2012).

Support for gender neutrality comes from research showing risk assessment tools originally developed with samples of men predict recidivism in women. One of the most commonly evaluated tools is the LSI-R (Andrews & Bonta, 2010). The LSI-R is developed in accordance with the PIC-R and thus assesses the Central eight risk factors. The LSI-R has been shown to accurately predict general and violent recidivism in sample of 411 women who had served either community-supervision or a sentence of two or more years in prison (Rettinger & Andrews, 2010). Additionally, the Big four risk factors were shown to explain the largest proportion of the variance in recidivism, which supports the empirical hierarchy of the Central eight risk factors with women (Rettinger & Andrews, 2010). Further, Coulson and colleagues found women had lower LSI-R scores than men and that the risk classifications were valid; the women classified as high-risk failed more on release compared to low-risk women (Coulson, Ilacqua, Nutbrown, Giulekas, & Cudjoe, 1996). In a large-scale meta-analytical review of 27 studies of the LSI-R, acceptable predictive validity was reported for the general use of the LSI-R with women (r = .35, 95% CI [.34. .36]; Smith et al., 2009).

Other empirical research has also shown male-derived tools to be predictively valid with women, supporting gender neutrality (see Flores, Lowenkamp, Smith, & Latessa 2006).

The 'typical' woman offender. Women offenders are considerably different from their male counterparts in terms of the type, severity and frequency of criminal behaviour (Caulfield, 2010). In New Zealand, the male-to-female ratio for persistent antisocial behaviour has been shown to be 10:1 (Moffitt & Caspi, 2001). This gender disparity in the commission of criminal behaviour is referred to as the gender gap in offending and has been replicated across culture, countries and sources of data (Casey, Day, Vess, & Ward, 2013; Steffensmeier & Allan, 1996). The gender gap also applies to the type of crimes that women and men engage in, the gap widens for violent crime as women are more likely to engage in crime involving alcohol, other drugs and/or property (Bloom, Owen, Covington, & Raeder, 2000). As a result of the difference in official crime statistics, being a male is a static risk factor for criminal behaviour (Nicholls et al., 2004).

In addition to differences in the context of criminal behaviour, women are qualitatively different from men, following distinct pathways into and out of the criminal justice system (Daly, 1992). A woman's criminality is often cited as stemming from abusive situations, which has a negative impact on her mental health and leads to alcohol and/or other drugs being used as maladaptive coping mechanism. Women are more often than men the primary caregivers of children prior to entering the criminal justice system; their status as primary caregivers is a unique challenge for this offender population. The gender-responsive perspective of risk assessment argues that women are a distinct offender population and questions the validity of actuarial assessment tools used with women, because they have largely been developed with men (Blanchette & Brown, 2006; Reisig, Holtfreter, & Morash, 2006). The perspective maintains that women's involvement in criminal behaviour is the result of their unique experiences and to produce reliable and accurate risk assessments gender should be considered (Blanchette & Brown, 2006).

The gender-responsive perspective. The gender-responsive literature is largely based on feminist theory, which broadly speaking stresses that female criminality is a result of the patriarchal and sexist socio-political system in which women live. The perspective advocates that actuarial tools are inappropriate for use with women because they fail to evaluate gender-responsive factors. Within the umbrella of gender-responsive factors are gender-specific and gender-salient risk factors. Gender-specific risk factors are the factors that are currently not considered to be major predictors of criminal behaviour. While gender-salient factors are those that are currently accepted within the gender-neutral Central eight;

however, they are advocated to be more strongly predictive of women's criminality relative to men.

Gender-responsive risk factors. One reason why gender-neutral tools are considered inappropriate to use with women is because they neglect or fail to appropriately weight factors considered pertinent to female criminality: gender-responsive factors. These proposed factors include: histories of victimisation, relationship dysfunction, mental health problems, substance abuse, self-efficacy and parental stress, (Wright, Salisbury, & Van Voorhis, 2007; Van Voorhis, Wright, Salisbury, & Bauman, 2010). However, to be considered important for risk assessment it is not enough to demonstrate the given experience is more prevalent in women offenders (which in some cases is well established, e.g., parental status); an empirical link between the gender-responsive factor and reoffending is required. To date, very few proposed gender-responsive factors have been empirically examined as risk factors. Below are some examples of commonly cited gender-specific (e.g., victimisation) and gender-salient (e.g., substance abuse) risk factors, which are examined in terms of their empirical relationship to criminal behaviour.

Victimisation. Victimisation has been defined as the experience of physical or sexual abuse in childhood, beating by an intimate partner in adulthood and criminal or sexual assaults at any time in one's life (McClellan, Farabee, & Crouch, 1997). Forms of victimisation are more common in offender all offender populations relative to community samples; however, victimisation, particularly sexual abuse, has been reported to be more prevalent among women than men and as a result is proposed as the starting point for women's criminal trajectories (Belknap & Holsinger, 2006; Daly, 1992; Smith et al., 2009). Despite qualitative evidence for the importance of victimisation in the development of women's criminality, empirical support for victimisation as a risk factor in women is equivocal. Some studies have shown no relationship between victimisation and recidivism (Lowenkamp, Holsinger, & Latessa, 2001; Scott et al., 2014), while others have shown victimisation increases criminal behaviour in women (Benda, 2005; Van Voorhis, Salisbury, Wright, & Bauman, 2008). Bonta and colleagues (1995) showed that victimisation, in the form of physical abuse as an adult was the only form of victimisation significantly associated with recidivism (Bonta, Pang, & Wallace-Carpetta, 1995). Unexpectedly, Bonta and collaborators found that, compared to women who did not report a history of physical abuse in adulthood, a significantly *lower* percentage of women who reported physical abuse in adulthood were recidivists.

Mental health difficulties. The prevalence of mental health problems and the type of difficulties experienced by male and female offenders (and community samples) are different (Belknap & Holsinger, 2006; Ruiz, Douglas, Edens, Nikolova, & Lilienfeld, 2012). Diagnoses of anxiety, depression, major mood disorders, and posttraumatic stress disorder are more common among female than male offenders (Belknap & Holsinger, 2006; Salisbury et al., 2009).

In a study comparing male and female substance-abusing offenders, women displayed greater levels of depressive and post-traumatic stress disorder (PTSD) symptoms in addition to borderline personality features (Ruiz et al., 2012). This presentation is consistent with other research suggesting women offenders suffer from mental health difficulties that are related to the high prevalence of trauma and abuse in the population. Benda (2005) showed stress, depression, fearfulness and suicidal thoughts/attempts were stronger predictors of women's recidivism than men's, while other research has shown mental health status is not a useful predictor of recidivism in women parolees (Scott et al., 2014).

Substance abuse. Early initiation into drug use is a risk factor and continued substance abuse is a maintaining factor of criminal behaviour in men and women. Substance abuse is one of the Central eight risk factors under the gender-neutral perspective (Andrews & Bonta, 2010). However, research has shown substance use precedes women's involvement in criminal activity, a pattern not replicated in men (Swan & Goodman-Delahunty, 2013). The gender-responsive literature considers substance abuse a gender-salient risk factor because of its high co-occurrence with other negative experiences such as victimisation and mental health (Blanchette & Brown, 2006). Research has shown substance abuse to be a maladaptive strategy used by women to cope with victimisation and mental illness (Chesney-Lind & Pasko, 2013).

Interpersonal relationships. Antisocial relationships are a risk factor shared across gender; however, interpersonal relationships—both antisocial and prosocial—are considered gender-salient risk factors in the gender-responsive literature. The feminist literature suggests differences in the socialisation processes of women and men result in women placing more weight on the relationships they form comparative to men. Brown and Blanchette (2006) suggest females will be less likely to engage in criminal behaviour if it threatens valued relationships, but only if that relationship is prosocial.

Sampson and Laub (1993) have posited that social bonding is a vital mechanism responsible for desistance from crime for all offenders. Benda (2005) investigated Sampson and Laub's life-course theory with male and female boot camp graduates over a 5-year follow up. Benda (2005) looked at whether satisfaction with life partners, friendships and employment predicted recidivism for the sample. Additionally, they looked at whether social bonding factors mediated the detrimental effect of victimisation and harmful feelings on recidivism. The study showed criminal partners were more powerfully predictive of women's recidivism than men's. In addition, satisfying prosocial relationships with romantic partners, friends, and children disproportionately facilitated women's desistance from crime (Benda, 2005).

Further, the gender-responsive literature considers women's status as primary care givers of children an important interpersonal factor influencing criminal activity. Research has shown women who do not have custody of their children are at increased risk for incurring new criminal charges within three years, particularly within the first 90 days of release from prison (Scott et al., 2014). Women who did not have custody of their children were almost four times more likely to commit a new offence than women without children. However, because Scott and colleagues did not control for static risk it is possible the result was due to the women without custody simply being a higher risk group.

Interacting negative life experiences. Empirical evidence linking the aforementioned gender-responsive factors with recidivism is equivocal. The gender-responsive literature draws heavily on the pathways research of Daly (1992) and thus proposes the predictive utility of these factors is not in isolation but in combination (Hollin & Palmer, 2006; Stewart, 2011). Some studies have looked at the interactive and cumulative affect of victimisation, mental health and substance abuse on recidivism.

Mullings, Marquart, and Diamond (2001) found a positive correlation between childhood maltreatment, drug use, and criminal justice involvement in a sample of women prisoners. Further, repeated modeling has found the three variables contribute differentially to women's and men's criminality (see Hollin & Palmer, 2006). A study by McClellan, Farabee, and Crouch (1997) showed women reacted to childhood trauma with more self-blame and depression than did males, which persisted through to adulthood and increased their susceptibility to substance abuse. The resulting misuse of substances and associated problems were more strongly predictive of female than of male recidivism. In a study of 101 federal female offenders, a gender-informed composite score of childhood emotional abuse and substance abuse had the strongest relationship with general and violent recidivism in women (Stewart, 2011).

Further, a complex association between women's dysfunctional interpersonal relationships and recidivism has been identified. A qualitative review of studies investigating relational factors in women suggested a dysfunctional intimate relationship alone is not a risk factor; however, when combined with other factors such as victimisation, mental health and substance abuse the risk increases. Similarly, lack of family support has been suggested to only affect women's recidivism when coupled with financial hardship, such as unemployment (Kreis, Schwannauer & Gillings, 2014).

Empirical support for the gender-responsive perspective. The gender-responsive literature criticises gender-neutral tools for misclassifying women (Nicholls et al., 2004). Because women have a lower base rate of reoffending the likelihood of false positives, that is, identifying a woman as high-risk when in fact she is not increases. Put simply, arguably there is more error associated with the assessment of a woman's risk than there is of a man's. For example, an empirical validation of a purely static risk scale with women offenders showed that although the total score predicted recidivism, the risk classifications (i.e., low, medium and high) failed to meaningfully differentiate between women in terms of the criminal reoffending rates observed (Bonta et al., 1995). The authors suggested past criminal behaviour might be less of a predictor for women compared to its robust relationship observed with men; however, the poor predictive power of the scale with women could have resulted from the small sample size.

As previously mentioned, only tentative empirical support exists for gender-responsive risk factors. However, research examining gender-neutral tools with women offenders provides some indirect support for the perspectives hesitation in using male-derived assessment tools with women. Research has shown different components of these tools predict recidivism for men and women, while other research has shown gender-responsive factors increase the predictive validity of gender-neutral tools with women.

Gender-neutral tools that have subdomains, such as the LSI-R, predict recidivism differently across men and women (Nicholls et al., 2004). For example, the predictive validity of LSI-R risk factors related to accommodation, education, work and relationships with friends, were more strongly correlated with general recidivism for men relative to women. Difficulties with emotional well-being was more strongly related to women's recidivism than men's and, relative to all risk domains examined in the LSI-R, difficulties with emotional well-being was more strongly predictive of general and violent recidivism in women compared to men (van der Knaap, Alberda, Oosterveld, & Born, 2012). A number of factors evaluated in the LSI-R have been shown to have poor predictive validity with women; for example, history of juvenile delinquency, weapons offences, offending with an associate and alcohol/drug abuse (Bonta et al., 1995).

Reisig, Holtfreter and Morash (2006) coded women offenders on Daly's (1992) pathways framework and subsequently examined the predictive validity of the LSI-R[8]. The results showed the LSI-R predicted recidivism for women who followed the economically motivated pathway, but failed to accurately predict recidivism for women on the gendered pathways (Reisig et al, 2006). The poor predictive validity of the LSI-R for women who followed the gendered pathway questions the ability of the Central eight risk factors to generalise to women. It is plausible this study, unlike previous studies, isolated the women who experienced cumulative adverse life experiences and suggests that for those women, the Central eight risk factors and subsequent male-derived risk assessment tools are not appropriate. A plausible explanation from the gender-responsive perspective for the poor generalisability of the LSI-R to women in this study is the tools' lack of sensitivity to gender-salient factors, such as substance abuse, and the neglect of gender-specific factors, such as victimisation.

Furthermore, the following two empirical studies make a case for the incremental predictive validity of gender-responsive risk factors. Van Voorhis and colleagues (2010) examined the incremental predictive validity of different combinations of gender-responsive supplements above the LSI-R in the prediction of different indices of recidivism in prison, parole and probation samples of women. They showed that in six of the eight samples studied, gender-responsive scales accounted for unique predictive variance above the gender-neutral models (Van Voorhis et al., 2010). Further, Blanchette and Taylor (2007) constructed an actuarial security classification tool using women offenders as both the construction and validation samples. The gender-responsive tool placed larger numbers of women at the lower level classifications and fewer at maximum security relative to an existing gender-neutral classification scheme, and was shown to predict institutional misconducts in women.

A recent study attempted to develop a fully gender-informed assessment tool from the ground up that could predict recidivism in released women offenders. However, contrary to research by Van Voorhis and colleagues (2010) and Blanchette and Taylor (2007), none of the gender-informed variables (e.g., victimisation and self-efficacy) increased the predictive capacity of gender-neutral risk factors (e.g., criminal history and employment). The final model consisted mostly of static criminal history variables, supporting the finding that, like

[8] Daly described five criminal trajectories that women follow: (1) Street women (2) Harmed and harming women, (3) Battered women, (4) Drug-connected women and (5) Economically motivated women. Four of the five pathways are considered gendered because they are hallmarked by gender-responsive factors. However, the economically motivated pathway is non-gendered because it is seen in male offenders.

with men, past behaviour is a reliable and robust predictor of future criminal behaviour in women (Zakaria et al., 2013).

Only one study to date has *directly* refuted the gender-neutral perspective in support of the gender-responsive perspective. Brown and Motiuk (2008) showed that of the risk factors that *significantly* predicted reoffending in men and women, 53% displayed gender-specificity; that is, predicted more strongly for one gender over another. The next section will identify some shortcomings of both perspectives reviewed above and suggest ways in which these can be addressed in order to advance the risk assessment of women offenders.

The extant research on women's risk assessment. Although overwhelmingly supported empirically, research from a gender-neutral perspective has almost exclusively focused on male offenders. In other words, gender-neutrality means 'developed on men and later applied to women', and never the reverse (Van Voorhis et al., 2010). The empirical validations that support gender-neutral tools with women have typically included disproportionately smaller samples of women offenders or have controlled for gender in the statistical analyses. These approaches are not considered statistically rigorous. Equal numbers of male and female offenders should be examined in separate analyses if the impact of gender in risk assessment is to be adequately investigated (Holtfreter & Cupp, 2007).

Because the gender-responsive literature is largely feminist based, the direct support for the perspective—with the exception of Brown and Motiuk, 2008—comes from research focused on women offenders only. In the absence of a male comparison group the findings, such as those by Van Voorhis and colleagues are not conclusive; it is possible the factors evaluated may have also produced enhanced predictive validity in male offender samples.

Gender-neutral tools are frequently being applied to women in practice; however, women are not the normalised population for these tools and, comparative to male offenders, little systematic research exists for the use of these tools with women (Holtfreter & Cupp, 2007). In the absence of gender-responsive risk tools, one approach to understanding the assessment of women's risk is to evaluate existing gender-neutral tools with samples of women offenders (Zakaria et al., 2013). As Flores and colleagues (2006) note, "classification systems must be validated to their specific offender populations" (p. 45). It follows then that gender-neutral tools—tools already being extensively used with women—should be empirically validated with women using appropriate methodology.

Introduction to the Current Research

Given the debate regarding the applicability of existing gender-neutral tools to women, the present research assesses the predictive validity of a gender-neutral assessment tool with a community-sentenced sample of women and a matched sample of men. This research examines the predictive validity of a structured dynamic risk assessment tool: The Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007; Serin, Mailloux, & Wilson, 2012). The DRAOR is a dynamic risk assessment tool used to facilitate the risk assessment and management of men and women in the community. Risk management in community settings is extremely important because, unlike institutional settings, real life opportunities for offending are constantly present and require regular monitoring.

In New Zealand, the largest proportion of women involved in the justice system serve community-based sentences. A 2013 "snapshot" of New Zealand's adult community-sentenced population showed women represented 22% of the total population (Department of Corrections, 2013). In 2013, the proportion of women serving community sentences was over four times the proportion serving prison sentences. A community-sentenced population of offenders was chosen for the research because relative to a parole sample it provided the most representative sample of women offenders in New Zealand. The offenders in the present research had served community supervision sentences, a category of community-based sentences. In New Zealand the length of a community supervision sentence can range between six months to two years (Sentencing Act, 2002). The sentence is designed to reduce the risk of further reoffending through the supervised rehabilitation and integration of the offender. A mandatory component of a community supervision sentence is the regular reporting to a probation officer.

This research was comparative, the fundamental aim being to determine whether the DRAOR, a gender-neutral tool, is an empirically valid tool for the prediction of breaches of sentence and criminal reconvictions with community-sentenced women and men. Additional aims of this research were to determine whether an up-to-date assessment of dynamic risk is a better predictor than an earlier assessment and, importantly for a case management tool like the DRAOR, whether change made on the DRAOR predicts future criminal behaviour. Finally, the research investigates whether the DRAOR captures unique variance in recidivism compared to the Risk of Re-conviction X Risk of Re-imprisonment (RoC*RoI Bakker, Riley, & O'Malley, 1999) a static actuarial tool. These later aims were intended to further the empirical and theoretical research on risk assessment of offenders.

Chapter 2: Method

Data

The data set used in the current research was drawn from a New Zealand Department of Corrections database of all adult offenders serving a community supervision sentence that began after 1 January 2011 and ended prior to 31 December 2013. The original data set included 1,100 randomly selected adult offenders: 550 women and 550 men. The database included offender demographic information, criminal history information, pre and post sentence static risk scores, recidivism data, and DRAOR assessment information including the dates of administration and scores.

Sample Preparation

The database provided by the Department of Corrections was fully anonymised archival data. As part of the anonymising process offenders' dates of birth were truncated to the first day of his or her respective birth months (e.g., if their birthday was 14 July it was recorded in the anonymised dataset as 1 July). To reduce the margin of error to +/- 15 days, each offender's date of birth was re-entered as the 15th day of his or her respective birth month. Offenders' age at the start of his or her sentence was rounded to the nearest year in the original data set (e.g., 27 years and 8 months rounded to 28 years). Each offender's age at the start of sentence was recalculated using the re-entered dates of birth (see above).

Exclusion/inclusion criteria. For this research each offender had a static risk score as measured by the RoC*RoI (Bakker et al., 1999: see measures section). Most offenders included in the sample had two RoC*RoI scores available; however, a number of the women and men had no RoC*RoI score available and were excluded from the research samples. For the offenders who had more than one RoC*RoI score available the score selected was based on two criteria. First, where possible the score selected had a completion date closest to the individual offender's sentence commencement date. However, if offenders had only one RoC*RoI score, then this score was used irrespective of its temporal relationship to his or her sentence commencement date. Secondly, the selected RoC*RoI score had to have a corresponding date of completion *before* any recorded reoffence date[9]. This criterion was important because the risk score produced by the RoC*RoI algorithm includes the offender's previous conviction history. Sixty-one women and 37 men were excluded from the research

[9] This criterion only impacted women and men who were convicted of a reoffence during his or her time-at-risk (see below).

samples due to the absence of a RoC*RoI score, or alternatively because the only score available was completed after a reoffence.

Recall that this research is a validation of the DRAOR (see measures section). In order to accommodate this aim, all offenders eligible for inclusion in the research were required to have a minimum of five DRAOR administrations during their community supervision sentence; those with four or fewer administrations were excluded from the research samples. This exclusion criterion was necessary to ensure that change in risk, as assessed by the DRAOR, could be analysed. Further, offenders who had their first DRAOR administration dated more than 1.5 months after their sentence commencement date were also excluded from the research sample. This criterion was set because for a small number of women and men, the first DRAOR assessment did not occur until five or six months following the commencement of their community supervision sentence. The large gap between sentence commencement date and the first DRAOR assessment was a problem for this research because it required a valid assessment of initial dynamic risk (see procedure section). As a dynamic risk tool the DRAOR should be administered regularly while an offender is on a community supervision sentence or parole. Consequently, offenders who had irregular DRAOR administrations were also excluded from the research samples. For the purpose of this research an irregular administration was defined as a gap larger than two months between any two administrations. As a result of the above exclusions, 283 offenders were removed from the research leaving data for 719 offenders—336 adult women and 383 adult men—available for inclusion in the final matched samples (see Appendix A for full details).

Matching the Samples

The key aim of this research was to explore the assessment of women's risk and to compare it to men. To facilitate this aim it was necessary to employ a method to minimise the differences between women and men on factors that may be predictive of criminal behaviour. A matching procedure was subsequently used to ensure that the women were as closely matched as possible to the men on a number of demographic and criminogenic variables. Propensity Score Matching (PSM; Rosenbaum & Rubin, 1984) was the statistical method used to produce the matched samples. PSM enables observational research as closely as possible to parallel the characteristics of random controlled studies by determining individual group membership that is conditional upon a number of observable variables (Austin, 2011).

Eight matching variables were entered as predictors in block 1 of a logistic regression model predicting gender. The logistic regression significantly predicted gender, meaning the two samples statistically differed on the eight matching variables (see Appendix B). The model accounted for 22-29% of the variance in gender and correctly classified 71% of offenders. The model generated individual propensity scores for the 719 eligible offenders.

Matched pairs were generated using the nearest neighbour optimal matching method (Austin, 2011). A caliper was set for the women offenders' propensity scores, meaning a man could be matched to a woman if his propensity score fell within + / - 0.008 of the woman's propensity score[10]. The optimal matching method ensured that the best-matched pairs were developed (e.g., the final man selected for a match to a woman had a propensity score that was closest to the assigned woman). The matching procedure was conducted without replacement, so that once a man had been matched to a woman that man could not be matched again. The matching process was carried out forwards: matching women with the smallest propensity scores first, and then reverse-matched: matching the women with the largest propensity scores first. This process helped to ensure that the finalised pairs were the optimal pairings. As a result of the matching procedure 202 women were successfully matched to a male comparison (Table 1 and Table 2).

An analysis of the matched women and men showed it was higher risk women who were successfully matched to lower risk men. The 134 women who were unable to be matched were an extremely low-risk group as estimated by the RoC*RoI (M = 0.10, SD = 0.09) with an average of 10 previous convictions. Ninety-five percent of the unmatched women had never been imprisoned, 62% had never been convicted of a violent offence and 74% had a non-violent index offence. The 181 unmatched men were a comparatively high-risk group compared to the matched men. Unmatched men had a mean RoC*RoI score of 0.44 (SD = 0.18), an average of 27 previous convictions, 38% had never been in prison and 21% of the unmatched men had at least two previous violent convictions. Just under half of the unmatched men had a non-violent index offence (49.2%) and approximately 40% had a violent index offence.

The success of PSM. To examine the success of the PSM procedure chi-square and independent-samples *t*-tests were conducted. As shown in Table 1 and Table 2 the propensity score matching procedure was successful. The women and men did not significantly differ

[10] This caliper was selected after a number of more liberal and conservative calipers were examined. The caliper of 0.008 produced the closely matched samples while balancing the side effect of reduced sample size if the caliper were smaller (e.g., 0.002).

from one another on any of the matching variables. The equivalence tests were supported by the Phi and Cohen’s *d* effect sizes, which were almost zero.

Table 1

Final Samples Categorical Matching Variables and Equivalence Test

	Women (n = 202)	Men (n = 202)	Tests of Equivalence	Phi (*Φ*)
			Chi-Square Analysis of Variance	
Ethnicity				
Māori	87 (43.1%)	89 (44.1%)	$\chi 2(3, n = 404) = 0.35, p = .950$	0.03
European	93 (46.0%)	88 (43.6%)		
Pacific Peoples	15 (7.4%)	17 (8.4%)		
Other	7 (3.5%)	8 (4.0%)		
Index Offence			$\chi 2(3, n = 404) = 0.28, p = .868$	0.03
Non-violent	134 (66.3%)	139 (68.8%)		
Violent	52 (25.7%)	48 (23.8%)		
Justice/admin	16 (7.9%)	15 (7.4%)		

Table 2

Final Samples Continuous Matching Variables and Equivalence Test

	Women (n = 202)	Men (n = 202)	Test of Equivalence	Cohen’s *d*
	M (*SD*)	*M* (*SD*)	Independent Samples t-test [95% CI]	
Age (years)	34.77 (*10.81*)	34.15 (*10.88*)	$t(402) = 0.57, p = .569$ [-1.51, 2.74]	0.00
Sentence length (days)	274.95 (*76.42*)	274.26 (*78.69*)	$t(402) = 0.09, p = .928$ [-14.48, 15.87]	0.00
RoC*RoI score[a]	.25 (*0.19*)	.25 (*0.17*)	$t(400) = -0.08, p = .933$ [-0.04, 0.03]	-0.02
Criminal history				
Number of previous convictions	17.73 (*15.72*)	17.91 (*19.70*)	$t(402) = -0.10, p = .920$ [-3.67, 3.31]	-0.00
Number of previous violent convictions	1.34 (*1.72*)	1.3 (*1.66*)	$t(402) = 0.24, p = .814$ [-0.29, 0.37]	0.02
Number of previous imprisonments	0.68 (*1.58*)	0.81 (*2.03*)	$t(402) = -0.71, p = .477$ [-0.48, 0.23]	-0.05

Note: [a] The Levene’s test for equality of variances was significant for the RoC*RoI (i.e. the variation in the RoC*RoI scores across women and men were significantly different) the equal variance *not* assumed values are reported.

Normality of Continuous Data

All continuous variables analysed in this research were examined for normality using skewness and kurtosis statistics. Additionally, histograms and diagnostic plots were examined. No corrections to normality were made for any of the variables in this research because the sample sizes were considered large enough (≥ 100) to accommodate any departures in normality (Tabachnick & Fidell, 2013).

Sample Characteristics

This research included two samples of community-sentenced adult offenders: a sample of 202 women and a matched comparison sample of 202 men. Because women and men were matched on key demographic and criminogenic variables a summary of their combined demographic characteristics follows.

The samples had a mean age of 34.5 years (*SD* = 10.84) and an average sentence length[11] of 274.6 days (*SD* = 77.5). The largest proportion of offenders in the sample identified as New Zealand European (44.8%) or Māori (43.6%), followed by a smaller proportion of Pacific Peoples (7.9%). The largest proportion (67.6%) of offenders committed a non-violent index offence and approximately a quarter (24.8%) were convicted of a violent offence and the smallest proportion (7.7%) were convicted of justice/administrative offences[12]. Overall, the sample was at low-risk of reimprisonment within the next five years as estimated by the RoC*RoI (*M* = 0.25, *SD* = 0.2) [13].

Measures

The Dynamic Risk Assessment for Offender Re-entry (DRAOR). The DRAOR was originally developed by Serin (2007) and the New Zealand adaptation is currently in its third version (Serin et al., 2012). The DRAOR is a structured risk assessment tool designed to facilitate the assessment of recidivism risk in the community in addition to guiding risk management and case planning of offenders (Yesberg & Polaschek, 2015). The DRAOR facilitates the aforementioned goals via the identification of an individual offender's dynamic risk and protective factors (Tamatea & Wilson, 2009).

[11] Sentence length includes the number of days between an offender's sentence start date and sentence end date.
[12] The index offences were coded into three categories: (1) Non-violent offences (e.g., included dishonesty and property offences), (2) Violent offences (e.g., assaults and grievous bodily harm offences), (3) Justice/administrative offences (e.g., breach of intensive supervision conditions, Sentence Act 2002)
[13] The RoC*RoI scores are categorised in this study as per the New Zealand Parole Board standard categories: 'Low' (0 ≥ 0.25), 'Low-moderate' (0.25 ≥ 0.5), 'Moderate' (0.5 ≥ 0.7), 'High' (0.7≥ 0.8), and 'Very high' (0.8+).

The DRAOR has been used in New Zealand community settings since April 2010 and comprises 19 items, divided into three subscales: the acute risk subscale, stable risk subscale and the protective subscale (see Table 3). The dynamic risk items included in the risk subscales were adapted from research on stable and acute risk factors for sexual offending (Hanson & Harris, 2000) to relate to general and violent reoffending. The acute risk factors of the DRAOR were developed to be proximal indicators of risk state (Douglas & Skeem, 2005), while the stable risk factors represent criminogenic needs, as referred to in the need principle of the RNR model (Andrews & Bonta, 2010). The protective factors of the DRAOR are internal strengths or external assets of the offender that are proposed to mitigate an offender's risk of reoffending (Tamatea & Wilson, 2009). The protective factors were incorporated from the desistance literature because of the increasing evidence of their correlation with parole and treatment success.

Table 3

The DRAOR

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

A probation officer can score an offender's current presentation on the DRAOR after each meeting. Each DRAOR item is scored on a three-point scoring system (0, 1, 2), the scores are allocated after an offender interview and gathering of collateral information (e.g., police files, and discussion with family members). For the stable and acute subscales a score of 0 indicates the item is not considered to be problematic for a given offender (i.e., "not a problem"), while a score of 2 is allocated when the item is presenting a considerable risk for reoffending (i.e., "definite problem"; Tamatea & Wilson, 2009). A score of 1 for an item indicates a slight or possible problem and allows the probation officer to be uncertain due to mixed or incomplete evidence. The acute and stable risk subscales incorporate seven and six items respectively, and the highest score an offender can receive on each is 14 and 12 respectively. The protective subscale is reverse scored, whereby a higher score signifies a greater degree of the protective factor. Thus, a score of 0 for an item in the protective subscale indicates it is not protective (i.e., “not an asset”) while a score of 2 signifies that the item is definitely present (i.e., “an asset”). Again, as with the stable and acute subscales a score of 1 for a protective item corresponds to the item being a slight or possible asset (Tamatea & Wilson, 2009). The protective subscale includes six items and the highest score an offender can receive is 12.

Existing research with the DRAOR. Since its development a number of studies have been conducted examining convergent, predictive and incremental predictive validity of the DRAOR with representative samples of New Zealand adult male offenders (Hanby, 2013; Tamatea & Wilson, 2009; Yesberg & Polaschek, 2015). Recently, the DRAOR has been evaluated in sample of general offenders from the United States (see Chadwick, 2014). In these studies the DRAOR subscales and the DRAOR total score (TS) been shown to be useful predictors for distinguishing men who are likely to be convicted of a new criminal reoffence (excluding breaches) from men who are not (AUC = .62; Yesberg & Polaschek, 2015). The DRAOR TS has shown to exhibit convergent validity with other dynamic risk instruments such as the Violence Risk Scale (r = .25; Yesberg & Polaschek, 2015) and to make a significant contribution when added to a static actuarial tool such as the RoC*RoI (Yesberg & Polaschek, 2015). However, to date only one study has directly examined the predictive validity of the DRAOR on matched samples of women and men (Yesberg, Scanlan, Serin, Hanby, & Polaschek, 2015).

The Risk of Re-conviction X Risk of Re-imprisonment. (RoC*RoI; Bakker et al., 1999). The RoC*RoI is an actuarial risk assessment tool developed in New Zealand and validated on two samples (where each validation comprised 24,000 offenders). The RoC*RoI expresses risk as the probability an offender will reoffend resulting in reimprisonment over a five year period in the community. For example, a RoC*RoI score of 0.63 indicates that the offender has a 63 percent likelihood of returning to prison within five years. Or, to put another way, of a sample of 100 offenders with the same score of 0.63 it would be expected that 63 of the 100 would be reimprisoned within five years. Because the offenders in this research were serving community sentences, not all had previously been imprisoned. Thus, for those offenders the RoC*RoI was predicting first time imprisonment.

The RoC*RoI requires no clinical judgment, as it is based on a computer algorithm that includes entirely static factors. The static factors consist of criminal history variables such as number of previous convictions and demographic factors (e.g., age; Bakker, et al., 1999). Analysis during development showed the RoC*RoI had moderate to high predictive validity (AUC = .76; Bakker, et al., 1999). More recent analysis has shown the RoC*RoI to have good predictive validity over three years post-release (Nadesu, 2007).

Procedure

Calculation of DRAOR TS. When evaluating risk tools that include protective factors, the convention in research is to subtract the protective score from the risk score to create a composite score. This convention is used because protective factors are conceptualised as having a direct risk-reducing impact on risk factors. The composite score is considered to be a representation of an offender's risk of recidivism accounting for protective factors (de Vries Robbé, de Vogel, Douglas, & Nijman, 2015).

This research followed convention and a DRAOR TS was calculated manually for each offender. The TS was calculated for by summing the acute and stable risk subscale scores and subtracting the protective subscale score. The TS can range from -12 to +26 and is an index of an offender's risk corrected for his or her available protective factors. A higher DRAOR TS is indicative of higher risk of recidivism, because of the disproportionate presence of risk factors and absence of protective factors.

Extraction of DRAOR scores. For this research two sets of scores on the DRAOR were extracted where available for each offender: the initial and the proximal scores. The initial DRAOR score was the offender's third DRAOR assessment following commencement of his or her community supervision sentence. The third DRAOR was used in this research because the scores are considered to be more reliable than those from the first assessment. By the third assessment the probation officer has had time to get to know the client and score the DRAOR items as they relate specifically to the client (Hanby, 2013).

The proximal DRAOR score came from the assessment just prior to a reoffence or, for offenders who did not commit a breach reoffence, from the last DRAOR assessment available on sentence. For offenders who convicted a criminal reoffence after their sentence end date or were not reconvicted, the proximal DRAOR was also the last assessment on sentence. The proximal DRAOR scores are considered time-dependent in this research because they are the second wave of dynamic assessment and occurred after the initial assessment. Both initial and proximal DRAOR scores came from assessments before the date of a reoffence and as a result were free from contamination of the actual offending behaviour.

Recidivism. Two recidivism indices were extracted from the New Zealand National Records Database on 13 June 2014. Recidivism outcomes were based on the date of the offence for which the offender was convicted. Breach recidivism was defined as an offender's first reconviction for a breach of community supervision conditions that occurred within their supervision sentence start and end dates. The breach reconviction outcome did *not* include other criminal offending and will henceforth be referred to as breaches. The second index of recidivism used was any new reoffences, which was defined as an offender's first new criminal conviction excluding convictions for breaches of community supervision. This outcome will henceforth be referred to as criminal reoffending. The two recidivism indices were recoded dichotomously for each offender 0 = no 1 = yes based on whether the reoffence occurred within the time-at-risk (see below).

Time-at-risk. In this research time-at-risk referred to the period of time over which it was possible for an offender to commit a reoffence. For breaches, time-at-risk was the length of an offender's community supervision sentence. Women's sentence lengths ranged from 180 to 693 days (M = 275 days, SD = 76.4 days) and men's from 180 to 686 days (M = 274.3 days, SD = 78.7 days). Because women and men were successfully matched on sentence length there was no difference between the two samples' time-at-risk to commit a breach reoffence (see Table 2).

For criminal reoffending, time-at-risk consisted of the length of the offender's community supervision sentence in addition to the length of time in the community post-sentence until the date of data extraction (13 June 2014). Women's time-at-risk post-sentence ranged from 168 to 1003 days (M = 537.2, SD = 207.7) and men's from 170 to 968 days (M = 527.2, SD = 219.5). The two samples' mean time-at-risk post-sentence did not significantly differ $t(402) = 0.47$, $p = .636$. Total time-at-risk (inclusive of sentence length and post-sentence follow-up) for women ranged from 382 to 1186 days (M = 812.2, SD = 205.4) and for men from 374 to 1240 days (M = 801.4, SD = 215.5). The total time-at-risk for criminal reoffending did not significantly differ between women and men $t(402) = 0.51$, $p = .608$.

Survival days. The survival days differed by offender. For offenders who were convicted of reoffence within the time-at-risk, his or her survival days were the days between his or her supervision sentence start date and the date of the reoffence. For offenders who were not convicted of a breach, survival days were the days between sentence start date and sentence end date (i.e. sentence length). While, for offenders who were not convicted of a criminal reoffence, survival days were the number of days between his or her sentence start date and the date of data extraction (13 June 2014).

Data Analysis

The statistical methods used in the present research are outlined in detail below. All data analysis in this research was performed on the IBM SPSS Statistics version 22.

Kaplan-Meier survival analyses. Kaplan-Meier analyses facilitate the examination of rate and frequency of recidivism controlling for varying follow up lengths. The Tarone-Ware statistic of equality is one of three tests that compute the weighted difference between the observed and expected number of recidivists at each time interval on the survival curve (Norušis, 2004). In this research the Tarone-Ware was used to determine whether significant differences existed between women and men in the rate and frequency of recidivism.

Cox regression. The Cox regression is a semi-parametric analysis that models the relationship between single or multiple predictors and an event (e.g., recidivism). The advantage of the Cox regression over other regression models is that it accounts for variation in time to the event (e.g., reoffence) and censoring (e.g., no reoffence; Tabachnick & Fidell, 2013). The hazard ratio is a parameter generated from the Cox regression analysis that predicts change in the hazard (i.e. recidivism) per one-unit change in the predictor variable (Lewis et al, 2012). The ratio is a "semi-parametric analysis in that no assumption is made about the shape of the hazard function, but assumptions are made about how covariates affect the hazard function" (Hanby, 2013, p. 61). The hazard ratio is an effect size of a given predictor's relationship with recidivism: a ratio greater than 1.00 indicates that increases in the predictor are associated with increases in recidivism while values less than 1.00 indicate that increases in the predictor are associated with decreases in recidivism (Lewis et al., 2012).

Area Under the receiver operating Curve (AUC). AUC analyses were computed using the X*Beta score generated from the Cox regression analyses. The standardised beta score represents the relationship between a single and/or multiple predictor variables and recidivism. The AUC provides an estimate of a predictor's discriminative accuracy. The AUC value equals the probability that a score drawn at random from one sample (e.g., recidivists) will be higher than that drawn from another sample (e.g., non-recidivists; Rice & Harris, 2005). The advantage of the AUC analysis is that unlike Cox regression, the analysis facilitates the direct comparison of a predictor's accuracy across two or more groups, because it allows for an unequal or low number of recidivist events (Eher et al., 2012). An AUC value can range from 0 to 1.00, where a value closer to 1 indicates the measure is able to distinguish recidivists from non-recidivists, and AUC values closer to 0.5 indicate the predictor is no better than chance at distinguishing recidivists from non-recidivists. In forensic literature, Rice and Harris (2005) have recommended the following qualitative interpretations of AUCs: 0.556 indicates low accuracy, 0.639 moderate accuracy, and 0.714 high accuracy. In this research, to compare the accuracy of predictors across two groups (i.e. women and men), examination of the associated 95% confidence intervals is required. If the

95% CI substantially overlap for the two samples under a given reconviction the accuracy of the given predictor will be concluded as comparable given the true AUC could be identical for the two samples. However, if there is no overlap, the accuracy of the predictor is considered to be significantly different across the two samples.

Interpretation of Analyses

This research will attempt to address the pitfalls of null hypothesis significance testing (NHST; Cumming, 2014) by supplementing the *p* values with standardised effect sizes (ES). The standardised effect sizes used in this thesis are presented in Table 4 and all qualitative interpretations with the exception of the Phi coefficient come from Cohen (1992). Cohen’s *d* values for mean difference between the two samples on continuous variables were calculated by dividing the mean difference score by the pooled standard deviation. Cohen’s *d* values for differences within each sample on continuous variables were calculated by dividing mean difference score by mean standard deviation. The eta-squared value was the effect size used in conjunction with ANOVA analyses, and was calculated by dividing the sum of squares between-groups by the total sum of squares. The Phi coefficient was the effect size used to examine the magnitude of the mean difference between the two samples’ categorical variables[14]. Taking the square root of the chi-square value and dividing it by the sample size calculated the Phi coefficient.

Confidence intervals (CI) associated with the hazard ratios (HR) were interpreted in this research to provide an estimate of precision. If the range of the confidence interval was small it indicates the estimate was likely the true effect. However, a confidence interval with a large range indicates the precision of the estimate is weak. If a confidence interval of a hazard ratio included 1.00 it is possible the predictor had no relationship with recidivism. Additionally, if the confidence interval of two or more hazard ratios overlaps substantially, it is likely there is no statistical difference between the respective predictor variables. The reverse also applies; when confidence intervals do not overlap between two or more hazard ratios, a statistical difference between the respective predictor variables is possible.

[14] The qualitative description for the Phi coefficient presented in Table 4 comes from Rea and Parker (2005, p. 189).

Table 4

Summary of Effect Sizes

Qualitative interpretation	Small	Medium	Large
Cohen's *d*	0.20	0.50	0.80
Phi (*Φ*)	0.20	0.40	0.80
Eta-squared	0.01	0.06	0.14
Qualitative interpretation	Weak	Moderate	Strong
r coefficient	0.10	0.30	0.50

Chapter 3: Results

The results are presented in two parts. Part I focuses on the predictive validity of the DRAOR for breaches and Part II for criminal reoffences.

Part I Breaches

The DRAOR was not designed to predict breaches; however, because breaches are prevalent among community-sentenced offenders, an examination of the DRAOR with this outcome had important practical implications. Recall breaches did not include other forms of criminal reoffences. Part I opens with key preliminary descriptive analyses and the subsequent sections detail in order the results of a series of statistical analyses that examined the five following research questions:

1. Does the DRAOR predict breaches for women and men?
    - 1.1 Does the initial DRAOR predict breaches for women and men? Are the predictive components and accuracy of the initial DRAOR comparable across gender?
    - 1.2 Does the proximal DRAOR predict breaches for women and men? Are the predictive components and accuracy of the proximal DRAOR comparable across gender?
2. Is the proximal DRAOR a better predictor of breaches than the initial DRAOR?
3. Do DRAOR scores change between the initial and proximal assessment? If so, in what direction?
    - 3.1 Do female and male recidivists differ in the amount of change made compared to their non-recidivist counterparts?
4. Does change on the DRAOR predict breaches for women and men? If so, does it do so as accurately with women as men?
5. Does the RoC*RoI predict breaches? If so, does the proximal DRAOR TS add incremental predictive validity above the RoC*RoI?

Base rates of breach reconvictions. Of the 202 women in the matched sample 58 were convicted of a breach, and of the 202 men 31 were convicted of a breach. To statistically compare the rate and frequency of breaches for the two samples controlling for the varying follow-up times a Kaplan-Meier survival analysis was produced.

The rate and frequency of breaches. In *Figure 1* the horizontal axis represents the survival days (see method). The vertical axis represents the proportion of women or men at a given time point who had not been convicted of a breach. The median survival time to breach was unable to be calculated for women because the survival proportion did not drop below 0.5. Because of this the mean survival times were reported for both samples.

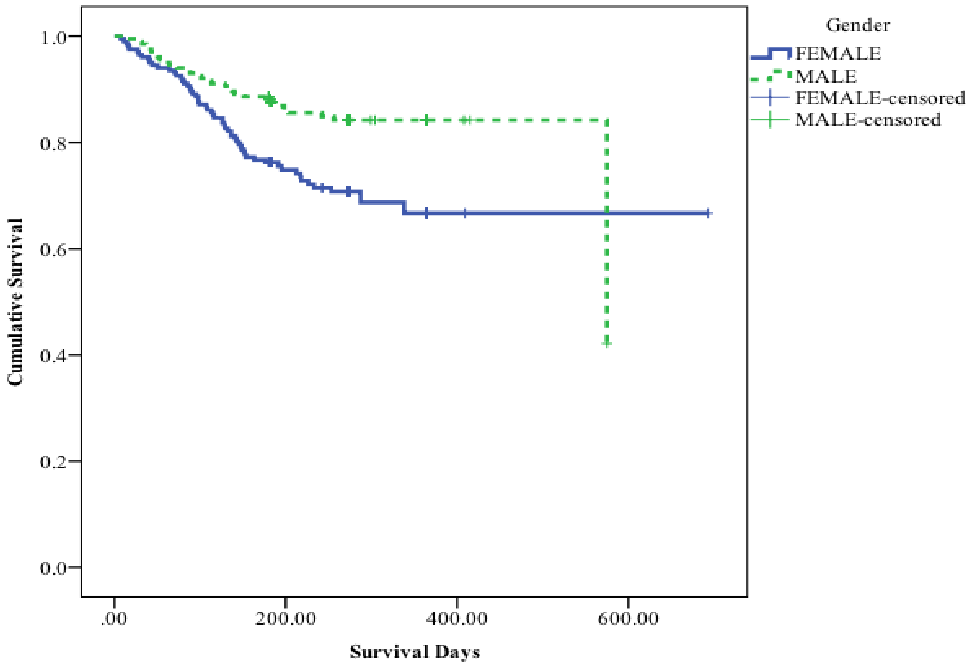

Figure 1 Survival curves of breaches for women and men

The mean survival time for women was 509.09 days (*SE* = 21.28), 95% CI [467.38, 550.81]. For men the mean survival time was 502.46 days (*SE* = 12.46), 95% CI [478.04, 526.89]. The survival analysis revealed a statistically significant difference between women and men in terms of their rate and frequency of breaches, with a Tarone-Ware statistic of equality, $\chi^2(1) = 10.549$, $p = .001$.

Inclusion criteria. For the subsequent analyses offenders from each sample were included only if they met two criteria.

1. If an offender was convicted of a breach, the breach was committed *after* their initial DRAOR assessment, and
2. Offenders had an initial and proximal DRAOR assessment on *different* days.

These inclusion criteria also applied to the offenders' matched pairs. For example, when a woman was removed due a violation of the above criteria her matched male was also excluded to ensure that the two samples remained comparable on the eight matching variables. As a result of this process 187 women and 187 men were included in all subsequent analyses in Part I.

Research question 1.1: Does the initial DRAOR predict breaches for women and men? Are the predictive components and accuracy of the DRAOR comparable across gender?

Together research question 1.1 and 1.2 were designed to understand whether the items included in the DRAOR were predictive of breaches for women and men. It is also intended to develop an understanding of how the DRAOR operated with the two samples.

Table 5

Initial DRAOR Scores and ANOVA Analysis for the Breach Outcome

	Female Recidivist (*n* = 46)	Female Non Recidivist (*n* = 141)	Male Recidivist (*n* = 23)	Male Non Recidivist (*n* = 164)	Analysis of Variance	Eta-squared
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)		
Acute	5.91^{a} (*2.54*)	5.11^{a} (*2.44*)	4.96 (*2.70*)	3.99^{b} (*2.32*)	$F(3, 370) = 10.03, p < .001$	0.08
Stable	6.20^{c} (*2.51*)	5.20^{d} (*2.24*)	5.04 (*1.61*)	5.26^{d} (*2.05*)	$F(3, 370 = 2.83, p = .039$	0.02
Protective	5.87^{e} (*2.31*)	6.77 (*2.20*)	6.65 (*1.85*)	6.92^{f} (*2.02*)	$F(3, 370) = 2.99, p = .031$	0.02
TS	6.24^{g} (*6.23*)	3.54^{h} (*5.41*)	3.35 (*4.88*)	2.33^{h} (*4.97*)	$F(3, 370) = 6.64, p < .001$	0.05

Note: Means with differing superscript were significantly different in post-hoc Tukey comparisons ($p < .05$).

Description of the samples' initial DRAOR scores. Table 5 presents the mean initial DRAOR subscale and DRAOR total score (TS) for women and men who were convicted of a breach reoffence and those who were not. A one-way between groups analysis of variance (ANOVA) and post-hoc Tukey tests were conducted to explore mean differences on initial DRAOR scores. The dependent variables were the four initial DRAOR scores and the independent variable was the four-level group variable: female recidivists, female non-recidivists, male recidivists, and male non-recidivists.

The ANOVA analysis revealed the four offender groups significantly differed from one another on all initial DRAOR scores. The female recidivist and non-recidivist groups had significantly higher initial acute scores than the male non-recidivist group; however, the female groups did not significantly differ from one another. The female recidivist group had a significantly higher initial stable score compared to the female non-recidivists and the male non-recidivists. However, the mean differences on the initial stable subscale between the female non-recidivists, male recidivists and male non-recidivists were not significantly different. The female recidivist group had a significantly lower initial protective score compared to the male non-recidivist group. The female recidivist group had a significantly higher initial TS compared to their non-recidivist counterparts and the male non-recidivist group; however, there was no significant difference in the initial TS of the female and male non-recidivist groups. The eta-squared effect size estimated the magnitude of the differences presented in Table 5 as small; however, the differences in the initial acute scores were considered medium.

Summary. The female recidivist group had higher initial DRAOR risk scores and lower initial protective scores compared to both of the non-recidivist groups. Because the survival analysis showed women were convicted of significantly more breaches it was expected that the initial DRAOR risk scores would be the highest for the female recidivist group and the protective scores the lowest, in other words, the initial DRAOR was sensitive to observed rates of breaches. Further, the initial stable and DRAOR TS were significantly different between the women who remained breach free and those who did not; however, the same sensitivity was not seen for men, suggesting for women the initial stable and DRAOR TS were sensitive to observed reoffending rates within the sample.

Univariate Cox regression models. To investigate whether initial DRAOR scores were in fact predictive of breaches as indicated by the ANOVA analysis a series of univariate Cox regression models were performed controlling for offenders' varying survival days. The univariate models investigated the strength of the predictive relationship between each initial DRAOR score and breaches. Following this a series of multivariate models were conducted to examine the differential predictive power of the initial DRAOR subscales with the two samples. Separate Cox regression analyses were performed for women and men. The initial DRAOR subscales and TS were entered individually (univariate) and in combination (multivariate) as the independent predictor variables. The criterion variable was the dichotomous breach reconviction variable and the time variable was survival days.

Recall that AUC values account for low or unequal number of recidivist events between the two samples; therefore, AUC values were produced to provide the necessary statistic to directly compare a significant predictor's discriminative accuracy across women and men. For a risk score (e.g., acute, stable, and TS) the AUC value was interpreted as the probability that a randomly selected recidivist would have a higher score than a randomly selected non-recidivist. Due to the reverse scoring of the protective subscale, the AUC value of the protective subscale was interpreted as the probability that a randomly selected recidivist would have a lower protective subscale score than a randomly selected non-recidivist.

Table 6

Univariate Cox Regression Models of the Initial DRAOR Predicting Breaches

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.12 (0.06)	4.06*	1.13 [1.00, 1.26]	0.58 [0.49, 0.68]	0.16 (0.08)	3.89*	1.17 [1.00, 1.37]	0.62 [0.48, 0.75]
Stable	0.16 (0.06)	6.60*	1.18 [1.04, 1.33]	0.60* [0.50, 0.70]	-0.03 (0.10)	0.08	0.97 [0.80, 1.18]	0.52 [0.41, 0.63]
Protective	-0.16 (0.07)	5.77*	0.85 [0.75, 0.97]	0.60* [0.51, 0.69]	-0.09 (0.11)	0.73	0.91 [0.74, 1.12]	0.55 [0.43, 0.67]
TS	0.08 (0.03)	8.48**	1.09 [1.03, 1.15]	0.61* [0.51, 0.70]	0.05 (0.04)	1.27	1.05 [0.97, 1.13]	0.57 [0.46, 0.69]

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

As shown in Table 6 all initial DRAOR scores were significant predictors of breaches for women; however, for men the only significant predictor was the initial acute subscale. The initial acute subscale predicted breaches for women $\chi^2(1, N = 187) = 4.05$, $p = .044$ and for men $\chi^2(1, N = 187) = 3.97$, $p = .046$. The initial stable score predicted breaches for women $\chi^2(1, N = 187) = 6.66$, $p = .010$, but not for men $\chi^2(1, N = 187) = 0.08$, $p = .772$. The initial protective score predicted breaches for women $\chi^2(1, N = 187) = 5.68$, $p = .017$, but not for men $\chi^2(1, N = 187) = 0.73$, $p = .394$. The initial TS predicted breaches for women $\chi^2(1, N = 187) = 8.44$, $p = .004$, but not for men $\chi^2(1, N = 187) = 1.28$, $p = .258$.

For women, each initial DRAOR subscale and the TS had a statistical relationship with breaches. In other words, the initial DRAOR scores were useful predictors of breaches. Recall, the size of the statistical relationship was estimated by the hazard ratio (HR) and the precision by the associated confidence interval (CI). The women's initial stable subscale had a HR of 1.18. This means that for every one-point increase on the initial stable score, the likelihood of a woman being convicted of a breach increased by 18%. However, although the CI associated with this ratio did not span 1.00 it only did so only marginally as indicated by the lower band of the CI. The lower and upper bands of the hazard ratio CI indicated the precision of the predicted increase in recidivism was lower compared to the smallest HR of 1.09 for the women's initial DRAOR TS, which had a CI that was more condensed and therefore more precise.

The protective subscale is reversed scored, so to interpret the HR I subtracted the ratio from one. This allowed the determination of the percentage decrease in the predicted likelihood of a breach reconviction per one-unit increase on the protective subscale. For women, the initial protective subscale HR was 0.85, which means that for every one-point increase on initial protective score, the likelihood of a woman being convicted of a breach decreased by 15%.

For men, expect for the initial acute subscale, the initial DRAOR scores were non-significant predictors. The HR for the men's initial acute subscale indicated that for every one-point increase on the acute subscale the likelihood of a man being convicted of a breach increased by 17%. However, because the lower band of the CI was 1.00, it was plausible that a one-point increase on the initial acute subscale could have no effect on the likelihood of a breach.

For women, all AUC values were significant with the exception of the acute subscale. This means the women's stable and protective subscales and TS not only had a significant predictive relationship with breaches; they were also better than chance at distinguishing recidivists from non-recidivists. For example, the women's stable subscale had an AUC value of 0.60, which meant that a randomly selected female recidivist was 60% more likely than a randomly selected female non-recidivist to have a higher initial stable score. For women, the AUC values indicated the stable, protective and TS had moderate discriminative accuracy (Rice & Harris, 2005). For women, although the stable subscale had the strongest relationship with breaches, as estimated by the HR, when taking into account base rates the AUC values indicated the four initial DRAOR scores had comparative discriminative accuracy.

Although the initial acute subscale was a significant predictor for women and men as estimated by the HR, when taking into account base rates of breaches, the AUC values were non-significant. This was not unexpected given the lower band of both hazard ratios CI were 1.00. The comparison of the AUC values across women and men showed the confidence intervals (CIs) substantially overlapped, thus the true accuracy of the initial DRAOR scores were likely comparable for women and men.

The relationship between initial DRAOR scores. To examine the strength and direction of the relationship between the initial DRAOR scores Pearson bivariate correlations were performed. As evident in Table 7 the initial acute and stable subscale scores were positively correlated with one another and the TS. The initial protective subscale scores, as expected, were negatively correlated with initial DRAOR risk scores for both samples. The size of the coefficients for the subscales ranged from 0.38 to 0.60 for women, and 0.23 to 0.56 for men. For women, the *r* coefficient indicated the strength of the relationship between the subscales ranged from moderate to strong. For men, the size of the relationship between the protective and stable subscales was considered weak. The strength of the relationship between the other DRAOR scores ranged from moderate to large. As evident in Table 7 the correlations between the two samples' subscales and the DRAOR TS were estimated to be large effects. This was also expected given the DRAOR TS was a composite of all three subscales.

The correlations shown in Table 7 suggest the initial DRAOR scores have convergent validity. Convergent validity is a subtype of construct validity and is used when the relationship between two or more theoretically related measures is empirically shown to be measuring the same construct (i.e. strongly correlated). Convergent validity was expected in this research given the three subscales were designed as independent measures of recidivism.

Table 7

Correlation Matrix of Initial DRAOR Scores for the Breach Outcome

	Women				Men			
	Acute	Stable	Protective	TS	Acute	Stable	Protective	TS
Acute	1	-	-	-	1	-	-	-
Stable	.47**	1	-	-	.45**	1	-	-
Protective	-.38**	-.60**	1	-	-.23**	-.56**	1	-
TS	.78**	.85**	-.80**	1	.75**	.85**	-.74**	1

** $p < .01$

Multicollinearity. Multicollinearity is a statistical phenomenon that refers to the situation where two or more predictors in a regression model are strongly correlated (Hair, Black, Babin, Anderson, & Tatham, 2006). As shown in Table 7 a number of initial subscales are highly linearly correlated ($r \geq 0.50$). The impact of multicollinearity is particularly problematic when the research question concerns how a set of variables influences an outcome. In a multivariate Cox regression model where predictors X and Y are highly correlated the impact of predictor Y on the outcome (recidivism in this research) is not independent of predictor X, which means there will be an imperfect estimate of the impact of predictor Y on recidivism. Re-sampling and/or obtaining a larger sample are suggested methods to remedy multicollinearity. However, because of the archival nature of the data and the strict matching and inclusion criteria these were not possible.

To determine the severity of multicollinearity on the subsequent multivariate regression models, multiple regression models were used to generate the Variance Inflation Factor (VIF). The VIF is a numerical index of the level of inflation between the predictor variables above that expected if there was no correlation (i.e. no multicollinearity; Lin, 2006). A VIF of ≤ 10.00 indicates that multicollinearity does not have a significant impact on the interpretation of the regression models (Lin, 2006; O'Brien, 2007). The VIF for all multivariate predictors in the research were ≤ 2.00. Thus all multivariate models in Part I and Part II of the results chapter were interpreted with a level of confidence had multicollinearity not existed. Additionally, the relationship between individual predictor variables and recidivism was considered independent of other predictor variables.

Multivariate Cox regression models. The multivariate analysis enabled the examination of whether the subscales in combination were predictive of breaches for women and men, as well as an examination of the individual subscales' differential predictive power. I entered the three subscale scores together in block 1 of a Cox regression model predicting breaches.

The model including all three subscales, predicted breaches for women $\chi^2(3, N = 187) = 8.54$, $p = .036$. As shown in Table 8 no individual subscale significantly contributed to the model independent of another; in other words, it was the combination of the three subscales together that contributed to the predictive power of the model. This result also indicated that each subscale explained a similar proportion of variance in breaches. This finding was not unexpected because each subscale was significantly correlated with each other, and each subscale in the univariate analyses was a statistically significant predictor. As evident in

Table 8 the hazard ratios (HRs) for the three subscales were similar and the confidence intervals of the HRs for each subscale substantially overlapped.

For men, the model predicted breaches $\chi^2(3, N = 187) = 7.88$, $p = .049$. As seen in Table 8 the acute subscale made a statistically significant unique contribution to the model after taking into account the contribution of the stable and protective subscales. This finding was expected because the acute subscale was the only individual subscale that had a statistically significant relationship with breaches in the univariate analyses. It indicated that the relationship between the initial acute subscale and breaches was robust. As shown in Table 8 the HR for the men's acute subscale in the multivariate model (1.26) was considerably higher than the HR for the acute subscale in the univariate model (1.17). It was likely the increased effect size of the acute subscale was the result of the shared variance between the subscales bolstering the strength of the relationship between the initial acute subscale and breaches.

For women, the AUC value was 0.61 which indicated that a randomly selected female recidivist was 61% more likely to score higher on the acute and stable subscales and lower on the protective subscale compared to a randomly selected non-recidivist, after accounting for the shared variance of the subscales. For men, the AUC value was 0.67, which indicated that a randomly selected male recidivist was 67% more likely than a randomly selected non-recidivist to have a higher initial acute score after accounting for the variance of the stable and protective subscales.

Although men had a slightly higher AUC value, it was concluded that both multivariate models had moderate accuracy because of the substantial overlap between the CIs associated with each sample's AUC value.

Table 8

Multivariate Model of Initial DRAOR Subscale Scores Predicting Breaches

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.06 (0.07)	0.78	1.06 [0.93, 1.21]		0.23 (0.09)	5.98*	1.26 [1.05, 1.51]	
Stable	0.10 (0.08)	1.50	1.10 [0.94, 1.29]	0.61* [0.52, 0.70]	-0.27 (0.14)	3.54	0.77 [0.58, 1.01]	0.67** [0.55, 0.79]
Protective	-0.09 (0.08)	1.25	0.91 [0.78, 1.07]		-0.17 (0.13)	1.69	0.85 [0.66, 1.09]	

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

Research question 1.2: Does the proximal DRAOR predict breaches for women and men? Are the predictive components and accuracy of the proximal DRAOR comparable across gender?

Research question 1.2 examined the predictive validity of the *proximal* DRAOR scores. Recall that the proximal DRAOR was the most-up-date assessment of offenders’ dynamic risk.

The relationship between the proximal DRAOR scores. Pearson bivariate correlations are presented in Table 9. All *r* coefficients were significant and in the expected direction. The strength of the relationship between the proximal DRAOR scores ranged from moderate (0.46) to strong (0.86). As was the case for the initial DRAOR scores the high correlation coefficients indicated convergent validity of the subscales, as was expected.

Table 9

Correlation Matrix of Proximal DRAOR Scores for the Breach Outcome

	Women				Men			
	Acute	Stable	Protective	TS	Acute	Stable	Protective	TS
Acute	1	-	-	-	1	-	-	-
Stable	.58**	1	-	-	.46**	1	-	-
Protective	-.53**	-.63**	1	-	-.49**	-.64**	1	-
TS	.83**	.87**	-.85**	1	.78**	.85**	-.86**	1

** $p < .01$

Description of the samples’ proximal DRAOR scores. Table 10 presents the mean proximal DRAOR scores for the four recidivist groups. As with the initial DRAOR scores an ANOVA was conducted to examine whether the four groups significantly differed on their proximal DRAOR scores.

Table 10

Proximal DRAOR Scores and ANOVA Analysis for the Breach Outcome

	Female Recidivist (n = 46)	Female Non Recidivist (n = 141)	Male Recidivist (n = 23)	Male Non Recidivist (n = 164)	Analysis of Variance	Eta-squared
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)		
Acute	5.76[a] (*2.88*)	3.86[b] (*2.37*)	4.57[ab] (*2.64*)	3.04[c] (*2.00)*	F (3, 370) = 14.30, p < .001	0.13
Stable	6.22[d] (*2.66*)	4.43[e] (*2.46*)	5.04 (*2.20*)	4.08[e] (*2.39*)	F (3, 370) = 9.61, p < .001	0.07
Protective	5.80[f] (*2.63*)	7.60[g] (*2.48*)	6.74 (*1.91*)	8.04[g] (*2.19*)	F (3, 370) =11.79, p < .001	0.09
TS	6.17[h] (*6.90*)	0.68[ij] (*6.03*)	2.87[ih] (*5.47*)	-0.92[j] (*5.44*)	F (3, 370) = 18.55, p < .001	0.13

Note: Means with differing superscript were significantly different in post-hoc Tukey comparisons (p < .05).

The post-hoc Tukey tests showed that the female recidivists had significantly higher proximal acute scores compared to female non-recidivists and male non-recidivists. Male recidivists also had a significantly higher proximal acute score compared to the male non-recidivist. However, there was no significant difference in proximal acute scores between female non-recidivists and male recidivists or between female and male recidivists. Female recidivists had a significantly higher proximal stable score compared to the female non-recidivist and the male non-recidivist groups. There were no other significant differences in proximal stable scores between the groups. Female recidivists had a significantly lower proximal protective score compared to female non-recidivists and male non-recidivists; there were no other significant differences in proximal protective scores between the groups. The female recidivists had significantly higher proximal TS compared to female non-recidivists and male non-recidivists. The male recidivist group also had a significantly higher proximal TS compared to the male non-recidivist group. There was no difference between the two recidivist groups proximal TS, the two non-recidivist groups proximal TS, or the female non-recidivists and the male recidivists proximal TS. The eta-squared estimated the differences in the proximal DRAOR scores to be medium in magnitude. The differences in the proximal acute and DRAOR TS were just below the cut off for a large effect size.

Summary. The ANOVA analysis of the proximal DRAOR scores parallels that of the initial DRAOR scores in that the female recidivist group had higher proximal DRAOR risk scores compared to the non-recidivist groups and lower protective DRAOR scores. The proximal DRAOR, like the initial DRAOR, was sensitive to the actual rate of breaches, as the women had higher DRAOR scores. However, unlike the analysis of the initial DRAOR scores, all four proximal DRAOR scores significantly differed between women who were convicted of a breach and those who were not. For the men, only the proximal acute and TS significantly differed between recidivists and non-recidivists. Unlike the initial DRAOR scores, the proximal DRAOR was unique in its sensitivity to recidivist and non-recidivists within each sample, this was particularly apparent for women.

Univariate models of the proximal DRAOR scores. The statistical models examined for the proximal DRAOR scores were identical to those for the initial DRAOR scores. A summary of the proximal DRAOR's predictive validity and discriminative accuracy with women and men are presented below.

Table 11

Univariate Models of the Proximal DRAOR Scores Predicting Breaches

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.24[a]	20.67***	1.28	0.69***	0.22[e]	7.82**	1.25	0.67**
	(0.05)		[1.15, 1.42]	[0.60, 0.78]	(0.08)		[1.07, 1.46]	[0.55, 0.79]
Stable	0.23[b]	15.97***	1.26	0.68***	0.14[f]	2.91	1.15	0.63*
	(0.06)		[1.13, 1.41]	[0.59, 0.77]	(0.08)		[0.98, 1.35]	[0.51, 0.74]
Protective	-0.22[c]	15.99***	0.81	0.69***	-0.21[g]	5.89*	0.81	0.67**
	(0.05)		[0.73, 0.90]	[0.60, 0.77]	(0.09)		[0.68, 0.97]	[0.57, 0.78]
TS	0.11[d]	25.65***	1.12	0.73***	0.09[h]	7.40**	1.09	0.70**
	(0.02)		[1.07, 1.17]	[0.64, 0.81]	(0.03)		[1.03, 1.16]	[0.59, 0.81]

[a] Model $\chi^2(1, N = 187) = 21.44$, $p < .001$
[b] Model $\chi^2(1, N = 187) = 16.62$, $p < .001$
[c] Model $\chi^2(1, N = 187) = 16.18$, $p < .001$
[d] Model $\chi^2(1, N = 187) = 25.65$, $p < .001$
[e] Model $\chi^2(1, N = 187) = 8.11$, $p = .004$
[f] Model $\chi^2(1, N = 187) = 2.96$, $p = .085$
[g] Model $\chi^2(1, N = 187) = 5.66$, $p = .017$
[h] Model $\chi^2(1, N = 187) = 7.46$, $p = .006$
* $p < .05$ ** $p < .01$ ** $p < .001$

As shown in Table 11, all individual proximal subscales and the TS significantly predicted breaches for women. The relationship between the proximal predictors and breaches was strongest for the proximal acute subscale as estimated by the HR of 1.28. For every one-point increase on the proximal acute subscale a woman was 28% more likely to be convicted of a breach. However, given the CI associated with the HR overlapped with the other subscales no definitive conclusions could be drawn about which subscale had the strongest relationship with breaches.

The acute, protective and TS all significantly predicted breaches for men. As for the women, the strongest relationship was between the proximal acute subscale and breaches. For men, every one-point increase on the proximal acute subscale the likelihood of being convicted of a breach increased by 25%. The CI for the HR of the proximal stable subscale spanned 1.00, which meant it did not significantly predict breaches for men. This was because it was plausible that the HR may have been 1.00, which would mean that changes on the subscale would have no effect on the likelihood of a breach.

For women, all three proximal subscales and the TS not only predicted the likelihood of breaches they also had moderate to high discriminative accuracy as estimated by significant AUC values, which ranged from 0.68 to 0.73. After taking into account the low base rate of breaches for men, the proximal acute, protective and TS were supported by significant AUC values that ranged from 0.67 to 0.70, indicating moderate discriminative accuracy. Although there was no significantly detectable relationship between the proximal stable subscale and breaches for men, when taking into account the low base rate of recidivist events in the male sample, the AUC value for the subscale was significant, albeit marginally (the lower band of the CI was 0.51).

As evident in Table 11 the proximal DRAOR scores were slightly more accurate for women as indicated by the AUC values. However, because the CIs associated with the values substantially overlapped, it was plausible that the true estimate of proximal DRAOR's accuracy may have been equal for women and men.

Multivariate models of proximal DRAOR scores. As shown in Table 12 the multivariate model, which included all proximal subscales in block 1 of the Cox regression predicted breaches for women $\chi^2(3, N = 187) = 26.35, p < .001$. The acute subscale was shown to provide a statistically significant unique contribution to the model after taking into account the shared variance of the stable and protective subscales, this indicated that the proximal acute subscale drove the predictive power of the proximal TS for women. This result was not unexpected given that the proximal acute subscale had the strongest relationship with breaches in the univariate analyses. It is important to note that the HRs for all three subscales in the multivariate model were lower than the corresponding HR for the subscale individually. This was likely the result of the shared variance between the subscales and the unexplained variance introduced by the stable and protective subscales.

For men, the multivariate model predicted breaches $\chi^2(3, N = 187) = 9.23, p = .026$. It was all three proximal DRAOR subscales in combination that contributed to the predictive power; in other words, no single subscale uniquely contributed over and above another. Although the acute subscale had the strongest relationship with breaches in the univariate analyses, the addition of the protective and non-significant stable subscale likely introduced considerable unexplained variance. Thus it is plausible the acute and protective subscales relationship with breaches in the multivariate model was undermined by the addition of unexplained variance. Each multivariate model had high discriminative accuracy as estimated by women's AUC value of 0.73 and men's 0.71. Analysis of the CIs associated with both AUC values showed identical upper arms and almost identical lower arms; thus due to the considerable overlap, it was unlikely the accuracy of the proximal multivariate models were different for women and men.

Table 12

Multivariate Model of the Proximal DRAOR Subscales Predicting Breaches

	Women[a]				Men[b]			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.15 (0.07)	4.46*	1.16 [1.01, 1.33]		0.17 (0.10)	2.93	1.19 [0.98, 1.44]	
Stable	0.11 (0.07)	2.29	1.12 [0.97, 1.29]	0.73*** [0.65, 0.82]	-0.02 (0.11)	0.04	0.98 [0.78, 1.22]	0.71** [0.61, 0.82]
Protective	-0.08 (0.07)	1.33	0.92 [0.81, 1.06]		-0.13 (0.13)	0.94	0.88 [0.68, 1.14]	

* $p < .05$ ** $p < .01$ ** $p < .001$

Summary: Research question 1. For women, the DRAOR was a robust predictor of breaches. Each initial acute subscale contributed equally to the predictive power of the initial DRAOR, while the proximal acute subscale drove the predictive power of the proximal DROAR. For men, the initial acute subscale emerged as the subscale driving the predictive power of the initial DRAOR; while all three proximal subscales in combination contributed to the predictive power of the proximal DRAOR. In general, the DRAOR was a poorer predictor of breaches for men; however, when taking into account the unequal base rates, the DRAOR had comparative accuracy across gender.

Compared to the initial DRAOR, the proximal DRAOR had larger hazard ratios and AUC values for both samples, suggesting the proximal DRAOR was a better predictor than the initial. But because the confidence intervals associated with the hazard ratios and AUC values overlapped no definitive conclusion could be drawn regarding the predictive superiority of the proximal DRAOR scores. The next research question examines this directly by conducting a series of incremental predictive validity analyses.

Research Question 2: Is the proximal DRAOR a better predictor of breaches than the initial DRAOR?

Relationship between initial and proximal DRAOR scores. Before the incremental validity analyses were performed, Pearson bivariate correlations were conducted to the relationship between the initial and corresponding proximal DRAOR scores for women and men. As shown in Table 13 all initial and proximal DRAOR scores were significantly positively correlated for both samples. The correlations of the two samples scores as estimated by r were strong; the lowest r value was 0.56.

Table 13

Correlation Matrix of Initial and Corresponding Proximal DRAOR Scores for the Breach Outcome

	Women				Men			
	Proximal acute	Proximal stable	Proximal protective	Proximal TS	Proximal acute	Proximal stable	Proximal protective	Proximal TS
Initial acute	.60**				.56**			
Initial stable		.68**				.56**		
Initial protective			.60**				.61**	
Initial TS				.68**				.60**

** $p < .01$

Incremental predictive validity of the proximal DRAOR scores. Research question two examines a different type of validity: incremental predictive validity. Only those results from the previous significant univariate results were used. A series of multivariate Cox regression analyses was performed to assess whether the proximal DRAOR scores contributed predictive power above that of initial DRAOR scores. In each model the proximal DRAOR score was entered alongside the corresponding initial DRAOR score as independent predictor variables. The criterion variable was the dichotomous breach reconviction variable and the time variable was survival days.

Recall from the univariate analyses that for men the initial and proximal stable, initial protective and initial DRAOR TS were not significant predictors of breaches: thus these were not included in the incremental analyses; however, all other pairings were included for both samples. See Appendix C for full details of excluded models.

For women, the proximal acute score contributed unique predictive power to the model when entered alongside the initial acute score predicting breaches $\chi^2(2, N = 187) = 21.73$, $p < .001$. As seen in Table 14, the addition of the proximal acute score rendered the initial acute score a non-significant predictor. As evidenced by the Wald value of 0.13 the proximal acute score was a better predictor of breaches for women as it accounted for almost all the variance that the initial acute score had previously and more. The proximal stable score accounted for all the predictive validity of the initial stable score for, women, as evidenced by the Wald value of 0.00, $\chi^2(2, N = 187) = 16.62$, $p < .001$. The result was similar with the addition of the women's proximal protective subscale $\chi^2(2, N = 187) = 16.18$, $p < .001$, and the proximal TS $\chi^2(2, N = 187) = 25.78$, $p < .001$.

For men, the proximal acute score contributed unique predictive power to the model when entered alongside the initial acute score $\chi^2(2, N = 187) = 8.88$, $p = .012$. The addition of the proximal acute score rendered the initial acute score a non-significant predictor.

The HRs of the incremental models are presented in Table 14. For women, the HRs for the initial DRAOR scores were reduced to 1.00 or below following the introduction of the proximal DRAOR scores. Such a finding supports the superior predictive power of the proximal DRAOR scores. In other words, once the proximal DRAOR score is considered the initial DRAOR scores have in most cases no statistical influence on the likelihood of breaches. For men, the addition of the proximal acute score did not reduce the observed HR of the initial score to 1.00 or below; however, the lower band of the CI went below 1.00; thus it is plausible that true HR for the initial acute subscale was 1.00.

Table 14

Incremental Predictive Validity of the Proximal DRAOR Scores Predicting Breaches

	Women			Men		
	β (SE)	Wald	Hazard Ratio [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]
Initial acute +	-0.03 (0.08)	0.13	0.97 [0.84, 1.13]	0.08 (0.09)	0.80	1.08 [0.91, 1.28]
Proximal acute	0.26 (0.06)	16.02***	1.29 [1.14, 1.47]	0.19 (0.09)	4.66*	1.21 [1.02, 1.44]
Initial stable +	0.00 (0.08)	0.00	1.00 [0.86, 1.17]			
Proximal stable	0.23 (0.07)	10.10**	1.26 [1.09, 1.45]			
Initial protective +	-0.00 (0.08)	0.00	1.00 [0.86, 1.16]			
Proximal protective	-0.21 (0.07)	10.27**	0.81 [0.71, 0.92]			
Initial TS +	-0.01 (0.03)	0.03	1.00 [0.93, 1.06]			
Proximal TS	0.12 (0.03)	17.21***	1.12 [1.06, 1.18]			

* $p < .05$ ** $p < .01$ *** $p < .001$

Summary: Research question 2. For women and men, although the respective initial and proximal DRAOR scores were predictive of breaches, when entered together the proximal DRAOR scores rendered the initial DRAOR scores non-significant predictors. In other words, the proximal DRAOR was a better predictor of breaches than the initial because it accounted for unique variance beyond that of the initial DRAOR score, and in some instances accounted fully for the variance recidivism that the initial DRAOR score was capturing. Because the proximal DRAOR assessment was the most up-to-date assessment of offenders’ dynamic risk, to find that the scores from this assessment outperformed the initial DRAOR scores fits with the theory of risk as a changeable construct.

Research questions three and four were conducted to further investigate the reasons behind the proximal DRAOR's enhanced predictive validity with women and men. The logical step forward was to analyse whether in fact the most up-to-date dynamic assessment (i.e. proximal) was different from the initial (i.e. third) assessment of dynamic risk. Research question three examined change on DRAOR risk between these two assessment points: the initial (i.e. Time 1) and proximal (i.e. Time 2).

Research Question 3: Do women and men's DRAOR scores change between the initial and proximal assessment? If so, in what direction?

Recall that the initial DRAOR scores come from an offender's third DRAOR assessment; the proximal DRAOR scores come from the assessment immediately prior to the breach, or if no reoffence, the last assessment on sentence. The initial and proximal DRAOR scores for women and men were compared using paired samples *t*-tests. As shown in Table 15 there was a statistically significant mean difference between women and men's initial and proximal DRAOR scores on all subscales and the TS. As shown in Table 15 all offenders had significantly lower mean risk scores and significantly higher mean protective scores at Time 2 compared to Time 1.

To look at the DRAOR's sensitivity to change in risk over time each offender had a DRAOR change score calculated. The change score was calculated by subtracting the proximal DRAOR score from the corresponding initial DRAOR score. For example, an initial DRAOR TS of 13 minus a proximal DRAOR TS of 10 equals a change score of +3. It is important to understand that the equation used to calculate change in this research examined the difference in DRAOR scores between two time points only. This calculation did not enable the detection of fluctuations in risk level that occurred on assessments within the elapsed time between the initial and proximal assessments.

The magnitude of change that offenders could make between the two assessments was dependent upon the size of the predictor. For example, the DRAOR TS includes 19 items meaning there was greater scope for change to be made when compared to the stable and protective subscales which each include 6 items. To account for these differences and to compare change across women and men the Cohen's *d* was calculated by dividing the mean change score by the standard deviation of the mean change. Recall a Cohen's *d* of 0.2 is considered a small effect, 0.5 a medium, and 0.8 a large effect.

The mean change scores for the two samples are presented in Table 15. Women and men had positive mean change scores for the risk subscales and the TS. A positive risk change score indicated offenders had significantly lower mean risk scores at the proximal assessment compared to the initial. For the protective subscale the mean change scores were negative, meaning offenders had significantly higher mean protective scores at the proximal assessment compared to their initial.

Table 15

Paired-Samples t-test of the Initial and Proximal DRAOR Scores for the Breach Outcome

Sample	Initial DRAOR *M* (*SD*)	Proximal DRAOR *M* (*SD*)	Paired Samples t-test	Mean change (*SD*)	Range	Cohen's *d*
Women						
Acute	5.30 (*2.48*)	4.33 (*2.63*)	$t(186) = 5.87, p < .001$	0.98 (*2.28*)	-4, +10	0.43
Stable	5.44 (*2.34*)	4.87 (*2.61*)	$t(186) = 3.92, p < .001$	0.57 (*2.00*)	-8, +7	0.29
Protective	6.55 (*2.26*)	7.16 (*2.63*)	$t(186) = -3.81, p < .001$	-0.62 (*2.20*)	-8, +9	0.28
TS	4.20 (*5.73*)	2.11 (*6.67*)	$t(186) = 5.90, p < .001$	2.17 (*5.03*)	-18, +17	0.43
Men						
Acute	4.11 (*2.38*)	3.22 (*2.14*)	$t(186) = 5.63, p < .001$	0.88 (*2.14*)	-10, +6	0.41
Stable	5.24 (*1.99*)	4.20 (*2.39*)	$t(186) = 6.80, p < .001$	1.04 (*2.09*)	-8, +8	0.50
Protective	6.89 (*2.00*)	7.88 (*2.19*)	$t(186) = -7.26, p < .001$	-0.99 (*1.86*)	-7, +8	0.53
TS	2.45 (*4.96*)	-0.45 (*5.57*)	$t(186) = 8.41, p < .001$	2.91 (*4.73*)	-26, +17	0.62

The effect size for the women's change scores ranged in magnitude from small (0.28) for the protective subscale to medium (0.43) for the acute and DRAOR TS; women made the greatest change on the acute and DRAOR TS. Compared to women, men made more change on the stable and protective subscales and TS as estimated by Cohen's *d* effect size. The Cohen's *d* indicated the men's change scores were of medium effect, with the most change being on the DRAOR TS as evidence by the Cohen's *d* of 0.62.

For women and men, each mean change score presented in Table 15 had a large standard deviation and range associated with it. These statistics indicated there was considerable variance in the amount of change and the direction of change made. For example, although as a group women had a mean change score for the acute subscale of + 0.98, which indicated a mean decrease in risk, the range showed the minimum as -4 and the maximum as +10. The minimum change score of -4 indicated that at least one woman had an increase of four points on the acute subscale across time. On the other end at least one other woman had a maximum change score of +10, which indicated she had a decrease of 10 points on the score between the initial and proximal assessments. Similarly for men, at least one man increased in risk between the initial and proximal assessment as indicated by the negative minimum range value associated with the risk change scores.

A further examination of change scores was warranted as a result of the large standard deviation and range scores observed for the whole sample. I hypothesised that it was plausible that those who reoffended in each sample made less change or in fact were responsible for the negative change scores. To investigate this hypothesis I looked at change made by women and men who were convicted of a breach and those who were not.

Research question 3.1: Do female and male recidivists differ in the amount of change made compared to their non-recidivists counterparts? *Figure 2* shows the initial and proximal DRAOR subscale and TS for the four offender groups. As shown in *Figure 2* the female recidivist group was the most risky of the four groups with the highest initial and proximal mean risk scores and the lowest initial and proximal mean protective scores. The male non-recidivist group was the least risky of the four groups with the lowest risk and highest protective scores at the proximal assessment and some overlap with male recidivists and female non-recidivists on initial DRAOR scores. The male recidivists' initial DRAOR stable and protective subscale scores appeared comparable to the two non-recidivist groups. The male recidivists' initial acute score as depicted in *Figure 2* appeared almost equal to the female non-recidivists' initial acute score. The male recidivists' initial DRAOR TS was slightly lower than the female non-recidivists' initial DRAOR TS. The male non-recidivists had the lowest initial and proximal acute and DRAOR TS scores compared to all other groups. However, the male non-recidivists' initial stable and protective score as illustrated in *Figure 2* appeared approximate to the respective scores for male recidivists and female non-recidivists.

The male recidivists' proximal acute, stable, and DRAOR total scores were higher than the scores for female non-recidivists and male non-recidivists. Compared to the male non-recidivists the female non-recidivists appeared to have higher proximal acute, stable and DRAOR total scores. The male non-recidivists had the highest proximal protective score followed by the female non-recidivists. In *Figure 2* the slope of each line indicates the magnitude and direction of the change in scores time 1 and time 2 for each of the four groups. Notice the flatter lines for the recidivist groups compared to the non-recidivist groups; a flatter line was indicative of less change.

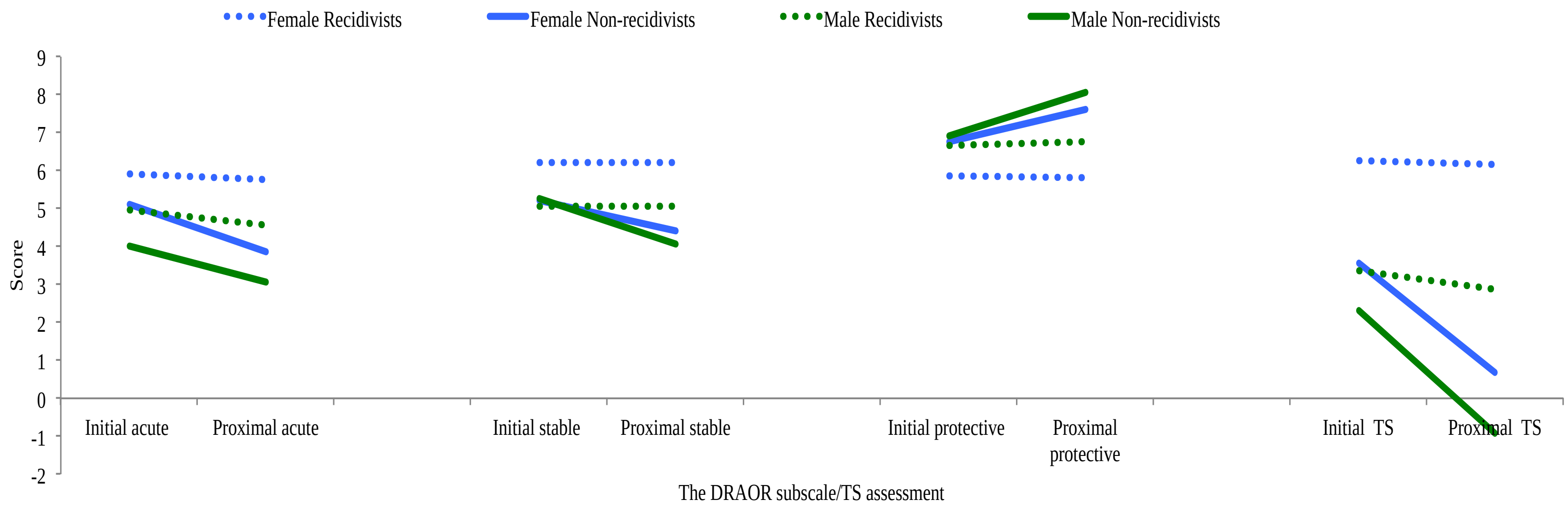

Figure 2 Initial and proximal DRAOR scores for women and men who were convicted of a breach and those who were not

Mean change scores for the four offender groups. In the following section, *Figure 3* and Table 16 are interpreted in relation to each other. *Figure 3* depicts the mean DRAOR change scores for the four offender groups. Recall a positive change value for the risk subscales indicated a lower risk score at Time 2, and a negative protective change score indicated a higher protective score at Time 2. The error bars in *Figure 3* represent the 95% CI associated with each mean change score. In *Figure 3* when the error bars do not overlap we expect the difference between the respective groups’ change scores to be significantly different; however, if the error bars overlap further analysis is required, such as an ANOVA.

Table 16 presents the results of an ANOVA analysis and post-hoc Tukey tests of change scores for the four groups presented in *Figure 3*. The dependent variables were the four change scores and the independent variable was the four-level group variable. The eta-squared statistic was computed to enable the magnitude of the difference in means between the groups to be qualitatively compared and described.

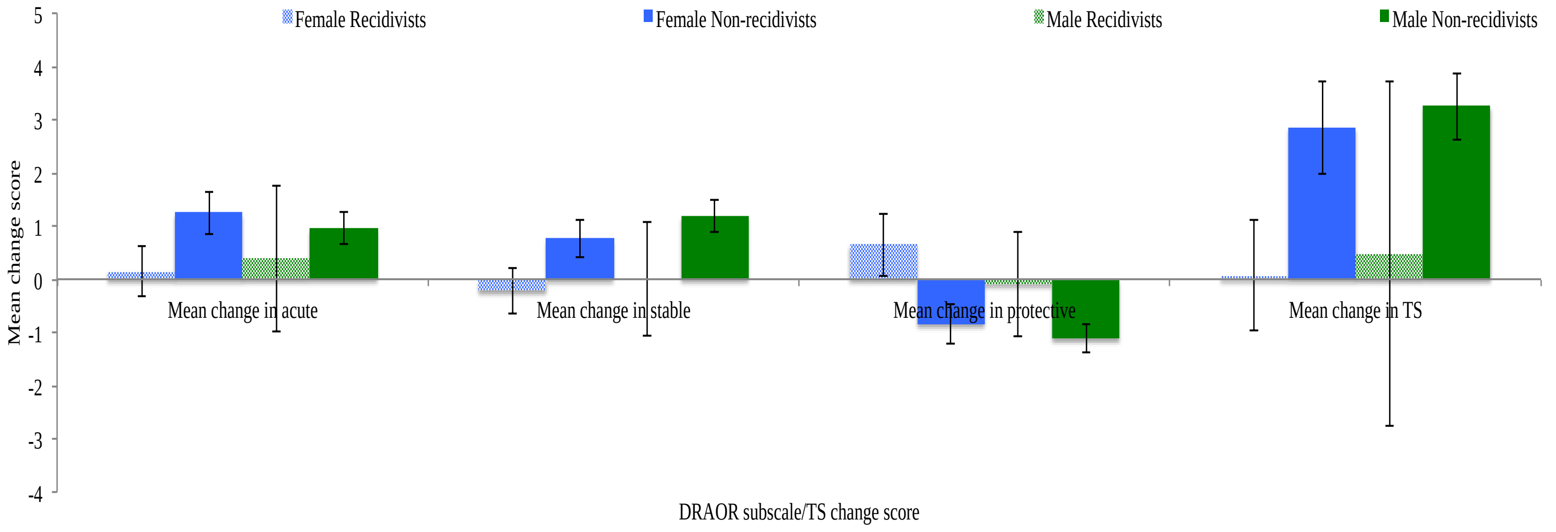

Figure 3 Mean change scores for women and men who were convicted of a breach and those who were not

Table 16

ANOVA of Mean DRAOR Change Scores for Women and Men who were Convicted of a Breach and those who were not

	Female Recidivist (n = 46)	Female Non Recidivist (n = 141)	Male Recidivist (n = 23)	Male Non Recidivist (n = 164)	Analysis of Variance	Eta-squared
	M (SD)	M (SD)	M (SD)	M (SD)		
Acute	0.15^{a} (*1.59*)	1.25^{b} (*2.41*)	0.39 (*3.17*)	0.95 (*1.96*)	$F(3, 370) = 4.47, p = .006$	0.03
Stable	-0.02^{c} (*1.44*)	0.77 (*2.12*)	0.00^{c} (*2.45*)	1.18^{d} (*1.20*)	$F(3, 370) = 7.58, p < .001$	0.04
Protective	0.07^{e} (*1.98*)	-0.84^{f} (*2.23*)	-0.09 (*2.27*)	-1.16^{f} (*1.77*)	$F(3, 370) = 5.15, p = .002$	0.04
TS	0.07^{g} (*3.50*)	2.86^{hi} (*5.27*)	0.48^{hg} (*7.48*)	3.25^{i} (*4.13*)	$F(3, 370) = 9.80, p < .001$	0.05

Note: Means with differing superscripts are significantly different ($p < .05$)

As shown in Table 16, the ANOVA revealed statistically significant differences between the four groups on all mean DRAOR change scores. *Figure 3* indicated the acute change scores were positive for the four offender groups, which meant irrespective of whether an offender was convicted of a breach or not, on average women and men, had lower acute risk scores at the proximal assessment. As expected based on the error bars presented in *Figure 3*, and as shown in Table 16 the female recidivist group made significantly less change on the acute subscale compared to the female non-recidivists. The mean acute change scores of the other groups were not statistically different.

As evident in *Figure 3* the female recidivists had a *negative* mean stable change score that indicated on average the women who were convicted of a breach had a higher stable score at the proximal assessment compared to the initial. The male recidivist group had a mean change score of zero; however, as expected the two non-recidivist groups had positive change scores. The ANOVA analysis confirmed the female and male recidivists stable change scores were statistically lower compared to the male non-recidivists; however, there were no statistically significant differences between the female and male recidivist groups.

As shown in *Figure 3* for the protective subscale, with the exception of the female recidivist group, all groups had negative mean protective change scores, which indicated higher protective factors at the proximal assessment. The female recidivist group; however, had a positive protective change score that indicated on average lower protective scores at the proximal assessment. As expected based on the error bars in *Figure 3* the female recidivist group made significantly less change on the protective subscale compared to the two non-recidivist groups; the two non-recidivist groups did not statistically differ.

Finally, as shown in *Figure 3*, all offender groups had positive DRAOR total change scores; however, the female recidivist group had a particularly small positive change score. The female recidivists had a significantly lower mean DRAOR total change score compared to female and male non-recidivists. The male recidivists had a significantly lower DRAOR total change score compared to their non-recidivist counterparts. There were no other significant differences between the four groups. As with the change scores for the subscales the statistical differences found for the DRAOR total change scores were expected given the error bars in *Figure 3*.

The above differences were significant as estimated at the $p < .05$ level. The eta-squared effect size was computed to indicate the magnitude of the mean differences. The values indicated that the size of the differences between the offender groups described above were small. The mean difference between the groups DRAOR total change score had an eta-squared of 0.5, which was the highest and just below the cut off for medium effect size.

Summary: Research question 3. The paired samples *t*-test showed the offenders' DRAOR scores changed over time. As expected, at the whole sample level both women and men had positive mean change scores and negative mean protective change scores. On further examination, at a more refined level the ANOVA results suggested the amount of change on the DRAOR was dependent not on whether you were a woman or man but on whether the offender was convicted of a breach. There were no statistical differences between recidivist groups or non-recidivist groups to suggest an effect of gender. Put simply, both female and male recidivists made less change compared to their non-recidivist counterparts. Research question three showed the DRAOR scores changed overtime, research question four examines whether the change on the DRAOR is empirically related to breaches.

Research Question 4: Does change on the DRAOR predict breaches for women and men? If so, does it do so as accurately with women as men?

The relationship between initial DRAOR scores and DRAOR change scores. The Pearson bivariate correlations presented in Table 17 show the initial DRAOR and DRAOR change scores were all significantly positively correlated, meaning higher initial DRAOR risk scores were associated with larger risk change scores. For the protective subscale, lower initial protective scores were associated with larger protective change scores. Put simply, the initial DRAOR scores determined the amount of change possible between the initial and proximal assessments. For women, the *r* coefficients estimated the strength of the relationship between the given DRAOR variables to be moderate. Likewise, for men, the relationships were estimated to be moderate in strength; however, the relationship between the initial acute and acute change score was estimated to be strong ($r \geq 0.5$).

Table 17

Correlation Matrix of Initial DRAOR Scores and DRAOR Change Scores for the Breach Outcome

	Women				Men			
	Change acute	Change stable	Change protective	Change TS	Change acute	Change stable	Change protective	Change TS
Initial acute	.39**				.56**			
Initial stable		.28**				.32**		
Initial protective			.31**				.36**	
Initial TS				.24**				.34**

** $p < .01$

Multivariate change models. In the subsequent analyses the predictive validity of the DRAOR change scores were examined. A series of multivariate Cox regression analyses were conducted separately for women and men. The DRAOR change scores were entered into all regression models alongside the corresponding initial DRAOR score. It was necessary to control for offenders' initial DRAOR scores for two reasons. Firstly, the score was the baseline assessment of risk that limited the amount of change an offender could make between the two DRAOR assessments. The positive correlations in Table 17 highlight this point. For example, an offender with an initial DRAOR TS of 17 had greater opportunity to show a reduction at the proximal assessment than an offender with an initial DRAOR TS of 3. Secondly, in some univariate models the initial DRAOR scores significantly predicted breaches. In all regression analyses the independent variables were the initial DRAOR score and the corresponding DRAOR change score. The criterion variable was the dichotomous breach reconviction variable and the time variable was survival days.

Recall from research question one that all four initial DRAOR scores predicted breaches for women; however, for men the only statistically significant predictor was the acute subscale. As shown in Table 18 the acute change score made an additional significant contribution to the prediction of breaches for women independent of the initial acute score $\chi^2(2, N = 187) = 21.73, p < .001$. In other words, for women, prediction of breaches was significantly improved by the addition of the acute change score. For men, the same result was shown; the acute change score made an additional significant contribution to the prediction of breaches independent of the initial acute score $\chi^2(2, N = 187) = 8.88, p = .012$.

The stable change score made an additional significant contribution to the prediction of breaches for women independent of the initial stable score $\chi^2(2, N = 187) = 16.62, p < .001$. However, for men, although the stable change score significantly contributed to the model, the final regression model did not predict breaches for men $\chi^2(2, N = 187) = 4.15, p = .126$.

The protective change score made an additional significant contribution to the prediction of breaches for women independent of the initial protective score $\chi^2(2, N = 187) = 16.18, p < .001$. For men, the protective change score significantly contributed to the model and the final regression model successfully predicted breaches $\chi^2(2, N = 187) = 5.99, p = .050$.

For women, the DRAOR total change score made an additional significant contribution to the prediction of breaches independent of the initial TS $\chi^2(2, N = 187) = 25.78, p < .001$. For men, the DRAOR total change score significantly contributed to the model and the final regression model predicted breaches $\chi^2(2, N = 187) = 7.48, p = .024$.

Because the mean change scores of women and men's acute, stable, and DRAOR TS were shown to be *positive* the hazard ratios presented in Table 18 are less than 1.00. Recall, hazard ratios less than 1.00 indicate a one-point increase on the predictor is associated with a percentage decrease in recidivism (Lewis et al., 2012). Thus, I subtracted each samples' hazard ratio for the risk change scores from one to calculate the percentage decrease in the likelihood of breaches per one-point increase on the risk change score (e.g., mean change score of 2 increasing to 3).

Recall that for both samples, the mean protective change scores were *negative*, the hazard ratios for the protective change scores presented in Table 18 are all larger than 1.00. Recall, hazard ratios larger than 1.00 indicate a one-point increase on the predictor is associated with a percentage increase in recidivism (Lewis et al., 2012). Thus the hazard ratios of the protective change scores were interpreted as follows: a one-point increase on the protective change (e.g., mean change of -2 increasing to -1) was associated with a percentage increase in the likelihood of a breach.

As shown in Table 18, for women, the HRs for the risk change scores ranged from 0.89 to 0.77, which translated to a estimated 11% to 23% reduction in the predicted likelihood of a woman being convicted of a breach reoffence for every one-point increase on the respective risk change score (i.e. positive change), taking into account the variance in initial risk scores. For women, the strongest relationship between a risk change score and breaches was for the acute change score; a one-point increase on the acute change score was shown to decrease the likelihood of a breach conviction by 23%.

For men, the HRs for acute and DRAOR total change scores ranged from 0.92 to 0.83, which translated to an estimated 8 to 17% reduction in the predicted likelihood of a man being convicted of a breach for every one-point increase on the respective risk change score, taking into account individual variance in initial risk scores. The strongest relationship between risk change scores and breaches for men was for the acute change score. In Table 18, although the stable change score was a significant variable in the men’s regression model, the overall model did not predict breach reconvictions. It was plausible the non-significant initial stable score introduced unexplained variance to the final model that disrupted the predictive relationship between the stable change score and breach reconvictions.

As shown in Table 18 the HRs for the protective change scores were 1.24 for women and 1.27 for men. The ratios indicated that a one-point increase on the protective change score was associated with an estimated 24% and 27% increase in the likelihood that a woman and man respectively would be convicted of a breach, taking into account the variance in initial protective scores. The HR showed the effect size was more precise for the female sample than the male, because the CIs were more condensed for the women’s HRs than the men’s.

The AUC values associated with the multivariate change models are shown in Table 18. After accounting for unequal base rates all change models had significant AUC values that ranged from 0.65 to 0.73; these values indicated moderate to high discriminative accuracy. For women, the DRAOR total change score had an AUC value of 0.73. The AUC value indicated that a randomly selected female recidivist was 73% more likely than a randomly selected female non-recidivist to have a lower DRAOR total change score, taking into account initial DRAOR TS. The female sample had slightly higher AUC values compared to the men. But because the CIs overlapped considerably, it was plausible the multivariate change model’s true level of accuracy was the same for women and men.

Table 18

Multivariate Model of DRAOR Change Scores Predicting Breaches Controlling for Initial DRAOR Scores

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Initial Acute	0.23 (0.07)	11.60**	1.26 [1.10, 1.43]	0.70*** [0.61, 0.79]	0.27 (0.10)	7.99**	1.31 [1.09, 1.58]	0.68** [0.56, 0.80]
Acute change	-0.26 (0.06)	16.02***	0.77 [0.68, 0.88]		-0.19 (0.09)	4.66*	0.83 [0.69, 0.98]	
Initial stable	0.23 (0.07)	11.73**	1.26 [1.10, 1.45]	0.68*** [0.59, 0.76]	0.06 (0.11)	0.28	1.06 [0.86, 1.31]	0.65* [0.53, 0.77]
Stable change	-0.23 (0.07)	10.10**	0.79 [0.70, 0.92]		-0.19 (0.09)	4.15*	0.83 [0.69, 0.99]	
Initial protective	-0.22 (0.07)	11.15**	0.81 [0.71, 0.92]	0.68*** [0.59, 0.77]	-0.18 (0.11)	2.66	0.84 [0.67, 1.04]	0.68** [0.57, 0.79]
Protective change	0.21 (0.07)	10.27**	1.24 [1.09, 1.41]		0.24 (0.10)	5.33*	1.27 [1.04, 1.55]	
Initial TS	0.11 (0.03)	16.26***	1.12 [1.06, 1.18]	0.73*** [0.65, 0.82]	0.09 (0.04)	4.19*	1.09 [1.00, 1.19]	0.70** [0.59, 0.81]
TS change	-0.12 (0.03)	17.21***	0.89 [0.84, 0.94]		-0.09 (0.04)	6.30*	0.92 [0.86, 0.98]	

* $p < .05$ ** $p < .01$ *** $p < .001$

Summary: Research question 4. For women, all DRAOR change scores significantly predicted breaches when controlling for the initial DRAOR scores. For men, all DRAOR change scores significantly predicted breaches controlling for initial DRAOR scores with the exception of the stable change score. The relationship between amount of change and breaches was stronger for the women as estimated by the hazard ratios. Finding the change scores predicted breaches supports the previous finding of the significant differences in change scores between recidivist and non-recidivist groups.

In research question one the initial DRAOR TS did not predict breach reconvictions for men in the univariate models; however, it was a significant variable in the multivariate change model as evident in Table 18. A possible explanation of this finding was that the initial DRAOR TS mediated the predictive power of the total change score. Such an explanation would not be unexpected given the initial DRAOR TS was integral in the change score calculation and limited the amount of change possible. A further examination of this potential mediation model was beyond the scope of this research. Furthermore, initial

DRAOR scores that were significant in the univariate models remained significant in the multivariate change models.

Research Question 5: Does the RoC*RoI predict breaches? If so, does the proximal DRAOR TS add incremental predictive validity above the RoC*RoI?

Predictive validity of the RoC*RoI. Separate Cox regression analyses were conducted for women and men to examine the predictive validity of the RoC*RoI[15]. The RoC*RoI was entered as an independent predictor variable in block 1 of the Cox regression model, the criterion variable was the dichotomous breach reconviction variable and the time variable was survival days.

Table 19

*Univariate Cox Regression Model of the RoC*RoI Predicting Breaches*

	Women			Men		
	β (SE)	Wald	Hazard Ratio [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]
RoC*RoI	0.98 (0.74)	1.76	2.68 [0.63, 11.45]	0.15 (1.15)	0.00	1.02 [0.11, 9.75]

As shown in Table 19 the RoC*RoI did not predict breaches for women $\chi^2(1, N = 187) = 1.77$, $p = .183$ or men $\chi^2(1, N = 187) = 0.00$, $p = .989$. The CI associated with women and men's HRs both spanned 1.00. For women, the initial and proximal DRAOR TS predicted breaches in the univariate analyses and the proximal DRAOR TS predicted breaches for men. Thus, by default the proximal DRAOR TS was a better predictor of breaches than the RoC*RoI and this was confirmed in a multivariate Cox regression model, see Appendix D for full details.

[15] Consistent with the previous analyses the sample used to answer research question 5 was the subset if each matched sample. As indicated by an independent samples *t*-test these samples did not significantly differ on RoC*RoI.

Results Part II: Criminal Reoffending

Part II two examines the same five research questions as Part I, but using criminal reoffending as the criterion variable. It was useful to examine the DRAOR with this second reoffending type because it enabled a true validation of the DRAOR's predictive validity, because as mentioned in the method the DRAOR is designed to predict criminal reoffending. Part II follows an identical structure to Part I: it opens with descriptive analyses of the base rate and survival analysis of criminal reoffending, followed by subsequent sections detailing the statistical analyses answering the five following research questions.

1. Does the DRAOR predict criminal reoffending for women and men?
    1.1 Does the initial DRAOR predict criminal reoffending for women and men? Are the predictive components and accuracy of the initial DRAOR comparable across gender?
    1.2 Does the proximal DRAOR predict criminal reoffending for women and men? Are the predictive components and accuracy of the proximal DRAOR comparable across gender?
2. Is the proximal DRAOR a better predictor of criminal reoffending than the initial DRAOR?
3. Do DRAOR scores change between the initial and proximal assessment? If so, in what direction?
    3.1 Do female and male recidivists differ in the amount of change made compared to their non-recidivist counterparts?
4. Does change on the DRAOR predict criminal reoffending for women and men? If so, does it do so as accurately with women as men?
5. Does the RoC*RoI predict criminal reoffending? If so, does the proximal DRAOR TS add incremental predictive validity above the RoC*RoI?

Base rates of criminal reoffending. Of the 202 women and 202 men in the matched samples 84 women, and 62 men were convicted of a criminal reoffence. To statistically compare the rate and frequency of criminal reconvictions between women and men controlling for varying follow-up time a Kaplan-Meier survival analysis was produced.

The rate and frequency of criminal reoffending. In *Figure 4* the horizontal axis represents the survival days (see method) and the vertical axis represents the proportion of women or men at a given time point who had not been convicted of a criminal reoffence

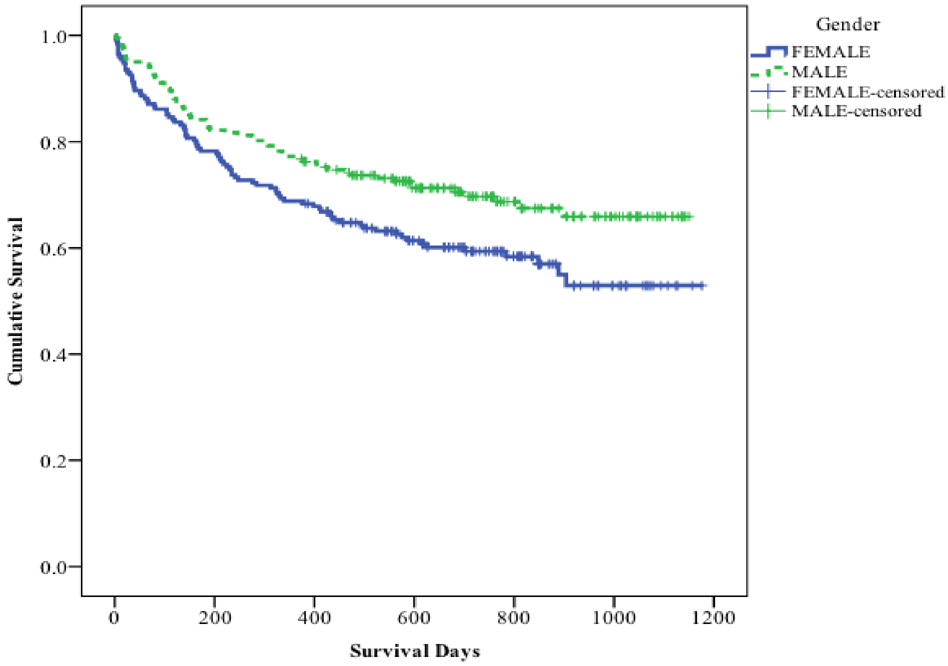

Figure 4 Survival curves of criminal reoffending for women and men

The mean survival time for women was 767 days (*SE* = 34.7), 95% CI [698.5, 834.6] and for men, 860 days (*SE* = 31.1), 95% CI [798.8, 920.9]. The survival analysis revealed a statistically significant difference between women and men in terms of their rate and frequency of criminal reoffences, Tarone-Ware statistic of equality, $\chi^2(1) = 5.081$, $p = .024$.

Inclusion criteria. Recall that matched offenders were included in the samples used to answer the research questions only if they met two criteria

1. If an offender was convicted of a criminal reoffence, the reoffence was committed *after* their initial DRAOR assessment, and
2. Offenders had an initial and proximal DRAOR assessment on *different* days.

Recall that these inclusion criteria also applied to the offenders' matched pairs. As a result, 175 women and 175 men were included in all subsequent analyses in Part II.

Research Question 1.1: Does the initial DRAOR predict criminal reoffending for women and men? Are the predictive components and accuracy of the DRAOR comparable across gender?

The rationale for research question one was identical to that detailed in Part I, namely, was the DRAOR a useful tool for the prediction of criminal reoffending in matched samples of community-sentenced women and men.

The relationship between initial DRAOR scores. The magnitude and the direction of the relationship between the samples' initial DRAOR subscales and total scores were examined using Pearson bivariate correlations. As shown in Table 20 all correlation coefficients were significant and in the expected direction. For both samples the initial acute and stable subscales were positively correlated with each other and the DRAOR TS. As expected the protective subscale was negatively correlated with the acute, stable, and DRAOR TS. The magnitude of the correlations ranged from weak to strong. The relationship between the men's acute and protective subscales was estimated as weak (0.21). However, the relationship between the women's and men's stable subscale and the DRAOR TS was estimated to be strong (0.85). In Table 20 the *r* coefficients ≥ 0.3 and their directions (negative vs. positive) indicated the scores had convergent validity.

Table 20

Correlation Matrix of Initial DRAOR Scores for the Criminal Outcome

	Women				Men			
	Acute	Stable	Protective	TS	Acute	Stable	Protective	TS
Acute	1	-	-	-	1	-	-	-
Stable	.45**	1	-	-	.42**	1	-	-
Protective	-.34**	-.56**	1	-	-.21**	-.60**	1	-
TS	.76**	.85**	-.78**	1	.73**	.85**	-.76**	1

** $p < .01$

Description of the samples' initial DRAOR scores. Table 21 presents the mean initial DRAOR subscale and TS for women and men who were convicted of a criminal reoffence and those who were not. An ANOVA analysis and post-hoc Tukey tests compared the mean initial DRAOR scores between the four groups. The dependent variables were the four initial DRAOR scores and the independent variable was the four-level group variable.

Table 21

Initial DRAOR Scores and ANOVA Analysis for the Criminal Outcome

	Female Recidivist (n = 63)	Female Non Recidivist (n = 112)	Male Recidivist (n = 82)	Male Non Recidivist (n = 132)	Analysis of Variance	Eta Squared
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)		
Acute	5.44[a] (*2.29*)	4.87[a] (*2.36*)	4.63 (*2.68*)	3.83[b] (*2.15*)	$F(3, 346) = 8.73, p < .001$	0.07
Stable	5.71 (*2.35*)	5.10 (*2.25*)	5.37 (*2.36*)	5.17 (*1.91*)	$F(3, 346) = 1.24, p = .296$	0.01
Protective	6.10[c] (*2.11*)	6.96[d] (*2.07*)	6.65 (*2.09*)	7.01[d] (*2.03*)	$F(3, 346) = 3.17, p = .025$	0.03
TS	5.06[e] (*5.19*)	3.01 (*5.39*)	3.35 (*5.28*)	1.99[f] (*4.83*)	$F(3, 346) = 5.15, p = .002$	0.04

Note. Means with differing superscript were significantly different in post-hoc Tukey comparisons ($p < .05$).

The results of the ANOVA analysis showed the four groups had significantly different initial acute, protective subscales and DRAOR TS. The female recidivist and female non-recidivists both had significantly higher initial acute scores compared to the male non-recidivists; however, the mean acute scores of the female groups were not significantly different. The female recidivists had significantly lower mean protective scores compared to the female and male non-recidivists; however, there was no significant difference in the non-recidivists’ mean protective scores. The female recidivists had significantly higher DRAOR TS compared to the male non-recidivists. The eta-squared estimated the magnitude of the differences presented in Table 21 to be small; however, the differences on the initial acute scores were moderate.

Summary. The female recidivists had the highest initial DRAOR risk scores and the lowest initial DRAOR protective scores. Because the survival analysis showed the female sample had been reconvicted of significantly more criminal reoffences than the male sample, this was anticipated. The ANOVA showed the initial protective subscale was sensitive to differentiation between women who had been reconvicted of criminal offences from those who were not; however, no other initial score for either sample showed the same sensitivity.

Univariate Cox regression models. Identical to Part I a series of univariate Cox regressions were produced to investigate whether the initial DRAOR scores were predictive of criminal reoffending as indicated by the ANOVA. The initial DRAOR subscale and TS were entered individually as the independent variables, the criterion variable was the criminal reconviction dichotomous variable and the time variable was survival days. As in Part I, AUC values were produced to facilitate the direct comparison of the models across women and men, accounting for the significant differences in the base rates of criminal reoffences.

Table 22

Univariate Models of the Initial DRAOR Scores Predicting Criminal Reoffending

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.07 (0.05)	1.70	1.07 [0.97, 1.19]	0.57 [0.48, 0.66]	0.14 (0.06)	4.87*	1.15 [1.02, 1.29]	0.59 [0.48, 0.69]
Stable	0.10 (0.06)	3.18	1.10 [0.99, 1.23]	0.58 [0.49, 0.67]	0.06 (0.08)	0.63	1.06 [0.91, 1.24]	0.52 [0.41, 0.62]
Protective	-0.20 (0.07)	8.35**	0.82 [0.72, 0.94]	0.60* [0.52, 0.69]	-0.08 (0.08)	1.22	0.92 [0.79, 1.07]	0.53 [0.43, 0.63]
TS	0.06 (0.02)	5.84*	1.06 [1.01, 1.11]	0.60* [0.52, 0.69]	0.06 (0.03)	3.24	1.06 [1.00, 1.12]	0.58 [0.48, 0.68]

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

As shown in Table 22 the acute subscale score did not predict criminal reoffending for women $\chi^2(1, N = 175) = 1.70$, $p = .193$; however, the acute score predicted criminal reoffending for men $\chi^2(1, N = 175) = 4.89$, $p = .027$. The stable score did not predict criminal reoffending for women $\chi^2(1, N = 175) = 3.19$, $p = .074$ or men $\chi^2(1, N = 175) = 0.63$, $p = .427$. The protective score predicted criminal reoffending for women $\chi^2(1, N = 175) = 8.21$, $p = .004$, but not for men $\chi^2(1, N = 175) = 1.22$, $p = .269$. The TS predicted criminal reoffending for women $\chi^2(1, N = 175) = 5.79$, $p = .016$ but not for men $\chi^2(1, N = 175) = 3.25$, $p = .071$.

The hazard ratios (HRs) for women's initial protective and DRAOR TS did not span 1.00, which indicated each predictor had a statistically significant effect on the likelihood of criminal reoffending. For women, the strongest relationship between initial DRAOR scores and criminal reoffending was for the protective subscale. The HR of 0.82 estimated that for every one-point increase on the initial protective subscale the likelihood of a woman being convicted of a criminal reoffence decreased by 18%. Although the initial acute and stable scores were non-significant predictors, given more statistical power they may have had a significant impact on the likelihood of criminal reoffending, because the lower band of confidence interval (CI) associated with the HR were only marginally below the 1.00.

For men, the only statistically significant predictor of criminal reoffending was the acute subscale. The HR for the acute subscale was 1.15, indicating that for every one-point increase on the initial acute subscale the likelihood of a man being convicted of a criminal reoffence increased by 15%.

As shown in Table 22, the AUC values for women's initial protective and DRAOR TS were significant at 0.60, which indicated these two univariate predictors had moderate discriminative accuracy. For example, a randomly selected female recidivist was 60% more likely than a randomly selected non-recidivist to have a higher initial DRAOR TS. However, for men, although the initial acute score was shown to have a statistically significant effect on criminal reoffending as estimated by the HR, after taking into account the lower number of recidivist events, the AUC value showed the predictor was not reliable when distinguishing recidivists from non-recidivists.

As evident in Table 22 the HR for women's and men's initial DRAOR TS were identical (1.06). The CIs associated with each HR only differed by 0.01 at the lower and upper bands. However, the DRAOR TS was predictive of criminal reoffending for women but not men. It was plausible this finding was the result of a higher base rate of criminal reoffending in the female sample. The AUC values support this theory as the CI associated with each AUC value substantially overlapped and thus after taking into account base rates it was possible the true accuracy of the initial DRAOR TS was the same for women and men.

Multivariate Cox regression models. Next a series of multivariate models of the three initial DRAOR subscales were produced to examine the differential predictive power of the subscales for women and men[16]. The multivariate model for women predicted criminal reoffending $\chi^2(3, N = 175) = 8.28, p = .041$. As shown in Table 23 the protective subscale made a statistically significant unique contribution to the model. The protective subscale was the only significant predictor in the univariate models; thus it was not unexpected to find the protective subscale uniquely contributed to the multivariate model. The finding indicated that although women's initial DRAOR subscale scores were related to each other, the protective and stable subscales particularly so, the protective subscale was the strongest predictor of criminal reoffending. This is because despite the shared variance with the acute and stable subscales it emerged as a statically unique predictor, indicating it likely drove the predictor power of the initial DRAOR TS with women.

For men, the multivariate model did not predict criminal reoffending $\chi^2(3, N = 175) = 5.61, p = .132$. However, as shown in Table 23 the initial acute subscale made a statistically significant unique contribution to the model. It was plausible the addition of the stable and protective subscales introduced unexplained variance into the final model that disrupted the relationship between the initial acute subscale and criminal reoffending.

Table 23

Multivariate Models of the Initial DRAOR Subscales Predicting Criminal Reoffending

	Women				Men			
	β (SE)	Wald	Hazard Ratios [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazards Ratios [95% CI]	AUC [95% CI]
Acute	0.01 (0.06)	0.01	1.01 [0.89, 1.14]		0.14 (0.07)	4.14*	1.15 [1.01, 1.31]	
Stable	0.02 (0.07)	0.06	1.02 [0.89, 1.17]	0.60* [0.52, 0.69]	-0.06 (0.11)	0.28	0.95 [0.77, 1.16]	0.62* [0.52, 0.72]
Protective	-0.18 (0.08)	5.54*	0.83 [0.71, 0.97]		-0.08 (0.10)	0.65	0.93 [0.77, 1.12]	

* $p < .05$

[16] For women and men, the only correlation high enough to introduce multicollinearity was between the initial stable and protective subscales (Table 20). Recall from Part I that all VIF values for the multivariate models in this research were ≤ 2.00. Thus, multicollinearity did not impact the interpretation of any of the multivariate models in Part II.

The AUC value for the female sample's multivariate model indicated moderate discriminative accuracy as estimated by the significant value of 0.60. For men, although the overall model was non-significant, when taking into account the low base rate of criminal reoffending the multivariate model had moderate discriminative accuracy, likely driven by the initial acute subscale.

Research question 1.2: Does the proximal DRAOR predict criminal reoffending for women and men? Are the predictive components and accuracy of the proximal DRAOR comparable across gender?

The relationship between the proximal DRAOR scores. Pearson bivariate correlations are presented in Table 24. All correlation coefficients were significant and in the expected direction for both samples. The proximal risk scores were positively correlated with one another and the DRAOR TS. As expected the proximal protective subscale scores were negatively correlated with the risk subscales and the TS. The strength of the relationship between proximal DRAOR scores ranged from moderate (0.42) to strong (0.87). As expected the *r* coefficients indicted convergent validity.

Table 24

Correlation Matrix of Proximal DRAOR Scores for the Criminal Outcome

	Women				Men			
	Acute	Stable	Protective	TS	Acute	Stable	Protective	TS
Acute	1	-	-	-	1	-	-	-
Stable	.47**	1	-	-	.45**	1	-	-
Protective	-.36**	-.62**	1	-	-.42**	-.60**	1	-
TS	.74**	.87**	-.82**	1	.74**	.86**	-.84**	1

** $p < .01$

Description of the samples' proximal DRAOR scores. Table 25 presents the mean proximal DRAOR scores for the four recidivist groups. As previously an ANOVA was conducted to examine whether the four groups significantly differed on their proximal DRAOR scores.

Table 25

Proximal DRAOR Scores and ANOVA Analysis for the Criminal Outcome

	Female Recidivist (n = 63)	Female Non Recidivist (n = 112)	Male Recidivist (n = 43)	Male Non Recidivist (n = 132)	Analysis of Variance	Eta Squared
	M (SD)	M (SD)	M (SD)	M (SD)		
Acute	4.48^{a} (*2.20*)	3.84^{a} (*2.24*)	3.79^{a} (*2.34*)	2.80^{b} (*1.73*)	$F(3, 346) = 11.98, p < .001$	0.09
Stable	5.38^{c} (*2.20*)	4.29^{d} (*2.60*)	4.67^{dc} (*2.83*)	3.73^{d} (*2.23*)	$F(3, 346) = 8.12, p < .001$	0.06
Protective	6.56^{e} (*2.18*)	7.67^{f} (*2.40*)	7.42^{ef} (*2.43*)	8.23^{f} (*2.19*)	$F(3, 346) = 7.84, p < .001$	0.06
TS	3.30^{g} (*4.87*)	0.47^{h} (*6.01*)	1.05^{gh} (*6.14*)	-1.70^{i} (*4.99*)	$F(3, 346) = 15.11, p < .001$	0.10

Note. Means with differing superscript were significantly different in post-hoc Tukey comparisons ($p < .05$).

The ANOVA analysis revealed the offender groups significantly differed on all four proximal DRAOR scores. Post-hoc Tukey tests revealed female recidivists, female non-recidivists, and male recidivists all had significantly higher proximal DRAOR scores compared to the male non-recidivists; however, there were no statistical differences between the aforementioned three groups. The female recidivists had significantly higher proximal stable scores compared to female non-recidivists and the male non-recidivists; however, there was no significant difference between any other groups' proximal stable scores. The female recidivists had significantly lower proximal protective scores compared to the female non-recidivists and also the male non-recidivists; there were no other significant differences between the groups' proximal protective scores. The female recidivists had a significantly higher proximal DRAOR TS compared to the female non-recidivists and the male non-recidivists. The male recidivists also had a significantly higher proximal DRAOR TS compared to the male non-recidivists; however, the female and male recidivists did not significantly differ on their proximal DRAOR TS. As shown in Table 25 the eta-squared estimated the magnitude of the differences between the four groups' proximal DRAOR scores to be moderate, while the magnitude of the differences between the groups' proximal DRAOR TS to be large.

Summary. As expected, the proximal DRAOR risk scores were the highest and the protective scores the lowest for the female recidivists; this pattern was consistent with the initial DRAOR scores. For women, the proximal stable, protective and DRAOR TS were significantly different between the female recidivist and female non-recidivists. For men, the proximal acute and DRAOR TS were significantly different between the recidivist and non-recidivist groups. These significant differences within each sample indicate the proximal

DRAOR scores were not only sensitive to observed rates of criminal reoffending between the two samples, but also within each sample. The magnitude of differences between the proximal DRAOR scores was larger than those between the initial DRAOR scores as estimated by the eta-squared effect size.

Univariate models of the proximal DRAOR scores. As in Part I a summary of the proximal DRAOR's predictive validity and discriminative accuracy with women and men are presented below.

As shown in Table 26 the stable, protective and DRAOR TS all predicted criminal reoffending for women in the univariate analyses. The strength of the relationship between predictors and criminal reoffending was strongest for the proximal stable and protective subscales. A one-point increase on either subscale was estimated to have a 16% effect on the likelihood of a woman being convicted of a criminal reoffence. Although for women, the acute subscale did not significantly predict criminal reoffending, the lower band of the hazard ratio CI was 0.99. Given greater statistical power it was plausible the subscale may have reached statistical significance.

Table 26

Univariate Models of the Proximal DRAOR Scores Predicting Criminal Reoffending

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.10[a] (0.06)	3.10	1.10 [0.99, 1.23]	0.58 [0.50, 0.67]	0.23[e] (0.07)	9.95**	1.26 [1.09, 1.45]	0.62* [0.52, 0.72]
Stable	0.15[b] (0.05)	8.75**	1.16 [1.05, 1.28]	0.63** [0.55, 0.71]	0.16[f] (0.06)	6.54*	1.17 [1.04, 1.33]	0.59 [0.48, 0.69]
Protective	-0.18[c] (0.06)	10.46**	0.84 [0.75, 0.93]	0.64** [0.55, 0.72]	-0.15[g] (0.06)	5.17*	0.86 [0.76, 0.98]	0.59 [0.49, 0.69]
TS	0.07[d] (0.02)	10.61**	1.08 [1.03, 1.13]	0.64** [0.56, 0.73]	0.09[h] (0.03)	10.48**	1.10 [1.04, 1.16]	0.63** [0.53, 0.73]

[a] Model $\chi^2(1, N = 175) = 3.11$, $p = .078$
[b] Model $\chi^2(1, N = 175) = 8.87$, $p = .003$
[c] Model $\chi^2(1, N = 175) = 10.48$, $p = .001$
[d] Model $\chi^2(1, N = 175) = 10.59$, $p = .001$
[e] Model $\chi^2(1, N = 175) = 10.21$, $p = .001$
[f] Model $\chi^2(1, N = 175) = 6.67$, $p = .010$
[g] Model $\chi^2(1, N = 175) = 5.17$, $p = .023$
[h] Model $\chi^2(1, N = 175) = 10.79$, $p = .001$

* $p < .05$ ** $p < .01$ *** $p < .001$

For men, all proximal DRAOR scores significantly predicted criminal reoffending in the univariate analyses. The predictor with the strongest relationship to criminal reoffending as estimated by the HR was the proximal acute subscale. A one-point increase on the subscale was estimated to have a 26% increase in the likelihood that a man would be convicted of a criminal reoffence.

For women, all statistically significant predictors were supported by significant AUC values that estimated moderate discriminative accuracy. For men, when taking into account the low base rate of recidivist events, the AUC values indicated reliable discriminative accuracy for the proximal acute and DRAOR TS only. Comparison of the univariate predictors directly across the two samples revealed that the CIs associated with the AUC values substantially overlapped. Thus I concluded that, when the unequal base rate of criminal reoffending was taken into account, the accuracy of the proximal DRAOR subscales and TS were likely comparable for women and men. This pattern has been consistently observed throughout the results thus far.

Multivariate models of proximal DRAOR scores. As shown in Table 27 the multivariate model including the three proximal subscales predicted criminal reoffending for women $\chi^2(3, N = 175) = 12.05, p = .007$.

Table 27

Multivariate Model of the Proximal DRAOR Subscales Predicting Criminal Reoffending

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Acute	0.01 (0.07)	0.02	1.01 [0.89, 1.15]		0.18 (0.08)	4.53*	1.19 [1.01, 1.40]	
Stable	0.08 (0.07)	1.46	1.09 [0.95, 1.24]	0.64** [0.55, 0.72]	0.09 (0.08)	1.22	1.09 [0.94, 1.27]	0.64** [0.54, 0.74]
Protective	-0.13 (0.07)	3.65	0.88 [0.77, 1.00]		-0.04 (0.09)	0.19	0.96 [0.82, 1.14]	

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

All proximal subscales in combination contributed to the predictive power of the model and this was expected given the univariate results. Although no one subscale significantly contributed to the model, the upper band of the CI for the protective subscale HR was 1.00. Thus it was plausible given greater statistical power, the proximal protective subscale would have contributed unique statistical power to the model.

For men, the multivariate model predicted criminal reoffending $\chi^2(3, N = 175) = 12.39, p = .006$. The proximal acute subscale contributed uniquely to the model above that of the stable and protective subscales. This finding was expected given the proximal acute subscale had the strongest relationship with criminal reoffending in the univariate analyses. It suggests the acute subscale drove the predictive power of the proximal DRAOR TS for men. Both multivariate models were supported by significant AUC values, which indicated that after taking into account the unequal base rate of criminal reoffending for women and men, the multivariate models had comparable moderate discriminative accuracy.

Summary: Research question 1. For both samples, the predictive validity of the DRAOR for criminal reoffending was mixed. The initial protective subscale emerged as the uniquely robust predictor for women, while for men; the only significant initial score was the acute subscale. However, for both samples, the proximal DRAOR scores were more robust predictors of criminal reoffending. For men, the acute subscale emerged as a uniquely predictive subscale. For women, all subscales had an equivalent relationship with criminal reoffending. The AUC values showed the DRAOR was as effective (or ineffective) in the prediction of criminal reoffending for women as for men.

As in Part I, research question one suggested the proximal DRAOR was a better predictor of criminal reoffending than the initial, because more proximal DRAOR scores significantly predicted criminal reoffending. Additionally, of the DRAOR scores that predicted at both the initial and proximal assessment, the hazard ratios and AUC values for the proximal scores were comparatively larger. However, the confidence intervals associated with the effect sizes overlapped, no definitive conclusions could be drawn regarding the predictive superiority of proximal DRAOR scores. Research question two directly examines the suggested predictive superiority of the proximal DRAOR score by conducting a series of incremental predictive validity analyses.

Research question 2: Is the proximal DRAOR a better predictor of criminal reoffending than the initial DRAOR?

Based on the significant results of the univariate analyses in research question one, the incremental predictive validity of the proximal DRAOR was examined. Recall that the women's initial acute and stable scores and their proximal acute did not predict criminal reoffending in the univariate analyses. For men, the initial stable, protective, and DRAOR TS did not predict criminal reoffending in the univariate analyses, and thus were not included in the incremental analyses (see Appendix C).

As evident in Table 28, the proximal protective score contributed unique predictive power to the model when entered with the initial protective score for women $\chi^2(2, N = 175) = 12.33, p = .002$. The addition of the proximal protective subscale rendered the initial protective score a non-significant predictor in the model. For women, this same result was shown for the addition of the proximal TS to the initial TS $\chi^2(2, N = 175) = 10.96$, $p = .004$.

For men, the proximal acute score contributed unique predictive power to the model when entered into the model with the initial acute score $\chi^2(2, N = 175) = 10.40$, $p = .006$. The proximal acute score rendered the initial acute score a non-significant predictor in the model.

Table 28

Incremental Predictive Validity of the Proximal DRAOR Predicting Criminal Reoffending

	Women			Men		
	β (SE)	Wald	Hazard Ratio [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]
Initial acute +				0.04 (0.07)	0.34	1.04 [0.91, 1.20]
Proximal acute				0.20 (0.09)	5.62*	1.23 [1.04, 1.45]
Initial protective +	-0.12 (0.08)	2.28	0.89 [0.76, 1.04]			
Proximal protective	-0.13 (0.06)	4.15*	0.88 [0.78, 1.00]			
Initial TS +	0.02 (0.03)	0.62	1.02 [0.97, 1.08]			
Proximal TS	0.06 (0.03)	5.69*	1.07 [1.01, 1.12]			

* $p < .05$

As shown in Table 28 the HRs indicated that when the proximal DRAOR score was considered the impact of a one-point increase on the initial DRAOR score had no effect on the likelihood of criminal reoffending. For women, the HR for the proximal DRAOR TS showed that for every one-point increase on the proximal DRAOR TS the likelihood of a woman being convicted of a criminal reoffence increased by seven percent, accounting for individual variation on the initial DRAOR TS.

Summary: Research question 2. For both samples, the proximal DRAOR scores were the best predictors of criminal reoffending, with the exception of the acute subscale for women. Research questions three and four were conducted to further investigate why the proximal DRAOR scores outperformed the initial scores in the prediction of criminal reoffending. Recall from Part I the third research questioned examined whether in fact the most up-to-date dynamic assessment (i.e. proximal) was different from the initial (i.e. third) assessment of dynamic risk.

Research question 3: Do DRAOR scores change between the initial and proximal assessment? If so, in what direction?

Recall that offenders' initial DRAOR assessment was the third assessment following their commencement of community supervision, and the proximal assessment occurred after the initial assessment (either the assessment most prior to a reoffence, or if no reoffence, the last assessment available on the offenders' sentence).

The initial and proximal DRAOR scores for women and men were compared using a paired samples *t*-test. As shown in Table 29 there were significant differences between initial and proximal DRAOR scores on all subscales and the TS for both samples. The two samples had significantly lower mean risk scores and significantly higher protective scores at Time 2 compared to Time 1. Recall that each offender had a DRAOR change score calculated to facilitate the examination of the DRAOR's sensitivity to change in risk overtime. Recall from Part I, the change scores were calculated for each offender by subtracting the proximal DRAOR score from the corresponding initial DRAOR score. The change scores for the two samples on all DRAOR scores are also presented in Table 29.

Table 29

Paired-Samples t-test of the Initial and Proximal DRAOR Scores for the Criminal Outcome

Sample	Initial Assessment *M* (*SD*)	Proximal Assessment *M* (*SD*)	Paired Samples *t*-test	Mean change (*SD*)	Range	Cohen's *d*
Women						
Acute	5.07 (*2.34*)	4.07 (*2.24*)	$t(174) = 5.59, p < .001$	1.01 (*2.38*)	-6, +10	0.42
Stable	5.32 (*2.30*)	4.69 (*2.51*)	$t(174) = 4.32, p < .001$	0.63 (*1.94*)	-4, +7	0.33
Protective	6.65 (*2.12*)	7.27 (*2.38*)	$t(174) = -3.86, p < .001$	-0.62 (*2.14*)	-8, +9	0.29
TS	3.75 (*5.39*)	1.49 (*5.78*)	$t(174) = 5.95, p < .001$	2.26 (*5.03*)	-16, +20	0.45
Men						
Acute	4.02 (*2.31*)	3.05 (*1.94*)	$t(174) = 6.99, p < .001$	0.98 (*1.85*)	-8, +7	0.53
Stable	5.22 (*2.03*)	3.96 (*2.42*)	$t(174) = 8.66, p < .001$	1.26 (*1.93*)	-4, +8	0.65
Protective	6.92 (*2.05*)	8.03 (*2.27*)	$t(174) = -8.10, p < .001$	-1.11 (*1.80*)	-7, +6	0.62
TS	2.33 (*4.96*)	-1.03 (*5.41*)	$t(174) = 10.89, p < .001$	3.35 (*4.07*)	-7, +17	0.82

Remember from Part I the magnitude of the change scores was dependent upon the size of the predictor. To account for the differences in the scope of change possible Cohen's *d* was computed and used to interpret the magnitude of each change score as well as compare the change across women and men. Recall a Cohen’s *d* of 0.2 is considered a small effect, 0.5 a medium, and 0.8 a large effect.

As shown in Table 29 the women and men had positive risk change scores, which indicated the offenders had significantly lower risk scores at the proximal assessment compared to the initial. In other words, their risk decreased over time. For the protective subscale both women’s and men’s change scores were negative, which indicated the offenders had significantly higher protective scores at the proximal assessment compared to the initial. In other words, the protective scores increased over time.

The effect size of change scores for women ranged from small (0.29) to medium (0.45); women made the greatest change as estimated by Cohen’s *d* on the DRAOR TS. As evident in Table 29, compared to women the Cohen’s *d* value of men’s change scores were considerably larger, indicating men made greater change over the period compared to women. For the men, the change scores ranged in magnitude from medium (0.53) to large (0.82). The men made the greatest change as estimated by Cohen’s *d* on the DRAOR TS.

As evident from Table 29 each mean change score had a large standard deviation and range associated with it. These statistics indicated there was considerable variance associated with the amount of change and the direction of change the two samples made. For example, although women had a mean change score for the stable subscale of + 0.63, which indicated a mean decrease on the stable subscale across time, the range showed the minimum as negative four and the maximum as positive seven. The minimum change score indicated that at least one woman had an increase of four points on the acute subscale between the two assessment points. At the other extreme at least one woman had a change score of positive seven, which indicated that she had a decrease of seven points on the stable subscale between the initial and proximal assessments. For men, at least one man increased in risk between the two DRAOR assessments as indicated by the range values.

Research question 3.1: Do female and male recidivists differ in the amount of change made compared to their non-recidivist counterparts? To explore the hypothesis that the large variation in change scores seen for the whole samples above was due to differences between women and men who were convicted of a criminal reoffence and those who were not, analysis of change made by the four offenders groups was conducted.

As illustrated in *Figure 5* the female recidivist group had the highest initial and proximal DRAOR risk scores and lowest DRAOR protective scores of all four groups. The male non-recidivists had the lowest initial and proximal acute scores; however, the female non-recidivists and the male recidivists had similar initial and proximal acute scores. The male-recidivists, female non-recidivists, and male non-recidivists as shown in *Figure 5* appeared to have approximately equivalent initial stable scores; however, the groups scores differed at the proximal assessment as the male recidivists stable scores appeared to be slightly higher than the two non-recidivist groups. The protective scores at the initial assessment, as evident in the figure, were approximately equal for the two non-recidivist groups; however, the male non-recidivists had the higher protective score of the two at the proximal assessment. The female non-recidivists’ protective score at the proximal assessment was estimated to be equal to the male recidivists score at the same assessment point. For the DRAOR TS, the male recidivists, as illustrated in the figure, had slightly higher initial and proximal DRAOR TS compared to the two non-recidivist groups. The slope of each line shown in *Figure 5* illustrates the magnitude and direction of change made by the four offender groups between the two DRAOR assessments. The slope of the lines for the stable, protective, and particularly the DRAOR TS were steeper for non-recidivists compared to recidivists. The steeper slopes were indicative of greater change scores.

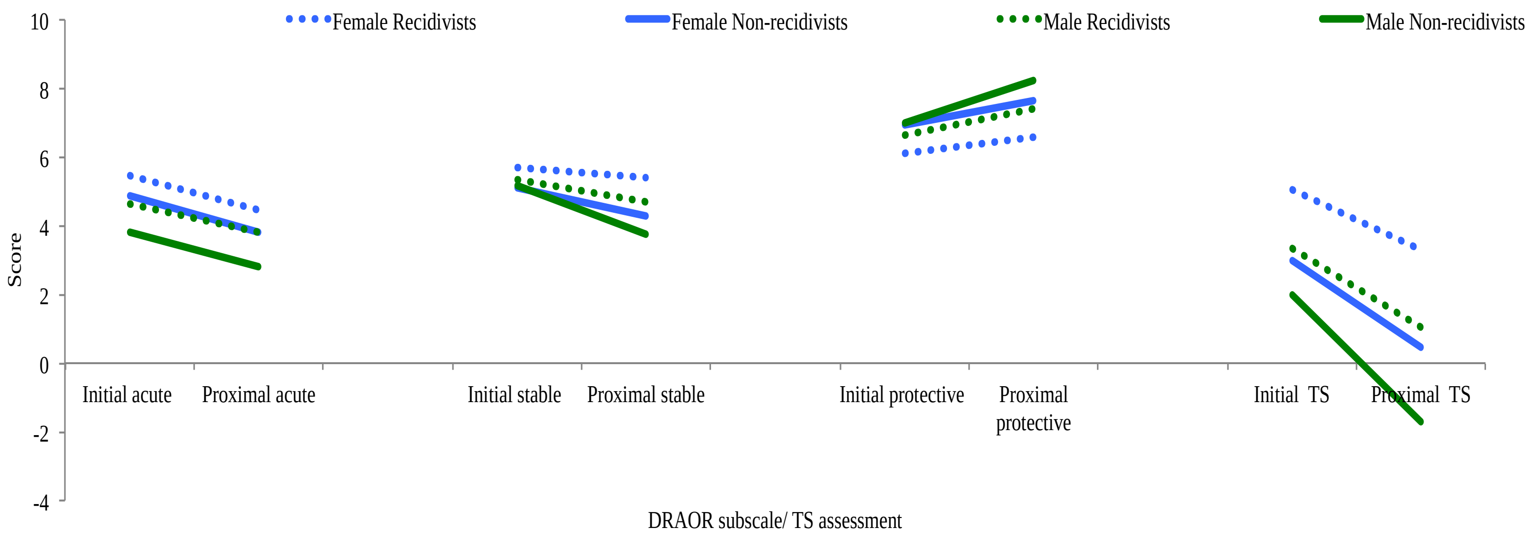

Figure 5 Initial and proximal DRAOR scores for women and men who were convicted of a criminal reoffence and those who were not

Mean change scores for the four offender groups. In the following section *Figure 6* and Table 30 are interpreted in relation to each other. *Figure 6* depicts the mean DRAOR change scores for the four offender groups. Recall, a positive change score for the risk subscales indicated a lower risk score at the proximal assessment, and a negative protective change score indicated higher protective score at the proximal assessment. As in Part I the error bars in *Figure 6* represent the 95% CI associated with each mean change score. In the figure, when the error bars do not overlap a significant difference between the respective groups' change scores was expected; however, if there is overlap further analysis is required, such as an ANOVA.

Table 30 presents the results of an ANOVA and post-hoc Tukey tests of the change scores of the four groups presented in *Figure 6.* As in Part I the dependent variables were the four change scores and the independent variable was the four-level group variable. The eta-squared effect size statistic was computed to enable the magnitude of the difference in means between groups to be qualitatively described and compared.

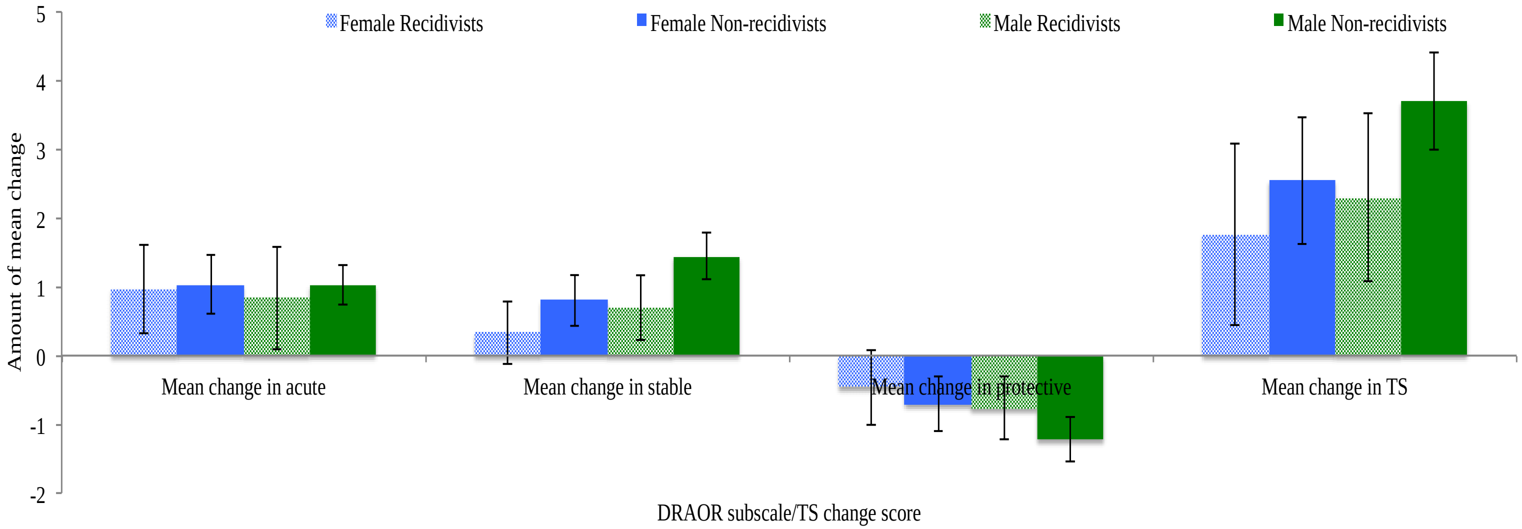

Figure 6 Mean change scores for those who were convicted of a criminal reoffence and those who were not

Table 30

ANOVA of Mean DRAOR Change Scores for Women and Men who were Convicted of a Criminal Reoffence and those who were not

	Female Recidivist (*n* = 63)	Female Non Recidivist (*n* = 112)	Male Recidivist (*n* = 43)	Male Non Recidivist (*n* = 132)	Analysis of Variance	Eta-squared
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)		
Acute	0.97 (*2.55*)	1.03 (*2.29*)	0.84 (*2.42*)	1.02 (*1.63*)	$F(3, 346) = 0.08$, $p = .970$	0.00
Stable	0.33[a] (*1.70*)	0.80[a] (*2.00*)	0.70 (*1.54*)	1.45[b] (*2.01*)	$F(3, 346) = 5.60$, $p = .001$	0.05
Protective	-0.46 (*2.17*)	-0.71 (*2.12*)	-0.77 (*1.49*)	-1.23 (*1.88*)	$F(3, 346) = 2.62$, $p = .051$	0.02
TS	1.76[c] (*5.23*)	2.54 (*4.92*)	2.30 (*3.96*)	3.70[d] (*4.07*)	$F(3, 346) = 3.08$, $p = .028$	0.03

Note: Means with differing superscripts are significantly different ($p < .05$)

As shown in Table 30, the ANOVA revealed significant differences between the four groups stable and DRAOR total change scores. *Figure 6* showed the female recidivists had the lowest change scores followed by the male recidivists. As shown in *Figure 6* all four groups had comparable positive mean acute change scores and as expected by the overlap of the error bars there were no significant difference between the groups' acute change scores. As shown in Table 30, the female recidivists and female non-recidivists had significantly lower mean stable change scores compared to the male non-recidivists; however, there was no significant difference between the female offender groups or the male offender groups, the significant differences were expected based on the error bars of *Figure 6*. As shown in *Figure 6*, all four groups had comparable negative mean protective change scores and each groups' error bar either overlapped fully or partially with the others, thus it was not unexpected to find the scores were not significantly different in post-hoc comparisons.

As shown in *Figure 6*, the two non-recidivist groups had the highest mean DRAOR total change scores and the female recidivist group had the lowest DRAOR total change score. As shown in Table 30 the female recidivists' DRAOR total change score was significantly lower compared to the male non-recidivist group, this difference was anticipated based on the error bars in *Figure 6*.

The above statistical differences were significant as estimated at the $p < .05$ level. The eta-squared effect size was computed to indicate the magnitude of the mean differences. As estimated by the eta-squared the mean differences between the groups' stable and DRAOR TS change score were considered small effects. However, the eta-squared of 0.5 for the stable change score was just below the cut off for a medium effect size.

Summary: Research question 3. For both samples, DRAOR scores changed over time; both women and men had positive risk change scores and negative protective change scores. A more refined analysis at the level of recidivist versus non-recidivist showed all offenders, irrespective of whether they had been convicted of a criminal reoffence, had lower DRAOR risk and higher protective scores at the proximal assessment. For the stable subscale, if you were a woman you were likely to make significantly less change than the male non-recidivists regardless of whether you were convicted of a criminal reoffence. The overall trend showed recidivists made less change on the DRAOR subscales compared to non-recidivists and the pattern was consistent across gender.

Research Question 4: Does change on the DRAOR predict criminal reoffending for women and men? If so, does it do so as accurately with women as men?

The relationship between initial DRAOR scores and DRAOR change scores. Pearson bivariate correlations are presented in Table 31.

Table 31

Correlation Matrix of Initial DRAOR Scores and DRAOR Change Scores for the Criminal Outcome

	Women				Men			
	Change acute	Change stable	Change protective	Change TS	Change acute	Change stable	Change protective	Change TS
Initial acute	0.56**				.59**			
Initial stable		.31**				.25**		
Initial protective			.38**				.31**	
Initial TS				.34**				.30**

** $p < .01$

As shown in Table 31 the initial DRAOR and DRAOR change scores were all significantly positively correlated. Meaning higher initial DRAOR risk scores were associated with larger risk change scores. For the protective subscale, this meant lower initial protective scores were associated with larger protective change scores. As in Part I, these correlations show the initial DRAOR scores determined the amount of change possible on the DRAOR. For women, the r coefficients estimated that strength of the relationship between the given DRAOR variables was moderate ($r \geq 0.3$). Likewise, for men, the relationships were estimated to be moderate in strength. The relationship between the women and men’s initial and change in acute scores were strong as estimated by the r coefficient.

Multivariate change models. In the subsequent analyses the predictive validity of DRAOR change scores for criminal reoffending were examined. As in Part I a series of multivariate Cox regression analyses were conducted separately for women and men. The initial DRAOR scores were entered into all models alongside the corresponding DRAOR change scores in order to control for offenders' initial risk scores. Recall it was necessary to control for offenders' initial DRAOR scores for two reasons. Firstly, as indicated by the positive correlations in Table 31 higher initial DRAOR scores meant higher change scores, in other words, the initial score limited the amount of change an offender could make between the two assessment points. Secondly, in research question one the univariate models showed that some initial DRAOR scores significantly predicted criminal reoffending. For example, recall the initial protective and DRAOR TS for women and the initial acute score for men predicted criminal reoffending.

In all multivariate change models the independent variables were the initial DRAOR score and the corresponding DRAOR change score. The criterion variable was the dichotomous criminal reconviction variable and the time variable was survival days.

The addition of the acute change score did not enable the final regression model to predict criminal reoffending for women $\chi^2(2, N = 175) = 3.45, p = .179$. For men, the acute change score made an additional significant contribution to the prediction of criminal reoffending independent of the initial acute score, which also remained significant $\chi^2(2, N = 175) = 10.40, p = .006$.

The initial stable score did not predict criminal reoffending in the univariate analyses for women. However, the stable change score significantly contributed to the regression model, which as a result predicted criminal reoffending for women $\chi^2(2, N = 175) = 8.89, p = .012$. Likewise, for the men, the initial stable score did not predict criminal reoffending in the univariate analyses; however, the stable change score significantly contributed to the model, which as a result predicted criminal reconvictions $\chi^2(2, N = 175) = 7.47, p = .024$.

For women, the protective change score provided an additional significant contribution to the prediction of criminal reoffending independent of the initial protective score $\chi^2(2, N = 175) = 12.33, p = .002$. For men, the initial protective subscale score did not predict criminal reoffending on its own. Despite the protective change score significantly contributing to the model, the overall model did not predict criminal reoffending $\chi^2(2, N = 175) = 5.37, p = .068$.

The initial DRAOR TS predicted criminal reoffending for women in the univariate analyses. For women, the DRAOR total change score made an additional significant contribution to the prediction of criminal reoffending independent of the initial TS $\chi^2(2, N =$

175) = 10.96, p = .004. For men, the initial TS did not predict criminal reoffending on its own. When the DRAOR total change score was added to the model it made an additional significant contribution to the model that enabled the final model to predict criminal reoffending for men $\chi^2(2, N = 175) = 10.99$, p = .004.

Table 32

Multivariate Model of DRAOR Change Scores Predicting Criminal Reoffending Controlling for Initial DRAOR Scores

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Initial acute	0.12 (0.07)	3.28	1.13 [0.99, 1.28]	0.60* [0.51, 0.68]	0.24 (0.08)	9.74**	1.28 [1.10 1.49]	0.62* [0.52, 0.72]
Acute change	-0.08 (0.06)	1.76	0.92 [0.82, 1.04]		-0.23 (0.09)	5.62**	0.81 [0.69, 0.97]	
Initial stable	0.14 (0.06)	6.01*	1.15 [1.03, 1.29]	0.63** [0.55, 0.71]	0.11 (0.08)	2.05	1.12 [0.96, 1.30]	0.62* [0.52, 0.72]
Stable change	-0.16 (0.07)	5.47*	0.85 [0.74, 0.97]		-0.22 (0.09)	6.39**	0.80 [0.67, 0.95]	
Initial protective	-0.25 (0.07)	11.54**	0.78 [0.68, 0.90]	0.65** [0.56, 0.73]	-0.13 (0.08)	2.76	0.88 [0.75, 1.02]	0.60 [0.50, 0.70]
Protective change	0.13 (0.06)	4.15*	1.14 [1.01, 1.29]		0.17 (0.08)	4.02*	1.18 [1.00, 1.39]	
Initial TS	0.09 (0.03)	9.46**	1.09 [1.03, 1.15]	0.64** [0.56, 0.73]	0.09 (0.03)	6.95**	1.09 [1.02, 1.16]	0.64** [0.54, 0.74]
TS change	-0.06 (0.03)	5.69*	0.94 [0.89, 0.99]		-0.11 (0.04)	7.14**	0.90 [0.83, 0.97]	

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

The HRs for the change models are presented in Table 32. For women, a one-point increase on the DRAOR total change score and the stable change score was associated with a reduction in the likelihood of criminal reoffending of between 6% and 15% respectively, after taking into account individual differences in initial DRAOR risk scores. Also for women, an increase on the protective change score was associated with a 14% increase in the likelihood of criminal reoffending, taking into account individual variance on the initial protective score. For men, a one-point increase on the risk change scores was associated with a reduction in the likelihood of criminal reoffending of 10% to 20%, after taking into account individual differences on initial DRAOR risk scores.

The AUC values supported the results of the multivariate regression models for women and indicated the models had moderate discriminative accuracy. For women, although neither the initial nor the acute change score significantly predicted criminal reoffending, the AUC value for the model was significant. This was a nuanced result that indicated when base rates were accounted for the model was better than chance at distinguishing recidivists from non-recidivists. However, it was the lowest AUC value for women and the lower band of the CI was only 0.01 above the chance level, which indicated it was plausible the accuracy was no better than chance. For the male sample, the AUC values supported the significant multivariate models. The values indicated the change models had moderate accuracy; however, caution is necessary given all the lower bands of the CI’s were only just above chance level.

The CIs associated with the AUC values for women and men considerably overlapped. Thus it was likely that when accounting for differences in base rates of criminal reoffending the accuracy of the multivariate change models was equivalent across gender.

Research Question 5: Does the RoC*RoI predict criminal reoffending? If so, does the proximal DRAOR TS add incremental predictive validity above the RoC*RoI?

Predictive validity of the RoC*RoI. To investigate the predictive validity of the RoC*RoI[17] for criminal reoffending two separate univariate Cox regression models were conducted. The RoC*RoI was the independent predictor variable, the criterion variable was the dichotomous criminal reconvictions variable and the time variable was survival days.

As shown in Table 34 the RoC*RoI predicted criminal reoffending for women $\chi^2(1, N = 175) = 11.34$, $p = .001$ and men $\chi^2(1, N = 175) = 20.29$, $p < .001$. The HR for the univariate models showed that for every 7 point increase (i.e. 0.07) on the RoC*RoI, a woman’s likelihood of being convicted of a criminal reoffence increased by 66%; for example, a woman’s RoC*RoI score increasing from 0.33 to 0.40. For men, a 26 point (i.e. 0.026) increase on the RoC*RoI was associated with an 89% increase in the likelihood of a man being convicted of a criminal reoffence.

[17] Consistent with the previous analyses the sample used to answer research question five was the subset of each matched sample. As indicated by an independent samples *t*-test these samples did not significantly differ on RoC*RoI.

The relationship between RoC*RoI and DRAOR TS. Pearson bivariate correlations were performed to examine the relationships between women and men's RoC*RoI scores and their initial and proximal DRAOR TS. As shown in Table 33 all RoC*RoI scores were significantly weakly correlated with the DRAOR TS; however, the correlation was not significant for women's RoC*RoI score and their initial DRAOR TS.

Table 33

*Correlation Matrix of RoC*RoI and Initial and Proximal DRAOR TS for the Criminal Outcome*

	Women		Men	
	Initial DRAOR TS	Proximal DRAOR TS	Initial DRAOR TS	Proximal DRAOR TS
RoC*RoI	.14	.24**	.28**	.22**

** $p < .01$

Incremental validity of the proximal DRAOR TS. The incremental predictive validity of the proximal DRAOR TS was assessed separately for women and men in a multivariate model. The initial DRAOR TS for each sample was included as a control for baseline dynamic risk. As previously shown, the proximal DRAOR TS has unique predictive capabilities as an up-to-date (i.e. time-dependent) assessment of dynamic risk.

As shown in Table 34 the multivariate model for women predicted criminal reoffending $\chi^2(3, N = 175) = 18.70, p < .001$. The RoC*RoI and the proximal DRAOR TS significantly contributed to the model independent of one another, controlling for the initial DRAOR TS[18]. This finding indicated that the proximal DRAOR TS was a useful predictor for criminal reoffending in addition to the RoC*RoI, as it accounted for unique variance in criminal reoffending. Likewise, for men, the multivariate model predicted criminal reoffending $\chi^2(3, N = 175) = 29.17, p < .001$. As with women, this result indicated that the proximal DRAOR TS was a useful predictor for criminal reoffending because it accounted for unique predictive variance above that accounted for by the RoC*RoI.

[18] Before the proximal DRAOR TS was added to the model a Cox regression model including only the RoC*RoI and the initial DRAOR TS in block 1 for each sample was produced. Both models were significant and the initial DRAOR TS was a non-significant variable in the model, see Appendix E for full details.

Table 34

*Univariate Model of the RoC*RoI Predicting Criminal Reoffending and Multivariate Model of the Incremental Predictive Validity of the Proximal DRAOR TS*

	Women			Men		
	β (SE)	Wald	Hazard Ratio [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]
RoC*RoI	2.04 (0.61)	10.98**	7.66 [2.30, 25.54]	3.29 (0.75)	19.10***	26.89 [6.15, 117.67]
RoC*RoI +	1.68 (0.64)	6.99**	5.37 [1.55, 18.68]	3.11 (0.78)	16.02***	22.48 [4.90, 103.20]
Initial TS +	0.02 (0.03)	0.32	1.02 [0.96, 1.08]	-0.60 (0.04)	1.85	0.94 [0.86, 1.03]
Proximal TS	0.06 (0.03)	4.11*	1.06 [1.00, 1.12]	0.11 (0.04)	7.39**	1.12 [1.03, 1.22]

* $p < .05$ ** $p < .01$ *** $p < .001$

Summary: Research question 5. For women and men, the RoC*RoI and proximal DRAOR TS were both shown to be useful tools to predict criminal reoffending, as each accounted for unique predictive variance. Because the initial DRAOR TS was non-significant both in the models presented in Table 34 and in model including only the RoC*RoI and the initial DRAOR TS, it is concluded the predictive validity of the proximal DRAOR TS was the result of its time-dependent property.

Chapter 4: Discussion

This research was the first empirical study to examine the predictive validity of the Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin et al., 2012) with matched samples of community-sentenced women and men. The research was comparative and examined the tool's predictive validity for two types of reconviction: breaches of sentence and criminal reoffending. The overarching aim of the research was to determine whether the DRAOR, a gender-neutral tool, was a valid tool for the prediction of reconvictions with women and men. Equally, this research was designed to empirically examine questions concerning the theory and practice of dynamic risk assessment. This discussion opens with an overview of the main research findings. Subsequently, the key research findings are reviewed with reference to relevant literature on the risk assessment of women. Next, the findings are discussed within the broader risk assessment literature. There is a discussion of the limitations and future directions of the research that is followed by a final conclusion.

Summary of Findings

The women offenders were matched to men on RoC*RoI scores and other criminal history variables. However, despite the matching, women had a significantly higher base rate of reconvictions relative to their matched male counterparts. The significant differences in the base rate of reconvictions indicated the static risk assessment tool, the RoC*RoI, may have under-classified women's risk level compared to men's.

The trend in scores on the dynamic risk assessment tool, the DRAOR, were in the anticipated direction: women and men who were reconvicted had higher DRAOR risk scores and lower protective scores compared to non-recidivists. All DRAOR scores at both assessment points were significantly different between the four recidivist groups, with the exception of the initial stable subscale under the criminal reconviction outcome. The trend in the DRAOR scores reached statistical significance for the female recidivists versus the male non-recidivists at both assessment points. However, significant differences between the DRAOR scores of men who were reconvicted compared to men who were not were less consistently found. Together, these results indicated the DRAOR scores were sensitive to the observed rates of reconvictions across and the two samples, and less consistently within each sample.

Research question one examined the comparative predictive validity of the DRAOR with the matched samples using a series of univariate and multivariate Cox regression models. The initial and proximal DRAOR total score (TS)—the composite score of the three

DRAOR subscales— predicted breach and criminal reconvictions for women. For men, only the proximal DRAOR TS predicted reconvictions. The DRAOR TS at both assessments had equivalent accuracy across gender and ranged from moderate to high (AUC 0.61- 0.73). A consistent pattern observed in the initial DRAOR scores was that the stable subscale did not predict criminal reconvictions for either sample. However, for the proximal DRAOR, all scores with the exception of the acute subscale for women and the stable subscale for men predicted reconvictions. Further, there were subtle differences in the predictive components of the DRAOR across gender and reconviction type.

Relative to the initial DRAOR scores, the proximal DRAOR scores were more consistently predictive of both reconviction types (research question two). In addition, the proximal DRAOR scores provided incremental predictive validity above the initial DRAOR scores, a pattern that was consistent across gender and reconviction type. Following from research question two, research questions three and four examined the dynamic nature of the DRAOR. Both samples’ DRAOR scores on average changed significantly in a positive direction between the initial and proximal assessments. Further, analysis at the recidivist versus non-recidivist level showed the amount of change made was dependent upon whether the offender was convicted of a reoffence, not on the offender’s gender.

On the whole change made on the DRAOR significantly predicted reconvictions for women and men (research question four). The change models had moderate and comparable predictive accuracy across the samples. Taken together, research questions two, three and four show that, for both samples, the DRAOR was able to document changes in risk over time and that change made on the DRAOR was an important predictor of future criminal behaviour.

Research question five examined the incremental predictive validity of the proximal DRAOR TS above a static risk estimate (the RoC*RoI). The RoC*RoI did not predict breach reconvictions for either sample; however, it did significantly predict criminal reconvictions. The proximal DRAOR TS added incremental predictive validity above the RoC*RoI for both samples. Research question five attests to the unique predictive validity of dynamic risk tools relative to static risk tools.

Theoretical and Practical Implications

The gender-neutral versus gender-responsive perspective. This research provided support for the gender-neutral perspective showing that the DRAOR— a tool informed by research with adult male offenders—was as good at predicting reconvictions for women as men. As a gender-neutral tool the DRAOR taps into the Central eight domains of risk. The current research showed that not only did the tool predict reconvictions with women but that the tool operated with women at a level comparable to men. This finding is consistent with research showing that the LSI-R performed just as well in the prediction of reconvictions for women as it did for men (Manchak, Skeem, Douglas, & Siranosian, 2009). Manchak and colleagues' study included a disproportionate number of male offenders (*N* = 1035) relative to female (*N* = 70). Because the present study included matched samples of women and men the research is a more methodologically rigorous test of gender neutrality.

For women, in the prediction of breach reconvictions, the DRAOR TS yielded an AUC value of 0.73, which is considerably larger than the accuracy of the LSI-R in a sample of women predicting nonviolent reconvictions (inclusive of technical violations, AUC 0.62; Folsom & Atkinson, 2007). The performance of the DRAOR in the prediction of criminal reconvictions for women was considered moderate (AUC 0.64) and slightly lower than that found for the LSI-R with women predicting general reoffending (exclusive of technical violations, AUC 0.67; Folsom & Atkinson, 2007). Taken together the findings suggest that the Central eight risk factors of the PIC-R model of criminal behaviour are relevant to the prediction of both female and male criminal behaviour (Andrews & Bonta, 2010). The above discussion was with reference to the DRAOR TS; however, this research also examined *how* the DRAOR predicted reconvictions across gender through the examination of the three DRAOR subscales, the findings of which are discussed below.

Although women and men were matched on risk related variables, the predictive components of the DRAOR were shown to be different. Consistent with the only other evaluation of the DRAOR with matched samples of women and men (Yesberg et al., 2015) the acute subscale drove the predictive power of the proximal DRAOR in the prediction of breaches. Three of the seven risk factors of the acute subscale are commonly referenced in the gender-responsive literature: substance abuse, relationship difficulties and living situation (Van Voorhis et al., 2010). Substance abuse is one of the most promising gender-responsive items; Andrews and colleagues (2012) have shown the substance abuse domain of the LSI-R to be more strongly predictive of recidivism in women than men. They concluded that the Big four risk factors of the PIC-R could be extended to the Big five for women, through the incorporation of substance abuse. Thus, a tentative explanation for the enhanced predictive validity of the acute subscale is its inclusion of gender-responsive factors.

The gender-responsive literature has highlighted that the formation and maintenance of interpersonal relationships is an integral factor relating to female criminality (Benda, 2005). Although protective factors are a relatively new addition to structured risk assessment tools; for women, the research suggests their incorporation is particularly welcome, because the initial protective subscale significantly predicted criminal reconvictions after controlling for the two risk subscales. The salience of the protective subscale for women is consistent with the research of McCoy and Miller (2013) that showed, in a matched sample of women and men, the level of perceived prosocial support, a protective factor in the DRAOR, more strongly predicted desistance in women relative to men. Thus, consistent with the extant research, the current findings suggest different components of a gender-neutral tool may be more or less important for predicting recidivism with women and men (see Holtfreter & Cupp, 2007; van der Knaap, 2012). A promising avenue to advance our understanding of how the DRAOR operates across gender could involve an evaluation of the DRAOR at the item level. Such analyses could unpack whether the gender-responsive factors present in the acute subscale of the DRAOR are accounting for the largest proportion of the variance in women's recidivism. Previous research has indicated the interpersonal relationship item of the acute subscale might be particularly important for women (Scanlan, 2013).

Although the findings of this research support the general premise of gender neutrality, they do not necessarily suggest the DRAOR, or gender-neutral tools more broadly, are the *best* tools for use with women offenders. The research showed the DRAOR predicts recidivism with a female sample at a level comparable to men. However, the research did not examine the predictive validity of proposed gender-informed risk factors. It is entirely possible the inclusion of these factors may have added incremental predictive validity above the DRAOR with women and/or men. Previous research has shown in a sample of women offenders that a series of gender-informed supplements enhanced the predictive power of an existing gender-neutral model (Van Voorhis et al., 2010). Nevertheless, because current practice is predicated on the assumption of gender neutrality, it is reassuring that the DRAOR is capable of operating comparatively well with women and men.

In addition to their predictive validity, the usefulness of the DRAOR and dynamic risk tools more generally, reside in their ability to identify areas of need for offenders (Manchak et al., 2009). As previously mentioned dynamic risk tools are frequently referred to, as *risk-need* tools because dynamic risk factors, as per the Risk, Need, Responsivity (RNR; Andrews et al., 199) framework, simultaneously predict risk of recidivism and identify areas of need to be targeted in intervention. In their recent article Monahan and Skeem (2014) highlight that some consider the combining of these two objectives in dynamic risk tools controversial. For women offenders, the combining of these two objectives may be particularly problematic, because although the Central eight risk factors have been shown to have predictive validity across gender, the importance of these factors in the development and maintenance of women's criminal behaviour remains unclear. In other words, it is not known whether these factors are as important in the treatment of women as they are in the prediction of women's risk, or whether gender-responsive factors such as victimisation and mental health difficulties are more important treatment targets for women offenders (Olver, Stockdale, & Wormith, 2014; Wright et al., 2007). The evidence supporting the use of dynamic risk tools to predict future criminal behaviour in women is well established; however, the evidence supporting the use of dynamic risk tools to identify intervention needs with the population is less developed.

The above discussion positioned the current research findings within the extant literature on the risk assessment of women offenders. Overwhelmingly this research supports the premise of gender neutrality; however, as discussed above gender neutrality as it applies to the need principle of the RNR model is less established. Future research is needed to further advance our understanding of how best to assess risk and importantly inform the treatment of female offenders. The second major aim of this research was to examine questions concerning the theory and practice of dynamic risk assessment and these findings are discussed below.

The dynamic nature of the DRAOR. In addition to the predictive validity of the DRAOR with women and men three key research findings attest to the dynamic nature of the tool. First, relative to initial DRAOR scores, proximal DRAOR scores were superior in the prediction of reconvictions. Second, women and men made significant positive change between the initial and proximal assessment for both types of reconviction. And third, change made on the DRAOR significantly predicted reconvictions comparatively well across gender.

With reference to third and fourth-generation risk tools Andrews and colleagues (2006) have stated, "reassessments will double and, perhaps, triple the outcome variance explained by intake assessments" (p. 16). The proximal DRAOR in this research significantly improved the prediction beyond that of initial assessment, the amount of additional variance explained was not determined; however, nevertheless it tentatively supports the statement of Andrews and colleagues. Further, the finding is consistent with some previous research; for example, Labrecque and colleagues (2014) showed LSI-R scores of adults on probation changed significantly between two assessment points and, although both the initial and re-assessment scores were predictive of rearrests, the re-assessment score had the strongest predictive power. In the current study, when proximal DRAOR scores were entered alongside the initial DRAOR scores in the incremental validity models, the initial scores were rendered non-significant; meaning there was almost no relationship between initial DRAOR scores and reconvictions. The redundancy of the initial DRAOR scores reinforces the highly fluid nature of dynamic risk across time and confirms that best practice guidelines for use of the DRAOR should involve regular updating of assessments over the period of supervision.

Dynamic risk assessment tools such as the DRAOR are theorised to assess risk state, which is an offender's fluctuating proclivity to engage in criminal behaviour at any given time (Douglas & Skeem, 2005). The assessment of an offender's risk state is dependent upon the capacity of dynamic risk factors to change over time. The research showed offenders made significant positive change between the initial and proximal assessment, which confirmed the DRAOR's sensitivity to change in risk over time. Furthermore, the amount of change made on the DRAOR significantly predicted reconvictions controlling for initial DRAOR scores. These findings are consistent with a study by Howard and Dixon (2013), which showed changes in dynamic risk factors were empirically linked to recidivism, after controlling for static and initial (i.e. time-invariant) dynamic risk.

Taken as a whole, the research recommends that dynamic risk assessment scores should not be considered in isolation, but instead in relation to one another, a conclusion supported by other studies of dynamic risk assessment tools (Schlager & Pacheco, 2011). The amount of change made on the DRAOR should be considered additional and of equal value for prediction purposes alongside a single assessment score. In addition, because change was found to be a significant predictor of reconvictions, probation officers should also monitor how much change an offender has made over time because change on the DRAOR captures unique variance in recidivism for women and men alike.

Dynamic risk assessment is an integral component of the RNR framework (Andrews et al., 1990). As previously mentioned third and fourth-generation risk tools, such as the DRAOR, inform both the risk and need principles. The risk principle states the risk of future criminal behaviour of *all* offenders can be predicted based on the structured evaluation of the Central eight risk factors, a contention supported in this research. Further, the principle states treatment resources should be targeted to the higher-risk offenders because a greater reduction in risk, and therefore criminal behaviour, is possible (Andrews & Bonta, 2010). The moderate to strong correlations observed in the research between the initial DRAOR scores and DRAOR change scores showed the amount of change made was, indeed, a function of baseline risk. Furthermore, unlike the incremental predictive validity models of the initial and proximal DRAOR scores, the initial DRAOR scores remained significant in the change models. As posited by the risk principle, the continued significance of the initial DRAOR scores in these models highlights that baseline risk continues to be an important predictor of criminal behaviour alongside change.

The need principle of the RNR model states that reductions in criminal behaviour will occur following the effective targeting of an offender's dynamic risk factors (Andrews & Bonta, 2010). It was beyond the scope of this research to establish whether the change offenders made on the DRAOR was as a result of targeted intervention delivered by probation officers. So although, the DRAOR is a fourth-generation risk tool designed for case management, this research did not determine whether it was being used this way. However, the findings do suggest that the risk and protective factors of the DRAOR satisfy the four criteria outlined by Kraemer and colleagues (1997) to be considered causal and thus useful for intervention: 1) the DRAOR predicted reconviction, (2) the DRAOR assessments preceded reconviction or censoring, (3) the scores significantly changed over time and (4) the amount of change predicted reconviction. It was necessary, and therefore encouraging, that the DRAOR satisfied the four criteria of Kraemer and colleagues because it suggests the DRAOR has promise as a useful case management tool.

Static versus dynamic risk tools. The predictive utility of second-generation (i.e. static) risk assessment tools is well established in the literature (Miller, 2006). However, what is less clear is whether dynamic risk tools provide incremental predictive validity above static risk tools. The lack of predictive validity for the RoC*RoI in the prediction of breach reconvictions with both samples was not surprising. Static risk tools, like the RoC*RoI, have a proven robust empirical relationship with future criminal behaviour. Recall the RoC*RoI algorithm comprises criminal history variables and was designed to predict serious reoffences that result in reimprisonment. However, because an offender in this research could incur a breach reconviction simply due to missing an appointment with their probation officer, breaches likely have little relationship with criminal intent, which arguably is what criminal history variables are a proxy for. The proximal DRAOR significantly predicted breach reconvictions for both samples and thus the DRAOR outperformed the RoC*RoI. Although the DRAOR was also not designed to predict breach reconvictions, the tool's capacity to do so reinforces the importance of dynamic risk tools that are grounded in theory (e.g., the PIC-R; Andrews et al., 2010).

The DRAOR demonstrated incremental predictive validity above the RoC*RoI in the prediction of criminal reconvictions for both samples. The dynamic risk assessment literature contends that because the factors tap into constructs outside of criminal history, dynamic risk tools should provide incremental predictive power above static tools; however, the empirical evidence for this contention is equivocal. For example, the incremental predictive validity of the DRAOR TS was consistent with the work of Hanson and colleagues, which showed, after controlling for differences in static risk, dynamic risk predictors continued to predict general and violent recidivism in sexual offenders (Hanson et al., 2007). However, the finding was at odds with the work of Caudy and colleagues, which showed dynamic risk factors failed to improve the prediction of criminal risk above static risk factors in a sample of offenders who were serving community supervision sentences (Caudy et al., 2013).

A speculative explanation for the equivocal findings in the extant literature is related to the timing of the dynamic assessment. In the present research the proximal DRAOR provided the incremental predictive validity above the RoC*RoI, but the initial DRAOR did not. Therefore, the incremental predictive validity of dynamic risk assessment tools may not necessarily reside in the risk factors themselves but the capacity of the factors to provide an up-to-date assessment of an offender's current level of risk, which the RoC*RoI and static risk tools more broadly are not capable of. Because the RoC*RoI remained significant in the combined model, static risk and dynamic risk factors are both valid for the prediction of future criminal behaviour. But importantly, these findings support the contention that in addition to their use as tools for monitoring and managing risk across time, dynamic risk assessment tools have the capacity to enhance risk predictions beyond that of purely static risk tools: a pattern that was replicated across gender in the current research.

The RoC*RoI used in the present research was completed prior to or at the beginning of an offender's community supervision sentence. An interesting finding in this research and one that has been replicated in the previous evaluation of the DRAOR with women (see Yesberg et al., 2015), was that although matched on RoC*RoI and other criminal history variables, women had a significantly higher base rate of reconvictions relative to men. In the present research, as indicated by the Cohen's *d* for the mean change scores, women made less positive change on the DRAOR compared to men. Thus, a possible explanation for the significant differences in the base rates of reconviction is that the male sample became less risky over the period of supervision. However, static risk tools, like the RoC*RoI, do not have the capacity to reflect change in risk over time and because of this, the reliance on static risk tools in community settings where risk level is particularly changeable is not recommended.

Furthermore, the mean RoC*RoI for the matched offenders in this research was 0.25, which translates to a predicted 25% of each sample being reimprisoned within five years.

Although this research examined reconvictions, not reimprisonment, which is a less serious criterion variable than reimprisonment, 42% of women and 31% of men were reconvicted of a criminal reoffence during an average of two and half years follow-up. Thus, consistent with the previous evaluation of the RoC*RoI with matched women and men (Yesberg et al., 2015), it appears the RoC*RoI may be misclassifying women compared with men. The RoC*RoI consists of a number of static risk factors, including gender. Being a male is weighted more heavily given the higher base rate of reoffending found among this population (Steffensmeier & Allan, 1996). Thus, a woman with a RoC*RoI score of 0.25 may potentially have a more extensive criminal history than a man with the same score. In other words, if gender were removed from the equation, women would likely be higher risk on the remaining factors. Thus, future matching procedures, if the research aims allow, should match the samples on criminal history variables alone, as opposed to a static actuarial estimate. The proposition that the RoC*RoI is misclassifying risk is speculative and will require further exploration given the normed outcome criterion of reimprisonment was not examined.

Protective factors. The incorporation of protective factors into fourth-generation risk assessment tools is intended to provide a more balanced appraisal of risk by shifting focus from what increases risk for criminal behaviour to what may mitigate or protect against an offender from future criminal behaviour (de Vries Robbè et al., 2013). In the present research, the protective subscale of the DRAOR was a consistent predictor of reconvictions at both assessment points for women; for men, the subscale significantly predicted both reconviction types at the proximal assessment only. Further, the proximal DRAOR TS, which represented an overall index of risk corrected for protective factors, was also a significant predictor of reconvictions for both samples. These results are consistent with research showing that protective factors, and a combined assessment of risk and protective factors, are valid predictors of reconvictions in men and women (see Jones et al., 2015; Rennie & Dolan, 2010). The findings suggest the protective subscale is an important component of the DRAOR and probation officers should ensure they focus equally on risk and protective factors when assessing an offender using the tool. The present research did not directly examine whether protective factors enhanced the predictive validity of the risk oriented subscales; however, future research should conduct hierarchical regression models to investigate whether the protective subscale of the DRAOR accounts for unique variance above that of the acute and stable risk subscales.

Limitations and Future Directions

Like all research, the present findings must be considered in light of some limitations. Firstly, due to the strict inclusion criteria and the matching procedure the final samples evaluated were low-risk. As highlighted in the method section, the matching process meant the higher-risk men were unable to be successfully matched to women and thus were removed from the research sample. Compared with the past examinations of the DRAOR with male offenders, the DRAOR underperformed with the men in this research. An explanation for this is likely due to the low-risk nature of the samples, the previous evaluations of the DRAOR with men have included high-risk samples (Yesberg & Polaschek, 2015). Furthermore, because this is the first study that has examined the DRAOR with offenders serving community-based sentences in New Zealand its interpretation requires caution and replication is needed.

Secondly, the research was limited by the operationalisation of change, as change in the present research was considered to be a linear increase or decrease between the initial and proximal assessments. However, this operationalisation fails to capture the fluctuations in change over time. Future research should attempt to examine the frequency of change across time and how those changes are related to reoffending. Multilevel growth modeling would be a viable statistical avenue for this task. In future research the time between the initial and proximal assessments should be controlled for, because more time between assessments may mean more opportunity to demonstrate change. In addition, research should examine whether this change is occurring as the result of intervention or naturally with the passage of time. If research can show change on the factors is the result of intervention then direct support for the DRAOR as a case management tool to guide probation officers in their intervention efforts with women and men serving community sentences will exist.

A key direction for future research would be to evaluate the performance of gender-informed factors above a gender-neutral tool or model with matched samples of women and men. This research would examine the gender-responsive perspective with adequate methodological rigor. The comparison group would enable direct appraisal of whether proposed gender-specific risk factors provide incremental predictive validity and importantly whether they do so uniquely for women. Promising gender-responsive factors would include victimization, mental health difficulties and potentially a composite of the aforementioned factors in addition to substance abuse. Such a composite would tap into the cumulative impact that the gender-responsive perspective advocates negative life experiences have on female criminality.

Given the samples were matched on static risk (i.e. low risk), the assumption of the risk principle, that the amount of change made will be greater for higher risk offenders relative to lower risk offenders was not directly examined in this research. Future research could examine the amount of change made as a function of risk bands (e.g., low, medium and high). Analysis of change using risk bands would provide a *direct* examination of whether higher risk offenders as per the risk principle make the greatest change, a finding that has been shown in other research (see Vose, Smith, & Cullen, 2013). Research of this kind, including women offenders, would be particularly useful given very few systematic studies have examined the principles of the RNR model with comparative samples of women and men.

The research has shown that up-to-date risk assessment outperforms an earlier risk assessment. Future research should examine how 'up-to-date' the risk assessment needs to be in order to retain the superior predictive validity evident in this research. Such information would have practical implications for probation officers; the determination of the maximum length of time allowed between risk assessments while maintaining the superior predictive validity of reassessment would be very useful given correctional resources (i.e. probation officers) are in high demand. To investigate this a prospective longitudinal research design would be required. A final area for future research would be to examine if and how community settings are conducive to positive change; for example, looking at the relationships that develop between the probation officer and their client.

Conclusion

The present research was the first empirical study to examine the predictive validity of the DRAOR with matched samples of community-sentenced women and men. Based on the present research the DRAOR is a valid risk assessment tool for the prediction of reconvictions in women and men. The DRAOR was shown to be a truly dynamic risk assessment tool capable of detecting changes in risk overtime. Best practice guidelines for use of the DRAOR and risk assessment tools more broadly, would ensure risk is regularly reassessed. Based on the evidence of the current research the amount of change made on the DRAOR is significantly related to criminal behaviour, which was an important finding because the DRAOR was developed as a fourth-generation case management tool to inform the assessment and intervention needs of offenders. The findings of this research suggest dynamic risk tools have unique predictive validity relative to their static actuarial counterparts. Thus, the capacity of dynamic risk factors to monitor changes in risk overtime is supplemented by their unique predictive validity and thus routine risk assessment should comprise a dynamic risk assessment.

Returning to the opening of this thesis, were a probation officer to assess Jennifer’s risk of recidivism using the DRAOR, a tool developed largely from research with adult men, the present research supports the gender-neutral assumption that the assessment would be as valid as with any male offender. However, for practice it is important to remember that gender neutrality means ‘developed on men and later applied to women’ and never the reverse (Van Voorhis et al., 2010). Thus, future research is required to determine if in fact gender-informed risk factors increase the predictive validity of a gender-neutral risk tool with samples of women and men.

References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. (5th ed.). New Providence, NJ: Anderson Publishing.

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

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17, 19-52. doi: 10.1177/0093854890017001004

Andrews, D. A., Guzzo, L., Raynor, P., Rowe, R. C., Rettinger, L. J., Brews, A., & Wormith, J. S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, *56*, 113-133. doi: 10.1177/0306624X10395716

Austin, P. C. (2011). An introduction to propensity score matching for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424. doi: 10.1080/00273171.2011.568786

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

Beech, A. R., & Craig, L. A. (2012). The current status of static and dynamic factors in sexual offender risk assessment. *Journal of Aggression, Conflict and Peace Research*, *4*, 169-185. doi: 10.1108/17596591211270671

Belknap, J., & Holsinger, K. (2006). The gendered nature of risk factors for delinquency. *Feminist Criminology*, *1*, 48-71. doi: 10.1177/1557085105282897

Benda, B. B. (2005). Gender differences in life-course theory of recidivism: A survival analysis. *International Journal of Offender Therapy and Comparative Criminology*, *49*, 325-342. doi: 10.1177/0306624X04271194

Bengtson, S., & Långström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse: A Journal of Research and Treatment, 19*, 135-153. doi: 10.1007/s11194-007-9044-5

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders: An integrative perspective*. England. John Wiley & Sons.

Blanchette, K., & Taylor, K. N. (2007). Development and field test of a gender-informed security reclassification scale for female offenders. *Criminal Justice and Behavior*, *34*, 362-379. doi: 10.1177/0093854806290162

Bonta, J., Pang, B., & Wallace-Carpetta, S. (1995). Predictors of recidivism among incarcerated female offenders. *The Prison Journal, 75,* 277-294. doi: 10.1177/0032855595075003002

Brown, S. L., Amand, M. D. S., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: a three-wave prospective study. *Law and human behavior*, *33*, 25-45. doi: 10.1007/s10979-008-9139-7

Brown, S. & Motiuk, L. (2008). Using dynamic risk factors to predict criminal recidivism in a sample of male and female offenders. Canadian Psychology/Psychologie Canadienne, 49 (2a), 298.

Casey, S., Day, A., Vess, J., & Ward, T. (2013). *Foundations of offender rehabilitation.* New York: Routledge.

Chadwick, N. (2014). *Validating the Dynamic Risk Assessment for Offender Re-entry (DRAOR) in a sample of U.S. probationers and parolees.* Retrieved from March 19 https://curve.carleton.ca/system/files/theses/31885.pdf

Campbell, M. A., French, S., Gendreau, P. (2009). The Prediction of Violence in Adult Offenders: A Meta-analytical comparison of instruments and methods of assessment. *Criminal Justice and Behaviour, 36*, 567-590. doi: 10.1177/0093854809333610

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*(6), 458-466. Retrieved from http://dx.doi.org/10.1016/j.jcrimjus.2013.08.004

Caulfield, L. (2010). Rethinking the assessment of female offenders. *The Howard Journal, 49,* 315-327. doi: 10.1111/j.1468-2311.2010.00625.x

Chesney-Lind, M., & Pasko, L. (2013). *Girls, Women, and Crime.* United States of America. Sage Publications.

Cohen, J. (1992). The power primer. *Psychological Bulletin, 112*(1), 155-159.

Coulson, G., Ilacqua, G., Nutbrown, V., Giulekas, D., & Cudjoe, F. (1996). Predictive utility of the LSI for incarcerated female offenders. *Criminal Justice and Behaviour, 23*(1), 427-439.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7-29. doi: 10.1177/0956797613504966

Daly, K. (1992). Women's pathways to felony court: Feminist theories of lawbreaking and problems of representation. *S. Cal. Rev. L. & Women's Studies*, *2*, 11-52.

de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. *International Journal of Forensic Mental Health, 10*, 171-177. doi: 10.1080/14999013.2011.600230

de Vries Robbé, M., de Vogel, V., & de Spa, E. (2011). Protective factors for violence risk in forensic psychiatric patients: A retrospective validation study for the SAPROF. *International Journal of Forensic Mental Health, 10,* 178-186. doi: 10.1080/14999013.2011.600232

de Vries Robbé, M. (2014). Protective factors: Validation of the Structured Assessment of Protective Factors for Violence Risk in Forensic Psychiatry (Unpublished doctoral thesis) Radboud University Nijmegen.

de Vries Robbé, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. Journal of Forensic Psychiatry & Psychology, 24, 440-457. doi: 10.1080/14789949.2013.818162

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 psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior,* 39, 53-61. doi: 10.1037/lhb0000089

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11,* 347-383. doi: 10.1037/1076-8971.11.3.347

Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexaul Abuse: A Journal of Research and Treatment, 24,* 5-28. doi: 10.1177/1079063211403164

Flores, A. W., Lowenkamp, C. T., Smith, P., & Latessa, E. J. (2006). Validating the Level of Service Inventory-Revised on a sample of federal probationers. *Federal Probation, 70*(2), 44-49.

Folsom, J., & Atkinson, J. L. (2007). The generalizability of the LSI-R and the CAT to the prediction of recidivism in female offenders. *Criminal Justice and Behavior, 34,* 1044-1056.
doi: 10.1177/0093854807300097

Garcia-Mansilla, A., Rosenfeld, B., & Nicholls, T. L. (2009). Risk assessment: Are current methods applicable to women? *International Journal of Forensic Mental Health*, *8,* 50-61.
doi: 10.1080/14999010903014747

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. R., & Tatham, R. L. (2006). *Multivariate data analysis.* (6 Eds). United States of America: Pearson Prentice Hall.

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

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

Hanson, R. K., Harris, A. J. R., Scott, T., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision*: *The Dynamic Supervision Project*. (Corrections Research User Report No. 2007-05). Ottawa, ON: Public Safety Canada.

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

Hollin, C. R., & Palmer, E. J. (2006). Criminogenic need and women offenders: A critique of the literature. *Legal and Criminological Psychology*, *11*, 179-195.
doi: 10.1348/135532505X57991

Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment the empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice*, *23*, 363-382. doi: 10.1177/1043986207309436

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: 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: 10.1177/0093854809335409

Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: the Kia Marama evaluation. *Sexual Abuse: A Journal of Research and Treatment*, *14*, 103-119.
doi: 10.1023/A:1014616113997

Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2015). Incorporating strengths into quantitative assessments of criminal risk for adult offenders. *Criminal Justice and Behavior, 42, 321-338.* doi: 10.1177/0093854814547041

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*(4), 337-343.

Kreis, M. K., Schwannauer, M., & Gillings, K. (2014). Relational Risk Factors for Reoffending in Women: A Systematic Review. *International Journal of Forensic Mental Health*, *13*, 381-393. doi: 10.1080/14999013.2014.962713

Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The Importance of Reassessment: How Changes in the LSI-R Risk Score Can Improve the Prediction of Recidivism. *Journal of Offender Rehabilitation*, *53*, 116-128. doi: 10.1080/10509674.2013.868389

Lewis, K., Olver, M. E., & Wong, S. C. (2012). 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*. Advance online publication.
doi: 10.1177/1073191112441242

Lin, F. (2008). Solving multicollinearity in the process of fitting regression model using the nested estimate procedure. *Quality and Quantity*, *42,* 471-426. doi: 10.1007/s11135-006-9055-1

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine*, *43*, S8-S23.
doi: 10.1016/j.amepre.2012.04.029

Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2001). Risk/need assessment, offender classification, and the role of childhood abuse. *Criminal Justice and Behavior*, *28*, 543-563. doi: 10.1177/009385480102800501

Manchak, S. M., Skeem, J. L., Douglas, K. S., & Siranosian, M. (2009). Does Gender Moderate the Predictive Utility of the Level of Service Inventory—Revised (LSI-R) for Serious Violent Offenders? *Criminal Justice and Behaviour*, *36*, 425-442.
doi: 10.1177/0093854809333058

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment*, *22*, 191-217. doi: 10.1177/1079063210366039

McClellan, D. S., Farabee, D., & Crouch, B. M. (1997). Early victimization, drug use, and criminality a comparison of male and female prisoners. *Criminal Justice and Behavior*, *24*, 455-476.
doi: 10.1177/0093854897024004004

McCoy, L. A., Miller, H.A. (2013). Comparing gender across risk and recidivism in nonviolent offenders. *Women & Criminal Justice*, *23*, 143-162. doi: 10.1080/08974454.2012.759054

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*(6), 767-782.
doi: 10.1002/bsl.728

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological review*, *100*(4), 674. Retrieved from http://search.proquest.com.helicon.vuw.ac.nz/psycinfo/docview/

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and psychopathology*, *13*(02), 355-375. Retrieved from http://gx4ej7nu5f.search.serialssolutions.com.helicon.vuw.ac.nz/directLink

Monahan, J., & Skeem, J. L. (2014). The evolution of violence risk assessment. *CNS spectrums*, *19*, 419-424. doi: 10.1017/S1092852914000145

Mullings, J. L., Marquart, J. W., & Diamond, P. M. (2001). Cumulative continuity and injection drug use among women: A test of the downward spiral framework. *Deviant Behavior, 22,* 211-238. doi: 10.1080/01639620151096074

Nadesu, A. (2007). *Reconviction patterns of released prisoners: A 36-months follow-up analysis.* Wellington, NZ: Department of Corrections.

Nicholls, T. L., Ogloff, J. R., & Douglas, K. S. (2004). Assessing risk for violence among male and female civil psychiatric patients: the HCR- 20, PCL: SV, and VSC. *Behavioral Sciences & the Law*, *22*, *127-158*. doi: 10.1002/bsl.579

New Zealand Department of Corrections (2013). *Trends in the offender population* (2013 calendar year). Retrieved from http://www.corrections.govt.nz/__data/assets/pdf_file/0012/738939/Trends_in_the_Offender_Population_2013_updated.pdf

Norušis, M. J. (2004). SPSS 13.0 Advanced statistical procedures companion. Upper Saddle River, NJ: Prentice Hall.

O'Brien, R. (2007). A caution regarding rules of thump for variance inflation factors. *Quality and Quantity, 41*, 673-690. doi: 10.1007/s11135-006-9018-6

Olver, M. E., Lewis, K., & Wong, S. C. (2013). Risk reduction treatment of high-risk psychopathic offenders: The relationship of psychopathy and treatment change to violent recidivism. *Personality Disorders: Theory, Research, and Treatment*, *4*, 160-167.
doi: 10.1037/a0029769

Olver, M., E, Stockdale, K., C, & Wormith, J., S. (2014). Thirty years of research on the Level of Service Scales: A meta-analytical examination of predictive accuracy and sources of variability. *Psychological Assessment, 26,* 156-176. doi: 10.1037/a0035080

Rea, L. M., & Parker, R. A. 1. (2005). Designing and conducting survey research. (3 Ed). San Francisco: Jossey-Bass.

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly*, *23,* 384-405. doi: 10.1080/07418820600869152

Rennie, C. E., & Dolan, M. C. (2010). The significance of protective factors in the assessment of risk. *Criminal Behaviour and Mental Health*, *20,* 8-22. doi: 10.1002/cbm.750

Rettinger, L. J., & Andrews, D. A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior*, *37,* 29-46.

doi: 10.1177/0093854809349438

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

Rosenbaum, P. R., & Rubin, D., B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*(387), 516-524. Retrieved from http://www.jstor.org/stable/2288398

Ruiz, M. A., Douglas, K. S., Edens, J. F., Nikolova, N. L., & Lilienfeld, S. O. (2012). Co-occurring mental health and substance use problems in offenders: Implications for risk assessment. *Psychological assessment*, *24*, 77-87. doi: 10.1037/a0024623

Salisbury, E. J., Van Voorhis, P., & Spiropoulos, G. V. (2009). The Predictive Validity of a Gender-Responsive Needs Assessment An Exploratory Study. *Crime & Delinquency, 55*, 550-585. doi: 10.1177/0011128707308102

Sampson, R. J., & Laub, J. H. (1993). *Crime in the making*. Cambridge, MA: Harvard Univeristy Press.

Scanlan, J. M. (2013). *Gender-neutral versus gender-specific risk assessment: A validation of the DRAOR with paroled New Zealand men and women* (Unpublished honours thesis).

Schlager, M. D., & Pacheco, D. (2011). An examination of changes in LSI-R scores over time: Making the case for needs-based case management. *Criminal Justice and Behavior, 38*, 541-553. doi: 10.1177/0093854811402300

Scott, C. K., Grella, C. E., Dennis, M. L., & Funk, R. R. (2014). Predictors of recidivism over 3 Years among substance-using Women released from Jail. *Criminal Justice and Behavior*, *41*, 1257-1289. doi: 10.1177/0093854814546894

Serin, R. C. (2007). *The Dynamic Risk Assessment of Offender Re-entry (DRAOR).* Unpublished user manual.

Serin, R. C., Mailloux, D. L., & Wilson, N. J. (2012). *The Dynamic Risk Assessment of Offender*
Re-entry (DRAOR). Unpublished user manual.

Sentencing Act 2002. Retrieved from http://www.legislation.govt.nz/act/public/2002/0009/latest/DLM135342.html

Skardhamar, T., & Savolainen, J. (2014). Changes in criminal reoffending around the time of job entry: A study of employment and desistance. *Criminology, 52, 263-291.* doi: 10.1111/1745-9125.12037

Smith, P., Cullen, F. T., & Latessa, E. J. (2009). Can 14,737 women be wrong? A meta-analysis of the LSI- R and recidivism for female offenders*. *Criminology & Public Policy, 8,* 183-208. doi: 10.1111/j.1745-9133.2009.00551.x.

Spice, A., Viljoen, J. L., Latzman, N. E., Scalora, M. J., & Ullman, D. (2012). Risk and protective factors for recidivism among juveniles who have offended sexually. *Sexual abuse: a journal of research and treatment ,25,* 347-369. doi: 10.1177/1079063212459086

Sjöstedt, G., & Grann, M. (2002). Risk assessment: What is being predicted by actuarial prediction instruments?. *International Journal of Forensic Mental Health, 1,* 179-183. doi: 10.1080/14999013.2002.10471172

Steffensmeier, D., & Allan, E. (1996). Gender and crime: Toward a gendered theory of female offending. *Annual Review of Sociology, 22,* 459-487. Retreieved from http://www.annualreviews.org/doi/pdf/10.1146/annurev.soc.22.1.459

Stewart, C. A. (2011). *Risk assessment of federal female offenders* (Doctoral dissertation). Retrieved from eCommons University of Saskatchewan.

Swan, A., C & Goodman-Delahunty, J. (2013). The relationship between drug use and crime among police detainees: Does gender matter? *International Journal of Forensic Mental Health, 12,* 107-115. doi: 10.1080/14999013.2013.787561

Tabachnick, B. G., 1936, & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson Education.

Tamatea, A., & Wilson, N. (2009). Dynamic Risk Assessment for Offender Re-entry (DRAOR): A pilot study. Unpublished government report.

The National Institute of Corrections, U.S. Department of Justice. (2002). Bloom, B., Owen, B., Covington, S., & Raeder, M. *Gender-responsive strategies: Research, Practice, and Guiding Principles for Women Offenders*.

Ullrich, S., & Coid, J. (2011). Protective factors for violence among released prisoners: Effects over time and interactions with static risk. *Journal of Consulting and Clinical Psychology, 79,* 381-390. doi: 10.1037/a0023613

van der Knaap, L. M., Alberda, D. L., Oosterveld, P., & Born, M. P. (2012). The predictive validity of criminogenic needs for male and female offenders: Comparing the relative impact of needs in predicting recidivism. *Law and human behavior, 36,* 413-422. doi: 10.1037/h0093932

Van Voorhis, P., Salisbury, E., Wright, E., & Bauman, A. (2008). Achieving accurate pictures of risk and identifying gender responsive needs: Two new assessments for women offenders. *University of Cincinnati Center for Criminal Justice Research, National Institute of Corrections, Washington DC*.

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women’s Risk Factors and Their Contributions to Existing Risk/Needs Assessment The Current Status of a Gender-Responsive Supplement. *Criminal Justice and Behaviour*, *37*, 261-288. doi: 10.1177/0093854809357442

Vose, B., Smith, P., & Cullen, F, T. (2013). Predicting validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior, 40,* 1383-1396. doi: 10.1177/0093854813508916

Walker, K., Bowen, E., & Brown, S. (2013). Psychological and criminological factors associated with desistance from violence: A review of the literature. *Aggression and Violent Behavior*, *18*, 286-299. doi: 10.1016/j.avb.2012.11.021

Ward, T., & Brown, M. (2004). The Good Lives Model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law*, *10*, 243-257. doi: 10.1080/10683160410001662744

Wong, S. C. P., & Gordon, A. E. (2000). *The Violence Risk Scale*. Unpublished Manuscript.

Wright, E. M., Salisbury, E. J., & Van Voorhis, P. (2007). Predicting the Prison Misconducts of Women Offenders The Importance of Gender-Responsive Needs. *Journal of Contemporary Criminal Justice*, *23*, 310-340. doi: 0.1177/1043986207309595

Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological bulletin*, *136*, 740-767.
doi: 10.1037/a0020473

Yesberg, J. A., & Polaschek, D. L. L. (2015). 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

Yesberg, J. A., Scanlan, J. M., Haby, L. J., Serin, R. C., & Polaschek, D. L. L. (2015). Predicting women’s recidivism: Validating a dynamic community-based ‘gender-neutral’ tool. *Probation Journal,* 1-16. doi: 10.1177/0264550514562851

Zakaria, D., Allenby, K., Derkzen, D. M., & Jones, N. (2013) Preliminary development of a dynamic risk assessment tool for women offenders: An examination of gender neutral and gender specific variables. Ottawa, ON: Correctional Service of Canada.

Appendices

Appendix A Demographic information for the female and male samples eligible for propensity score matching

A.1 *Descriptive Information of the Female and Male Samples Prior to PSM*

	Women (*n* = 336)	Men (*n* = 383)
Index Offence		
Non-violent	233 (69.3%)	288 (59.5%)
Violent	79 (23.5%)	122 (31.9%)
Justice/ admin	22 (6.5%)	27 (7.0%)
Other	2 (0.6%)	0
Unknown	0	6 (1.6%)
Ethnicity		
Māori	169 (50.3%)	163 (42.6%)
European	141 (42.0%)	160 (41.8%)
Pacific Peoples	16 (4.8%)	41 (10.7%)
Other	10 (3.0%)	19 (5.0%)
	M (*SD*)	M (*SD*)
Age	34.2 (*11.0*)	33.5 (*10.7*)
Sentence length (days)	265.1 (*72.2*)	271.6 (*73.0*)
RoC*RoI score	.19 (*0.2*)	.34 (*0.20*)
Criminal history		
Number of previous convictions	14.7 (*11*)	22.1 (*22.4*)
Number of previous violent convictions	1.0 (*1.5*)	2.1 (*2.5*)
Number of previous imprisonments	.45 (*1.3*)	1.4 (*2.7*)

Appendix B Propensity score matching

B.1 *The Eight Matching Variables used for PSM*

1. Age at Sentence Commencement Date
2. Sentence length in days
3. Ethnicity
4. Index offence
5. RoC*RoI score
6. Total number of previous convictions
7. Total number of previous imprisonments
8. Total number of previous violent convictions

B.2 *PSM Logistic Regression Model*

Matching Variable	β (SE)	Wald (*df*)	Odds ratio [95% CI]
Index Offence			
Violent	-19.91 (15729.1)	.000 (*1*)	.000
Non-violent	-19.92 (15729.10	.000 (*1*)	.000
Justice/admin	-19.97 (15729.1)	.000 (*1*)	.000
Other	-40.91 (32134.1)	.000 (*1*)	
Unknown		.03 (*4*)	
Ethnicity			
Māori	-1.32 (.45)	8.62 (*1*)	.267* [.110, .644]
European	-.98 (.45)	4.77 (*1*)	.375* [.156, .905]
Pacific Peoples	-.13 (.54)	.059 (*1*)	.877 [.306, 2.518]
Other		18.2 (*3*)	
Age	.02 (.01)	3.34 (*1*)	1.02 [.999, .1035]
Sentence Length (days)	.00 (.00)	.63 (*1*)	1.00 [.999, 1.003]
RoC*RoI	5.62 (.71)	61.8 (*1*)	274. 61* [67.72,1113.58]
Previous convictions	-.04 (.01)	18.04 (*1*)	.962* [.945, .979]
Previous violent convictions	.26 (.06)	17.23 (*1*)	1.298* [1.147, .1.468]
Previous imprisonments	.16 (.01)	3.39 (*1*)	1.175 [1.023, 1.349]

Note:
Model $\chi^2(1, N = 719) = 176.67, p < .001$

Appendix C Predictive validity of the proximal DRAOR and the non-significant initial DRAOR scores for women and men

C.1 *Multivariate Model of the Proximal DRAOR Scores and Non-significant Initial DRAOR Scores Predicting the Likelihood of Breaches*

	Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Initial + proximal stable [a]	-0.13	1.34	0.88	
	(0.12)	4.15*	[0.70, 1.10]	0.65*
	0.19		1.21	[0.53, 0.77]
	(0.09)		[1.01, 1.45]	
Initial + proximal protective [b]	0.06	0.22	1.06	
	(0.12)	5.33*	[0.84, 1.33]	0.68**
	-0.24		0.79	[0.57, 0.79]
	(0.10)		[0.65, 0.97]	
Initial + proximal TS [c]	-0.00	0.00	1.00	
	(0.04)	6.33*	[0.92, 1.08]	0.70**
	0.09		1.09	[0.59, 0.81]
	(0.04)		[1.02, 1.17]	

Note:
[a] Model $\chi^2(2, N = 187) = 4.15, p = .126$
[b] Model $\chi^2(2, N = 187) = 5.99, p = .050$
[c] Model $\chi^2(2, N = 187) = 7.48, p = .024$

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

C.2 *Multivariate Model of the Proximal DRAOR Scores and Non-significant Initial DRAOR Scores Predicting the Likelihood of Criminal Reoffending*

	Women				Men			
	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]	AUC [95% CI]
Initial + proximal acute	0.04 (0.06) [a] 0.08 (0.06)	0.37 1.76	1.04 [0.92, 1.17] 1.09 [0.96, 1.22]	0.60* [0.51, 0.68]				
Initial + proximal stable	-0.02 (0.08) [b] 0.16 (0.07)	0.06 5.47*	0.98 [0.85, 1.14] 1.17 [1.03, 1.34]	0.63** [0.55, 0.71]	-0.11 (0.11) [c] 0.22 (0.09)	1.08 6.39*	0.90 [0.73, 1.10] 1.25 [1.05, 1.48]	0.62* [0.52, 0.72]
Initial + proximal protective					0.04 (0.10) [d] -0.17 (0.08)	0.14 4.02**	1.04 [0.86, 1.26] 0.85 [0.72, 1.00]	0.60~ [0.50, 0.70]
Initial + proximal TS					-0.03 (0.04) [e] 0.11 (0.04)	.0.35 7.14**	0.98 [0.90, 1.06] 1.12 [1.03, 1.21]	0.64** [0.54, 0.74]

Note:
[a] Model $\chi^2(2, N = 175) = 3.45, p = .179$
[b] Model $\chi^2(2, N = 175) = 8.89, p = .012$
[c] Model $\chi^2(2, N = 175) = 7.47, p = .024$
[d] Model $\chi^2(2, N = 175) = 5.37, p = .068$
[e] Model $\chi^2(2, N = 175) = 10.99, p = .004$

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

Appendix D The RoC*RoI and breaches

D.1 *Correlation Matrix of Women's and Men's RoC*RoI and Initial and Proximal DRAOR TS*

	Women		Men	
	Initial DRAOR TS	Proximal DRAOR TS	Initial DRAOR TS	Proximal DRAOR TS
RoC*RoI	.17*	.25**	.30**	.25**

D.2 *Multivariate Model of the RoC*RoI Predicting the Likelihood of Breaches*

	Women[a]			Men[b]		
	β (SE)	Wald	Hazard Ratio [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]
RoC*RoI	0.01 (0.82)	0.00	1.01 [0.20, 5.00]	-0.21 (1.25)	0.03	0.81 [0.07, 9.34]
Initial TS	-0.01 (0.03)	0.03	1.00 [0.93, 1.06]	0.00 (0.05)	0.00	1.00 [0.92, 1.10]
Proximal TS	0.12 (0.03)	16.71***	1.12 [1.06, 1.19]	0.09 (0.04)	6.13*	1.09 [1.02, 1.17]

Note:
[a] Model $\chi^2(3, N = 187) = 25.79, p < .001$
[b] Model $\chi^2(3, N = 187) = 7.49, p = .058$

* $p < .05$ *** $p < .001$

Appendix E Multivariate model of the RoC*RoI and initial DRAOR TS predicting criminal reoffending

E.1 *Multivariate Model of the RoC*RoI and Initial DRAOR Predicting the Likelihood of Criminal Reoffending*

	Women[a]			Men[b]		
	β (SE)	Wald	Hazard Ratio [95% CI]	β (SE)	Wald	Hazard Ratio [95% CI]
RoC*RoI	1.82 (0.62)	8.68**	6.18 [1.84, 20.73]	3.12 (0.79)	15.54***	22.52 [4.79, 105.94]
Initial TS	0.05 (0.02)	3.80	1.05 [1.00, 1.10]	0.02 (0.03)	0.53	1.02 [0.96, 1.09]

Note:
[a] Model $\chi^2(2, N = 175) = 15.47, p < .001$
[b] Model $\chi^2(2, N = 175) = 21.08, p < .001$

** $p < .01$