Does Within-Treatment Change on Dynamic Risk Factors and Psychopathology Predict Recidivism for Men Graduating from Tai Aroha?

A thesis

submitted in fulfilment

of the requirements for the degree

of

Master of Social Sciences in Psychology

at

The University of Waikato

by

ARDEN EMERSON TANNER-DEMPSEY

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Supervisor: Professor Devon Polaschek

2023

Abstract

Dynamic risk factors are modifiable aspects of justice-involved people and their environment that can change over time and are associated with criminal behaviour. Because they are considered changeable, they are therefore typically the focus of rehabilitative treatment (Bonta & Andrews, 2024). Research shows that correctional rehabilitation that targets dynamic risk factors effectively reduces recidivism, and psychometric tools assessing these factors generally show improvements during treatment. However, there is a missing link; current research lacks consistent evidence linking changes on dynamic risk factors during treatment to recidivism. Relatedly, there are few studies assessing within-treatment changes in psychopathology and how they may affect recidivism, though they are commonly assessed alongside dynamic risk factors in correctional settings.

This study, involving 164 people with violent offence convictions who completed the Tai Aroha community treatment programme, aimed to address this missing link. A battery of tests measured pre- and post-programme included dynamic risk factor measures, the Anger Disorders Scale (ADS), Psychological Inventory of Criminal Thinking Styles (PICTS), Criminal Attitudes to Violence Scale (CAVS), and the Millon Clinical Multiaxial Inventory (MCMI-III and IV) which is designed to assess personality and clinical psychopathology. Standardised residual change scores were used to account for variance in pre-treatment scores, and Cox regression hazard ratios to test prediction of general and violent recidivism.

Most subscales for dynamic risk factor measures and the MCMI-III and IV showed substantial improvements during treatment, however, this did not translate to predicting the probability of recidivism. While several hazard ratios indicated effects in a positive direction, none were significant when adjusted for multiple comparisons, and controlling for actuarial risk did not improve predictive accuracy. There was some evidence of increased socially desirable responding which increased at the end of treatment, this included responses on the MCMI-III and IV consistent with the “Normal Quartet.” Interpretations of these findings and both theoretical and empirical implications are discussed.

Keywords: dynamic risk factors, recidivism, RNR model, violent offending, within-treatment change

Acknowledgements

First, I would like to extend my thanks to the tāne who have participated in the Tai Aroha programme, past and present, without whom this research would not be possible. Thank you to the clinical staff at Tai Aroha, in particular Glen Kilgour who helped with data emergencies and Emily Norman who supported me to attend a pōwhiri at the Tai Aroha whare and explain my research to the residents. Many thanks also to Ara Poutama Aotearoa for allowing me access to this valuable data. I would also like to acknowledge the financial support from scholarships provided by the University of Waikato, and from the Māori and Psychology Research Unit. These funds were a lifeline that allowed me to focus on my thesis and clinical work.

A huge thank you to the many women who have supported me through my academic journey. This includes the Crime Lab members, particularly Lisa, Apriel, Sam, and the indomitable Devon, I've learnt a great deal from each of you. Prof Devon: thank you for the excellent and sometimes ruthless feedback, your level-headedness in times of crisis, and the many opportunities and encouragement over the last few years. A version of you is permanently installed in my mind, reminding me to think like a scientist, use accurate language, and take breaks.

Most of all, my deepest appreciation is reserved for my dear partner Soph, and my unconditionally loving mother Trish. Your support has made my journey to becoming a clinical psychologist possible. Now that my thesis is complete, equality can return to the division of labour in the household.

Finally, no thesis is truly complete without acknowledging the dead pets along the way. Wesley and Ursula, thank you for faithfully keeping my lap warm while I was up late analysing data.

Table of Contents

List of Figures and Tables

Figures

Tables

Tables in the Appendix

Introduction

It is interesting to note that the task of establishing robust empirical support for dynamic risk factors, a fundamental concept used within correctional psychology, is referred to by some researchers as the 'search for the holy grail' due to their assumed function eluding strong empirical support (Baglivio et al., 2018; Serin et al., 2013). It is remarkable too, that a decade on from that term being employed, research is yet to come to a clear resolution. The current study aims to contribute to this research endeavour and broader efforts to improve the effectiveness of correctional rehabilitation for justice-involved people.

The following introduction presents an overview of extant literature related to dynamic risk factors with particular attention to their significance in rehabilitative treatment, and the questions researchers are addressing regarding their assumed causal relationship to recidivism. Beginning with a brief historical overview of correctional rehabilitation and the importance of risk measurement, we[1] then outline the development of the influential Risk Needs Responsivity (RNR) model and its theoretical components, highlighting dynamic risk factors as central constructs supported by a large evidence base. Following this we examine psychometric tools designed to assess dynamic risk factors and challenges associated with assessment, then explore methods researchers use to determine whether these factors change during rehabilitative treatment including an overview of common statistical approaches. Next, we provide an overview of extant literature exploring whether changes on dynamic risk factors predict reductions in recidivism, focusing on inconsistent findings in different justice-involved populations. Finally, we introduce Tai Aroha, the correctional rehabilitation programme where our research participants completed treatment. These sections aim to provide the reader with a general understanding of the evidence, theory, criticism, and importance of dynamic risk factors in correctional treatment, particularly regarding their presumed predictive role in recidivism.

[1] While the work in this thesis is entirely my own, the active 'we' pronoun acknowledges my being part of a research lab made up of a community of academics, and university students from whom I received advice and who informed my approach.

From 'Nothing Works' to 'What Works'

Throughout modern history, societies have used varied methods of reducing crime, often employing cruel and unusual punishments (Hollin, 2012). More compassionate methods have also been employed; for instance, in Britain the advent of rehabilitation for prisoners can be traced back to the 18th century, coinciding with broader social and welfare policy reform (Morris & Rothman, 1998). These methods can be considered attempts at criminal justice. However, until relatively recently, they have been used without a clear understanding of their effectiveness at truly reducing crime. As noted by Polaschek (2012) "the scientific study of criminal justice interventions has a short history", emerging only in the mid-20th century due to advancements in scientific methods, especially in data gathering and statistical analysis (p.1). Anstiss (2003) attributed the beginning of robust empirical research to a period in the 1970s described as a "fierce struggle between proponents of punishment and proponents of rehabilitation" (p. 84). On one hand, proponents of punishment argued for stricter sentences and the use of incarceration as a deterrent, emphasizing the need to protect society from criminal behaviour (sometimes through the use of capital punishment). On the other hand, proponents of rehabilitation advocated for a more humane and restorative approach, focusing on addressing the root causes of criminal behaviour and reintegrating justice-involved people into society as productive, law-abiding citizens. This debate intensified in the late 1970s as the faulty idea that rehabilitation was ineffective became prominent among academics and politicians across the Western world (Cullen, 2013). As a result, correctional practices trended towards more punitive methods, and investment in rehabilitative approaches decreased. This trend was influenced by Martinson's (1974) influential critique labelled 'Nothing Works!' which extolled the failure of correctional rehabilitation to deliver real reductions in offending. Following a review of 231 United States-based evaluations Martinson (1974) stated: "With few and isolated exceptions, the rehabilitative efforts that have been reported so far have had no appreciable effect on

recidivism[2]" (p.25). This article and others expressing similar sentiments (e.g., Von Hirsch et al., 1976) were widely read among policymakers, leading to a scaling down of many rehabilitative programmes (Anstiss, 2003; Fortune & Heffernan, 2019). Perhaps due to Aotearoa New Zealand's relative isolation, Martinson's ideas did not take hold as they did in other parts of the world, but they were still influential (Ward et al., 2006).

As Gendreau and Ross (1987) said in their review of offender rehabilitation "it's easier to be critical than be correct" (p.394); a reanalysis of Martinson's (1974) research some years later revealed large methodological and statistical flaws, and he himself would come "almost full circle" in his interpretation prior to his death in 1980 (Sarre, 2001, p. 41). This historic episode shows the importance of iterative empirical research that does not fall victim to overconfident claims. In response to the 'Nothing Works!' claim, the 'What Works' literature began to emerge with the aim of establishing an evidence base for the rehabilitative approach using robust statistical methods. By the late 1980's evaluation studies on offender rehabilitation gradually began to establish that rehabilitative treatment did, on average, reduce recidivism (Bonta & Andrews, 2017). A finding that has since been strongly established across different offender populations (e.g., Beaudry et al., 2021; Lipsey & Cullen, 2007; Papalia et al., 2019). The next step was to create a robust correctional rehabilitation framework, one which would use this growing evidence base to provide an explanation for why some interventions worked and others didn't. Beginning in earnest in the 1970s Canadian psychologists, James Bonta, Donald A. Andrews, and Paul Gendreau most notable among them, would take up this challenge. These psychologists began distilling variable theories of criminal offending, correctional treatment, and psychology, creating three[3] psychological principles: risk, needs, and responsivity (Polaschek, 2012). Bonta and colleagues' (1990b) subsequent meta-analytical review provided the opportunity to test the authors' newly theorised framework.

[2] Recidivism is a binary outcome (i.e., people either do or do not) and can refer to a reoffence, reconviction, or reimprisonment after a previous offence.

[3] The earliest published work on the RNR model included a fourth core principle, professional override, but this received little research attention and was subsequently dropped (Bonta & Andrews, 2024).

Comparing 154 interventions involving justice-involved youth and adults, Bonta and colleagues (1990b) found not only that criminal sanctioning (i.e., the punitive approach) actually had a negative effect on recidivism, but crucially, that rehabilitative treatments that conformed to their risk, needs, and responsivity principles significantly outperform those that do not[4]. Following this, the risk–need–responsivity (RNR) model was developed.

Risks-Needs-Responsivity

In 1994 the first edition of the Psychology of Criminal Conduct (PCC) outlining the RNR model was published; the RNR model emerged from the 'nothing works/what works' debate effectively synonymous with the 'what works' literature, becoming the foremost framework for offender rehabilitation used internationally (Polaschek, 2012). Since the initial 1990 meta-analysis, the databank underpinning RNR has expanded significantly with 225 studies and 374 tests included in the most recent seventh edition of PCC (Bonta & Andrews, 2024). The model has been subjected to robust and consistent critique (e.g., Fortune & Heffernan, 2019; Newsome & Cullen, 2017; Polaschek, 2012) but has maintained its position as the premier rehabilitative framework. Of course, promising alternative rehabilitative frameworks such as the Good Lives Model (See Ward & Brown, 2004) and Desistance model (See Maruna & Farrall, 2004) are used in some jurisdictions, but the RNR model continues to hold a dominant position in North America, Australia, and here in Aotearoa New Zealand (Cullen, 2013).

Before we move to literature more specific to our research, a brief overview of the core ideas of RNR is worthwhile. First, an explanation of the model's three eponymous principles: *risk*, *needs*, and *responsivity*. The *risk* principle states that the intensity and focus of treatment should be matched with the risk level of the offender, the *need* principle (more on this later), that treatment should target changeable factors which are associated with reductions in recidivism, and

[4] The most recent results provided by Bonta and Andrews (2024) report criminal sanctions (e.g., imprisonment, fines, or community service) associated with a minor increase in recidivism (r = −0.03) compared to a decrease for treatment with human services (r = 0.12). This effect increases (r = 0.28) when services fully adhere to RNR principles (p.31).

responsivity, which includes both general and specific levels. General responsivity holds that treatment should utilise evidence-based treatment strategies (e.g., cognitive-behavioural techniques, and skill building). Specific responsivity emphasises the adaptation of treatment to the individual needs of the offender, considering factors such as ethnicity, gender, personal strengths, and motivations, to best affect change.

The most recent edition of PCC includes 12 additional principles (15 in total) which cover three domains: *core* (key clinical principles); *overarching* (treatment is most effective when psychologically informed); and *organizational* (community-based programmes with supportive management are best). While the *need* principle which directly relates to dynamic risk factors is most relevant to this study, the *Structured Assessment* principle[5] (principle 11) is also applicable, specifically part a) "…employ structured and validated assessment instruments" (Bonta & Andrews, 2024, p. 187). This principle is important as our research has a strong focus on the predictive validity of several frequently used assessment tools.

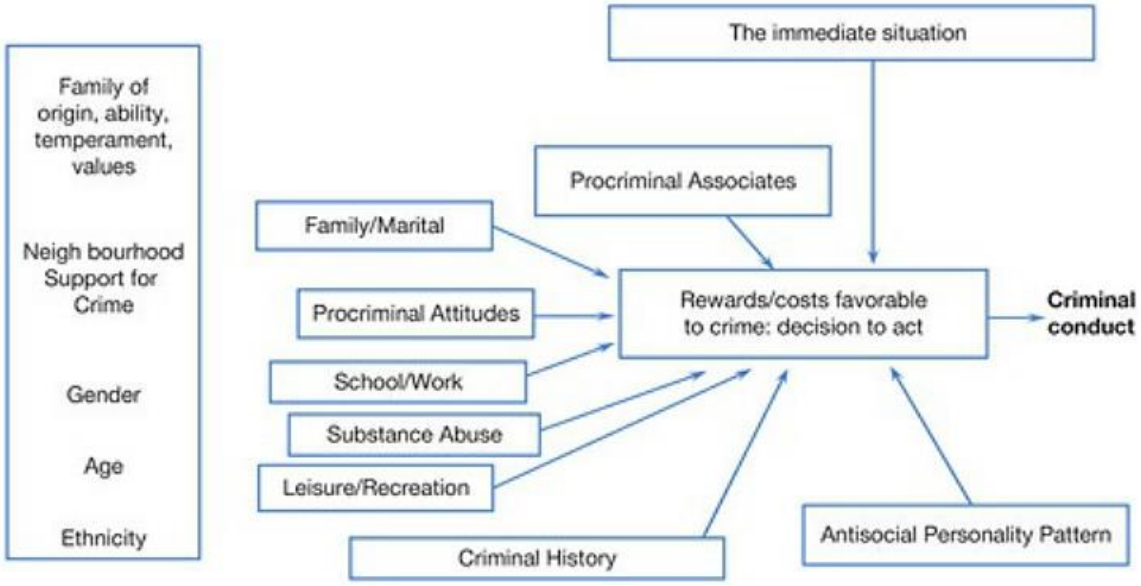
Underlying the RNR principles set out by Bonta and Andrews (2017) is their General Personality and Cognitive Social Learning (GPCSL) theory, which can be thought of as the "body of empirical, theoretical, and practical work" of the RNR model (Polaschek, 2012, p. 3). As the title suggests, personality factors (e.g., impulsiveness, hostility) and social learning (e.g., modelling, imitation, and reinforcement of behaviour) are key components in the propensity for and maintenance of criminal conduct (the GPCSL can be viewed in schematic form in Figure 1). While the model considers many factors related to offending, it places particular salience on the 'Central Eight' risk factors which are statistically linked to offending behaviour (Bonta & Andrews, 2017). These Central Eight risk factors or *criminogenic needs* include: 1) Criminal History, 2) Procriminal associates (or companions), 3) Procriminal Attitudes, 4) Antisocial Personality Patterns, 5) Education/Employment, 6) Family/Marital, 7) Substance Abuse, and 8) Leisure/Recreation.

[5] Principle 11 emphasises the *assessment* of treatment related factors rather than *reassessment* which is a primary focus of this study.

According to Bonta and Andrews (2016), “chances of illegal conduct for an individual increase dramatically as the number of the Central Eight factors increase” (p. 54). Seven of these factors, except Criminal History, are characterised by their changeability and are therefore considered dynamic risk factors (more on this shortly). Criminal history is a static risk factor alongside other variables such as ethnicity, sex, and age, which are related to offending but tend to be immutable and therefore have more limited treatment utility (Bonta & Andrews, 2017).

Figure 1

A General Personality and Cognitive Social Learning Perspective (Bonta & Andrews, 2017)



Dynamic Risk Factors

To say that dynamic risk factors are important to the RNR model and correctional psychology more broadly would be an understatement; Fortune and Heffernan (2019) argue they are “the most widely used concept in forensic and correctional practice” (p.662). When measured dynamic risk factors are used by clinicians as targets of rehabilitative treatment, the evidence that determines risk status, and importantly, they can inform sentencing and parole decisions including release from prison (Beggs, 2010; Serin et al., 2016). So, what exactly are dynamic risk factors? They are modifiable aspects of justice-involved people and their environment that can change over time and are associated with criminal behaviour and changes in the risk of recidivism (Douglas & Skeem,

2005). Because they are considered changeable they are therefore typically the focus of psychological treatment that adheres to RNR principles (Andrews et al., 1990), with research finding that programmes that target dynamic risk factors are more effective at reducing

Table 1

Seven Major Dynamic Risk Factors with Non-criminogenic Need Factors for Reference

Major Dynamic risk factors/ Criminogenic Needs	**Indicators**	**Intervention goals**
Antisocial personality pattern	Impulsive, adventurous pleasure seeking, restlessly aggressive and irritable	Build self-management skills, teach anger management
Procriminal attitudes	Rationalisations for crime, negative attitudes towards the law	Counter rationalisations with prosocial attitudes; improve prosocial identity
Social supports for crime	Criminal friends, isolation from prosocial others	Replace procriminal friends and associates with prosocial friends and associates
Substance abuse	Abuse of alcohol and/or drugs	Reduce substance abuse, enhance alternatives to substance use
Family/marital relationships	Inappropriate parental monitoring and disciplining, poor family relationships	Teaching parenting skills, enhance warmth and caring
School/work	Poor performance, low levels of satisfaction	Enhance work/study skills, nurture interpersonal relationships within the context of work and school
Prosocial recreational activities	Lack of involvement in prosocial recreational/leisure activities	Encourage participation in prosocial recreational activities, teach prosocial hobbies and sports
Non-criminogenic needs		
Self-esteem	Poor feelings of self-esteem, self-worth	Promote self-compassion and consistent self-esteem
Psychological distress	Anxious, feeling blue	Enhance coping skills and positive emotions
Major mental disorder	Schizophrenia, bipolar disorder	Relapse management and psychoeducation
Physical health	Physical deformity, nutrient deficiency	Increase nutritional balance, including meal planning and dieting

Adapted from (Bonta & Andrews, 2007)

recidivism than those that do not (Baglivio et al., 2018; Bonta & Andrews, 2017; Smith et al., 2009). dynamic risk factors, which are sometimes referred to using the term *criminogenic needs*, are contrasted with non-criminogenic needs. Non-criminogenic needs (e.g., mental illness, or self-esteem) are also changeable and may be a secondary focus of treatment but are not associated with reductions in criminal behaviour for most people[6] (Bonta & Andrews, 2017). Some non-criminogenic needs may therefore be better conceptualised as responsivity factors in the RNR model such that treatment may be adjusted based on these factors to enhance engagement (Kingston et al., 2015). A list of the 7 major dynamic risk factor variables linked to offending adapted from Bonta and Andrews (2007) can be viewed in Table 1. This includes indicators for each dynamic risk factor and intervention goals, as well as non-criminogenic factors for reference.

A concept closely related to dynamic risk factors, and applied in psychological treatment literature more broadly, is *psychological mechanisms*. Psychological mechanisms "offer a type of causal explanation of mental states and behaviour, often with reference to underlying processes, systems, activities, or entities" (Koch & Cratsley, 2020, p. 4146). Common examples of psychological mechanisms include worry maintaining anxiety symptoms (Borkovec et al., 2004), or avoidance behaviours maintaining depressed mood (Jacobson et al., 2001). While not suggesting that dynamic risk factors currently meet standards to be considered psychological mechanisms, they can be understood as factors similar to psychological mechanisms that maintain criminal behaviour rather than maintaining psychopathology. There may also be some overlap between the two concepts.

The primary evidence for the association of dynamic risk factors and recidivism provided by Andrews and Bonta (2017) in their PCC textbook is a large-scale meta-analysis of Level of Service Inventory (LSI) scales by Olver and colleagues (2014). A total of 137,931 justice-involved people across 128 different studies using various LSI scales between 1981-2012 were included in the study. The LSI scales adhere to RNR principles and include measurement of the central eight risk factors

[6] This is a general finding and that does not apply to all justice-involved people. Perhaps a better way to put this is that programmes that predominantly target non-criminogenic needs are not associated with reduced recidivism for most people.

and additional dynamic risk factors (Bonta & Andrews, 2024). The study showed that overall, each of the Central Eight based dynamic risk factors predicted both general and violent recidivism, although the scales were less consistent for predicting sexual recidivism. Overall, effect sizes were mostly small, with substance abuse and criminal associates the strongest predictors for both general and violent recidivism. The authors noted considerable heterogeneity in the magnitude of effects between studies and samples were mostly North American which may limit the generalisability of these findings to other populations. One New Zealand sample was included which supported the general findings but specifically pertained to justice-involved youth. Additionally, the family of LSI scales, which include the LSI-Revised (Andrews & Bonta, 2000) and the Youth level of service/case management inventory (YLS/CMI; Hoge & Andrews, 2011) are closely affiliated with the developers of the RNR model which raises questions about possible allegiance effect[7] and/or publication bias[8]. The majority of researchers who contributed to Olver and colleagues' meta-analysis (2014) appear to be independent, so this is unlikely to be a serious issue. However, exploring alternative forms of evidence and varied approaches to operationalizing the measurement of dynamic risk factors remains valuable.

A recent meta-analysis by Yukhnenko and colleagues (2020) analysed the relationship between dynamic risk factors and recidivism involving 15 studies and 246,608 adults on community sentences (community sentences tend to be given for less serious offences). Each of the included dynamic risk factors (referred to as modifiable risk factors) included in the study was associated with recidivism. These included six general dynamic risk factors such as substance misuse and employment problems which broadly mapped on to central eight risk factors. The authors concluded that the strength of the association of dynamic risk factors and recidivism was equivalent to static risk factors (e.g., age, gender, criminal history).

[7] Allegiance effect refers to a potential bias in research due to a researcher's alignment with a particular intervention, instrument, or theory (Boccaccini et al., 2017).

[8] Publication bias refers to when certain outcomes such as nonsignificant findings influencing the likelihood of a study's publication. There is some evidence of publication bias in criminal justice research (Kim, 2022).

A similar meta-analysis by Eisenberg and colleagues (2019) looked at dynamic risk factors and recidivism for forensic outpatients with a combined sample of 116,982 justice-involved adults. Forensic outpatients live in the community either under intensive supervision and/or actively receiving rehabilitative outpatient treatment as a condition of their sentence. Eisenberg and colleagues (2019) specifically examined central eight risk factors and provided additional support for Bonta and Andrews' (2017) RNR model, finding that each was associated with both general and violent recidivism. Other factors included in the analysis, self-esteem and non-cooperation with treatment were not associated with general recidivism, and financial problems were not associated with violent recidivism. Finally, Assink and colleagues (2015) conducted a meta-analysis on dynamic risk factors and different offending trajectories among justice-involved youth. The outcome of interest was whether these youth either engaged in adolescent-limited offending, desisting from crime as they aged, or a life-course persistent pattern of offending. All dynamic risk factors included in the study were associated with life-course persistent delinquency with the largest effect found for Aggression. A separate study with 505 adolescents found that poor anger management and low empathy predicted both general and violent recidivism (Lawing et al., 2017).

Antisocial Personality Pattern is an identified Central Eight dynamic risk factors consistently linked with recidivism (Bonta & Andrews, 2017; Eisenberg et al., 2019; Olver et al., 2014c; Yukhnenko et al., 2020). APP, which describes a pattern of personality and behaviour, is distinct from Antisocial Personality Disorder, which is a diagnosed disorder, but there is meaningful overlap. There is some evidence that other personality disorders and constructs may also relate to offending behaviour and are worth outlining. However, while mental disorders are understood to be prevalent among justice-involved people, as previously stated, mental disorders are generally considered non-criminogenic needs as they are considered not associated with reoffending (Bonta & Andrews, 2024). However, in a meta-analysis of justice-involved people with mental disorders[9], Bonta and

[9] In general, having a mental disorder does not predict recidivism any better than not having one (Bonta et al., 2014; Kingston et al., 2015).

colleagues (2014) found that having any personality disorder diagnosis predicted both general and violent recidivism almost as effectively as antisocial personality disorder. A similar pattern was also observed among people with sexual offence convictions (Kingston et al., 2015); however, both studies used a general personality disorder variable, creating ambiguity regarding the connection between specific personality disorders and recidivism. There is some evidence that borderline personality disorder is also associated with violent recidivism (Biedermann et al., 2023). Using the Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon et al., 1997) personality assessment, Antisocial personality has been linked to recidivism (Carbajosa et al., 2017; Molina-Coloma et al., 2022) as well as Borderline, Aggressive-sadistic (Molina-Coloma et al., 2022), Psychopathology, and Antisocial/narcissistic personality profile clusters (Sissons, 2013). In a meta-analysis of studies using the Personality Assessment Inventory (PAI; Morey, 1991), Gardner and colleagues (2015) found that while Antisocial Features, Aggression and the Violence Potential Index were the strongest predictors of recidivism and violent behaviour, Borderline features and Dominance were also somewhat associated.

Assessing Dynamic Risk Factors

In the previous section, we focussed on dynamic risk factors as variables that are theoretically and, importantly, statistically linked to recidivism. We now turn to the psychometric tools that are commonly used to measure them. The wide variety of tools used to measure dynamic risk factors has been noted by multiple authors (e.g., Douglas & Skeem, 2005; Klepfisz et al., 2016; Papalia et al., 2020), with over 400 different tools measuring the risk of violence and offending more broadly (Singh et al., 2014). Yet these tools are similar in many ways. The sources of information that establish dynamic risk factors scores tend to be either self-reported by justice-involved people or based on clinician observations and are typically represented by numerical scores suitable for statistical analysis (Polaschek, 2019). Researchers have called attention to issues in self-reported data such as social desirability bias (Beggs, 2010; Cording et al., 2016; Heffernan et al., 2019); which will be discussed later in this section. But there are also potential issues with clinician-rated

measures. For instance, perceptions of change may be influenced by confirmation bias when the same raters conduct both pre-and post-treatment assessments and are aware of the treatment details (Papalia et al., 2020). There may also be possible bias due to clinicians having personal involvement with individuals and their treatment outcomes (Yesberg & Polaschek, 2019).

Many of the prominent dynamic risk factor measurement tools are considered third-generation risk assessment tools, focussing on measuring dynamic risk factors alongside static factors and emphasising person-specific treatment needs (Cording et al., 2016). There are four generations of risk assessment tools, which have evolved alongside new evidence and theory (for a review see Olver & Wong, 2019). The first generation of tools relied on unstructured clinical judgement which tended to introduce bias and poor reliability between assessors, but subsequent generations introduced actuarial risk measurement based on static offence data, providing a more objective statement of risk (Mills et al., 2011). Third-generation tools assess both actuarial and dynamic factors that are used to inform treatment targets, while the most recent fourth-generation tools are designed to link assessment data with treatment, guiding case formulation and risk management efforts (Olver & Wong, 2019).

A number of commonly used risk assessment tools specifically assess Central Eight dynamic risk factors domains (Olver & Wong, 2019); in fact, de Ruiter and Hildebrand (2022) argue that the mainstream correctional paradigm strongly encourages the use of assessment tools that include Central Eight risk factors due to their empirical status. The two most common measurement instruments used to assess dynamic risk factors do just that, these are the previously discussed Level of Service Inventory scales (e.g., LSI-R Andrews & Bonta, 2000) which are said to be the most used risk assessment tools internationally (Bonta & Andrews, 2017), and the Violence Risk Scales (VRS; Wong & Gordon, 1999-2003). Both tools are considered third generation. The LSI-R includes ten different dynamic risk factors, mapping closely onto central seven risk factors outlined by Bonta and Andrews (2024), while the VRS includes 20 more varied dynamic risk factors. Other composite measures include the Short-Term Assessment of Risk and Treatability (START; Webster et al., 2006),

the Violence Risk Appraisal Guide (V-RAG; Harris et al., 1993), the Historical Clinical Risk Management-20 (HCR-20; Douglas et al., 2014), and the Dynamic Risk Assessment of Offender Re-entry (DRAOR; Serin et al., 2007) which is used by Ara Poutama Aotearoa to assist in monitoring people on community sentences (Muirhead, 2016). Most of the tools have been found to perform roughly equivalently at predicting recidivism (Yang et al., 2010), and there is some evidence that random combinations of items provide similar accuracy to formally constructed tools. An innovative study by Kroner and colleagues (2005) compared the predictive accuracy of four risk assessment tools including the PCL-R, LSI-R, and VRAG with 'Coffee Can[10] measures' which each included 13 items randomly derived from the combined measures, finding no statistical difference between them. The authors argued that improvements were needed in the conceptualisation of risk assessments and the specific constructs they purport to measure.

Another method of assessing dynamic risk factors uses measurement tools that measure individual dynamic risk factor constructs, often in greater depth. These tools tend to be self-reported rather than clinician-administered, and some may measure dynamic risk factors incidentally i.e., they tap into dynamic risk factor constructs without this being their stated purpose. While a comprehensive review of these measures is beyond the scope of this study, examples include the Anger Disorders Scale (ADS; DiGiuseppe & Tafrate, 2004) or Novaco Anger Scale (NAS; Novaco, 1994) to measure anger-related problems, the Criminal Sentiments Scale (CSC; Andrews & Wormith, 1984) for procriminal attitudes, and relatedly the Psychological Inventory of Criminal Thinking scale (PICTs; Walters, 1995). Other tools like the Criminal Attitudes to Violence Scale (CAVs; Polaschek et al., 2004) or Rape Myth Acceptance Scale (MRA; Burt, 1980) assess dynamic risk factors related to specific types of offending (violent offending, and sexual offending, respectively). Personality assessment tools are commonly used alongside dynamic risk factors tools in correctional settings (Edens et al., 2018). These include the Minnesota Multiphasic Personality Inventory-2 (MMPI-2;

[10] The Coffee Can method is sometimes used metaphorically to describe a statistical process which randomises items from various tools. However, the method used by Kroner and colleagues (2005) used items written on paper that were "shaken vigorously" in an empty 1 kg coffee container (p.364).

Butcher et al., 2001), Personality Assessment Inventory (PAI; Morey, 1991) and Millon Clinical Multiaxial Inventory-IV (MCMI-IV; Millon et al., 2015).

Often dynamic risk factor tools are used within a battery of tests informed by an offender's case formulation (Cording et al., 2016). These batteries can comprise multiple tests and be difficult to interpret. Consequently, some researchers (e.g., Allan et al., 2007; Beggs & Grace, 2010; Olver et al., 2014a) have used statistical data reduction techniques such as Principal Component Analysis (PCA) to summarise and interpret data which can pertain to overlapping dynamic risk factor constructs.

Assessment issues

A major consideration for risk assessment tools is their cross-cultural validity. This is an important issue in Aotearoa New Zealand, where tools developed overseas may not be specifically validated, particularly for Māori who are overrepresented in the criminal justice system[11] (Ara Poutama Aotearoa New Zealand Department of Corrections, 2023; Statistics New Zealand Tatauranga Aotearoa, 2023). While the Central Eight dynamic risk factors domains have been identified across cultures (Gutierrez et al., 2017; Olver et al., 2014c), there is evidence that dynamic risk factor assessment tools are more accurate in predicting recidivism among European individuals than other ethnic groups including indigenous populations (Olver et al., 2014c; Singh et al., 2011; Wilson & Gutierrez, 2014). Conversely, recent research in Aotearoa showed stronger predictive accuracy for Māori justice-involved people (Coulter et al., 2023), and Olver and colleagues (2014c) found that ethnic minorities scored higher than non-minorities on most Central Eight subscales, though this did not equate to differences in predictive accuracy. Cross-cultural issues have led some researchers to deem dynamic risk factor measurement tools potentially biased against some minority populations and not appropriate for Indigenous populations globally (Allan et al., 2006; Schmidt et al., 2020, 2022; Shepherd & Anthony, 2018) or in Aotearoa New Zealand (Webb, 2018)

[11] Māori make up 52.8% of the prison population but 17.3% of the National population.

citing their Eurocentric and individualistic assumptions. While work clearly needs to be done to improve the cross-cultural validity of risk assessment tools, there is some consideration of these issues by practitioners and researchers. For example, the *Code of Ethics For Psychologists Working in Aotearoa/New Zealand* (2002) and the *American Psychological Association Ethical Principles of Psychologists and Code of Conduct* (2002) require psychologists who are using assessment tools not validated on the tested population to acknowledge these limitations when reporting on them.

Assessing impression management and socially desirable responding in justice-involved populations is important for putatively mitigating response bias and ensuring the integrity of psychological assessments, particularly for self-report measures (Hildebrand et al., 2018). Common measures include the Paulhus Deception Scale (PDS; previously known as the Balanced Inventory of Desirable Responding – BIDR; Paulhus, 1998) and the Marlowe-Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960). Personality assessments such as the MCMI-3 (Millon et al., 1997) also include validity scales designed to assess response bias like portraying themselves overly positively (desirability), overly negatively (debasement), or in an open manner (disclosure).

Research examining the impact of impression management on the measurement of dynamic risk factors has found mixed results with divergent interpretations. A meta-analysis looking at studies using the PDS/BIDR scales (Paulhus, 1998) found higher scores on impression management and self-deception were associated with lower scores on dynamic risk factors (Hildebrand et al., 2018). The authors propose two different interpretations of this finding: either increased impression management and self-deception reduces dynamic risk factors or respondents who make fewer attempts to appear socially desirable may indicate a "disregard for social conventions rather than response bias" (Hildebrand et al., 2018, p. 166). The latter point may relate particularly to people with antisocial personality or psychopathy who tend to violate social norms more readily than others. This accords with Mills and Kroner's (2006) finding that high impression management scores were associated with lower antisocial attitudes among people with violent offence convictions, and Brown and colleagues (2009) who found that people who scored higher on impression management

were less likely to recidivate in the follow-up period. Stevens and colleagues (2016) also found that socially desirable responding was associated with lower scores on dynamic risk factors and was unrelated to sexual recidivism among people with sexual offence convictions. Another study found that impression management increased significantly as treatment progressed (Grant et al., 2004). These findings support a notion posited by Uziel (2010) that evidence of socially desirable responding among offender populations may indicate a positive trait which Uziel refers to as 'interpersonally oriented self-control'. However, other authors see increased impression management following treatment as understandable efforts to showcase treatment success and secure benefits like parole that can introduce bias to findings (Heffernan et al., 2019; Juarez & Howard, 2018). They suggest that these variables should be controlled for in calculations. Clearly, there are several different interpretations and somewhat mixed findings on how impression management may relate to the assessment of dynamic risk factors. In any case, these variables provide valuable insights beyond what is measured by factors directly targeted in treatment.

Within-treatment Change

In light of the central significance of dynamic risk factors, a large amount of scholarship has been directed toward both conceptual and empirical criticism. In 2016, the prominent forensic psychology journal *Psychology, Crime & Law* dedicated a double issue to the question: "Dynamic Risk Factors: What role should they play in the explanation, assessment and rehabilitation of offenders" (Ward & Fortune, 2016). In an editor's introduction to the special issue (later released as an edited book) Gannon (2016) stated that dynamic risk factors "appear to have shaped worldwide assessment and treatment practices in the absence of significant constructive critique" (p.1). In short, most critiques (e.g., Cording et al., 2016; Heffernan et al., 2019; Thornton, 2016; Ward, 2016) have concentrated on the *causal* status dynamic risk factors are assumed to have concerning recidivism, which researchers argue is not empirically justified.

A key distinction made is between the concepts of prediction and causality. Hanson and colleagues (2009) explain the necessary steps for causality to be established: each factor must a)

predict recidivism b) be capable of change, and c) that change must be demonstrated to change recidivism outcomes. It is the latter two points regarding change on dynamic risk factors which have not been well established (Douglas & Skeem, 2005; Heffernan et al., 2019). Instead of assessing change, most research statistically linking dynamic risk factors to recidivism has examined them at a single point in time, essentially treating them as static factors and providing no information about whether dynamic risk factor change over time and if that change, as the Needs principle suggests, leads to a lower likelihood of recidivism (Serin et al., 2016).

So, how have researchers attempted to tackle these questions? First, in order to establish evidence of change, dynamic risk factors must be *reassessed*. This means that scores must be measured more than once over a period of time where change can realistically occur. The difference between scores on these time points represents dynamic risk factor change. This is generally referred to as *intermediate treatment outcome* (Papalia et al., 2020), *intra-individual change* (e.g., Serin et al., 2013), or *within-treatment change* (e.g., Beggs, 2010). Within-treatment change is the preferred term for this study, as our focus is on changes occurring during rehabilitative treatment. Common practice when estimating within-treatment change (usually to measure the effectiveness of an intervention) is to assess people twice, once at intake and once upon graduating from a treatment programme (Polaschek, 2017), but a small number of studies have monitored change on three or more occasions (e.g., Brown et al., 2009; Hanson et al., 2007; Yesberg & Polaschek, 2014). However, collecting data for these multi-wave studies tends to have practical constraints making it uncommon.

A scenario provided by Polaschek (2019) to demonstrate dynamic risk factors within-treatment change is an intervention for controlled drinking aimed at reducing alcohol abuse for an offender (an identified Central Eight dynamic risk factor). The chosen measurement tool may measure relevant dynamic risk factors before and after the intervention to examine whether the offender learned relevant skills to reduce alcohol consumption, and to what extent, as well as possible changes in attitudes toward binge drinking and associated problematic behaviour. This

within-treatment change data may then be used to estimate the likelihood that an offender will commit an alcohol-related reoffence. Evidence of reduced dynamic risk factors during treatment can also potentially result in parole decisions such as early release from prison (Beggs, 2010; Polaschek, 2019), or reduced risk state (Douglas & Skeem, 2005).

Measuring Change

The method of estimating within-treatment change used in the majority of studies uses the raw score difference between two time points. The magnitude of these change scores tends to be specific to respective scales and these scores are usually unstandardized (Olver et al., 2015). This means that change scores will differ depending on the tool unless additional statistical manipulation is performed. Researchers such as Beggs (2010) and Serin and colleagues (2013) note the importance of controlling pre-treatment dynamic risk factors scores which can account for differences in baseline risk that individuals had before any intervention occurred. For example, high-risk individuals may have very high scores at the beginning of treatment and therefore have the "greatest scope for improvement" (Daffern et al., 2019, p. 398). Controlling these pre-treatment variables means that an observed effect can be more confidently attributed to change, rather than pre-existing differences. This approach is also important as it avoids treating someone with more change but a higher post-treatment score as lower risk compared to someone with no change but a lower score.

A more sophisticated method for achieving this is using residualised change scores (RCZ), which is preferred by a growing number of researchers who have generally found promising (e.g., Beggs & Grace, 2011; Higgs et al., 2020; Olver et al., 2014b; Wakeling & Barnett, 2014). The residualised change scores represent the portion of change that is not explained by pre-treatment scores, aiming to isolate the true change in the variable of interest. Simply put, using residuals derived from linear regression that represent the unexplained portion of change, RCZ estimate what observed change would be if everyone in the sample had equivalent scores at the beginning of treatment (Daffern et al., 2019). Alternative methods to change scores include *Clinically significant*

change, and the *Reliable Change Index* (e.g., Higgs et al., 2020; Mastromanno et al., 2018). Clinically significant change (CSC) is usually granted if treatment targets are reached or if scores move from the 'dysfunctional' range to a 'functional' range (Polaschek, 2017). Few studies have used the clinically significant change method when analysing dynamic risk factors, and results have generally been nonsignificant (Barnett et al., 2013; Klepfisz et al., 2014). The Reliable Change Index (RCI) method suggested by Jacobson and Truax (1992) uses standard deviations as a standardised way of assessing the magnitude of change. Again, studies using this method are relatively rare and have not yielded compelling results (e.g., Higgs et al., 2020; Mastromanno et al., 2018).

After assessing change, the next step involves using statistical methods to test if changes on dynamic risk factors can be linked to reductions in recidivism. As Brown and colleagues (2009) stated, "Establishing a gold standard for the statistical analysis of change data in relation to recidivism is one of the most formidable obstacles facing correctional research" (p. 42). Despite advancements and some increased consistency in statistical approaches, a universally accepted 'gold standard' remains elusive in the field. However, several predictive statistical approaches are used most commonly, each with strengths and weaknesses that are more appropriate for certain conditions.

There are broadly two types of suitable methods: those that assess *discrimination*, examining how well predictor variables discriminate between binary outcomes (i.e., recidivism or not); and *calibration*, which compares expected results (i.e., from a scales published norms) and observed scores (Helmus & Babchishin, 2017). Methods of discrimination are mostly used in research assessing dynamic risk factors and recidivism and commonly used methods are worth outlining in some detail (Singh et al., 2011). Area Under the Curve analysis (AUC; Mossman, 1994) measures discriminatory accuracy and is used most commonly due to its interpretability and consistency with other areas of research like behavioural and medical research (Casey, 2016; Singh, 2013). Regression models are also frequently used, with Logistic regressions (McCullagh & Nelder, 1989), used to predict the probability of recidivism, and Cox Proportional Hazards Models (referred to as Cox

regressions; Cox, 1972), are used to predict the hazard of an event (i.e., recidivism) occurring at a specific time point within a defined period, making them particularly relevant for studies with temporal considerations (Hanson, 2022). An advantage of regression models is that they allow covariates (e.g., static risk factors) to be controlled, which is not a feature of AUC analysis. In summary, there are a range of statistical methods which are suitable for different data and research aims. Therefore, it may be more realistic for researchers to aim for consensus on a set of gold-standard approaches, rather than a single method.

Evidence Linking Change to Recidivism

An important study, which refers to intra-individual change on dynamic risk factors as the 'holy grail' due to its importance to the field of criminal justice research, is a meta-analysis conducted by Serin and colleagues (2013). Overall, the review of 53 studies involving mostly justice-involved men provided some marginal support that change on dynamic risk factors may predict recidivism. Only 23 studies reported both change scores and recidivism and just 27% of reported effect sizes in these studies were significant. An earlier review of 22 studies by Beggs (2010) looked at within-treatment change in men with sexual offence convictions. At this time only three reports (two from the Kia Marama treatment programme) included recidivism as an outcome variable and overall results were inconclusive. Both Beggs (2010) and Serin and colleagues (2013) noted the paucity of research and inconsistent findings, a claim repeated by many subsequent researchers (e.g., Juarez & Howard, 2022; Papalia et al., 2020).

There has been an increase in research in more recent years that has arguably demonstrated a somewhat more robust link between dynamic risk factor change and recidivism. Davies and colleagues (2023) attribute this, in part, to a growing number of published studies utilizing the VRS (Wong & Gordon, 1999-2003). The VRS and associated tools attend to RNR principles and are designed to measure change on dynamic risk factors before and after correctional rehabilitation treatment, making them ideal for this research. Several studies using the VRS and VRS: Sexual Offence version have found a consistent association between dynamic risk factor change and

reduced recidivism (e.g., Beggs & Grace, 2010; Coupland & Olver, 2020; Olver et al., 2021; Olver et al., 2013; Sowden, 2013). However, a smaller number of studies have found mixed results. Lewis and colleagues (2012) found that dynamic risk factor change scores predicted reduced recidivism but post-treatment scores performed approximately as well. Two studies, one including 82 men convicted of violent offences by O'Brien and Daffern (2017), and another on 123 high-risk justice-involved men by Yesberg and Polaschek (2019) both found that VRS change scores did not predict either violent recidivism or reimprisonment. Change on a separate measure of victim empathy was predictive in the former study (O'Brien & Daffern, 2017), but no dynamic risk factors was predictive in the later study (Yesberg & Polaschek, 2019). A more recent study by Higgs and colleagues (2020) assessing 227 men convicted of violent offences found that VRS change only predicted one of four recidivism outcomes (general offending at 3-year follow-up).

More recent review studies have focussed on specific offender populations. A meta-analysis by van den Berg and colleagues (2018) looked at 52 studies assessing dynamic risk factors and recidivism on men with sexual offence convictions that were published until late 2014. While only 12 studies examined dynamic risk factor change; overall, they predicted sexual recidivism, violent recidivism, and general recidivism, with small effect sizes. The authors noted that the observed changes in dynamic risk factors contributed only minimally to the overall reduction in recidivism, calling for more valid tools that measure the underlying correlates of offending. A study by Nunes and colleagues (2016) published after the inclusion date, involving 146 men with sexual offence convictions, found no association between change scores or pre- and post-treatment scores and recidivism using a measure of sexual offence supporting cognitions. Banse and colleagues (2013) reviewed 24 studies measuring pro-criminal attitudes and recidivism (a Central Eight dynamic risk factor). Most studies were on sex offender samples, with large variations in outcomes and methodologies and few that included change data. The authors concluded that overall, there was no solid empirical evidence that change on pro-criminal attitudes predicted recidivism. This is not necessarily against expectations, however; Bonta and Andrews (2024) suggest that reductions in

recidivism are more likely when multiple dynamic risk factors are targeted, rather than a single factor. A meta-analysis by Papalia and colleagues (2020) investigated dynamic risk factors and recidivism for men with violent offence convictions in 22 studies with control groups published before 2017. While change scores were not compared with recidivism outcomes in this study, averaged results showed small to moderate reductions on several dynamic risk factors that are known correlates of violent offending (trait anger, impulsivity, social problem solving, and general social skill), but a non-significant result for changes in antisocial cognitions.

Other research published since these reviews are worth briefly outlining. A novel study employing a more complex methodology was conducted by Baglivio and colleagues (2018). The study titled *The search for the holy grail* assessed 1678 mostly male justice-involved adolescents after release from residential treatment. Participants received treatment based on their three highest dynamic risk factors and were compared to a statistically matched sample who received nonspecific treatment. After statistically matching samples on demographic and initial scores, the dynamic risk factors focussed treatment group showed reduced recidivism compared to the matched controls (with a 17% relative difference). Though change scores were not compared with recidivism outcomes in this study, significant change was demonstrated on several dynamic risk factor variables for the treated group. Mastromanno and colleagues (2018) conducted a study of 48 justice-involved people treated in a secure forensic mental health service. Change on dynamic risk factors measured by the HCR-20 (Douglas et al., 2014) did not predict violent or general recidivism, with one dynamic risk factor concerning risk management worsening over the course of treatment. This contrasts with a prior study by De Vries Robbé and colleagues (2015) also using the same tool, which found a significant association. A large study by Howard and van Doorn (2018) that looked at antisocial attitudes among 1,858 justice-involved people who had completed forensic treatment programs also found no association with change and recidivism. Again, this contrasted with a previous study which found significant results using the same measure (Kroner & Yessine, 2013). Berman and colleagues (2019) assessed 776 justice-involved people involved in a CBT-based

programme using four measurement tools that broadly map onto dynamic risk factors. While all measures changed over the course of treatment, only positive changes on social skills were associated with reduced recidivism. Juarez and Howard (2022) assessed 2,337 men with violent offence convictions involved in a different CBT-based programme, this time assessing antisocial attitudes and criminal associates. Results were mixed, of the four antisocial factors measured, two (violence and antisocial intent) were significantly associated with general recidivism, but the other two (entitlement or antisocial associates) were not. No factors predicted violent recidivism.

Change on anger and aggression as a specific dynamic risk factor appears to be understudied given it is commonly addressed in rehabilitative programs (i.e., anger management interventions; Bonta & Andrews, 2017). This is somewhat surprising considering the increasingly established link between aggression and violent behaviour (See Coid et al., 2018; Hardin et al., 2022; Skeem et al., 2006). However, anger is a complex and multifaceted variable and within-treatment change may only be indirectly related to offending behaviour making it challenging to accurately detect potential changes during treatment. Serin and colleagues' (2013) meta-analysis found inconclusive results on eight reviewed studies that measured anger, with no clear link to recidivism outcomes. A study by Higgs and colleagues (2020) mentioned earlier on 227 justice-involved men who completed a violence prevention programme found that change on anger (as measured by the Aggression Questionnaire) predicted general recidivism at three years but not five years follow-up and did not predict violent recidivism. A small study by Klepfisz and colleagues (2014) on 42 justice-involved people found trait anger was not linked to recidivism and did not change significantly over the course of treatment. Authors from both studies noted that further research is needed to understand anger as a potential dynamic risk factor (Higgs et al., 2020; Klepfisz et al., 2014). However, current evidence suggests that improvements in anger during treatment do not result in meaningful reductions in recidivism.

Overall, while research continues to develop in a positive direction, there is a considerable amount more work needed in order to establish whether there is a consistent and robust link

between changes on dynamic risk factors and recidivism. It appears that the large variation in assessment tools complicates the ability to synthesise extant literature on the topic, with the predictive validity of some measures (particularly the VRS) outpacing other tools, particularly older or less sophisticated tools which may be less suitable to this research. Several authors have also highlighted the necessity for more research in community settings, given that the majority of studies have been conducted in prison environments (Baglivio et al., 2017; Davies et al., 2023; Lloyd et al., 2020).

The Current Research

This study uses data collected from men who have participated in Tai Aroha, a community treatment programme for men with violent offence convictions deemed at high-risk of reoffending and generally on home detention sentences (Kilgour, 2015). Tai Aroha is operated by the New Zealand Department of Corrections (Ara Poutama Aotearoa) Psychological Services and is located in a residential area in Hamilton, New Zealand. Tai Aroha runs over 16 weeks and includes approximately 300 hours of active psychological treatment (Batley et al., 2021). The programme uses a rolling group format which has a capacity of 10 residents, with residents at various stages of treatment at any one time. In 2010 the programme received the name 'Tai Aroha' (King, 2012); Tai Aroha in te reo (Māori language) can be translated to 'tide of love' reflecting the Kaupapa (purpose) of the programme which uses a supportive group-based approach with elements of tikanga (Māori cultural protocol). In staking the claim for the importance of such a programme Polaschek and Dixon (2001) stated:

> "There is much evidence that many forms of adult pathology relate to a history of abuse, especially at the hands of men. If one wanted to instigate one mental illness prevention program, then targeting male violence would possibly be the single most significant one" (p.1).

Perhaps the most straightforward description of Tai Aroha residents is provided by Kilgour (2015): "they are tough and tough to manage" (p.2). Residents must generally be deemed at high-

risk of both reconviction and reimprisonment (over .7 RoC*RoI[12]) and have a history of violent offences (King, 2012). However, in certain cases, men may be accepted with a lower risk level following a formal assessment with a psychologist (Kilgour, 2015).

The origins of Tai Aroha date back to 1987 when it operated under its previous iteration, the Montgomery House Violence Prevention Programme (Berry, 1998). The Montgomery House programme was formed as part of the Violence Prevention Project (VPP) which sought to explore psychological treatments to reduce recidivism among high-risk men with violent offence convictions (Polaschek & Dixon, 2001). But there have been major changes over the years. Tai Aroha has operated as a therapeutic community (see De Leon, 2000) with a foundation in social learning theory (Bandura, 1976) since 2010, using group-based processes (e.g., mutual self-help) to develop new skills and address addiction problems among residents in a supportive peer group (Kilgour, 2015; King, 2012). A more recent development has been the introduction of the cognitive self-change model (Bush et al., 2016) which is designed to reduce offending by targeting anti-social attitudes and 'risky' thinking (i.e., thoughts that may lead to offending) and promoting prosocial alternatives (Batley et al., 2021). Tai Aroha is strongly informed by the cultural needs of its residents, reflecting the 74-85% of participants who identify as Māori (Kilgour, 2015; King, 2012). Kaupapa Māori processes such as whakawhanaungatanga (establishing relationships), karakia (traditional prayer), and kapa haka (performance) are integrated into the programme, with residents encouraged to explore their cultural identity in dedicated modules (Batley et al., 2021).

Most relatedly to this study, the programme "attend[s] closely to the principles of Risk, Need and Responsivity", with the correctional framework overarching its various rehabilitation processes (Kilgour, 2015, p. 6). The RNR principles and theories set out by Bonta and Andrews (2017) are not only attended to by the Tai Aroha treatment protocol but underpin rehabilitative service delivery by Ara Poutama Aotearoa (New Zealand Department of Corrections) more broadly (Ryan, 2018).

[12]RoC*RoI (Bakker et al., 1999) is a composite risk measure of two metrics: RoC represents Risk of ReConviction, whereas RoI represents the Risk of Imprisonment. The asterisk (*) means 'multiplied by'. RoC*RoI is a risk assessment tool widely used by Ara Poutama Aotearoa.

Aims of this Study

As the title indicates, the general aim of this study is to answer the question: Does Within-Treatment Change on Dynamic Risk Factors and Psychopathology Predict Recidivism for Men Graduating from Tai Aroha? Our main focus is on examining dynamic risk factors in relation to recidivism, this builds on a large but overall inconclusive research base, with many researchers who have called for additional high-quality research (e.g., Eisenberg et al., 2019; Papalia et al., 2019; Serin et al., 2013). Our secondary focus is on psychopathology. We know relatively little about potential within-treatment change on measurement tools designed to assess psychopathology, which includes personality pathology, and how change on these factors may affect recidivism. Given that there is relatively little research regarding psychopathology, we did not want to be restrictive in selecting specific variables. We also acknowledge that it is uncommon that psychopathology is reassessed for groups within correctional rehabilitation programmes, this is evidenced by the limited research on the topic outlined previously. In line with this, our research questions investigating dynamic risk factors and psychopathology are broad. Though not explicitly stated as a research question, we also recognise the importance of assessing social desirability in correctional research and aim to integrate this analysis where feasible. We have four research questions that guide this study. The first two investigate within-treatment change, and the second two expand on that, looking at whether change scores predict recidivism:

1. Is there evidence of significant[13] within-treatment change on tools designed to measure dynamic risk factors?
2. Is there evidence of significant within-treatment change on a tool designed to measure psychopathology?

[13] Statistical significance is usually achieved when below a predetermined significance level which is conventionally set at $\alpha = .05$ (i.e., less than a 1 out of 20 chance). A result that meets this criterion means that the observed result is considered unlikely to have occurred by chance. However, more conservative significance levels with higher thresholds may be applied in certain cases.

3. Do within-treatment change scores on tools designed to measure specific dynamic risk factors predict the probability of general or violent recidivism occurring over a follow-up period?
4. Do within-treatment change scores on a tool measuring psychopathology predict the probability of general or violent recidivism occurring over a follow-up period?

Method

Participants

Data for this study combined two datasets with an initial total of 283 justice-involved men convicted of violent offences and deemed at high-risk of recidivism. These men participated in the Tai Aroha treatment programme between August 2010 and June 2021. The first dataset (*n* = 89) covered the 2010-2014 period and was also used for a programme effectiveness evaluation (Kilgour, 2015); the second dataset covered the remaining period and had a larger sample size (*n* = 194). After 118 programme noncompleters (including five repeaters who later completed the programme), and one duplicate case were removed, 164 (57.9% of the initial total) Tai Aroha programme graduates remained in the final dataset. Demographics can be viewed on Table 2. To summarise, the vast majority of people were Māori, with an average age of 31.2 years (*M* = 31.17, *SD* = 7.21). Six out of ten were gang affiliated (62.2%), and most were in a relationship with 56.7% either married, in a de facto relationship or partnered. In terms of criminal history, the average RoC*RoI (described in the next section) was .71 (*SD* = .10) which is in the high-risk band (i.e., .7 or over) and consistent with the stated programme entry requirement for Tai Aroha (Kilgour, 2015).

Table 2

Demographics for Tai Aroha Graduates (N = 164)

Variable	*n*	%
Ethnicity		
Māori	144	87.8
Pasifika	4	2.4
NZ European	14	8.5
Unknown	2	1.2
Primary gang affiliation		
Mongrel mob	31	18.9
Black power	25	15.2
Crips	12	7.3
Other	34	20.7
None	62	37.8
Relationship status[†]		
Married/de facto	24	14.6
Partnered	69	42.1
Single/unknown	71	43.3

Note. [†] Relationship status recorded at departure from the programme.

Measures

Anger Disorder Scales

The Anger Disorder Scales (ADS) is a 74-item self-report measure that assesses clinically dysfunctional anger in adults (DiGiuseppe & Tafrate, 2004). It is typically used to assist practitioners in designing individualised treatment plans. The ADS is a multidimensional tool designed to categorise different levels of pathological anger (mild, moderate, severe) using 18 subscales in five anger domains: provocations, arousal, motives, cognitions, and behaviour. T-scores above 65 in the ADS indicate clinical issues related to anger duration, frequency, and intensity, based on the ADS manual's normative sample (DiGiuseppe & Tafrate, 2004). The ADS has high internal consistency for the total scale (α = .91). Higher-order factor scales range from α = .91 to .95 and for the 18 subscales range from α = .70 to .92. The ADS has shown some minimal evidence of within-treatment change in previous small studies (Fuller et al., 2010; Henwood et al., 2018; Moulden et al., 2020). To the best of our knowledge, there are currently no studies investigating the predictive validity of the ADS and recidivism outcomes.

Criminal Attitudes to Violence Scale

The Criminal Attitudes to Violence Scale (CAVS) is a 20-item self-report measure designed to assess criminal attitudes to violence (Polaschek et al., 2004). Participants rate each item on a 5-point Likert scale from strongly disagree (1) to strongly agree (5), and items are summed for a single total score. The CAVS aims to evaluate various aspects of an individual's attitudes towards violence, including their acceptance of violent behaviour as a solution to problems, their endorsement of aggression as a means of achieving goals, and their overall tolerance for violence in different contexts. It is designed to assist in understanding how an individual perceives and justifies violent actions, which can be helpful for determining appropriate interventions to reduce the risk of future violent behaviour. The CAVS has been shown to have high internal consistency (α = .95) among people with violent offence convictions (Polaschek et al., 2004). Within-treatment change on the

CAVS has been demonstrated (Calliss, 2021; Polaschek et al., 2010; Sissons, 2013); but no studies to date have tested the CAVS ability to predict recidivism.

Psychological Inventory of Criminal Thinking Styles

Psychological Inventory of Criminal Thinking Styles (PICTS) is an 80-item self-report measure designed to assess criminal thinking (Walters, 1995). Participants rate items on a 4-point Likert scale from disagree (1) to strongly agree (4). The PICTS measures various cognitive patterns and thought processes that are commonly associated with criminal behaviour and includes 13 subscales in total. These include eight thinking styles: Mollification (Mo); Cutoff (Co); Entitlement (En); Power Orientation (Po); Sentimentality (Sn); Superoptimism (So); Cognitive Indolence (Ci); and Discontinuity (Ds). The PICTS also includes two validity scales (Confusion and Defensiveness) and three content scales: General Criminal Thinking (GCT); Proactive Criminal Thinking (P); and Reactive Criminal Thinking (R) scores. The GCT score sums a respondent's scores on all thinking style items. The P score, derived from four subscales (Mo, En, Po, and So) evaluates whether respondents have a planned/calculated approach to crime. The R score, based on three subscales (Co, Ci, and Ds) reflects impulsive approaches to crime. Internal consistency is moderate for the thinking styles scales, high for the P and R content scales, and very high for the overall GCT scale (Walters, 2011). The predictive validity for the PICTS and recidivism has generally been modest (Walters, 2002; Walters, 2011; Walters & Lowenkamp, 2016). Several studies have shown within treatment change on the PICTS (Banse et al., 2013; Walters et al., 2002). Walters and Cohen (2016) found evidence that increases on the GCT content scale predicted increased recidivism, though this was not an intervention study.

Treatment Readiness, Responsivity, and Gain: Short Version

The TRRG:SV (Serin et al., 2005) is a clinician administered 24-item measure with items separated into those that measure readiness for and responsivity to treatment, and gain made from treatment (administered at the end of treatment) on a four-point scale. Readiness refers to an offender’s willingness to engage, while responsivity involves aspects of an offender's interpersonal style (e.g., callousness or use of intimidation) that may hinder the treatment process. The TRRG:SV

subscales have demonstrated high internal consistency (α = .92-95) and change on readiness and responsivity has successfully predicted sexual and violent recidivism among 180 people with sexual offence convictions (Sowden, 2013).

Millon Clinical Multiaxial Inventory

Two versions of the Millon Clinical Multiaxial Inventory (MCMI) were used during the evaluated period; MCMI-III (Millon et al., 1997) and MCMI-IV (Millon et al., 2015). The MCMI measures are designed to identify personality disorders, and psychopathology in clinical populations in line with the Diagnostic and Statistical Manuals (American Psychiatric Association, 2013). The MCMI-III has 175 items and contains 14 personality scales and 10 clinical syndrome scales. The MCMI-IV has 195 items measuring 15 personality scales (one additional scale), and 10 clinical syndrome scales. Respondents rate items as either yes or no. Both versions also have three validity scales (Disclosure, Desirability, and Debasement). Base rate (BR) scores range from 0-115 on personality and clinical syndrome scales. With BR scores between 75-84 indicating problematic pathology, and over 85 indicating serious pathology that is likely to impair functioning. A BR score of 60 is the median based on a normative clinical sample. On average MCMI-III personality and clinical syndrome scales have moderate internal consistency (Millon et al., 1997) and the MCMI-IV personality and clinical syndrome scales have high internal consistency (Millon et al., 2015). According to the MCMI-IV manual, the two editions have large intercorrelations with their paired subscales ranging from .76 to .92 (Millon et al., 2015). A small number of studies assessed within-treatment change on the MCMI-III, but no studies were identified for the MCMI-IV. Importantly, an evaluation of the Tai Aroha programme using the first dataset from the present study, found within-treatment change on all scores except the Narcissistic, Antisocial, Sadistic and Drug-Dependence subscales. Within-treatment change was also demonstrated among high-risk justice-involved people using MCMI-III derived psychopathology clusters (Sissons, 2013). Importantly, an evaluation of the Tai Aroha programme utilising the same sample as the first dataset in this study found within-treatment change on all scores, except Narcissistic, Antisocial, Sadistic, and Drug-Dependence

subscales (Kilgour, 2015). Similar findings were observed among high-risk justice-involved individuals using MCMI-III-derived psychopathology clusters (Sissons, 2013). Another source of evidence comes from a study of drug users undergoing methadone treatment which found change on several subscales among light drug users at an 18-month follow-up, but not for heavy users (Calsyn et al., 2000). Psychopathology clusters 'violent-antisocial' and 'borderline/dysphoric' comprised of various MCMI subscales have been shown to predict recidivism (Carbajosa et al., 2017). Though not assessing change, Halon (2001) coined the term "normal quartet" to describe a pattern of elevated scores on Desirability, Histrionic, Narcissistic, and Compulsive personality traits associated with socially desirable reporting.

Violence Risk Scale

The VRS (Wong & Gordon, 1999-2003) has six static and 20 dynamic risk items and is designed to assess violence risk, inform treatment planning, and evaluate changes during treatment. Items the clinician rated on a four-point scale from not related (0) to significantly related (3) to violence. Dynamic risk items are summed to create a dynamic total score, static items to create a static total score, and a combined total score. The VRS has good internal consistency across subscales (Wong & Gordon, 2006). As outlined fulsomely in the previous section, the VRS has demonstrated predictive validity for within-treatment change and recidivism in various studies (e.g., Beggs & Grace, 2010; Coupland & Olver, 2020; Olver et al., 2021; Olver et al., 2013; Sowden, 2013). Refer back to this section for reference.

Risk of (re)conviction x Risk of (re)imprisonment

The Roc*RoI (Bakker et al., 1999) is an actuarial risk assessment tool developed for use on the Aotearoa New Zealand correctional population and is routinely used by Ara Poutama Aotearoa. Using criminal history and demographic data, a Roc*RoI score between 0-1 is computed by an algorithm. Scores indicate an offender's probability of recidivism, with risk bands from low risk (0.0 - 0.29), medium risk (0.3 -0.69), and high-risk (0.7-0.99). The Roc*RoI is considered a robust and valid predictor of recidivism and is commonly a factor in parole board decisions (Johnston, 2021). It also

determines programme eligibility; as previously stated, Tai Aroha requires a .7 (high-risk) classification.

Recidivism Outcomes

General and violent convictions incurred following the completion of the Tai Aroha programme were the outcomes of interest for this study. The date that recidivism data was recorded before being extracted was 15/08/2023. General recidivism was all reconvictions recorded in the non-violent offence category (e.g., drug-related, theft, breach of sentence conditions) or the violent offence category (e.g., assault, family violence, robbery). Violent recidivism was any reconviction in the violent offence category. Recidivism was coded as a binary outcome (occurred or not), and days until the first reconviction from the date of programme completion were also recorded. Reconviction was used as the most appropriate outcome instead of reoffence or reimprisonment which were also included in the received dataset. The rationale for this was that reoffences rarely result in reconviction, potentially introducing additional measurement error. Likewise, reimprisonment may underestimate recidivism events as most convictions do not result in imprisonment. However, it is acknowledged that each of these criteria underestimates true rates of recidivism (Hanson et al., 2013).

Procedure

Ara Poutama Aotearoa (New Zealand Department of Corrections) provided anonymised archival datasets as password-protected Excel spreadsheets, one for the 2010-2014 period and the other until 2021. These data were collected by people qualified to administer the relevant measures, including Tai Aroha psychologist staff members responsible for each man or graduate student psychometricians under staff supervision. At the commencement of the programme, demographic details were gathered, and pre-programme psychometric tools were completed. For post-programme data, psychometric tools were completed if the programme was completed successfully. These data were then collated by programme staff and transferred into an Excel spreadsheet. Identifying details were anonymised and record numbers were used to identify individual cases. We

then cleaned and organised datasets and combined them before exporting the data to SPSS. When potential errors or missing data were identified, Ara Poutama data analysts were contacted, and data were amended. Datasets with psychometric scores and those with recidivism data were provided separately but we matched these using participant record numbers. The combined dataset was then screened for duplication of participants using the SPSS Identify Duplicate Cases wizard function using demographic and crime history details.

Next, the MCMI-III (Millon et al., 1997) and IV (Millon et al., 2015) versions were combined to simplify the analysis and increase sample size and statistical power. To our knowledge, only one previous study has combined the MCMI-III and IV, an unpublished PhD thesis by Chipollini (2022) that did not make specific accommodations for the minor scoring differences between the two versions (for reference of scoring differences see Table 3 in Sellbom et al., 2022). There do not appear to be any guidelines for combining the measures either provided by the instrument publishers or other authors. However, a table with intercorrelations for base rate scores is provided in the MCMI-IV manual. Therefore, the two versions were combined without adjustment except for excluding the Turbulent personality scale which was introduced in the MCMI-IV. The psychometric properties of both the MCMI-III and IV are acknowledged to be compromised, and those of the combined measure are unknown. That is, it is likely that the validity and reliability of the combined scales are diminished compared to what is reported in each version's respective manual.

Ethics Approval

The Research and Evaluation Steering Committee approved this research and use of secure data in June 2022 on behalf of Ara Poutama Aotearoa, NZ Department of Corrections. The University of Waikato Human Research Ethics Committee then provided approval [HREC(Health)2022#26] in August 2022. Both committees recommended consultation with a cultural supervisor due to the majority of participants being Māori and the programme itself incorporating Māori cultural practices.

Missing Data

An initial screening of psychometric data showed incomplete or missing data for various variables and tests, with more data missing from post-programme results meaning some change scores could not be calculated. Little's (1988) Missing Completely at Random (MCAR) test showed that data was missing at random, $\chi^2 = 1766.36$, DF = 1775, $p = .55$. A significance level of greater than .05 indicates that data was missing completely at random. Ten missing cases in the RoC*RoI and VRS Static risk measurements were replaced with the series mean.

Analysis Plan

IBM Statistical Package for the Social Sciences (SPSS) version 29 was used to compute and analyse data. First, descriptive statistics were calculated for recidivism outcomes to determine the rate of general and violent recidivism in the follow up period. We then calculated sample characteristics (shown in the methods section) and for actuarial risk assessments measured before treatment for the RoC*RoI and VRS Static subscales. The outcome, or dependent variable, was defined as any conviction occurring between the completion date of the Tai Aroha programme and the date of the first general (any reconviction) or violent reconviction (for recidivists) or the censor date (for non-recidivists).

Cox regressions (Cox, 1972) were deemed the most suitable analysis for our dataset and research aims for several reasons. While Cox regressions were briefly described in the introduction, our reasons for using this method are worth outlining in some detail. While Area Under the Curve (AUC; Mossman, 1994) analysis is most commonly used in similar studies (Casey, 2016), Cox regressions are better suited for time-to-event data with high variability in follow-up time (See Hanson, 2022). This is because Cox regressions make no assumptions about the distribution of survival times, making it robust in the presence of variability and potential outliers in the data. This method also accounts for censoring, which refers to cases 'lost' in the follow-up period not due to recidivism but instead due to limited follow-up. They have been applied in many similar studies (e.g.,

Berman, 2004; Brown et al., 2009; Cuevas et al., 2019; Howard & van Doorn, 2018; Howard & Dixon, 2013; Klepfisz et al., 2022; O'Brien & Daffern, 2017; Olver et al., 2014a; Walters & Cohen, 2016). As already noted, follow-up times were highly variable in our study, ranging from eight months to nine years. Cox regression was preferred to using AUCs with fixed follow-up periods (e.g., 3 years or 5 years) which are favoured by some researchers (e.g., Coupland & Olver, 2020; Higgs et al., 2020) in order to maximise usable data by preventing the exclusion of cases with follow-up time ending prior to the fixed period. Cox regressions also allow for the examination of covariate effects such as actuarial risk, which cannot be added to AUC calculations. Cox regression is a semiparametric model meaning there are fewer statistical assumptions than comparable methods of prediction (Hanson, 2022; Harrell, 2015). Unlike General Linear Regression models, Cox regressions do not require predictors to have normal distribution, homoscedasticity, or linearity. The proportional hazard assumption, commonly assessed through methods such as plotting Schoenfeld residuals, was not fulfilled in this analysis (Hanson, 2022). This was due to the feasibility of calculating and reporting the large number of variables included in this study.

A series of Cox regressions were first performed to assess if sample characteristics (converted to binary variables where appropriate) predicted recidivism. This was repeated for actuarial risk scales. Next, we set about answering our first two research questions. Raw change scores were calculated for tools designed to measure dynamic risk factors and psychopathology. Scores were calculated conventionally as the post-treatment score minus the pre-treatment score (Juarez & Howard, 2022). Paired-sample t-tests were then used to assess whether scores were significantly different between time points, and effect sizes were calculated using Cohen's *d* (Cohen, 1988) to assess the magnitude of these differences. To account for the large number of comparisons (34 in total) Bonferroni corrections (IBM, 2020) were used to interpret significance tests to reduce the chance of type I errors (i.e., false positives). This was calculated by dividing the conventional α of .05 by the number of subscales within each measure. Subscales that did not demonstrate statistically significant change at the adjusted level were omitted from subsequent analyses.

Next, following a number of researchers (e.g., Beggs & Grace, 2011; Higgs et al., 2020; Olver et al., 2014b) pre-treatment scores were regressed onto dynamic risk factor change scores. The resulting residuals, known as residualised change scores (RCZ), were saved for each subscale across all dynamic risk factor measures, as well as for the MCMI III and IV. These RCZ scores for retained subscales were used in all subsequent analyses. To address the final two research questions, a series of Cox regressions were conducted to evaluate the predictive capacity of RCZ change scores for each dynamic risk factor subscale, followed by separate analyses for MCMI III and IV subscales. Lastly, these analyses were repeated controlling for actuarial risk in line with other researchers (e.g., Allan et al., 2007; Beggs & Grace, 2010).

Results

Recidivism Rates

First, descriptive statistics for recidivism outcomes were calculated. The average follow-up period for assessing recidivism outcomes was four and a half years (*M* = 4.47; SD = 2.39), ranging considerably between eight months after programme graduation to nine years. Over that period, 89.0% (*n* = 145 out of 163) of programme completers had at least one general reconviction on their record, whereas just over half (53.4%; *n* = 87 out of 163) had a violent reconviction. For general recidivism, participants took an average of 430 days (*SD* = 481.3) to their first reconviction, whereas they took more than twice as long, 930 days (*SD* = 792.1), to their first violent reconviction. A relatively small number of offences were recorded for completers in the first six months following the programme. For general recidivism, 36.8% (*n* = 60 out of 163) had been convicted within the first six months and 9.8% (*n* = 16 out of 163) for violent recidivism.

Sample Characteristics

Demographic characteristics are summarised in the previous section in Table 2. To briefly reiterate, most men were Māori (87.8%), with an average age of 31.2 years (*M* = 31.17, *SD* = 7.21). Most were gang affiliated (62.2%), and more than half (56.7%) were in a romantic relationship of some kind. In order to detect whether any of these characteristics effected recidivism outcomes, a series of Cox Regressions were performed due to variability in follow-up times. First, relationship status (either single or in a relationship), and gang membership (either gang member or not) were converted to binary categories to be used as indicator variables. Ethnicity was not analysed given the ethnic homogeneity of the sample. Relationship status did not statistically significantly predict general, Exp(B) = .77, 95% CI [.56 – 1.08], *p* = .13, or violent recidivism, Exp(B) = .78, 95% CI [.51 – 1.19], *p* = .25, but scores trended toward lower recidivism. Gang membership was nonsignificant for general, Exp(B) = 1.09, 95% CI [.77 – 1.52], *p* = .63, and violent recidivism, Exp(B) = .81, 95% CI [.53 – 1.25], *p* = .34, but trended toward higher recidivism. Age was also nonsignificant for general, Exp(B) = .98, 95% CI [.95 – 1.01], *p* = .18, and violent recidivism, Exp(B) = .98, 95% CI [.95 – 1.01], *p* = .18. As

no effects on recidivism were observed, these demographic variables were not statistically controlled in subsequent analyses.

Actuarial Risk Assessments

We then analysed pre-treatment risk assessment scores. For the total sample, the average RoC*RoI score was .71 (*SD* = .1). For the VRS scores recorded at the beginning of treatment, the average total dynamic score was 41.21 (*SD* = 7.23), the average static score was 13.55 (*SD* = 2.5), and combined average total score was 55.11 (*SD* = 7.56). As previously outlined, both the VRS (Wong & Gordon, 1999-2003) and RoC*RoI (Bakker et al., 1999) have previously demonstrated robust predictive validity. The RoC*RoI score did not statistically significantly predict general, Exp(B) = 3.23, 95% CI [.60 – 17.80], *p* = .18, or violent recidivism, Exp(B) = 1.24, 95% CI [.16 – 9.51], *p* = .83. The general recidivism hazard ratio was notably high but imprecise, as reflected by the extreme ranges in confidence intervals. Results for the VRS static total score were also nonsignificant predicting for general, Exp(B) = 1.12, 95% CI [.98 – 1.12], *p* = .20, or violent recidivism, Exp(B) = .98, 95% CI [.90 – 1.07], *p* = .66; as were VRS dynamic total scores for general, Exp(B) = 1.00, 95% CI [.98 – 1.03], *p* = .43, or violent recidivism, Exp(B) = .99, 95% CI [.96 – 1.02], *p* = .53.

RQ 1: *Is there evidence of significant within-treatment change on tools designed to measure specific dynamic risk factors?*

Next, we investigated our first research question. Absolute change scores (post-treatment score minus pre-treatment score) were calculated for dynamic risk factor measures. Table 3 and Table 4 show descriptive and inferential statistics, with the ADS in Table 3, and the PICTS, CAVS, and TRRG:SV displayed together in Table 4. Paired-sample t-tests were used to assess whether scores on variables were significantly different across time points. To account for the large number of comparisons (34 in total) Bonferroni corrections (IBM, 2020) were used to interpret significance tests to reduce the chance of type I errors (i.e., false positives). This was calculated by dividing the conventional α of .05 by the number of subscales within each dynamic risk factor measure. For example, for the ADS with 20 subscales, the adjusted alpha was calculated as .05/20 = .0025. Raw

scores were used for ADS subscales because T-scores for individual cases were not provided in our dataset. Average T-scores are provided for reference but were calculated from the sample mean raw scores using the ADS technical manual (DiGiuseppe & Tafrate, 2004).

Across the four different dynamic risk factors measurement tools, all except two changes between time points were statistically significant at the adjusted significance level. Two ADS subscales, Relational aggression (p = .32), and Tension Reduction (p = .13), did not show evidence of within-treatment change and were omitted from subsequent analyses. Cohen's d effect sizes for significant changes are too numerous to report individually but range from small (d = .20) to large (d = .80) in magnitude for most variables according to standard conventions (Cohen, 1988).

For the Positive Impression subscale of the ADS, a decreased mean score indicates more attempts to appear in a positive light or "faking good," with scores under 15 indicating likely socially desirable responding (DiGiuseppe & Tafrate, 2004). In this study, Positive Impression scores reduced significantly from pre-programme to post-programme but, on average, participants did not respond in a socially desirable manner. However, a notable shift was observed when using the clinical cut-off points with 13.6% (n = 22 out of 162) scoring below the clinical threshold before the programme, compared to 32.7% (n = 50 out of 153) after the programme. McNemar's test revealed statistically significant differences in proportions between these time points, $\chi^2(1) = 16.757$, $p < .001$. Notably, the majority of those who responded in a socially desirable manner pre-programme also did so post-programme (63.6%; n = 14 out of 22). While socially desirable responding is more accurately treated as a continuous rather than a binary variable, these findings are still notable.

Table 3

Comparisons of Psychometric Results for the ADS Pre- and Post-programme with Effect Sizes

	Measures	Pre-programme *M* (*sd*)	*T*-score *	Post-programme *M* (*sd*)	*T*-score*	*t*	Degrees of freedom (*df*)	*P*	Pre/post raw change	*d*
ADS										
Behaviours	Brooding	14.25 (4.44)	56	11.05 (3.53)	47	7.70	150	<.001	-3.20	0.63
	Verbal Aggression	12.64 (5.82)	62	9.49 (4.01)	53	6.49	150	<.001	-3.15	0.53
	Physical Aggression	5.40 (2.91)	61	3.81 (1.41)	54	6.87	150	<.001	-1.59	0.56
	Relational Aggression	4.27 (2.05)	53	4.01 (3.40)	53	1.00	150	.32	-0.26	0.08
	Passive Aggression	11.75 (5.22)	63	8.75 (4.22)	54	7.00	150	<.001	-3.0	0.57
	Indirect Aggression	8.46 (3.76)	66	6.86 (3.23)	57	5.10	150	<.001	-1.6	0.42
Motives	Revenge	11.97 (5.43)	64	8.58 (4.48)	54	7.13	150	<.001	-3.39	0.58
	Tension Reduction	9.75 (3.02)	57	9.10 (3.50)	54	2.18	150	.13	-0.65	0.18
	Coercion	10.23 (3.87)	55	8.34 (3.10)	49	6.21	150	<.001	-1.89	0.51
Cognitions	Suspiciousness	10.74 (3.91)	59	7.75 (2.95)	49	10.05	150	<.001	-2.99	0.82
	Resentment	9.15 (4.09)	54	7.13 (3.27)	48	6.72	150	<.001	-2.02	0.55
	Rumination	10.57 (4.37)	64	7.96 (3.92)	54	6.91	150	<.001	-2.61	0.56
	Impulsivity	7.13 (3.62)	65	4.97 (2.90)	55	6.53	150	<.001	-2.16	0.53
Arousal	Physiological Arousal	9.77 (4.29)	62	8.55 (3.71)	58	3.68	150	<.001	-1.22	0.30
	Duration of Anger	10.87 (4.35)	65	9.05 (4.66)	60	4.65	150	<.001	-1.82	0.38
	Episode Length	6.05 (2.65)	53	4.71 (1.87)	49	5.83	149	<.001	-1.34	0.48
P	Hurt/Social Rejection	15.68 (4.96)	54	13.01 (4.91)	47	6.44	150	<.001	-2.67	0.52
	Scope of Anger	10.97 (3.29)	55	9.00 (3.41)	49	6.49	150	<.001	-1.97	0.53
	Positive Impression [A]	20.16 (5.21)	-	16.78 (5.36)	-	7.30	150	<.001	-3.38	0.59
	Total score	43.75 (11.67) [†]	65	34.14 (11.22)	53	9.93	153	<.001	-9.61	0.63

Note. Subscale cluster labels on left-most column where applicable. Raw scores were used for ADS calculations; *ADS T-scores calculated from average sample scores not from individual participant scores; P = Provocations; ADS = Anger Disorders Scale.

[A] Lower score indicates more socially desirable responding.

Table 4

Comparisons of Psychometric Results for the PICTS, CAVS, and TRRG:SV Pre- and Post-programme with Effect Sizes

Measures		Pre-programme *M* (*sd*)	Post-programme *M* (*sd*)	*t*	Degrees of freedom (*df*)	*p*	Pre/post raw change	*d*
PICTS								
Thinking styles	Mollification (Mo)	60.61 (12.45) ††	53.35 (12.23) †	7.11	151	<.001	-7.26	0.57
	Entitlement (En)	61.81 (12.96) ††	56.32 (12.17) †	5.45	151	<.001	-5.49	0.44
	Power Orientation (Po)	60.44 (14.97) ††	53.28 (12.76) †	6.10	151	<.001	-7.16	0.49
	Sentimentality (Sn)	54.77 (10.27) †	50.18 (10.73) †	5.07	151	<.001	-7.59	0.41
	Superoptimism (So)	59.86 (12.68) †	55.58 (12.16) †	4.00	151	<.001	-4.28	0.32
	Cutoff (Co)	65.39(10.27) ††	56.54 (11.41) †	8.58	151	<.001	-8.85	0.69
	Cognitive Indolence (Ci)	59.79 (9.66) †	51.48 (9.62) †	10.41	151	<.001	-8.31	0.84
	Discontinuity (Ds)	61.40 (10.46) ††	54.57 (9.92) †	8.51	151	<.001	-6.83	0.68
Content Scales	Reactive Criminal Thinking (R)	63.71(10.51) ††	54.19 (10.76) †	10.19	151	<.001	-9.52	0.82
	Proactive Criminal Thinking (P)	63.75(12.66) ††	58.81 (12.35) †	5.07	151	<.001	-4.94	0.41
	General Criminal Thinking (GCT)	61.69 (10.56) ††	52.22 (10.72) †	10.76	151	<.001	-9.47	0.86
CAVS								
	Total Score	55.34 (18.95)	38.56 (17.32)	9.97	153	<.001	-16.78	0.80
TRRG: SV								
	Readiness	15.71 (2.94)	19.29 (3.18)	-11.70	100	<.001	3.57	-1.16
	Responsivity	14.16 (2.85)	18.77 (3.74)	-11.58	99	<.001	4.61	-1.41

Note. Subscale cluster labels on left-most column where applicable. T-scores were used for PICT; † Within the average clinical range; †† within the high clinical range. Both TRRG:SV and CAVS use raw scores. PICTS = Psychological Inventory of Criminal Thinking; CAVS = Criminal Attitudes to Violence Scale; TRRG:SV = Treatment Readiness, Responsivity, and Gain: Short Version.

Turning now to Table 4, results for the PICTS showed average T-scores mostly in the high clinical range (60-69) for pre-programme scores, but lowered significantly into the normal clinical range (40-59) for all post-programme scores (Walters, 2006). No average scores were in either the low (<40) or very high (>70) ranges. Of the content scales, the highest scores and least relative change between time points were observed for the Proactive Criminal Thinking subscale which relates to a planned/calculated approach to crime. CAVS change scores indicate that men agreed with fewer violent attitudes. For the TRRG:SV, higher post-programme scores reflect improvements in Readiness and Responsivity, both with large effect sizes. Note that sample was smaller for this scale due to only two thirds of post-programme scores being completed. The third subscale for the TRRG:SV, treatment gain, was excluded from analysis as it is only recorded post-treatment.

RQ 2: ***Is there evidence of significant within-treatment change on a tool designed to measure psychopathology?***

We then assessed change on psychopathology scores on the combined MCMI-III and IV measures, performing the same set of analyses as in the previous research question. These results are displayed on Table 5. Bonferroni corrections were also applied as 27 individual tests were performed, with a revised standard of $\alpha = .0019$ (.05/27) to reach significance.

Scores on the majority of MCMI subscales improved over the course of treatment. All changes were statistically significant with the exception of the Narcissistic subscale which did not demonstrate change so was removed from subsequent analyses. There were also several within-treatment changes observed for subscale average scores relative to clinical thresholds. As stated in the method section, higher scores on the MCMI indicate increased presence or severity of psychopathology, with scores between 60-75 considered subclinical, between 75-84 indicating problematic pathology, and scores over 85 indicating severe pathology (Millon et al., 2015). For the pre-programme mean scores, two subscales were in the problematic range, Alcohol, and Antisocial, with the later on the cusp of severe pathology. Drug dependence was in the severe pathology range. For the post-programme mean scores, only Antisocial, and Drug dependence were in the

problematic range, indicating that these tend to be the most serious and persistent problems experienced by the Tai Aroha men. While noteworthy, these findings only reflect average scores within the sample. Consequently, individuals may have scored within the clinical range, even for subscales with averages below the clinical threshold for psychopathology. However, addressing this matter further is beyond the scope of this study. In any case, these results do not justify the exclusion of specific subscales with average scores below clinical significance from further analyses.

The Histrionic and Compulsive clinical personality patterns, and the Desirability Validity subscale significantly increased over treatment. Increases on the Histrionic subscale indicate more interpersonal confidence, self-assuredness, and behaviour that serves to elicit praise and validation (Grossman & Amendolace, 2017). The compulsive personality pattern relates to a conflict between feelings of hostility toward others and a fear of social disapproval (Grossman & Amendolace, 2017). This results in overt passivity but underlying emotions like anger and frustration. In the context of individuals with histories of violent offending completing group-based therapy with expectations of compliance and self-regulation, it is not surprising that scores on this subscale increase during treatment.

For the Desirability validity scale on the MCMI, it appears that people made efforts to be more socially desirable at the end of treatment. This is also consistent with scores on the ADS Positive Impression subscale. However, as discussed previously, higher scores in Socially Desirable responding may be a positive indication (e.g., evidence of 'interpersonally oriented self-control'; Uziel, 2010). It is notable that scores on the two other validity subscales on the MCMI, Disclosure and Debasement, decreased while Desirability increased. This indicates that at the end of treatment, there is a significant reduction in self-disclosure and less inclination to admit negative traits or behaviours, but an increase in efforts made to appear socially desirable.

As previously outlined, Sellbom (2022) describes a ‘normal quartet’ pattern of impression management which has been found for both the MCMI-III (Halon, 2001) and MCMI-IV (Eastin et al., 2022). The ‘normal quartet’ comprises increased scores on the Desirability, Histrionic, Narcissistic,

and Compulsive subscales. subclinical range for pre-programme scores, and all showed significant positive within-treatment changes except Narcissistic. Desirability and Narcissistic were elevated within the subclinical range for pre-programme score. While Desirability, Narcissistic, and Histrionic, were all elevated for post-programme, Compulsive was under the threshold for both time points though scores significantly increased at post-programme.

Pearson's bivariate correlations were then performed on the 'normal quartet' for both pre- and post-programme scores and are displayed on Table 6. Pre-programme scores were highly correlated, with effects ranging for small to large, and all except Compulsive and Narcissistic were statistically significantly correlated at the post-programme but with relatively smaller effects ranging from small to medium. While the 'normal quartet' subscales scores were generally higher at post-programme as already noted, they were associated more clearly at pre-programme.

Table 5

Correlations for the 'Normal Quartet' for the MCMI III and IV subscales

Variables	1	2	3	4	5	6	7
Pre-programme	-						
1 Desirability							
2 Histrionic	.68**	-					
3 Narcissistic	.35**	.38**	-				
4 Compulsive	.52**	.32**	.22**	-			
Post programme					-		
5 Desirability	.47**	.30**	.08	.30**		-	
6 Histrionic	.35**	.56**	.07	.15	.59**		-
7 Narcissistic	.13	.15	.47**	.06	.17*	.16*	
Compulsive	.24**	.03	.04	.47**	.49**	.18*	.12

Note. * p < .05; ** p < .01.

RQ 3: Do within-treatment change scores on tools designed to measure specific dynamic risk factors predict the probability of general or violent recidivism occurring over a follow-up period?

For the next phase of the analysis, we investigated whether dynamic risk factor change scores predicted recidivism. First, change scores were adjusted for pre-treatment differences using RCZ

scores derived from Linear Regression models. Using these revised scores, a series of Cox regressions were then performed using residualised dynamic risk factor change scores to predict general recidivism and violent recidivism. While statistical assumptions have been suggested, few researchers test for them as they are complex and difficult to adjust for. Bonferroni corrections were also calculated based on the number of individual tests in each measure analysed.

Hazard ratios were calculated for Cox regressions as a measure of the magnitude and direction of effect. All dynamic risk factor variables that demonstrated statistically significant change in the previous analyses were included, and results are shown on Table 7. A visual overview of the table reveals a large number of nonsignificant results that appear imprecise with mostly wide confidence intervals. The hazard ratios predominantly straddle the value 1.00, suggesting that the dynamic risk factor change scores make no discernible statistical contribution to the likelihood of the recidivism event. However, of all 33 dynamic risk factor subscales contained in the four multi-subscale measurement tools, only the RCZ for the Discontinuity PICTS subscale significantly predicted general recidivism. Change on Discontinuity was associated with a small reduction in the likelihood of general recidivism (p = .047) at the conventional but not at the Bonferroni adjusted alpha threshold (α = .005 for the PICTS).

A second series of Cox regressions were performed, this time controlling for actuarial risk measures using both the VRS static risk and RoC*RoI scores. These data can be viewed on Table S1 in the appendix. As reported earlier, these variables did not demonstrate an independent effect on recidivism, however, it is a standard convention performed by researchers to control for criminal risk using actuarial risk measures such as these. Discontinuity remained significant at the standard level, Exp(B) = 3.23, 95% CI [.60 – 17.80], p = .18, but not at the Bonferroni corrected threshold. There were no additional significant results, but some minor variations in hazard ratios and confidence intervals can be seen.

Table 6

Comparisons of Psychometric Results for the Combined MCMI-III and IV on Pre- and Post-programme with Effect Sizes

	Subscales	Pre-programme *M* (*sd*)	Post-programme *M* (sd)	t	Degrees of freedom (df)	*p*	Pre/post raw change	Cohen's *d*
Validity Scales	X Disclosure	74.35 (15.31)†	60.57 (18.26)†	8.97	151	<.001	-13.78	0.73
	Y Desirability	64.08 (15.92)†	72.39 (11.21)†	-7.00	154	<.001	8.31	-0.56
	Z Debasement	57.43 (18.50)	39.61 (21.09)	9.90	154	<.001	-17.82	0.80
Clinical personality patterns	1 Schizoid	60.32 (16.48)†	47.45 (22.27)	7.31	154	<.001	-12.87	0.59
	2A Avoidant	58.73 (24.25)	40.67 (28.25)	7.57	154	<.001	-18.06	0.61
	2B Depressive	58.29 (25.12)	38.33 (28.15)	8.95	154	<.001	-19.96	0.72
	3 Dependent	60.18 (22.34)†	41.75 (27.48)	8.96	154	<.001	-18.43	0.72
	4 Histrionic	50.50 (18.46)	59.85 (14.25)†	-7.64	154	<.001	9.35	-0.61
	5 Narcissistic	69.01 (19.38)†	68.26 (19.43)†	.66	154	.51	-0.75	0.05
	6A Antisocial	84.66 (14.37)††	76.88 (15.08)††	5.99	154	<.001	-7.78	0.48
	6B Sadistic	70.86 (13.20)†	62.11 (18.84)†	5.28	154	<.001	-8.75	0.42
	7 Compulsive	46.46 (18.69)	55.14 (16.51)	-5.94	154	<.001	8.68	-0.48
	8A Negativistic	65.49 (20.00)†	46.98 (26.87)	7.39	154	<.001	-18.51	0.59
	8B Masochistic	62.65 (20.17)†	43.75 (28.31)	7.99	153	<.001	-18.9	0.64
Severe	S Schizotypal	62.40 (20.55)†	44.00 (27.09)	8.16	154	<.001	-18.4	0.66
	C Borderline	62.74 (19.36)†	45.34 (25.85)	7.87	154	<.001	-17.4	0.63
	P Paranoid	69.12 (18.98)†	54.92 (23.20)	6.73	154	<.001	-14.2	0.54
Clinical syndromes	A Anxiety	69.83 (28.61)†	46.77 (35.20)	7.96	154	<.001	-23.06	0.64
	H Somatoform	37.05 (26.64)	22.13 (24.24)	6.78	154	<.001	-14.92	0.54
	N Bipolar/ Manic	71.64 (14.96)†	61.35 (21.53)†	6.49	154	<.001	-10.29	0.52
	D Dysthymia	51.65 (24.70)	31.17 (23.81)	9.99	154	<.001	-20.48	0.80
	B Alcohol	82.26 (14.35)††	72.97 (17.25)†	7.44	154	<.001	-9.29	0.60
	T Drug Dependence	90.42 (15.54)†††	84.96 (15.31)†††	4.64	154	<.001	-5.46	0.37
	R PTSD	58.11 (27.88)	37.98 (29.85)	8.24	154	<.001	-20.13	0.66
Severe	SS Thought Disorder	56.09 (22.05)	38.01 (26.08)	8.18	154	<.001	-18.08	0.66
	CC Major Depression	40.95 (32.08)	21.61 (25.57)	7.84	154	<.001	-19.34	0.63
	PP Delusional	63.89 (17.63)†	48.83 (27.00)	6.76	154	<.001	-15.06	0.54

Note. Subscale cluster labels on left most column. MCMI scores are presented as base rate (BR) scores; † within subclinical range; †† within problematic pathology range; ††† within the serious pathology range. MCMI = Millon Clinical Multiaxial Inventory.

Table 7

Predictors of General and Violent Recidivism with Standardised Residualised Change Scores for dynamic risk factors Measures Using Cox Regressions

			General Recidivism		Violent Recidivism	
	Variables	***n***	**HR**	**95% CI**	**HR**	**95% CI**
ADS						
Behaviours	Brooding	152	1.05	[.90 – 1.23]	.89	[.72 – 1.09]
	Verbal Aggression	152	1.03	[.88 – 1.20]	.90	[.73 – 1.12]
	Physical Aggression	152	.99	[.83 – 1.20]	.85	[.67 – 1.07]
	Passive Aggression	152	1.04	[.88 – 1.24]	1.00	[.81 – 1.23]
	Indirect Aggression	152	.98	[.82 – 1.17]	.96	[.75 – 1.22]
M	Revenge	152	1.08	[.91 – 1.28]	1.04	[.83 – 1.29]
	Coercion	152	1.00	[.95 – 1.05]	.91	[.74 – 1.12]
Cognitions	Suspiciousness	152	1.05	[.88 – 1.25]	.89	[.99 – 1.23]
	Resentment	152	1.00	[.84 – 1.19]	1.04	[.84 – 1.29]
	Rumination	152	.99	[.83 – 1.18]	.92	[.73 – 1.16]
	Impulsivity	152	.95	[.80 – 1.12]	.96	[.76 – 1.21]
Arousal	Physiological Arousal	152	.99	[.82 – 1.18]	.91	[.72 – 1.14]
	Duration of Anger	152	1.10	[.92 – 1.32]	1.02	[.83 – 1.27]
	Episode Length	152	1.13	[.95 – 1.35]	.94	[.76 – 1.17]
P	Hurt/Social Rejection	152	1.06	[.90 – 1.25]	1.02	[.82 – 1.26]
	Scope of Anger	152	1.16	[.97 – 1.38]	1.01	[.80 – 1.28]
	Positive Impression	152	1.15	[.96 – 1.39]	1.06	[.85 – 1.32]
	Total score	152	1.08	[.91 – 1.27]	.97	[.78 – 1.19]
CAVS						
	Total Score	152	.92	[.78 – 1.08]	.96	[.79 – 1.18]
PICTS						
Thinking styles	Mollification (Mo)	152	.96	[.81 – 1.14]	.97	[.78 – 1.19]
	Entitlement (En)	152	.88	[.74 – 1.05]	1.03	[.83 – 1.28]
	Power Orientation (Po)	152	.93	[.80 – 1.09]	.93	[.80 – 1.09]
	Superoptimism (So)	152	.95	[.81 – 1.11]	1.01	[.80 – 1.26]
	Sentimentality (Sn)	152	.96	[.81 – 1.13]	.86	[.69 – 1.08]
	Cutoff (Co)	152	.90	[.76 – 1.07]	.90	[.76 – 1.07]
	Cognitive Indolence (Ci)	152	.90	[.75 – 1.08]	.85	[.68 – 1.06]
	Discontinuity (Ds)	152	.85*	[.73 – 1.00]	.86	[.70 – 1.07]
Content scales	Reactive Criminal Thinking (R)	152	1.02	[.86 – 1.20]	.97	[.79 – 1.20]
	Proactive Criminal Thinking (P)	152	.95	[.79 – 1.13]	1.03	[.82 – 1.29]
	General Criminal Thinking (GCT)	151	.88	[.75 – 1.04]	.87	[.71 – 1.08]
TRRG: SV						
	Readiness	152	.90	[.69 – 1.18]	.96	[.88 – 1.05]
	Responsivity	152	.93	[.71 – 1.22]	.98	[.91 – 1.05]

Note. Subscale cluster labels on the left-most column where applicable. M = Motives; P = Provocations; HR = hazard ratio; ADS = Anger Disorders Scale; PICTS = Psychological Inventory of Criminal Thinking; CAVS = Criminal Attitudes to Violence Scale; TRRG:SV = Treatment Readiness, Responsivity, and Gain: Short Version. *$p < .05$.

RQ 4: ***Do within-treatment change scores on a tool measuring psychopathology predict the probability of general or violent recidivism occurring over a follow-up period?***

We next addressed our final research question. As with the previous change score calculations, RCZ change scores were utilised for each MCMI subscale. Cox regressions were calculated for pre-treatment controlled MCMI variables which can be viewed in Table 8. Again, Bonferroni corrections were also applied as 26 individual tests were performed for the MCMI. A more conservative $\alpha = .0019$ was therefore required to reach statistical significance. A second series of Cox regressions were performed, this time controlling for actuarial risk measures using both the VRS static risk and RoC*RoI scores. These data can be viewed on Table S2 in the appendix.

The vast majority of hazard ratios were near the 1.00 point with wide confidence intervals, indicating very little effect of change on these variables either increasing or decreasing the hazard of recidivism. Change on only one subscale, the Sadistic ($B = .74$, $p < .009$) clinical personality pattern, predicted recidivism at the unadjusted alpha level, but the effect was not retained at the adjusted alpha threshold ($\alpha = .0019$). Sadistic personality is characterized by a disregard for others' feelings and an active orientation toward inflicting pain on others, which tends to be experienced as a source of pleasure (Grossman & Amendolace, 2017). It is noteworthy that a significant effect was not observed for the Antisocial personality pattern, despite its overlap with the Central Eight dynamic risk factors Antisocial Personality Pattern and similarities to the Sadistic personality pattern. This is particularly interesting given the larger pre-post change treatment effect size indicated in Table 2.

Table 8

Predictors of General and Violent Recidivism with Standardised Residual Change Scores for Combined MCMI-III and IV Using Cox Regressions

			General Recidivism		Violent Recidivism	
	Variables	***n***	**HR**	**95% CI**	**HR**	**95% CI**
MCMI-III and IV						
Validity scales	X Disclosure	152	1.00	[.83 – 1.20]	.98	[.75 – 1.27]
	Y Desirability	155	1.02	[.86 – 1.20]	.96	[.77 – 1.20]
	Z Debasement	155	1.06	[.88 – 1.28]	.89	[.70 – 1.13
Clinical personality patterns	1 Schizoid	155	.93	[.78 – 1.11]	.94	[.71 – 1.12]
	2A Avoidant	155	1.03	[.87 – 1.23]	1.00	[.80 – 1.26]
	2B Depressive	155	1.06	[.87 – 1.28]	.93	[.73 – 1.19]
	3 Dependent	155	1.02	[.87 – 1.21]	.89	[.71 – 1.11]
	4 Histrionic	155	1.00	[.84 – 1.18]	1.12	[.89 – 1.42]
	6A Antisocial	155	.90	[.75 – 1.08]	.90	[.70 – 1.14]
	6B Sadistic	155	.85	[.70 – 1.03]	**.74***	[.59 – .93]
	7 Compulsive	155	1.04	[.86 – 1.24]	.94	[.75 – 1.18]
	8A Negativistic	155	.96	[.80 – 1.15]	.87	[.68 – 1.09]
	8B Masochistic	154	.98	[.83 – 1.17]	.90	[.72 – 1.14]
Severe	S Schizotypal	155	.88	[.73 – 1.06]	1.03	[.80 – 1.31]
	C Borderline	155	1.08	[.91 – 1.29]	.94	[.74 – 1.20]
	P Paranoid	155	.91	[.76 – 1.10]	.95	[.76 – 1.20]
Clinical syndromes	A Anxiety	155	1.07	[.89 – 1.29]	.94	[.75 – 1.18]
	H Somatoform	155	1.12	[.93 – 1.36]	.93	[.75 – 1.17]
	N Bipolar/ Manic	155	1.02	[.85 – 1.22]	.96	[.76 – 1.22]
	D Dysthymia	155	1.03	[.85 – 1.24]	.97	[.77 – 1.23]
	B Alcohol	155	1.00	[.83 – 1.20]	.90	[.71 – 1.15]
	T Drug Dependence	155	1.07	[.89 – 1.27]	1.03	[.80 – 1.32]
	R PTSD	155	1.03	[.85 – 1.25]	1.03	[.81 – 1.30]
Severe	SS Thought Disorder	155	1.01	[.84 – 1.20]	.92	[.73 – 1.16]
	CC Major Depression	155	1.17	[.97 – 1.41]	.96	[.78 – 1.19]
	PP Delusional	155	1.07	[.89 – 1.28]	1.12	[.88 – 1.43]

Note. Subscale cluster labels on left-most column. HR = hazard ratio; MCMI = Millon Clinical Multiaxial Inventory. * $p < .001$.

Discussion

A significant amount of theoretical discussion has critiqued the central importance that dynamic risk factors or criminogenic needs have within correctional rehabilitation despite their questionable empirical status (e.g., Cording et al., 2016; Heffernan et al., 2019; Thornton, 2016; Ward, 2016). To resolve this, scholars have persistently called for more high-quality research investigating whether within-treatment change on tools that measure dynamic risk factors is reliably linked to reductions in recidivism, given the current inconsistency in findings (Eisenberg et al., 2019; Papalia et al., 2019; Serin et al., 2013). The primary objective of this study was to contribute to this research endeavour. A secondary aim was to investigate whether within treatment change predicted recidivism for psychopathology, which pertains to factors generally considered non-criminogenic needs. While often assessed and treated alongside dynamic risk factors by correctional psychologists, psychopathology has received less research attention. We investigated these objectives among justice-involved men convicted of violent offences who completed Tai Aroha, an intensive 16-week violence prevention programme. Four research questions guided our study:

1. Is there evidence of significant within-treatment change on tools designed to measure dynamic risk factors?
2. Is there evidence of significant within-treatment change on a tool designed to measure psychopathology?
3. Do within-treatment change scores on tools designed to measure specific dynamic risk factors predict the probability of general or violent recidivism occurring over a follow-up period?
4. Do within-treatment change scores on a tool measuring psychopathology predict the probability of general or violent recidivism occurring over a follow-up period?

Key Findings

Within-treatment change scores were broadly statistically significant and mostly of a moderate or large magnitude across tools assessing both dynamic risk factors and psychopathology,

but this did not translate to predicting the probability of recidivism. For within-treatment change scores, of the 63 predictor variables analysed for both dynamic risk factors and psychopathology subscales, all except three (two subscales from the ADS, and one from the MCMI III and IV) showed significant changes over the course of treatment. Overall, these results indicate that men made substantial progress during treatment on a range of personality characteristics and crime-related cognitive and emotional factors, but there was no evidence that this effected criminal behaviour once treatment ended.

Comparisons with Previous Research

We did not find evidence that the assessed dynamic risk factor variables function as mechanisms of change influencing recidivism. Of course, as other authors have suggested (e.g., van den Berg et al., 2018), this does not preclude other unmeasured latent factors potentially changing during treatment and effecting recidivism that may be detected using other measurement tools. However, this finding accords with a body of evidence which has failed to demonstrate a clear link between dynamic risk factor change and recidivism (e.g., Banse et al., 2013; Berman et al., 2019; Higgs et al., 2020; Howard & van Doorn, 2018; Klepfisz et al., 2014; Mastromanno et al., 2018; Nunes et al., 2016; O'Brien & Daffern, 2017; Yesberg & Polaschek, 2019). Consistent with our findings, absent of a link to recidivism, many studies have also found within-treatment change on dynamic risk factors (Berman et al., 2019; Higgs et al., 2020; Howard & van Doorn, 2018; Mastromanno et al., 2018; O'Brien & Daffern, 2017). Therefore, it is not unusual for measures assessing dynamic risk factors to improve during treatment without this change being associated with recidivism.

There were several inconsistencies with previous research for specific measures. Existing research has demonstrated within-treatment change on PICTS, including when compared to a control group (Walters et al., 2002). Our findings are consistent with this Walters and colleagues (2002), but not with Walters and Lowenkamp (2016) who found that increased PICTS scores over time were linked to higher recidivism, although participants in this study did not receive active treatment. There has been limited previous research using the CAVS measure, with no studies

identified linking change to recidivism. However, our findings are consistent with previous studies demonstrating within-treatment change (Calliss, 2021; Polaschek et al., 2010; Sissons, 2013). However, improvements were notably smaller than those found by Polaschek and colleagues (2010). Previous research using the ADS has shown some evidence of within-treatment change in small studies (Fuller et al., 2010; Henwood et al., 2018; Moulden et al., 2020), but no studies assessing prediction of recidivism. Our findings showed significant change on a greater number of subscales than demonstrated by Moulden and colleagues (2020) but results are difficult to compare due to our study interpreting raw scores rather than T-scores. Our findings regarding increased Positive Impression scores at post-programme are consistent with Henwood and colleagues (2018).

For the TRRG:SV, which was the only clinician administered measure used in this study, our results were consistent with Sowden (2013) in terms of demonstrating within-treatment improvements, but not for the prediction of recidivism. However, Sowden's (2013) research was conducted with a different population, people convicted of sexual offences, who received treatment for a longer period (6 to 8 months) meaning treatment effects were more likely to be more enduring.

Findings for the MCMI III and IV were somewhat difficult to compare with the limited previous research. Nevertheless, within-treatment changes align broadly with findings from Sissons (2013), who studied a similar justice-involved population. Additionally, our results were partially consistent with Calsyn and colleagues (2000), although their research on MCMI change specifically pertained to drug users undergoing medical addiction treatment, so is not directly comparable. Our findings regarding the lack of predictive capacity of the MCMI III and IV contrasts with research by Molina-Coloma and colleagues (2022), though this study was on people convicted of domestic violence.

Our findings also provide partial support for the 'Normal Quartet' outlined by Halon (2001) comprised of elevations in the Desirability, Histrionic, Narcissistic, and Compulsive personality traits. Desirability, Histrionic, and Narcissistic subscales all increased at post-programme, but the Compulsive subscale was slightly below the clinical cut-off suggesting a 'Normal Triplet' rather than a

'Normal Quartet.' Halon's (2001) findings, along with the more recent findings by Eastin and colleagues (2022), were both based on assessments of parental fitness. This is a distinct population compared to men convicted of violent offenses. While both populations have clear incentives for coming across well, the assessed parents had no reported criminal history, implying potentially different motivations for impression management. Unlike Brown and colleagues (2009) we did not find evidence of an association with measures related to impression management and a lower likelihood to recidivate.

Interpretation of Findings

There were several broad ways to interpret the overall findings. First, it is possible that the specific dynamic risk factors measures used in these studies do not tap into the crime-related constructs they purport to measure. In other words, they have poor construct validity. This is a concern raised by multiple previous researchers (Cording et al., 2016; Heffernan et al., 2019; Ward & Fortune, 2016).

A simpler explanation is that the failure to predict recidivism can be explained by a lack of effectiveness in the programme. Results from the Kilgour's (2015) evaluation of the Tai Aroha programme which analysed the first dataset in the present study concluded there was no evidence that completion reduced recidivism when compared to matched controls. We did not conduct analyses to compare possible differences in the predictive capacity of change scores between the two datasets, however, these were unlikely given the programme has not undergone major revisions over the combined assessed period. Another point is that while the programme is intensive, a 16-week period is relatively short compared to other programmes with the goal of rehabilitating people convicted of violent offending (Batley et al., 2021). It is also possible that treatment effects were transient, and potentially coupled with the therapeutic community and Tai Aroha accommodation. Therefore, upon reintegration, where their environment is less supervised and likely to include more triggers that can lead to criminal activity, treatment effects may diminish (Howard & van Doorn, 2018).

The predictive capacity of the PICTS subscale Discontinuity, and the MCMI Sadistic personality pattern warrants some discussion. Despite being nonsignificant, these elevations were noteworthy due to having more pronounced effects compared to the other assessed subscales. With unadjusted significance levels, of the 116 Cox regression analyses performed across all measures, only these two change scores predicted either general or violent recidivism. The Discontinuity construct refers to a lack of consistency in ones thinking and actions which can result in difficulties following through on commitments and good intentions. An example of an item for the Discontinuity subscale is "I start out with the best of intentions, but then something happens" (Walters, 2011, p. 34). Therefore, it appears that when participants improve their ability to behave consistently with prosocial thoughts and beliefs during the programme, this may result in a slight reduction in general recidivism. Sadistic personality is characterized by a disregard for others' feelings and an active orientation toward inflicting pain on others, which tends to be experienced as a source of pleasure (Grossman & Amendolace, 2017). It is possible that rehabilitative treatment focussing on anti-social cognitions within a group format may enhance empathy and insight into unhelpful motivations for behaviour. Notably, no effect was found for the Antisocial personality pattern, which has similarities to the Sadistic personality (Grossman & Amendolace, 2017). One possible reason for this is that treatment components may more effectively target motivations for behaviour rather than assist in managing impulsivity, which is a key feature of antisocial personality.

Our findings revealed that self-reports showed heightened levels of social desirability at the conclusion of treatment. Researchers have called attention to this pattern (Beggs, 2010; Cording et al., 2016; Heffernan et al., 2019). Olver and colleagues (2014a) put this well: "Many self-report measures can be quite transparent and face valid, and offenders were understandably motivated to present themselves in a favourable light at the end of treatment, which may also contribute to the inflation of change scores" (p.230). Juarez and Howard (2018) also noted that offenders might be more prone to impression management and other response biases upon completing treatment, aiming to demonstrate successful treatment gains or secure benefits such as parole. Motivations for

offenders to participate in treatment often include external gains, such as obtaining parole, as emphasized by several authors (Banse et al., 2013; Howard, 2016; Mann et al., 2010). The prospect of parole may be contingent on presenting evidence of positive treatment progress (Howard & van Doorn, 2018). For the Desirability validity scale on the MCMI, it appears that people made efforts to be more socially desirable at the end of treatment. This is also consistent with scores on the ADS Positive Impression subscale. However, as discussed previously, higher scores in Socially Desirable responding may be a positive indication (e.g., evidence of 'interpersonally oriented self-control'; Uziel, 2010). It is notable that scores on the two other validity subscales on the MCMI, Disclosure and Debasement, decreased while Desirability increased. This indicates that at the end of treatment, there is a significant reduction in self-disclosure and less inclination to admit negative traits or behaviours, but an increase in efforts made to appear socially desirable.

A final point; though we did not make specific hypotheses for actuarial risk tools (VRS Static, and RoC*RoI), the uniformity of nonsignificant results is surprising given results from previous research (e.g., Bakker et al., 1999; Wong & Gordon, 2006). For the RoC*RoI this may be due to low variability in scores (i.e., the standard deviation of RoC*RoI was .1, suggesting very similar scores within the sample) making it difficult to detect recidivism outcomes. The homogeneity likely is a result of the specified .7 or above RoC*RoI score programme entry requirement meaning most participants enter the programme with similar risk levels.

Implications

Our research contributes to the iterative process of empirical research which appears to be moving toward assessment and research techniques better able to reach a common scientific understanding of dynamic risk factors. The overarching goal increasing our understanding of dynamic risk factors is to assist researchers and practitioners in the correctional treatment space to improve service delivery and effectiveness of rehabilitation programmes. Our research also contributes to an important body of research with justice-involved people treated in the community

rather than in prison settings. The comparative lack of this kind of research has been highlighted by multiple authors (Baglivio et al., 2017; Davies et al., 2023; Lloyd et al., 2020).

None of the dynamic risk factors measured in this study demonstrated the steps required for causality to be achieved as outlined by Hanson and colleagues (2009). Each factor must a) predict recidivism b) be capable of change, and c) that change must be demonstrated to change recidivism outcomes. Of these, only b), the factors demonstrating changeability, was found in this study. This falls short of dynamic risk factors or psychopathology variables being considered causal factors for recidivism. As many authors have stated (e.g., Cording et al., 2016; Heffernan et al., 2019; Thornton, 2016; Ward, 2016), there is considerable evidence that overall, the causal status of dynamic risk factors as they are currently measured, appears to be increasingly unjustified. This research adds an incremental amount of data in support of this stance.

Strengths and Limitations

The availability of a large battery of psychometrics allowed for the assessment of a wide range of dynamic risk factor constructs, some with limited previous research. For instance, this appears to be the first study assessing whether change on the ADS predicts recidivism. A major methodological decision was to include all available dynamic risk factor subscales in lieu of presenting fewer subscales and/or total scores. While this meant that a more conservative significance level was adopted, it also allowed greater content coverage for tools often designed to be interpreted at the subscale level rather than higher-order scales. However, there were some limitations from the data received. For instance, no T-scores were provided for the ADS and scores for TRRG:SV were missing for about a third of the sample, limiting our ability to analyse these results more fully.

Cross-cultural validity was an important assessment issue and highly relevant to this study given the mostly Māori sample. However, the ethnic homogeneity meant that analysing ethnicity as a variable was not practical. Another limitation of this is that receiving cultural supervision for this research was unfortunately not practical due to time constraints.

While the total sample was a moderate size, there were several variables which appeared close to statistical significance which may have potentially reached significance with increased statistical power. It is not possible to know whether true effects would have been detected had this been the case.

Future Research

This is somewhat tautological given the framing of this study, but more high-quality research is required to continue our understanding of dynamic risk factors and their relationship to recidivism. Relevant to our findings, Hanson (2022) states that large confidence intervals typically indicate imprecision and a need for more research. This was very much the case for many of our study variables. Aside from this, while our findings on personality and psychopathology variables using the MCMI III and IV contributed to this endeavour, additional research is required to understand whether these factors are changeable and relevant to either dynamic risk factors or perhaps non-criminogenic needs.

More investigation into self-reporting bias including impression management is necessary as current thinking appears to treat it as either a potentially encouraging sign or a factor biasing results and requiring statistical controlling. While it is certainly possible the concept fits both criteria in different contexts, the topic would benefit from additional scholarship to work towards a more consistent approach when measuring within-treatment change among justice-involved people.

Conclusions

This study sought to investigate whether change scores on psychometric tools designed to measure dynamic risk factors and psychopathology were predictive of recidivism among graduates of an intensive violence prevention programme. Overall, while we did find evidence of significant within-treatment change across various tools, some of which have few published studies, these changes were not predictive of recidivism. Factors such as inflated within-treatment change scores possibly due to impression management and self-report bias, and a lack of programme effectiveness may contribute to these results. However, our findings do not support the notion that dynamic risk

factors are causal mechanisms of change that influence recidivism. The field of correctional psychology's 'holy grail'. according to Serin and colleagues (2013), remains elusive. While this study is by no means the final word on this topic, it contributes to an important empirical evidence base.

References

Allan, A., Dawson, D., & Allan, M. M. (2006). Prediction of the risk of male sexual reoffending in Australia. *Australian Psychologist*, *41*(1), 60-68.

Allan, M., Grace, R. C., Rutherford, B., & Hudson, S. M. (2007). Psychometric assessment of dynamic risk factors for child molesters. *Sexual Abuse*, *19*(4), 347-367.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (Vol. 5). https://doi.org/https://doi.org/10.1176/appi.books.9780890425596

Andrews, D. A., & Bonta, J. (2000). *The level of service inventory-revised*. Multi-Health Systems Toronto.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990a). Classification for Effective Rehabilitation. *Criminal Justice and Behavior*, *17*(1), 19-52. https://doi.org/10.1177/0093854890017001004

Andrews, D. A., & Wormith, J. (1984). *Criminal sentiments and criminal behavior: A construct validation*. Programs Branch, Ministry of the Solicitor General.

Andrews, D. A., Zinger, I., Hoge, R., Bonta, J., Gendreau, P., & Cullen, F. (1990b). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, *28*(3), 369-404.

Anstiss, B. (2003). Treatment at Reducing Re-Offending? *New Zealand journal of psychology*, *32*(2), 85.

Ara Poutama Aotearoa New Zealand Department of Corrections. (2023). *Quarterly prison statistics: March 2023*. Retrieved from https://www.corrections.govt.nz/resources/statistics/quarterly_prison_statistics/prison_stats_march_2023

Assink, M., Van Der Put, C. E., Hoeve, M., De Vries, S. L. A., Stams, G. J. J. M., & Oort, F. J. (2015). Risk factors for persistent delinquent behavior among juveniles: A meta-analytic review. *Clinical Psychology Review*, *42*, 47-61. https://doi.org/10.1016/j.cpr.2015.08.002

Baglivio, M. T., Wolff, K. T., Howell, J. C., Jackowski, K., & Greenwald, M. A. (2018). The search for the holy grail: Criminogenic needs matching, intervention dosage, and subsequent recidivism among serious juvenile offenders in residential placement. *Journal of Criminal Justice*, *55*, 46-57.

Baglivio, M. T., Wolff, K. T., Piquero, A. R., Howell, J. C., & Greenwald, M. A. (2017). Risk Assessment Trajectories of Youth During Juvenile Justice Residential Placement. *Criminal Justice and Behavior*, *44*(3), 360-394. https://doi.org/10.1177/0093854816668918

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

Bandura, A. (1976). *Social learning theory*. Englewood cliffs Prentice Hall.

Banse, R., Koppehele-Gossel, J., Kistemaker, L. M., Werner, V. A., & Schmidt, A. F. (2013). Pro-criminal attitudes, intervention, and recidivism. *Aggression and Violent Behavior*, *18*(6), 673-685. https://doi.org/10.1016/j.avb.2013.07.024

Barnett, G., Wakeling, H., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime & Law*, *19*(1), 85-110. https://doi.org/10.1080/1068316x.2011.607820

Batley, J., Kilgour, G., & Smith, M. (2021, 7 September 2021). *Group Therapy - High-risk Violent Offending*. Psychology of Criminal Conduct Lecture.

Beaudry, G., Yu, R., Perry, A. E., & Fazel, S. (2021). Effectiveness of psychological interventions in prison to reduce recidivism: a systematic review and meta-analysis of randomised controlled trials. *The Lancet Psychiatry*, *8*(9), 759-773. https://doi.org/10.1016/s2215-0366(21)00170-x

Beggs, S. (2010). Within-treatment outcome among sexual offenders: A review. *Aggression and Violent Behavior*, *15*(5), 369-379. https://doi.org/https://doi.org/10.1016/j.avb.2010.06.005

Beggs, S. M., & Grace, R. C. (2010). Assessment of Dynamic Risk Factors: An Independent Validation Study of the Violence Risk Scale: Sexual Offender Version. *Sexual Abuse*, *22*(2), 234-251. https://doi.org/10.1177/1079063210369014

Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: a comparative validity study. *Journal of Consulting and Clinical Psychology*, *79*(2), 182.

Berman, A. H. (2004). The Reasoning and Rehabilitation program: Assessing short-and long-term outcomes among male Swedish prisoners. *Journal of Offender Rehabilitation*, *40*, 85.

Berman, A. H., Gajecki, M., Morien, P., & Priestley, P. (2019). Measuring Psychological Change and Predicting Recidivism Following the Swedish One-to-One Program [Brief article]. *Frontiers in Psychiatry*, NA. https://link.gale.com/apps/doc/A606999918/HRCA?u=anon~2046ff2d&sid=googleScholar&xid=d55da7ac

Berry, S. (1998). An evaluation of the Montgomery House Violence Prevention Program. *Hamilton, NZ: Psychological Service, Department of Corrections*.

Biedermann, L., Eher, R., Rettenberger, M., Gaunersdorfer, K., & Turner, D. (2023). Are mental disorders associated with recidivism in men convicted of sexual offenses? *Acta Psychiatrica Scandinavica*, *148*(1), 6-18. https://doi.org/10.1111/acps.13547

Boccaccini, M. T., Marcus, D., & Murrie, D. C. (2017). Allegiance effects in clinical psychology research and practice. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 323-339.

Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation*, *6*(1), 1-22.

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct*. Routledge.

Bonta, J., & Andrews, D. A. (2024). *The psychology of criminal conduct* (Seventh edition. ed.). Routledge.

Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior*, *19*(3), 278-287. https://doi.org/https://doi.org/10.1016/j.avb.2014.04.014

Borkovec, T. D., Alcaine, O. M., & Behar, E. (2004). Avoidance theory of worry and generalized anxiety disorder.

Brown, S. L., St. Amand, M. D., & Zamble, E. (2009). The dynamic prediction of criminal recidivism: A three-wave prospective study. *Law and Human Behavior*, *33*(1), 25-45. https://doi.org/10.1007/s10979-008-9139-7

Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of personality and social psychology*, *38*(2), 217.

Bush, J., Harris, D. M., & Parker, R. J. (2016). *Cognitive self change: How offenders experience the world and what we can do about it*. John Wiley & Sons.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W., & Kaemmer, B. (2001). MMPI-2: Manual for administration, scoring, and interpretation (Rev. ed.). *Minneapolis, MN: University of Minnesota*.

Calliss, J. B. (2021). *Attitudes to Violence and Difficulties in Emotion Regulation as Dynamic Risk Factors for Physical Aggression and Violence Risk Among Violent Offenders*

Calsyn, D. A., Wells, E. A., Fleming, C., & Saxon, A. J. (2000). Changes in Millon Clinical Multiaxial Inventory Scores Among Opiate Addicts as a Function of Retention in Methadone Maintenance Treatment and Recent Drug Use. *The American Journal of Drug and Alcohol Abuse*, *26*(2), 297-309. https://doi.org/10.1081/ada-100100606

Carbajosa, P., Catalá-Miñana, A., Lila, M., & Gracia, E. (2017). Differences in treatment adherence, program completion, and recidivism among batterer subtypes. *The European Journal of Psychology Applied to Legal Context*, *9*(2), 93-101. https://doi.org/10.1016/j.ejpal.2017.04.001

Casey, S. (2016). Dynamic risk and sexual offending: the conundrum of assessment. *Psychology, Crime & Law*, *22*(1-2), 104-123. https://doi.org/10.1080/1068316x.2015.1111366

Chipollini, M. (2022). Are Narcissism and Entitlement Associated with Higher Rates of Violence and Unsuccessful Treatment Completion?

Cohen, J. (1988). Statistical power analysis for the behavioural sciences New York. *NY: Academic*.

Coid, J. W., Kallis, C., Doyle, M., Shaw, J., & Ullrich, S. (2018). Shifts in positive and negative psychotic symptoms and anger: effects on violence. *Psychological Medicine*, *48*(14), 2428-2438.

Cording, J. R., Beggs Christofferson, S. M., & Grace, R. C. (2016). Challenges for the theory and application of dynamic risk factors. *Psychology, Crime & Law*, *22*(1-2), 84-103. https://doi.org/10.1080/1068316x.2015.1111367

Coulter, D. J., Lloyd, C. D., & Serin, R. C. (2023). Psychometric Properties of a Risk Tool Across Indigenous Māori and European Samples in Aotearoa New Zealand: Measurement Invariance, Discrimination, and Calibration for Predicting Criminal Recidivism. *Assessment*, *0*(0), 10731911231153838. https://doi.org/10.1177/10731911231153838

Coupland, R. B. A., & Olver, M. E. (2020). Assessing Dynamic Violence Risk in a High-Risk Treated Sample of Violent Offenders. *Assessment*, *27*(8), 1886-1900. https://doi.org/10.1177/1073191118797440

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187-202.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of consulting psychology*, *24*(4), 349.

Cuevas, C., Wolff, K. T., & Baglivio, M. T. (2019). Dynamic risk factors and timing of recidivism for youth in residential placement. *Journal of Criminal Justice*, *60*, 154-166.

Cullen, F. T. (2013). Rehabilitation: Beyond Nothing Works. *Crime and Justice*, *42*(1), 299-376. https://doi.org/10.1086/670395

Daffern, M., Klepfisz, G., Sweller, T., & Day, A. (2019). Measuring the intermediate effects of offense-focused intervention on offenders. *The Wiley international handbook of correctional psychology*, 395-409.

Davies, S. T., Lloyd, C. D., & Polaschek, D. L. L. (2023). Does Risk Change Simply Indicate Current Risk for Recidivism or Something More? A Review of the Predictive Value of Intraindividual Change. *Criminal Justice and Behavior*, *50*(8), 1102-1119. https://doi.org/10.1177/00938548231174905

De Leon, G. (2000). The therapeutic community: Theory, model, and method.

de Ruiter, C., & Hildebrand, M. (2022). Violence Risk Assessment: Research and Practice. *Clinical Forensic Psychology: Introductory Perspectives on Offending*, 479-514.

De Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: predicting reductions in postdischarge community recidivism. *Law and Human Behavior*, *39*(1), 53.

DiGiuseppe, R., & Tafrate, R. C. (2004). *Anger disorders scale (ADS): Technical manual*. Toronto: Multi-Health Systems.

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-clinical-risk management-20, version 3 (HCR-20V3): development and overview. *International Journal of Forensic Mental Health*, *13*(2), 93-108.

Douglas, K. S., & Skeem, J. L. (2005). VIOLENCE RISK ASSESSMENT: Getting Specific About Being Dynamic. *Psychology, Public Policy, and Law*, *11*(3), 347-383. http://dx.doi.org/10.1037/1076-8971.11.3.347

Eastin, S. M., R. Brandwein, D., R. Marks, D., Safran, R., & Giordano, K. (2022). The MCMI-IV: a new normal quartet for parental fitness litigants. *Journal of personality assessment*, *104*(1), 57-63.

Edens, J. F., DeMatteo, D., Kelley, S. E., & Arnold, S. (2018). Personality assessment in forensic psychology. In *The SAGE handbook of personality and individual differences: The science of*

personality and individual differences. (pp. 548-572). Sage Reference. https://doi.org/10.4135/9781526451163.n26

Eisenberg, M. J., Van Horn, J. E., Dekker, J. M., Assink, M., Van Der Put, C. E., Hendriks, J., & Stams, G. J. J. M. (2019). Static and Dynamic Predictors of General and Violent Criminal Offense Recidivism in the Forensic Outpatient Population: A Meta-Analysis. *Criminal Justice and Behavior*, *46*(5), 732-750. https://doi.org/10.1177/0093854819826109

Fortune, C.-A., & Heffernan, R. (2019). The psychology of criminal conduct: a consideration of strengths, weaknesses and future directions. *Psychology, Crime & Law*, *25*(6), 659-674. https://doi.org/10.1080/1068316x.2018.1560445

Fuller, J. R., Digiuseppe, R., O'Leary, S., Fountain, T., & Lang, C. (2010). An Open Trial of a Comprehensive Anger Treatment Program on an Outpatient Sample. *Behavioural and Cognitive Psychotherapy*, *38*(4), 485-490. https://doi.org/10.1017/s1352465810000019

Gannon, T. A. (2016). Editor's Introduction to the Special Issue. *Psychology, Crime & Law*, *22*(1-2), 1-1. https://doi.org/10.1080/1068316x.2015.1116878

Gardner, B. O., Boccaccini, M. T., Bitting, B. S., & Edens, J. F. (2015). Personality Assessment Inventory scores as predictors of misconduct, recidivism, and violence: A meta-analytic review. *Psychological Assessment*, *27*(2), 534.

Gendreau, P., & Ross, R. R. (1987). Revivification of rehabilitation: Evidence from the 1980s. *Justice Quarterly*, *4*(3), 349-407.

Grant, B. A., Kunic, D., MacPherson, P., McKeown, C., & Hansen, E. (2004). The high intensity substance abuse program (HISAP): results from the pilot programs. *Correctional Service of Canada, Ottawa*.

Grossman, S. D., & Amendolace, B. (2017). *Essentials of MCMI-IV assessment*. John Wiley & Sons.

Gutierrez, L., Helmus, L., & Hanson, R. K. (2017). *What we know and don't know about risk assessment with offenders of indigenous heritage*. Public Safety Canada Ottawa, Canada.

Halon, R. L. (2001). The Millon Clinical Multiaxial Inventory-III: The normal quartet in child custody cases. *American Journal of Forensic Psychology*, *19*(1), 57-76.

Hanson, R. K. (2022). *Prediction statistics for psychological assessment*. American Psychological Association.

Hanson, R. K., Babchishin, K. M., Helmus, L., & Thornton, D. (2013). Quantifying the relative risk of sex offenders: Risk ratios for Static-99R. *Sexual Abuse*, *25*(5), 482-515.

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

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), 1.

Hardin, K. M., Contreras, I. M., Kosiak, K., & Novaco, R. W. (2022). Anger rumination and imagined violence as related to violent behavior before and after psychiatric hospitalization. *Journal of Clinical Psychology*, *78*(9), 1878-1895.

Harrell, J. F. E. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (Second edition. ed.). Springer Nature. https://doi.org/10.1007/978-3-319-19425-7

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, *20*(4), 315-335.

Heffernan, R., Wegerhoff, D., & Ward, T. (2019). Dynamic risk factors: Conceptualization, measurement, and evidence. *Aggression and Violent Behavior*, *48*, 6-16.

Helmus, L. M., & Babchishin, K. M. (2017). Primer on Risk Assessment and the Statistics Used to Evaluate Its Accuracy. *Criminal Justice and Behavior*, *44*(1), 8-25. https://doi.org/10.1177/0093854816678898

Henwood, K. S., Browne, K. D., & Chou, S. (2018). A randomized controlled trial exploring the effects of brief anger management on community-based offenders in Malta. *International Journal of Offender Therapy and Comparative Criminology*, *62*(3), 785-805.

Higgs, T., Olver, M. E., Nunes, K., & Cortoni, F. (2020). Risk relevance of psychometric assessment and evaluator ratings of dynamic risk factors in high-risk violent offenders. *Legal and Criminological Psychology*, *25*(2), 219-236. https://doi.org/10.1111/lcrp.12173

Hildebrand, M., Wibbelink, C. J. M., & Verschuere, B. (2018). Do impression management and self-deception distort self-report measures with content of dynamic risk factors in offender samples? A meta-analytic review. *International Journal of Law and Psychiatry*, *58*, 157-170. https://doi.org/10.1016/j.ijlp.2018.02.013

Hoge, R. D., & Andrews, D. A. (2011). *Youth level of service/case management inventory 2.0 (YLS/CMI 2.0): User's manual*. Multi-Health Systems Toronto, Ontario, Canada.

Hollin, C. R. (2012). A short history of corrections: The rise, fall, and resurrection of rehabilitation through treatment. In J. A. Dvoskin, J. L. Skeem, R. W. Novaco, & K. S. Douglas (Eds.), *Using Social Science to Reduce Violent Offending* (Vol. American Psychology-Law Society Series).

Howard, M. V., & van Doorn, G. (2018). Within-treatment change in antisocial attitudes and reoffending in a large sample of custodial and community offenders. *Law and Human Behavior*, *42*(4), 321.

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*(3), 163.

IBM, S. S. (2020). The Calculation of Bonferroni-Adjusted P-Values. In.

Jacobson, N. S., Martell, C. R., & Dimidjian, S. (2001). Behavioral activation treatment for depression: returning to contextual roots. *Clinical Psychology: Science and Practice*, *8*(3), 255.

Jacobson, N. S., & Truax, P. (1992). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research.

Johnston, P. (2021). Assessing risk of re-offending: Recalibration of the Department of Corrections' core risk assessment measure. *Practice: The New Zealand Corrections Journal*, *8*(1).

Juarez, T., & Howard, M. (2018). Assessing offender change over treatment: The influence of treatment context on self-reported antisocial attitudes. *Manuscript in preparation*.

Juarez, T., & Howard, M. V. A. (2022). Self-Reported Change in Antisocial Attitudes and Reoffending Among a Sample of 2,337 Males Convicted of Violent Offenses. *Criminal Justice and Behavior*, *49*(1), 3-19. https://doi.org/10.1177/00938548211013576

Kilgour, G. (2015). *Tai Aroha Evaluation Project Report* (0478180691).

Kim, B. (2022). Publication bias: A "bird's-eye view" of meta-analytic practice in criminology and criminal justice. *Journal of Criminal Justice*, *78*, 101878.

King, L. (2012). *Tai Aroha-the First Two Years: A Formative Evaluation of a Residential Community Based Programme for Offenders*. Department of Corrections.

Kingston, D. A., Olver, M. E., Harris, M., Wong, S. C. P., & Bradford, J. M. (2015). The Relationship between Mental Disorder and Recidivism in Sexual Offenders. *International Journal of Forensic Mental Health*, *14*(1), 10-22. https://doi.org/10.1080/14999013.2014.974088

Klepfisz, G., Daffern, M., & Day, A. (2016). Understanding dynamic risk factors for violence. *Psychology, Crime & Law*, *22*(1-2), 124-137. https://doi.org/10.1080/1068316x.2015.1109091

Klepfisz, G., Lloyd, C. D., Day, A., & Daffern, M. (2022). Increasing client motivation ratings across violence rehabilitation are promising predictors of reduced post-custody recidivism. *Psychology, Crime & Law*, 1-23. https://doi.org/10.1080/1068316x.2022.2108422

Klepfisz, G., O'Brien, K., & Daffern, M. (2014). Violent Offenders' Within-Treatment Change in Anger, Criminal Attitudes, and Violence Risk: Associations with Violent Recidivism. *International Journal of Forensic Mental Health*, *13*(4), 348-362. https://doi.org/10.1080/14999013.2014.951107

Koch, U., & Cratsley, K. (2020). Psychological Mechanisms. In (pp. 4145-4154). Springer International Publishing. https://doi.org/10.1007/978-3-319-24612-3_1562

Kroner, D. G., Mills, J. F., & Reddon, J. R. (2005). A Coffee Can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk. *International Journal of Law and Psychiatry*, *28*(4), 360-374. https://doi.org/https://doi.org/10.1016/j.ijlp.2004.01.011

Kroner, D. G., & Yessine, A. K. (2013). Changing risk factors that impact recidivism: In search of mechanisms of change. *Law and Human Behavior*, *37*, 321-336. https://doi.org/10.1037/lhb0000022

Lawing, K., Childs, K. K., Frick, P. J., & Vincent, G. (2017). Use of structured professional judgment by probation officers to assess risk for recidivism in adolescent offenders. *Psychological Assessment*, *29*(6), 652.

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*, *20*(2), 150-164.

Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annu. Rev. Law Soc. Sci.*, *3*, 297-320.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, *83*(404), 1198-1202.

Lloyd, C. D., Hanson, R. K., Richards, D. K., & Serin, R. C. (2020). Reassessment improves prediction of criminal recidivism: A prospective study of 3,421 individuals in New Zealand. *Psychological Assessment*, *32*, 568-581. https://doi.org/10.1037/pas0000813

Martinson, R. (1974). What works?-Questions and answers about prison reform. *The public interest*, *35*, 22.

Maruna, S., & Farrall, S. (2004). Desistance from crime: A theoretical reformulation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *43*(2), 171-194.

Mastromanno, B., Brookstein, D. M., Ogloff, J. R., Campbell, R., Chu, C. M., & Daffern, M. (2018). Assessing change in dynamic risk factors in forensic psychiatric inpatients: Relationship with psychopathy and recidivism. *The Journal of Forensic Psychiatry & Psychology*, *29*(2), 323-336.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall.

Millon, T., Davis, R. D., & Millon, C. (1997). *Millon Clinical Multiaxial Inventory III (MCMI-III)*. National Computer Systems.

Millon, T., Grossman, S., & Millon, C. (2015). *Millon Clinical Multiaxial Inventory-IV (MCMI-IV)*. Pearsons.

Mills, J. F., & Kroner, D. G. (2006). Impression Management and Self-Report Among Violent Offenders. *Journal of Interpersonal Violence*, *21*(2), 178-192. https://doi.org/10.1177/0886260505282288

Mills, J. F., Kroner, D. G., & Morgan, R. D. (2011). *Clinician's guide to violence risk assessment*. Guilford Press.

Molina-Coloma, V., Salaberría, K., & Pérez, J. I. (2022). A comparative study between recidivism offenders and non-recidivism offenders in a prison sample. *Anuario de Psicología Jurídica*, *32*(1), 33-39.

Morey, L. C. (1991). *Personality assessment inventory*. Psychological Assessment Resources Odessa, FL.

Morris, N., & Rothman, D. J. (1998). *The Oxford history of the prison: The practice of punishment in Western society*. Oxford University Press, USA.

Mossman, D. (1994). Assessing predictions of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*(4), 783.

Moulden, H. M., Mamak, M., & Chaimowitz, G. (2020). A preliminary evaluation of the effectiveness of dialectical behaviour therapy in a forensic psychiatric setting. *Criminal Behaviour and Mental Health*, *30*(2-3), 141-150. https://doi.org/10.1002/cbm.2158

Muirhead, J. (2016). Risky business: Evaluating the dynamic risk assessment for offender re-entry for use with New Zealand youth.

Newsome, J., & Cullen, F. T. (2017). The risk-need-responsivity model revisited: Using biosocial criminology to enhance offender rehabilitation. *Criminal Justice and Behavior*, *44*(8), 1030-1049.

Novaco, R. W. (1994). Anger as a risk factor for violence among the mentally disordered. In *Violence and Mental Disorder: Developments in Risk Assessmen* (pp. 21-59).

Nunes, K. L., Pettersen, C., Hermann, C. A., Looman, J., & Spape, J. (2016). Does change on the MOLEST and RAPE scales predict sexual recidivism? *Sexual Abuse*, *28*(5), 427-447.

O'Brien, K., & Daffern, M. (2017). Treatment Gain in Violent Offenders: The Relationship Between Proximal Outcomes, Risk Reduction and Violent Recidivism. *Psychiatry, Psychology and Law*, *24*(2), 244-258. https://doi.org/10.1080/13218719.2016.1209804

Olver, M., Nicholaichuk, T., & Wong, S. (2014a). The predictive and convergent validity of a psychometric battery used to assess sexual offenders in a treatment programme: An 18-year follow-up. *Journal of Sexual Aggression*, *20*(2), 216-239.

Olver, M. E., Beggs Christofferson, S. M., & Wong, S. C. P. (2015). Evaluation and Applications of the Clinically Significant Change Method with the Violence Risk Scale-Sexual Offender Version: Implications for Risk-Change Communication. *Behavioral Sciences & the Law*, *33*(1), 92-110. https://doi.org/https://doi.org/10.1002/bsl.2159

Olver, M. E., Kelley, S. M., Kingston, D. A., Beggs Christofferson, S. M., Thornton, D., & Wong, S. C. P. (2021). Incremental Contributions of Static and Dynamic Sexual Violence Risk Assessment: Integrating Static-99R and VRS-SO Common Language Risk Levels. *Criminal Justice and Behavior*, *48*(8), 1091-1110. https://doi.org/10.1177/0093854820974400

Olver, M. E., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C. P. (2014b). A psychometric examination of treatment change in a multisite sample of treated Canadian federal sexual offenders. *Law and Human Behavior*, *38*, 544-559. https://doi.org/10.1037/lhb0000086

Olver, M. E., Lewis, K., & Wong, S. C. P. (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*(2), 160-167. https://doi.org/10.1037/a0029769

Olver, M. E., Stockdale, K. C., & Wormith, S. J. (2014c). Thirty years of research on the Level of Service Scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, *26*(1), 156.

Olver, M. E., & Wong, S. C. (2019). Offender risk and need assessment: Theory, research, and applications. In D. L. Polaschek, A. Day, & C. R. Hollin (Eds.), *The Wiley international handbook of correctional psychology* (pp. 461-475).

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. (2019). A meta-analytic review of the efficacy of psychological treatments for violent offenders in correctional and forensic mental health settings. *Clinical Psychology: Science and Practice*, *26*(2), e12282.

Papalia, N., Spivak, B., Daffern, M., & Ogloff, J. R. P. (2020). Are Psychological Treatments for Adults With Histories of Violent Offending Associated With Change in Dynamic Risk Factors? A Meta-Analysis of Intermediate Treatment Outcomes. *Criminal Justice and Behavior*, *47*(12), 1585-1608. https://doi.org/10.1177/0093854820956377

Paulhus, D. L. (1998). *Paulhus deception scales (PDS): the balanced inventory of desirable responding-7: user's manual*. Multi-Health Systems North Tanawanda, NY.

Polaschek, D. L., Bell, R. K., Calvert, S. W., & Takarangi, M. K. (2010). Cognitive-behavioural rehabilitation of high-risk violent offenders: Investigating treatment change with explicit and implicit measures of cognition. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *24*(3), 437-449.

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology*, *17*(1), 1-17. https://doi.org/10.1111/j.2044-8333.2011.02038.x

Polaschek, D. L. L. (2017). Reporting change. In S. Brown, E. Bowen, & D. S. Prescott (Eds.), *The Forensic Psychologist's Report Writing Guide* (pp. 91-102). Routledge.

Polaschek, D. L. L. (2019). Treatment change. In R. D. Morgan (Ed.), *The Sage Encyclopedia of Criminal Psychology.*

Polaschek, D. L. L., Collie, R. M., & Walkey, F. H. (2004). Criminal attitudes to violence: Development and preliminary validation of a scale for male prisoners. *Aggressive Behavior*, *30*(6), 484-503. https://doi.org/10.1002/ab.20081

Polaschek, D. L. L., & Dixon, B. G. (2001). The violence prevention project: The development and evaluation of a treatment programme for violent offenders. *Psychology, Crime & Law*, *7*(1), 1-23. https://doi.org/10.1080/10683160108401779

Ryan, J. (2018). Editorial. *Practice: The New Zealand Corrections Journal*, *6*(2).

Sarre, R. (2001). Beyond 'what works?'A 25-year jubilee retrospective of Robert Martinsons famous article. *Australian & New Zealand Journal of Criminology*, *34*(1), 38-46.

Schmidt, S., Heffernan, R., & Ward, T. (2020). Why we cannot explain cross-cultural differences in risk assessment. *Aggression and Violent Behavior*, *50*, 101346.

Schmidt, S., Heffernan, R., & Ward, T. (2022). Why Dynamic Risk Factors Cannot Be Applied Universally: Their Normative Nature and the Importance of Cultural Awareness in Risk Assessment and Intervention. In *Challenging Bias in Forensic Psychological Assessment and Testing* (pp. 52-68). Routledge.

Sellbom, M., Flens, J., Gould, J., Ramnath, R., Tringone, R., & Grossman, S. (2022). The Millon Clinical Multiaxial Inventory-IV (MCMI-IV) and Millon Adolescent Clinical Inventory-II (MACI-II) in legal settings. *Journal of personality assessment*, *104*(2), 203-220.

Serin, R., Kennedy, S., & Mailloux, D. (2005). Manual for the Treatment Readiness, Responsivity, and Gain Scale: Short Version (TRRG: SV). *Unpublished manuscript. Carlton University*.

Serin, R., Mailloux, D., & Wilson, N. (2007). The dynamic risk assessment for offender re-entry (DRAOR). *Unpublished user manual*.

Serin, R. C., Chadwick, N., & Lloyd, C. D. (2016). Dynamic risk and protective factors. *Psychology, Crime & Law*, *22*(1-2), 151-170. https://doi.org/10.1080/1068316x.2015.1112013

Serin, R. C., Lloyd, C. D., Helmus, L., Derkzen, D. M., & Luong, D. (2013). Does intra-individual change predict offender recidivism? Searching for the Holy Grail in assessing offender change. *Aggression and Violent Behavior*, *18*(1), 32-53.

Shepherd, S. M., & Anthony, T. (2018). Popping the cultural bubble of violence risk assessment tools. *The Journal of Forensic Psychiatry & Psychology*, *29*(2), 211-220. https://doi.org/10.1080/14789949.2017.1354055

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, *31*(1), 8-22.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., & Grann, M. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, *13*(3), 193-206.

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499-513.

Sissons, M. K. (2013). *The Clinical Utility of Personality Profiles of High-Risk Offenders in Rehabilitation* Open Access Te Herenga Waka-Victoria University of Wellington].

Skeem, J. L., Schubert, C., Odgers, C., Mulvey, E. P., Gardner, W., & Lidz, C. (2006). Psychiatric symptoms and community violence among high-risk patients: A test of the relationship at the weekly level. *Journal of Consulting and Clinical Psychology*, *74*(5), 967.

Smith, P., Gendreau, P., & Swartz, K. (2009). Validating the Principles of Effective Intervention: A Systematic Review of the Contributions of Meta-Analysis in the Field of Corrections. *Victims & Offenders*, *4*(2), 148-169. https://doi.org/10.1080/15564880802612581

Sowden, J. (2013). *Examining the relationship of risk, treatment readiness, and therapeutic change to recidivism in a sample of treated sex offenders* University of Saskatchewan Saskatoon, Canada].

Statistics New Zealand Tatauranga Aotearoa. (2023). *Māori population estimates: At 30 June 2023*. https://www.stats.govt.nz/information-releases/maori-population-estimates-at-30-june-2023

Stevens, C. D. J., Tan, L., & Grace, R. C. (2016). Socially desirable responding and psychometric assessment of dynamic risk in sexual offenders against children. *Psychology, Crime & Law*, *22*(5), 420-434. https://doi.org/10.1080/1068316X.2015.1120868

Thornton, D. (2016). Developing a theory of dynamic risk. *Psychology, Crime & Law*, *22*(1-2), 138-150. https://doi.org/10.1080/1068316x.2015.1109092

Uziel, L. (2010). Rethinking Social Desirability Scales:From Impression Management to Interpersonally Oriented Self-Control. *Perspectives on Psychological Science*, *5*(3), 243-262. https://doi.org/10.1177/1745691610369465

van den Berg, J. W., Smid, W., Schepers, K., Wever, E., van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*, *30*, 179-191. https://doi.org/10.1037/pas0000454

Von Hirsch, A., Incarceration, C. f. t. S. o., & Gaylin, W. (1976). *Doing justice: The choice of punishments*. Hill and Wang New York.

Wakeling, H., & Barnett, G. (2014). The relationship between psychometric test scores and reconviction in sexual offenders undertaking treatment. *Aggression and Violent Behavior*, *19*(2), 138-145.

Walters, G. (2006). The psychological inventory of criminal thinking styles (PICTS) professional manual. *Allentown, PA: Center for Lifestyle Studies*.

Walters, G. D. (1995). The psychological inventory of criminal thinking styles: Part I: Reliability and preliminary validity. *Criminal Justice and Behavior*, *22*(3), 307-325.

Walters, G. D. (2002). The Psychological Inventory of Criminal Thinking Styles (PICTS): A Review and Meta-Analysis. *Assessment*, *9*(3), 278-291. https://doi.org/10.1177/1073191102009003007

Walters, G. D. (2011). *Crime in a psychological context: From career criminals to criminal careers*. Sage.

Walters, G. D., & Cohen, T. H. (2016). Criminal thought process as a dynamic risk factor: Variable-and person-oriented approaches to recidivism prediction. *Law and Human Behavior*, *40*(4), 411.

Walters, G. D., & Lowenkamp, C. T. (2016). Predicting recidivism with the Psychological Inventory of Criminal Thinking Styles (PICTS) in community-supervised male and female federal offenders. *Psychological Assessment*, *28*(6), 652-659. https://doi.org/10.1037/pas0000210

Walters, G. D., Trgovac, M., Rychlec, M., DiFazio, R., & Olson, J. R. (2002). Assessing change with the Psychological Inventory of Criminal Thinking Styles: A controlled analysis and multisite cross-validation. *Criminal Justice and Behavior*, *29*(3), 308-331.

Ward, T. (2016). Dynamic risk factors: scientific kinds or predictive constructs. *Psychology, Crime & Law*, *22*(1-2), 2-16. https://doi.org/10.1080/1068316x.2015.1109094

Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law*, *10*(3), 243-257.

Ward, T., Day, A., & Casey, S. (2006). Offender Rehabilitation Down Under. *Journal of Offender Rehabilitation*, *43*(3), 73-83. https://doi.org/10.1300/j076v43n03_05

Ward, T., & Fortune, C.-A. (2016). From dynamic risk factors to causal processes: A methodological framework. *Psychology, Crime & Law*, *22*(1-2), 190-202.

Webb, R. (2018). Rethinking the utility of the risk factors and criminogenic needs approaches in Aotearoa New Zealand. *Journal of Global Indigeneity*, *3*(1), 1-21.

Webster, C. D., Nicholls, T. L., Martin, M. L., Desmarais, S. L., & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START): The case for a new structured professional judgment scheme. *Behavioral Sciences & the Law*, *24*(6), 747-766.

Wilson, H. A., & Gutierrez, L. (2014). Does one size fit all? A meta-analysis examining the predictive ability of the Level of Service Inventory (LSI) with Aboriginal offenders. *Criminal Justice and Behavior*, *41*(2), 196-219.

Wong, S. C., & Gordon, A. (1999-2003). *Violence risk scale (VRS)*. Regional Psychiatric Centre and University of Saskatchewan.

Wong, S. C., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law*, *12*(3), 279.

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*(5), 740.

Yesberg, J. A., & Polaschek, D. L. L. (2014). Using Information From the Violence Risk Scale to Understand Different Patterns of Change:An Exploratory Investigation of Intensively Treated Life-Sentenced Prisoners. *Journal of Interpersonal Violence*, *29*(16), 2991-3013. https://doi.org/10.1177/0886260514527824

Yesberg, J. A., & Polaschek, D. L. L. (2019). How Does Offender Rehabilitation Actually Work? Exploring Mechanisms of Change in High-Risk Treated Parolees. *International Journal of Offender Therapy and Comparative Criminology*, *63*(15-16), 2672-2692. https://doi.org/10.1177/0306624x19856221

Yukhnenko, D., Blackwood, N., & Fazel, S. (2020). Risk factors for recidivism in individuals receiving community sentences: a systematic review and meta-analysis. *CNS Spectrums*, *25*(2), 252-263. https://doi.org/10.1017/s1092852919001056

Appendix One: Supplemental Tables

Table S1

*Predictors of General and Violent Recidivism with Standardised Residualised Change Scores for dynamic risk factors Measures Using Cox Regressions and Controlled for VRS static risk and RoC*RoI*

			General Recidivism		Violent Recidivism	
	Variables	***n***	**HR**	**95% CI**	**HR**	**95% CI**
ADS						
Behaviours	Brooding	152	1.00	[.84 – 1.20]	.88	[.70 – 1.11]
	Verbal Aggression	152	.97	[.82 – 1.15]	.88	[.71 – 1.11]
	Physical Aggression	152	1.00	[.83 – 1.21]	.87	[.68 – 1.11]
	Relational Aggression	152	1.01	[.86 – 1.18]	1.07	[.88 – 1.31]
	Passive Aggression	152	1.00	[.83 – 1.19]	1.04	[.83 – 1.29]
	Indirect Aggression	152	.98	[.82 – 1.18]	.88	[.71 – 1.11]
M	Revenge	152	1.04	[.87 – 1.25]	1.08	[.86 – 1.36]
	Coercion	152	1.00	[.84 – 1.19]	.97	[.77 – 1.21]
Cognitions	Suspiciousness	152	1.04	[.86 – 1.25]	1.02	[.81 – 1.29]
	Resentment	152	.98	[.81 – 1.18]	1.07	[.85 – 1.33]
	Rumination	152	.98	[.82 – 1.18]	.97	[.77 – 1.22]
	Impulsivity	152	.94	[.80 – 1.11]	1.00	[.79 – 1.25]
Arousa l	Physiological Arousal	152	.97	[.81 – 1.17]	.90	[.71 – 1.13]
	Duration of Anger	152	1.10	[.90 – 1.32]	1.04	[.84 – 1.30]
	Episode Length	152	1.10	[.91 – 1.31]	.98	[.78 – 1.23]
P	Hurt/Social Rejection	152	1.02	[.86 – 1.22]	1.04	[.84 – 1.30]
	Scope of Anger	152	1.10	[.91 – 1.33]	1.01	[.78 – 1.30]
	Positive Impression	152	1.13	[.93 – 1.37]	1.11	[.88 – 1.41]
	Total score	152	.98	[.80 – 1.20]	1.00	[.78 – 1.28]
CAVS						
	Total Score	152	.99	[.82 – 1.20]	.99	[.79 – 1.23]
PICTS						
Thinking styles	Mollification (Mo)	152	1.02	[.82 – 1.27]	.95	[.73 – 1.24]
	Entitlement (En)	152	.93	[.76 – 1.14]	1.02	[.78 – 1.34]
	Power Orientation (Po)	152	.94	[.78 – 1.13]	1.02	[.78 – 1.33]
	Superoptimism (So)	152	.96	[.79 – 1.15]	1.00	[.76 – 1.31]
	Sentimentality (Sn)	152	1.06	[.99 – 1.47]	1.17	[.93 – 1.08]
	Cutoff (Co)	152	.91	[.76 – 1.08]	.99	[.79 – 1.24]
	Cognitive Indolence (Ci)	152	.94	[.77 – 1.15]	.87	[.67 – 1.14]
	Discontinuity (Ds)	151	.83*	[.70 – .98]	.86	[.62 – 1.04]
Conten t scales	Reactive Criminal Thinking (R)	152	1.08	[.90 – 1.29]	.97	[.76 – 1.22]
	Proactive Criminal Thinking (P)	152	1.01	[.82 – 1.23]	1.03	[.78 – 1.35]
	General Criminal Thinking (GCT)	152	.88	[.74 – 1.04]	.86	[.68 – 1.08]
TRRG: SV						
	Readiness	152	1.03	[.82 – 1.29]	.90	[.69 – 1.18]
	Responsivity	152	.99	[.81 – 1.20]	.92	[.70 – 1.20]

Note. Subscale cluster labels on left-most column where applicable. M = Motives; P = Provocations; HR = hazard ratio; ADS = Anger Disorders Scale; PICTS = Psychological Inventory of Criminal Thinking; CAVS = Criminal Attitudes to Violence Scale; TRRG:SV = Treatment Readiness, Responsivity, and Gain: Short Version. *$p < .05$.

Table S2

*Predictors of General and Violent Recidivism of Standardised Residual Change Scores for Combined MCMI-III and IV Using Cox Regressions Controlled for VRS Static Risk and RoC*RoI*

			General Recidivism		Violent Recidivism	
	Variables		**HR**	**95% CI**	**HR**	**95% CI**
MCMI-III and IV						
Validity scales	X Disclosure	152	.91	[.74 – 1.12]	1.02	[.77 – 1.35]
	Y Desirability	155	1.05	[.89 – 1.25]	.97	[.77 – 1.22]
	Z Debasement	155	.98	[.81 – 1.20]	.92	[.72 – 1.19
Clinical personality patterns	1 Schizoid	155	.91	[.77 – 1.08]	.92	[.73 – 1.17]
	2A Avoidant	155	1.02	[.85 – 1.23]	1.05	[.83 – 1.33]
	2B Depressive	155	.95	[.79 – 1.14]	.96	[.75 – 1.23]
	3 Dependent	155	.91	[.72– 1.16]	.91	[.72 – 1.16]
	4 Histrionic	155	1.03	[.86 – 1.22]	1.11	[.87 – 1.43]
	5 Narcissistic	155	.88	[.73 – 1.05]	.92	[.71 – 1.19]
	6A Antisocial	155	.90	[.75 – 1.08]	.90	[.75 – 1.08]
	6B Sadistic	155	.85	[.70 – 1.03]	**.74****	[.58 – .95]
	7 Compulsive	155	1.17	[.96 – 1.45]	.99	[.77 – 1.30]
	8A Negativistic	155	.89	[.73 – 1.09]	.91	[.70 – 1.18]
	8B Masochistic	154	.99	[.82 – 1.19]	.98	[.76 – 1.25]
Severe	S Schizotypal	155	.86	[.71 – 1.04]	1.05	[.82 – 1.36]
	C Borderline	155	1.02	[.84 – 1.23]	.94	[.72 – 1.21]
	P Paranoid	155	.90	[.75 – 1.09]	.99	[.78 – 1.25]
Clinical syndromes	A Anxiety	155	1.04	[.85 – 1.26]	.99	[.78 – 1.26]
	H Somatoform	155	1.04	[.85 – 1.27]	.94	[.75 – 1.18]
	N Bipolar/ Manic	155	.95	[.78 – 1.15]	.92	[.71 – 1.19]
	D Dysthymia	155	.98	[.81 – 1.20]	1.05	[.82 – 1.33]
	B Alcohol	155	.91	[.75 – 1.11]	.90	[.68 – 1.18]
	T Drug Dependence	155	1.02	[.84 – 1.24]	1.05	[.80 – 1.38]
	R PTSD	155	.96	[.78 – 1.18]	1.03	[.81 – 1.33]
Severe	SS Thought Disorder	155	.98	[.81 – 1.18]	.95	[.74 – 1.23]
	CC Major Depression	155	1.08	[.88 – 1.32]	.94	[.74 – 1.20]
	PP Delusional	155	1.06	[.87 – 1.28]	1.11	[.86 – 1.46]

Note. Subscale cluster labels on left-most column. HR = hazard ratio; MCMI = Millon Clinical Multiaxial Inventory. * p < .05. ** *p* <.001.