Examining Trajectories of Change on Risk and Protective Factors among White and Black Men Offenders on Community Supervision in Iowa

by

Nicholas Chadwick

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University
Ottawa, Ontario

Abstract

Relying on the results of validated risk assessments is vital to evidence-based decision making in corrections. Advancements in the approach to risk assessment has seen an emphasis on measuring dynamic risk factors alongside protective factors, as both are expected to be useful for identifying treatment targets and measuring changes in risk over time. Despite these advancements, empirical evidence of change on dynamic risk and protective factors remains limited. Utilizing a sample of 3,976 White and Black men offenders on community supervision with a minimum of three waves of assessment, this thesis tested a three-step analytic approach to establish whether: (a) the Dynamic Risk Assessment for Offender Re-entry (DRAOR) measured the constructs of dynamic risk and protective factors consistently over time, (b) there was evidence of trajectories of within-offender change throughout community supervision, and (c) individual trajectories of change predicted recidivism above and beyond baseline DRAOR score and static risk. Results indicated that the DRAOR was measuring the same constructs in the same way over repeated assessments. Change was observed across each of the DRAOR domains, indicating that on average, offenders were expected to decrease in stable and acute risk over time, and increase in their protect score. Variations in the rates of change were unrelated to static risk, age, and race, but change on each DRAOR domain was partially explained by scores on the remaining domains. Change trajectories were significant predictors of revocations of community supervision after considering the effects of initial score, static risk, and age, but change trajectories were not related to new convictions. The findings indicate that the DRAOR is sensitive to change and

information on an offender's level of change should inform case management decisions, such as referrals for intervention. Although findings provide preliminary theoretical support for dynamic risk and protective factors, future research should consider an item-level analysis to identify whether there are specific factors that can be targeted to best support reintegration efforts. Further, increased attention on implementation efforts is needed to gain a thorough understanding of how the DRAOR is used in practice and how it can be optimized moving forward.

Acknowledgements

First, I want to express my gratitude to my supervisor, Dr. Ralph Serin. I am truly grateful for the years of mentorship that you provided, which were filled with support, advice, understanding, and opportunities. I am appreciative of the constructive feedback that I received from my committee members, Dr. Shelley Brown and Dr. Julie Blais, throughout the development of this research. I also wish to thank Dr. Ryan Labrecque, my external examiner, and Dr. Katharine Kelly, my internal examiner, for contributing their knowledge to strengthen the overall quality of this document.

This project would not have been possible without the support of Iowa Department of Corrections. In particular, I want to thank the members of the Research Department at the Central Office for their role in preparing and updating the datasets and always being willing to answer my questions.

Finally, the unwavering support, patience, and encouragement from my family and friends made accomplishing this milestone possible. Thank you for being a constant source of inspiration.

Table of Contents

List of Tables

List of Figures

Examining Trajectories of Change on Risk and Protective Factors among Men Offenders on Community Supervision

In Canada and the United States, the majority of the correctional population is supervised in the community (Kaeble & Cowhig, 2018; Statistics Canada, 2019). For example, across the United States, an estimated 6.6 million individuals were supervised by adult correctional systems in 2016, of these, 4.5 million were supervised in the community (Kaeble & Glaze, 2016). These statistics underscore the importance of developing and enhancing assessment tools used primarily for this population, as even relatively minor improvements can have a substantial impact. Offender risk assessment used at this stage of the criminal justice process (e.g., post-sentencing) is critical for determining appropriate supervision intensity (e.g., frequency of contact standards), programming targets and necessary dosage, and evaluating whether management and intervention efforts (e.g., supervision conditions, referrals to correctional interventions) continue to be appropriate by monitoring changes in risk. In order to fulfill the aforementioned functions, the assessment tool utilized must be theoretically relevant, sensitive to change, and reassessed scores need to be related to recidivism. Previously, predicting recidivism was the primary purpose of risk assessment. This meant that items that were assessed were largely static, unchangeable, factors (e.g., age at first arrest, number of convictions).

Although recidivism prediction remains an important consideration, especially to ensure public safety, relying on static factors precludes the ability to inform other important risk management decisions. As a result, there has been a growing interest in

developing and researching risk instruments that assess criminogenic items expected to change. A challenge with many of these studies is that the methodologies employed limit the information that is able to be gleaned from risk reassessments. Many studies utilize two time point designs, where a difference score is calculated and the relationship with recidivism is examined (e.g., Lewis, Olver, & Wong, 2012; Vose, Lowenkamp, Smith, & Cullen, 2009). Although change may be observed between two time points, methodological limitations prevent partitioning true change from change that may be due to measurement error. Additionally, patterns of change are unable to be examined unless 3 or more assessments are conducted (Brown, St. Amand, & Zamble, 2009; Singer & Willett, 2003). As a result, the knowledge generated from the majority of research conducted on purported dynamic risk and protective factors has been limited by methodology. The current study aims to enhance the understanding of dynamic risk and protective factors by applying rigorous statistical techniques to multi-wave reassessment data.

The current dissertation focuses on examining trajectories of change among offenders supervised in the community and scored on the Dynamic Risk Assessment for Offender Re-entry (DRAOR; Serin, 2007), which assesses a variety of stable and acute dynamic risk and protective factors. The relevance of static risk, age, race, and remaining DRAOR domains as covariates predicting heterogeneity in the trajectories of change (i.e., both initial score and rates of change) is explored. Change trajectories are then incorporated into prediction models to determine whether knowledge of trajectories of change over time enhances the prediction of recidivism. Previous work on

the DRAOR (namely Davies, 2019; Hanby, 2013; Lloyd, 2015; Lloyd, Hanson, Richards, & Serin, in press) investigated similar questions. However, Lloyd (2015) and Davies (2019) primarily focussed on the timing of reassessment and whether the most proximal assessment predicted imminent recidivism. Hanby (2013) examined trajectories of change on the DRAOR for recidivists versus non-recidivists but did not examine how scores on one domain may predict changes in scores on another. Relying on data from the US, this study seeks to extend the previous work in New Zealand and enhance the field's understanding of how dynamic risk and protective factors change throughout the course of supervision, and how that change may be relevant for case management efforts.

This dissertation begins with a review of the leading theoretical models for criminal behaviour and desistance from crime. Conceptual and definitional issues associated with dynamic risk and protective factors are highlighted, followed by a review of leading assessment tools that incorporate these factors. Finally, empirical evidence pertaining to measuring changes in risk over time is examined, with a particular focus on the various methodologies utilized.

Theoretical Underpinnings of Crime

Understanding the mechanisms that underlie the initiation, continuation, and eventual cessation from criminal behaviour is central to effective risk assessment and offender management. Without an understanding of why individuals engage in criminal behaviour, offender management efforts can easily be misguided. Andrews and Bonta's work on the psychology of criminal conduct (PCC) has been instrumental in shaping the

field's understanding of criminal behaviour. The leading theories for criminal behaviour are rooted in the social learning theory framework. Specifically, Bonta and Andrews (2017) argue that a General Personality and Cognitive Social Learning (GPCSL) theory best captures the various components that influence an individual's propensity to commit crime, while acknowledging that there are multiple paths to criminal behaviour. The Central Eight risk/need factors (Bonta & Andrews, 2017) form the major correlates of crime. These include criminal history, procriminal attitudes, procriminal associates, antisocial personality pattern, family and/or marital issues, school or employment issues, substance abuse, and leisure or recreation issues. Importantly, all but criminal history represent dynamic risk factors, allowing for changes in risk over time to be monitored, while also identifying targets for intervention to reduce reoffending. GPCSL theory emphasizes that procriminal attitudes and associates are the most salient influences on criminal behaviour.

GPCSL argues that individuals have agency over their behaviour, which is often goal-directed. As a result, identifying the anticipated consequences associated with a given behaviour is fundamental to understanding an individual's likelihood to commit crime. Within the GPCSL framework, consequences (rewards and costs) associated with a specific behaviour influence the likelihood of a similar behaviour happening again in the future. Consequences are often signalled (Bonta & Andrews, 2017), which means that they are the result of experiences that associate the behaviour with the rewards or costs. These signalled consequences can be informed by previous behaviour or by influential models (i.e., individuals that you respect or to whom you can relate). The

schedule, or density, of consequences is another important component that influences the probability of a given behaviour occurring. When rewards for prosocial behaviour outweigh the potential rewards associated with criminal behaviour, prosocial behaviour is likely the result. Under this theory, choosing a prosocial alternative to crime does not depend as much on motivation to do so, but rather that the costs associated with the criminal behaviour are greater than the rewards. The GPCSL highlights the interplay between individual factors (e.g., Central eight, cultural values) and environmental (e.g., geographical, political) factors when considering the availability and salience of rewards and costs for a given behaviour.

Overall, the GPCSL outlines a fluid process, whereby calculated rewards and costs ultimately influence a given behaviour. This assertion has received considerable research support, indicating that procriminal attitudes demonstrate one of the strongest relationships with reoffending (Bonta & Andrews, 2017; Gendreau, Little, & Goggin, 1996; Walters & Cohen, 2016). The GPCSL provides the foundation for the core criminogenic needs consistently identified as being associated with criminal behaviour. However, the criminogenic needs do not represent an exhaustive list of factors involved in the commission and eventual cessation (i.e., desistance) from crime. As a result, it is important to consider how alternative theories view crime and desistance to gain a more comprehensive understanding of the various factors that may relate to sustaining criminal behaviour. The discussion turns next to prominent theories of offender change and desistance in an effort to understand how these dynamic needs might change over time and contribute to leaving crime. It is important to note that leading scholars

maintain that the process of entering crime and leaving crime are dissimilar experiences, not merely opposite processes (Serin & Lloyd, 2009).

As discussed, the GPCSL posits that prosocial behaviour will prevail when the costs of procriminal behaviour outweigh the benefits. Healy (2010) interviewed a small sample of repeat offenders (*n* = 73) on probation to uncover the underlying factors associated with desistance and persistent offending. Largely, the results provided support for the GPCSL, in that desisters demonstrated improvements on a variety of factors that intuitively would shift the rewards for prosocial behaviour as opposed to crime. Specifically, Healy (2010) found that primary desisters (*n* = 45), defined as those who were self-reported to be crime free for a month, demonstrated significantly lower levels of pro-criminal attitudes, were less likely to endorse criminal thinking styles, and evidenced fewer general social problems in their life, compared to persistent offenders (*n* = 28). Despite being based on a small sample over a very short period of remaining crime free, these results provide preliminary support for the role of criminal thinking patterns in predicting primary desistance. Although these correlates of desistance appear to mirror typical risk factors (e.g., procriminal attitudes vs. prosocial attitudes), many desistance theories tend to emphasize different underlying mechanisms that might explain the observed changes in behaviour relative to GPCSL. An example of this is the theory of cognitive transformation (Giordano, Cernkovich, & Rudolph, 2002).

Within the desistance literature, the theory of cognitive transformation (Giordano et al., 2002) emphasizes the individual's effort to sustain a different way of life. This theory maintains that cognitive shifts are a necessary component of the

transformation process. The authors distinguish between four types of cognitive transformations that support the desistance process. The first requires an individual to be open to change; the second involves being receptive to experiencing a particular event (e.g., becoming employed); the third involves envisioning and implementing a "replacement self"; and lastly, the fourth cognitive change involves altering how the individual views deviant behaviour. According to this theory, desistance is relatively complete when the individual no longer supports the notion that deviant behaviour is acceptable. Giordano and colleagues (2002) found that events associated with social control theories (e.g., marriage, employment) were not directly related to desisting from crime. Rather, consistent themes emerged from qualitative interviews with offenders that demonstrated the importance of the cognitive transformations discussed above. The authors theorized that control processes, referred to as "hooks", vary in their ability to promote change, and the desistance process is unlikely to be sustained until the individual begins transforming their identity and cognitions. Although placing a greater emphasis on the role of the individual's identity in shaping behaviour, the primary tenants of this theory complement those of the GPCSL. Specifically, the idea that once an individual has transformed their cognitive self so that they no longer identify as deviant, parallels reward contingencies associated with a given behaviour. If an individual identifies as prosocial, the costs associated with returning to crime and re-transforming their identity would likely outweigh the potential benefits.

Rather than predominately focusing on the cognitive transformations that an individual navigates while leaving a criminal lifestyle, the life-course theory of crime

perspective views the initiation of crime and desistance as occurring within a social context. It posits that the process of exiting crime depends on variations in social control or social bonds (e.g., prosocial relationship with an adult; Laub & Sampson, 2001). Similar to reward contingencies in the GPCSL, the life-course framework views meaningful life events (e.g., work, marriage) as mechanisms that shift the opportunities for crime. Although this framework deemphasizes the importance of an individual's decision to commit crime, there are similarities with the GPCSL regarding the acquisition of social controls that shift the rewards to prosocial behaviour.

Approaching the initiation and cessation of criminal behaviour with a multi-pronged theoretical approach can assist in understanding the relevance of assessing a variety of dynamic risk and strength factors. Examining how these various factors interact with each other might also allow for improved differentiation between those who are initiating a process of exiting crime versus those who are still entrenched in a criminal lifestyle. As discussed later, each of these theories are represented in the various dynamic risk and protective factors that are the primary focus of the current study.

Conceptualizing Dynamic Risk and Protective Factors

Although the practical benefits of assessing dynamic risk factors are largely uncontested, some argue that the theoretical foundations for these factors has not been adequately established, ultimately affecting their utility (e.g., Harris & Rice, 2015; Ward, 2017). From a practical perspective, including dynamic risk factors is appealing, in that psychologically meaningful information is obtained that can be used to inform

treatment targets and monitor progress over time (i.e., through treatment or supervision). However, there remains ambiguity surrounding what exactly constitutes a truly dynamic risk factor. For some, dynamic factors have become synonymous with criminogenic needs (e.g., Andrews & Bonta, 2010), which are expected to reduce overall risk when improvements are noted. Others have argued that theoretically dynamic risk variables should additionally be able to identify precipitants of recidivism (Yang & Mulvey, 2010). Identifying precipitants of recidivism refers to the timing of the assessment, highlighting that there should be a proximal association between change on dynamic risk factors and imminent recidivism (Skeem & Monahan, 2011; Serin, Chadwick, & Lloyd, 2016).

Relatedly, Kraemer and colleagues (1997) outlined that increases in a given risk factor should be associated with an increased likelihood to experience criminal behaviour. In addition to classifying a factor as dynamic, Skeem and colleagues (Douglas & Skeem, 2005; Skeem & Mulvey, 2002) posit that consideration must be given to *risk status* and *risk state* to accurately represent an offender's overall level of risk. *Risk status* primarily focuses on static factors that are unable to change over time. These factors provide little utility in treatment planning and evaluation, but prove useful for differentiating between who is at risk for future crime (Serin et al., 2016). Alternatively, *risk state* involves an assessment of dynamic factors that are expected to change and be related to changes in the propensity to reoffend. These factors are often assessed in conjunction with static items to provide a comprehensive assessment of offender risk (Douglas & Skeem, 2005).

A framework for the expected temporal stability of dynamic factors has also been developed, which categorizes items as either stable or acute dynamic (Hanson & Harris, 2000). Stable dynamic items are not expected to change rapidly (e.g., procriminal attitudes), whereas acute dynamic factors represent items expected to fluctuate over short periods of time (e.g., intoxication, anger). As a result, it is hypothesized that intervention efforts should be directed at stable dynamic factors to yield sustained benefits, while acute dynamic factors likely serve as an important flag for supervising officers, as they are thought to signal the imminence of reoffending (Serin et al., 2016).

In addition to considering dynamic risk factors, there has been recent advocacy to include factors thought to represent positive attributes or resources that predict a decreased involvement in crime (Polaschek, 2017). These are commonly referred to as protective or promotive factors, or more generally known as strengths (Farrington, 2003; Jones, Brown, Robinson, & Frey, 2015). It is thought that the inclusion of strength factors can contribute to a more comprehensive understanding of criminal behaviour and potentially identify mechanisms associated with the initiation and continuation of desistance (Serin & Lloyd, 2009; Serin, Lloyd, & Hanby, 2010). However, these strength-based factors have been inconsistently conceptualized and measured. Specifically, the field has debated two interrelated issues pertaining to protective factors: (a) whether protective factors are conceptually distinct from risk factors, and (b) how protective factors interact with risk to reduce the likelihood of reoffending.

In the context of risk factors, protective factors have often been operationalized as either representing the opposite of a risk factor (i.e., the inverse) or the absence of a

risk factor (e.g., Borum, Bartel, & Forth. 2006; Rice & Harris, 2015). Protective factors considered to be the inverse of a risk factor can be scored as either a vulnerability or a strength, but not both simultaneously. Intelligence has often been used as an example of a factor that could be considered protective at high levels, while low levels could increase risk. Conversely, others have argued that protective factors are either non-linearly related or independent from risk factors (Polaschek, 2017). Farrington (2007) refers to a protective factor that exists without an opposing risk factor as a *free-standing* protective variable. Under this definition, a high level of the protective factor may predict low involvement in crime, but lower levels are unrelated to the likelihood of future crime. A third model, considers protective factors to be *buffering* if the factor weakens the expression of risk factors when they are strongly present, but are irrelevant, or the effect is diminished, when risk factors are weakly present (Farrington, Ttofi, & Piquero, 2016; Fougere & Daffern, 2011; Lodewijks, de Ruiter, & Doreleijers, 2010; Lösel & Farrington, 2012; Jones et al., 2015; Polaschek, 2017; Ullrich & Coid, 2011). Polaschek (2017) uses the example of high religious commitment to highlight this effect; at high levels of risk, a strong commitment to religion may reduce the likelihood of criminal behaviour, but it would not have the same effect (or could be irrelevant) at lower levels of risk.

Jones and colleagues (2015) present an appealing empirical framework for how strength-based factors relate to criminal behaviour. They use *strength* as an umbrella term that refers to a positive aspect of an individual's life that has the potential to buffer risk for criminal behaviour. *Promotive factors* are defined as variables that

demonstrate a negative relationship with criminal behaviour, regardless of the level of risk (Farrington, 2003). In order for a factor to be termed *protective*, Jones and colleagues (2015) posit that the variable must decrease the risk for criminal behaviour in higher risk individuals to a greater extent than in lower risk groups.

Relatedly, Serin, Chadwick, and Lloyd (2016) highlight the importance of considering an individual's level of agency when examining the relationship between strengths and recidivism. Merely possessing strength factors is insufficient to conclude that they will have a protective effect, but rather, the individual needs to believe they have the ability to choose an adaptive, non-criminal response when faced with an opportunity for crime. This position reiterates the role of identity and agency in the process of desistance from crime (Lloyd & Serin, 2012), and maintains that understanding potential strength factors must also account for the offender's attitudes towards crime and desistance (Serin, Chadwick, & Lloyd, 2016). Figure 1 presents a hypothesized conceptual model detailing how risk and protective factors are related and how they come together to impact the likelihood and imminence of criminal behaviour. This model posits that protective factors offset dynamic risk factors, and that all factors are filtered through internal mechanisms such as agency, identity, and hope. This then informs the likelihood and imminence of criminal behaviour. The current study is unable to directly test the proposed relationships between dynamic risk, protective factors, and the internal mechanisms, however results from the current study are intended to shed light on how offenders change on a series of dynamic risk and protective factors, and how that change may help differentiate between recidivists and nonrecidivists.

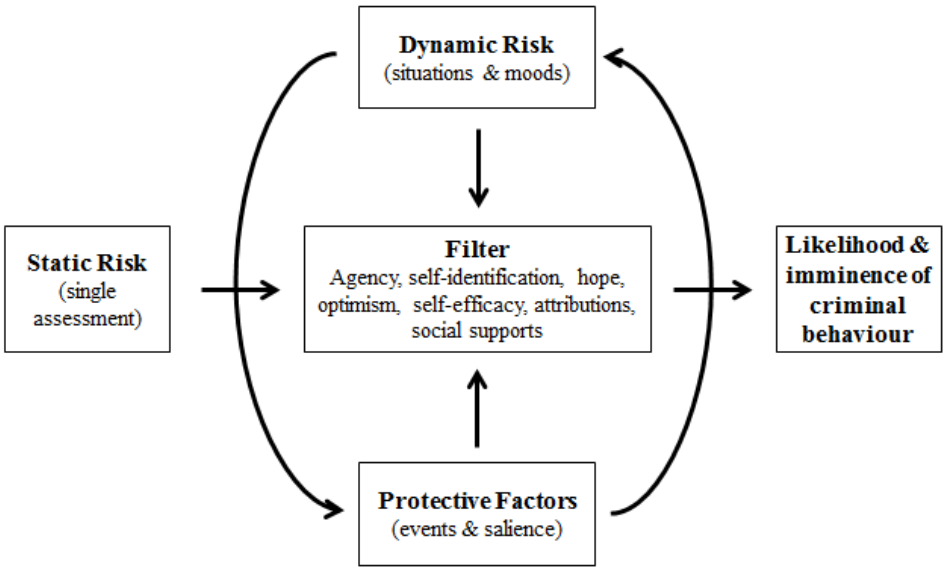

Figure 1. Model of dynamic risk and protective factors impacting criminal behaviour. Adapted from Serin, Chadwick, and Lloyd (2016).

Although the underlying conceptual framework for strength-based factors remains under development, they are increasing being included in risk assessment tools. A recent review of risk assessment tools (or standalone strength based tools) that measure offender strengths was conducted to assess whether the empirical evidence for each tool supported that the included strengths were operating as promotive or protective factors (Wanamaker, Jones, Brown, 2018). In their review of 48 studies that examined the relationship between recidivism and the strength component of a given tool, Wanamaker and colleagues (2018) highlighted that, despite often identifying strengths as protective factors, only a minority of the studies explored whether the

components exerted a true buffering effect. However, there was considerable support that the strength components were acting as promotive factors.

As the conceptual framework for risk and protective factors continues to develop, Polaschek (2016) recommends that these factors should not be investigated separately, but rather, emphasis should be placed on understanding how they work together. Preliminary results from a longitudinal study of dynamic risk and protective factors (New Zealand Parole Project) among high-risk offenders supervised in the community in New Zealand provides support for this recommendation. Pre-release scores of dynamic risk and protective factors predicted risk and quality of life in the community (made up of supervising officer's ratings of risk relevant domains), both of which predicted desistance at 12 months post-release (Polaschek, 2016). This suggests that a comprehensive assessment of both factors can enhance our understanding of why some offenders return to crime.

Despite the promising headway related to the inclusion of protective factors when predicting risk for criminal behaviour, it is important to acknowledge that the field has not unanimously embraced the notion that protective factors are independent from risk. Rather, some key scholars in the area of risk assessment strongly assert that protective factors are simply the inverse of risk factors, and do not need to be independently assessed as a result (Baird, 2009; Harris & Rice, 2015). Harris and Rice (2015) highlighted that increases in predictive accuracy, or the incremental predictive validity that has been observed by including protective factors could simply be that the assessment was capturing distinct risk factors (that are positively worded) that were not

otherwise being assessed. Similarly, Ward (2017) demonstrated concern that the field is moving too quickly to embrace the notion of protective factors, without having developed a strong theoretical understanding of how these factors inform behaviour. Ward argues that the practical implications associated with protective factors will be limited until a thorough theoretical model for dynamic risk and protective factors is developed. Jones and colleagues (2015), however, highlight that if the inclusion of these factors is adding practical utility, in that they are incrementally improving the prediction of outcome, then concern regarding how a strength factor interacts with risk should not hinder further research.

It is important to highlight that strength factors included in the DRAOR are termed protective. Although preliminary research has not supported that the domain score interacts with risk (e.g., Chadwick, 2014; Smeth, 2013), the term *protect* is used throughout this dissertation to align with the initial terminology utilized in the DRAOR. As Wanamaker and colleagues' (2018) review highlighted, previous research on the DRAOR has demonstrated that the protect domain operates as promotive factors, in that the domain score has been consistently negatively associated with recidivism.

Correlates of Desistance

As mentioned above, protective factors are often viewed as constructs that are expected to be related to the process of desistance. As a result, it is not surprising that many of the correlates of desistance are represented in research involving protective factors. Given that desistance appears to be a complex interaction between internal and external events (LeBel, Burnett, Maruna, & Bushway; Serin & Lloyd, 2009), adopting a

comprehensive assessment of these types of factors provides the opportunity to enhance our understanding of the relevance of risk and protective factors. Although the importance of individual and environmental predictors of desistance is an issue that remains unsolved (Farrington, 2007), empirical evidence is accumulating for certain correlates of desistance. One of the most robust correlates of desistance is age (e.g., Maruna, 2001; van Mastrigt & Farrington, 2009). The age-crime curve highlights the relationship between age and crime, such that as age increases, the frequency of criminal behaviour decreases (Blokland, Nagin, & Nieuwbeerta, 2005; Piquero et al., 2001; Sampson & Laub, 2003). Specifically, criminal behaviour peaks in late adolescence or early adulthood and tends to show a sharp decrease after age 30. The exact mechanism behind the relationship between aging and decreasing crime remains unknown, but some hypothesize that the process of maturing places greater emphasis on family and relationships, while others indicate it may simply be that older individuals no longer possess the physical agility to commit crime (Maruna, 2001).

Being engaged and satisfied in marriage (Blokland et al., 2005; Farrington, 1995; Maume, Ousey, & Beaver, 2005; Sampson & Laub, 2005) and having prosocial bonds are also related to desistance (Giordano, Cernkovich, & Holland, 2003). Good relationships are thought to protect against recidivism by providing a sense of purpose and direction, making the rewards for prosocial behaviour outweigh the rewards for crime (Giordano et al., 2003). Obtaining a quality job (Laub & Sampson, 2003; Paternoster, Bachman, Kerrison, O'Connell, & Smith, 2016; Stouthamer-Loeber, Wei, Loeber, & Masten, 2004; Uggen, 2000) and abstaining from substance use (Hussong et al., 2004) are other

external factors that appear to be related to desistance. Much like the other external desistance factors, it appears that the psychological meaning associated with these events is more relevant to the study of desistance (Lloyd & Serin, 2012). This is exemplified with the effect of employment on desistance, where employment that provides a sense of achievement and satisfaction is most likely to support desistance, rather than just having a job (Uggen, 2000). Similarly, it appears that abstaining from substance use is more than merely avoiding use, but rather involves an identity transformation that no longer involves using substances (HaggÅrd, Gumpert, & Grann, 2001).

Given that it appears the psychological meaning of many of these events is the key ingredient for understanding the offender change process, measuring internal psychological constructs is critical. Examples of some of these internal constructs include hope, motivation, identity, and outcome expectancies. Having hope and motivation has been associated with reductions in risk and eventual desistance (Martin & Stermac, 2010; Moulden & Marshall, 2005). Further, the importance of not having a criminal identity has been underscored (Maruna, 2001; Paternoster et al., 2016). For example, Maruna (2001) found that ex-offenders often classified their previous criminal behaviour as not reflecting their true self, whereas persistent offenders felt that they were doomed for failure. Although considerably less research has been conducted on internal mechanisms associated with desistance, it appears that agency (i.e., the belief that one has the capacity to control thought process, motivation, and action), outcome expectancies, and attributions about behaviour are relevant constructs for desistance

(Lloyd & Serin, 2012). Although an empirical foundation for these correlates of desistance is continuing to grow, there is improvement needed in the measurement of these constructs, particularly related to measuring within-individual changes and desistance.

Measuring Within-Individual Change

Various measurement techniques (e.g., self-report) of relevant constructs have been examined to attempt to uncover the relevance of intra-individual change and recidivism. Measures developed to assess psychological constructs (e.g., Criminal Sentiments Scale-Modified (CSS-M); Shields & Simourd, 1992; Measure of Criminal Attitudes and Associates (MCAA); Mills & Kroner, 2001; Psychological Inventory of Criminal Thinking Styles (PICTS); Walters, 1995) have been utilized to examine the relationship with recidivism. Results generally support the use of these measures, as evidenced by the significant relationship between total scores and recidivism (e.g., Walters, 2012; Walters & Lowenkamp, 2016). However, there is a scarcity of research examining whether these measures are sensitive to change, and whether changes over time are related to recidivism (Serin, Lloyd, Helmus, Derkzen, & Luong, 2013). A recent investigation of the General Criminal Thinking (GCT) score from the PICTS involved examining the impact of including scores from a second assessment in predicting new arrests for 35,147 male and 5,254 female offenders supervised in the community across US federal districts (Walters & Cohen, 2016). Results indicated that the inclusion of the second administration of the PICTS in a Cox proportional hazards regression incrementally added to the prediction of time to new arrest, while controlling for age,

criminal history, race, ethnicity, and initial GCT score. Those who demonstrated increases in GCT scores from time 1 to time 2 had higher rates of rearrest compared to those with no change or decreases, across all levels of initial GCT scores.

In an effort to identify the factors that represent the best candidates for exploring the process of offender change, Serin and colleagues (2013) conducted a review of studies that linked individual change scores on criminogenic needs to recidivism. Their review yielded 53 studies that examined the relationship between change across three broad treatment domains, including cognitive, violence, and substance misuse, and recidivism. Results indicated that decreases in criminogenic risk factors such as procriminal attitudes and associates, family problems, and negative emotion demonstrated large effects ($d > 0.50$) with recidivism within the cognitive treatment domain. These findings were largely replicated across the substance misuse and violence treatment domains. However, not all change scores related to these criminogenic factors demonstrated a significant association with recidivism. This suggests that scales used to assess these constructs may require revision, or may be relevant for use in some contexts or samples, but irrelevant for others.

Kroner and Yessine (2013) investigated whether individual change (pre-post treatment) across a series of self-report measures of criminal attitude was evident, and whether change was related to recidivism for 182 offenders supervised in the community who had completed a cognitive behavioural treatment program. Results indicated that a small proportion of the sample demonstrated change on scores after treatment (ranging from -2% to 14% across the various measures). An examination of

the change in scores and recidivism suggested that only the MCAA associates scale demonstrated a significant negative correlation (r = -.16) with a binary recidivism (yes/no) outcome. Change scores across the other measures of criminal attitude (i.e., MCAA, CSS-M, and Pride in Delinquency (PID); Sheilds & Whitehall, 1991) were not related to binary recidivism or the number of new offences. The general lack of an association between change on these self-report measures of psychological constructs and recidivism have encouraged some to consider whether information gleaned from dynamic risk assessments can better function as a measure of change.

Measures of Dynamic Risk and Protective Factors

As the conceptual framework for dynamic risk and protective factors continues to be refined, risk measures that incorporate these factors have surfaced. Interested readers are encouraged to refer to Serin and colleagues' (Serin, Lloyd, & Chadwick, 2016; Serin, Chadwick, & Lloyd, 2016) recent work for a thorough discussion of dynamic risk and protective items across contemporary assessments. What is noteworthy is that there appears to be relative consistency in the items measured, with most incorporating primarily internal factors (e.g., self-regulation, criminal orientation), with less focus on situational or external events (e.g., employment, gang membership; Serin, Chadwick, & Lloyd, 2016). Serin, Chadwick, and Lloyd (2016) argue that given the consistency across measures, it is likely most productive to future research to focus on how these items are defined and measured, rather than searching for additional items to include. Generally, these new measures have been developed for use with specific samples (e.g., forensic clients, juvenile offenders) and have not been applied to general adult offenders.

Notably, three measures have emerged for use with general offenders in corrections, these include the Inventory of Offender Risk, Needs, and Strengths (IORNS; Miller, 2006), the Service Planning Instrument (SPIn; Orbis, 2003), and the DRAOR. Additionally, the Short-Term Assessment of Risk and Treatability (START; Webster, Martin, Brink, Nicholls, & Middleton, 2004) and the Structured Assessment of Protective Factors for violence risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009) have received considerable research attention among adult forensic populations. A description of these instruments that incorporate dynamic risk and protective factors follows, along with a summary of the predictive utility of each scale.

The START is an SPJ approach that provides a framework for structuring the assessment of dynamic risk and protective factors. The START incorporates 20 items scored both as a vulnerability or an asset on a 3-point scale (0 – absent, 1 – possibly present, and 2 – present). Final risk estimates are then formulated for offender outcomes of violence against others, self-harm, suicide, unauthorized leave, substance abuse, self-neglect, and victimization. Desmarais, Nicholls, Wilson, and Brink (2012) found that total START strength and vulnerability scores predicted any inpatient aggression well (*AUC* = .76, *AUC* = .79, respectively). Additionally, including strength and vulnerability scores added incremental validity above the Historical Clinical Risk Management-20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997). Similar findings were obtained among a sample of 252 outpatient forensic clients, where total strengths, total vulnerability, historical items on the HCR-20, and the overall summary risk rating

represented the optimal prediction model (*AUC* = .62) of violent and criminal outcomes within 3 months of the START assessment (Troquete et al., 2015).

The SAPROF was developed as a dynamic addition to the HCR-20. The instrument consists of 17 items (2 static factors and 15 protective factors) that are assessed to identify potential targets for intervention. Each item is assessed on a 0 (not present and not protective) to 2 (clearly present and clearly protective) point scale. Items are grouped into 3 subscales, representing Internal (e.g., self-control, coping), Motivational (e.g., work, attitudes toward authority), and External (e.g., living circumstances, external controls) factors. Protective factors included in the SAPROF are defined as any characteristic, either within the individual or the environment, that reduces the risk of violence. This conceptualization aligns with promotive factors, as discussed by Jones and colleagues (2015) and Farrington (2003). In a retrospective validation de Vries Robbé, de Vogel, and de Spa (2011) found that SAPROF total scores were significantly related violent reoffending at 1, 2, and 3 year follow-ups (*rpb* = -.35, *rpb* = -.38, *rpb* = -.35). Results have also suggested that when combined with a dynamic measure of risk, the inclusion of protective factors among the SAPROF improved the overall prediction of recidivism (de Vries Robbé, de Vogel, & Douglas, 2013). Further, results from retrospectively coding patient files at the start of treatment and at the end of treatment suggested that scores on the SAPROF increased over time and change scores predicted violent recidivism at one year (*AUC* = .78; De Vries Robbé, de Vogel, Douglas, & Nijman, 2015).

The IORNS is a self-report questionnaire that includes 130 items that assess static risk, dynamic need, and protective strengths. Each item is responded to as either true or false. An initial validation study indicated that the domain scores of the IORNS were significantly related to LSI-R total risk scores (static r = .38, dynamic r = .43, and protective strength r = -.20; Miller, 2006). An 18-month follow-up on 111 general offenders indicated that the total scores (static + dynamic - protective strength) were able to differentiate between offenders who were not returned to custody versus those who were returned multiple times. Low base rates precluded the analysis of the predictive accuracy of the IORNS domain scores for this sample. Although preliminary results highlight that the IORNS has practical utility, it is important to differentiate this measure from the other leading assessments. Given that it relies on a series of true/false questions, it may be that change information is not as refined as what is provided through a rating scale.

The SPIn represents a comprehensive risk assessment tool that incorporates a series of risks, needs, and strengths. The SPIn includes a full version (90 items) and a Pre-Screen (35 items), which assesses risk, needs, and strengths. The SPIn Pre-Screen is used for triaging and classification purposes, and thus is appropriate for examining the predictive accuracy. The SPIn items are scored on a 6-point scale and responses are summed together to inform classification as low, moderate, and high for risk and strength separately. Total scores for the SPIn Risk and Strength domains have been demonstrated to be significantly related to recidivism (OR = 1.02, OR = 0.89, respectively) over an 18-month follow-up (Jones et al., 2015).

The DRAOR is a purely dynamic assessment protocol that does not include static risk considerations. The DRAOR includes 19 items that assess dynamic risk and protective factors. Dynamic risk items are organized according to the rate in which they are expected to change (i.e., stable—months, acute—hours or days), and protective items are conceptualized as being independent from risk factors. Although items are classified as *protective*, research on the DRAOR has largely failed to support that their effect is heightened for higher risk offenders (e.g., Chadwick, 2014; Smeth, 2013). The DRAOR has been implemented in New Zealand, and empirical findings from this population suggest that the subscale scores have predictive utility (*AUC*s ranging from .67 to .71). Research from US samples of offenders on community supervision has also supported the predictive utility of the DRAOR, although the magnitude of the effect tends to be smaller (*AUC*s ranging from .58 to .62 for general offenders). Encouragingly, results consistently support that the DRAOR significantly improves the prediction of recidivism above and beyond static risk (e.g., Chadwick, 2014; Hanby, 2013; Lloyd, 2015).

Dynamic Risk, Protective Factors, and Recidivism

Differing methodologies used to assess purported dynamic risk or protective factors has also contributed to the lack of clarity surrounding these issues. Some studies have employed single time-point designs while others utilized multi-wave studies to examine the benefit of multiple assessments over time. Single-wave studies are useful in that they can assess the relationship between a dynamic risk or protective factor and offender outcome, but they lack the ability to contribute to understanding whether the

factor is truly dynamic (i.e., scores can change and that change is associated with recidivism). Examining change scores between two assessments provides some indication that scores are fluctuating, but these largely fail to incorporate measurement error, thus clouding the confidence that true change has occurred. One promising technique, known as the Reliable Change Index (Jacobson & Truax, 1991), has emerged in corrections research designs that utilize difference scores (e.g., Kroner & Yessine, 2013; Olver, Beggs Christofferson, & Wong, 2015). Briefly, this technique incorporates measurement error in the form of scale reliability, to determine the magnitude of observed change that would be required to determine that it is not solely the result of error. This technique is often paired with examining clinically significant change, which categorizes participants' initial scores as either within dysfunctional or functional ranges on a measure, and then examines whether the change observed is large enough to sufficiently conclude that they have changed categories. Although this technique has applied benefits, particularly when evaluating treatment success, it falls short in being able to examine patterns of change over time. Multi-wave (i.e., 3 or more assessments) study designs represent the gold standard for modelling change over time (Brown et al., 2009; Singer & Willett, 2003), in that measurement error can be partitioned from observed change and patterns of change can be investigated, rather than focussing solely on difference scores. An important element that also needs to be considered in the examination of change across repeated assessments is whether the same constructs are measured in the same way over time (i.e., measurement invariance). The following section reviews empirical studies of either dynamic risk or protective factors that

utilized one of the aforementioned designs, with particular emphasis on multi-wave studies.

Single-wave studies. Single wave studies provide a snapshot of the potential utility of assessing factors anticipated to change. Although the importance of incorporating change cannot be assessed, these types of studies can provide support for why it is useful to consider other factors that are typically not included in standard risk assessment. An example of this is Zamble and Quinsey's (1997) comprehensive retrospective analysis of 311 offenders who reoffended after being released from prison. A small sample of offenders (*n* = 36) who did not reoffend after release from prison served as a control group in this study. Offenders were interviewed to understand their circumstances prior to committing a new offence. Participants were specifically asked about their living arrangements, employment, their leisure time, and relationships, in addition to responding to questionnaires that assessed depression, anger, anxiety, and substance use. Results indicated that compared to those offenders who remained in the community, reoffenders lived unsettled lives with frequent moves, frequent unemployment, and unstable relationships. When offenders were asked to recall potential problems that immediately preceded their most recent offence, the most commonly cited problems included interpersonal conflicts, substance abuse, and financial problems. Further, reoffenders noted that feelings of frustration, depression, anxiety, and anger were all common during the month preceding the reoffence. Overall it appeared that those who reoffended demonstrated poor coping skills compared to those offenders who successfully remained in the community. More recent studies have

supported these initial findings, suggesting that recidivists often demonstrate higher scores on dynamic risk factors as compared to offenders who successfully remain crime free (e.g., Brown et al., 2009; Howard & Dixon, 2013; Miller, 2006; Schlager, & Pacheco, 2011; Simourd, 2004).

To enhance our understanding of the effect of protective factors on violent reoffending, Ullrich and Coid (2011) conducted a study on 800 violent offenders released into the community. Dynamic protective factors were developed based on a review of the literature and included items that captured social support, emotional support, use of leisure time, accommodations, and financial and employment stability. Slightly less than a third of the sample (30%) violently reoffended throughout the 5.3 year average follow-up. Results indicated that 5 of the 15 assessed protective factors demonstrated a relationship with violent reoffending after release. These items assessed social support (OR = 0.36), emotional support (OR = 0.39), spare time spent with family or friends (OR = 0.26), involvement in religious activities (OR = 0.41), and closeness to others (OR = 0.35; Ullrich & Coid, 2011). The effect of these protective factors on the time to violent reoffending was also examined. Interestingly, social support and spare time spent with family and friends remained consistently related to outcome across time. Emotional support was significant for the first 2 years after release, and closeness to others was significant in the 2nd year, 3rd year, and after 3 years. This highlights the potential differential relevance of specific protective factors over time. It is important to note that the scores on the protective items were collected just prior to release and later in the community, but changes on the items were not

considered. The relationship between protective items and time to reoffending could differ depending on if changes on the protective items are incorporated. An additional strength of this study was that it examined the relationship between the individual protective item and reoffending, rather than aggregating the protective items together. In doing so, it moved the field closer to identifying specific factors associated with desistance.

Ullrich and Coid (2011) also investigated whether the protective effects of the 5 items differed across static risk levels. Results from a logistic regression examining the independent effects of static risk and the protective items indicated that time spent with family and friends retained a significant independent effect on reoffending. All other significant items from the univariate models lost their significance. No significant interactions between the levels of static risk and each protective item were noted. This suggested that the protective items were operating as promotive factors, as defined by Jones and colleagues (2015).

Jones and colleagues (2015) investigated the predictive utility of assessing strength factors alongside risk and need factors. A sample of Canadian offenders serving supervision orders (*n* = 3,656) were assessed on the SPIn Pre-Screen assessment (Orbis Partners, 2003). The SPIn Pre-Screen is a shortened version of a full risk assessment that is used primarily for triaging and classification purposes, and therefore is appropriate for evaluating predictive accuracy. The Pre-Screen includes 11 strength items (e.g., stable accommodations) and 39 risk (e.g., age at first arrest) and need items (e.g., relationships with antisocial peers). The results from a hierarchical logistic regression indicated that

risk (OR = 1.02) and strength scores (OR = 0.89) were significantly related to committing a new offence during the 18 month follow-up. These promotive effects were replicated in sub-analyses that considered gender and Indigenous ancestry. Additionally, an interaction term consisting of total risk and strength score was significant for the entire sample, and separately for males, females, and non-Indigenous offenders. This suggested that the aggregate strength domain differentially impacted the relationship between risk and outcome, such that high strength scores were particularly effective in reducing the recidivism rates for higher risk cases. This study provides a strong foundation for the predictive utility of incorporating strengths alongside risks. Additionally, this is one of the only studies with adult, general offenders, to find evidence that aggregate strength scores exerted a protective effect. Although this study provided unique insight into how strengths may operate alongside dynamic risk factors, particularly regarding predicting recidivism, given that only one assessment was used, the effect of change on these factors was unable to be explored.

An examination of dynamic risk and protective factors among 3 samples of juvenile violent offenders from the Netherlands produced similar results to Jones and colleagues (2015). Participants were assessed on the SAVRY at various stages of the justice system, namely pretrial, institution, and pre-release (n_1 = 111, n_2 = 66, and n_3 = 47; Lodewijks et al., 2010). The SAVRY considers static and dynamic risk and protective factors. Protective items include prosocial involvement, social support, attachment to prosocial others, attitude towards intervention, commitment to school, and resilient personality (e.g., cognitive skills, calm mood, self-esteem). The relationship between

SAVRY scores and outcome (violent recidivism for two of the samples, and violent institutional conduct for the other) was of primary interest to the researchers. Results indicated that the protective and dynamic scores were significantly related to outcome across all three samples (*AUC*s ranged from .69 to .81 for the dynamic scale, and .72 to .87 for the protective scale; Lodewijks et al., 2010). Interestingly, the static scores were not significantly related to outcome across any of the samples. The inclusion of protective scores significantly improved the prediction of outcome, above and beyond dynamic risk scores, for 2 of the samples, while the protective scale approached significance in the last sample. When protective factors were present for high risk offenders, there was a marked difference in recidivism rates (e.g., recidivism for sample 1 high risk, no protective factors = 40%, whereas high risk, one or more protective factors present = 6%). Although the sample size was small and a more rigorous test of the interaction could have been examined (e.g., including an interaction term in the logistic regression), these results provide preliminary support for a buffering effect of protective factors on risk and highlight the potential for increased predictive validity when assessing protective factors in addition to static and dynamic risk factors.

Although the results from these studies are encouraging, it is important to reiterate that the results were obtained from a single assessment of dynamic risk or protective factors. As a result, one cannot confidently conclude that these items are indeed dynamic factors. One promising consideration is that the assessment of these additional factors tends to improve the prediction of criminal outcome. This suggests that regardless of whether the items are changing, the inclusion of these additional

factors appears to be worthwhile from a recidivism prediction perspective, which can have important implications for initial case management decisions (e.g., frequency of contact) and resource allocation.

Difference scores. Results from studies that examine difference scores on dynamic risk factors have generally supported the notion that recidivists tend to evidence less improvement on risk, relative to non-recidivists (e.g., Chadwick, Serin, & Lloyd, 2015; Cohen & VanBenschoten, 2014; Cohen, Lowenkamp, & VanBenschoten, 2016; Howard & Dixon, 2013; Labrecque, Smith, Lovins, & Latessa, 2014; Lewis, Olver, & Wong, 2012; Olver et al., 2015; Miller, 2006; Raynor, 2007). As a result, the inclusion of dynamic risk, and change on those factors between assessments, has been found to enhance prediction models that otherwise exclusively focus on static risk. A recent examination of change scores for 363 offenders scored on the DRAOR in Iowa suggested that the addition of change incrementally added to the prediction of recidivism, such that a one-unit increase in change was related to a 14% increase in the likelihood of recidivism, after controlling for baseline dynamic risk (Chadwick et al., 2015).

Additionally, it appears that it is relevant to consider the individual's starting risk level in conjunction with their level of raw change (Cohen et al., 2016; Vose et al., 2013; Olver et al., 2017). Vose and colleagues (2013) found that moving from one risk level to the next (e.g., moving from high risk to moderate) for previously identified high risk offenders corresponded to a 16% reduction in the recidivism rate (Vose et al., 2013). Importantly, after controlling for relevant covariates (e.g., race, age, gender, risk, and supervision category), multivariate logistic regression analyses revealed that change on

the LSI-R was a significant predictor of recidivism. However, the change in effect size was minimal (R_N^2 = .04), which suggests that a more refined measurement strategy, including more frequent assessments or other measures, may be required to obtain a more substantial effect. Similar findings were obtained when difference scores on the Post Conviction Risk Assessment (PCRA; Administrative Office of the U.S. Courts, 2011) were examined among a sample of 32,647 offenders on federal supervision (Cohen et al., 2016). Scores from two assessments completed approximately 9 months apart, on average, were examined. Results indicated that decreases in overall PCRA score for offenders initially scored as high or moderate, or low/moderate risk, were associated with lower odds of arrest within 12 months, whereas reductions among those already considered low risk were not significant. The absence of an effect among low risk offenders could represent a floor effect, however, in that low risk offenders do not have as much room to decrease their total scores further. Nonetheless, the results highlight that PCRA total scores change between assessments and that total change is related to the odds of rearrest. Additionally, the effect of change is refined when the initial risk level is considered in combination with the observed change.

Olver and colleagues (2017) further demonstrated the importance of augmenting risk information with change scores. Without considering change, individuals with the same post-treatment score will be considered the same level of risk (assuming all else is equal). Recent results suggest that incorporating change scores can enhance the precision of estimated recidivism rates, with individuals who have demonstrated change obtaining lower 5-year sexual recidivism rates than compared to

offenders with the same post-treatment score but having not evidenced change, putatively throughout treatment (Olver et al., 2017).

One issue common to the studies discussed above is that the factor structure of the risk assessment was not investigated over time. Observed fluctuations in risk scores between assessments could be the result of the structure of the risk assessment changing over time (e.g., measurement error changes over time). Bergeron and Miller (2013) applied an analytic approach that tested that the structure of the risk assessment was consistent over time. Their results suggested that the structure of the IORNS remained constant over time, providing support that any observed change was due to true change, rather than the measure changing over time. When modelling the latent scores for the IORNS pre- and post-treatment, results from a structural equation model suggested that the scores were significantly different in the expected direction (i.e., dynamic risk decreased, strength scores increased). Importantly, this was one of the first studies to address the concern of the potential for invariant measurement properties influencing the observed change. As a result, it advanced the understanding of how dynamic and protective scores may truly change before and after treatment. However, given that there were only two measurement occasions, patterns of change over time, including the rate of change was unable to examined. In order to understand the patterns of change over time, multi-wave approaches are needed.

Multi-wave investigations of change. Various methodological approaches have been applied to investigating whether dynamic factors change across multiple measurement occasions. Lloyd (2015; see also Lloyd et al., in press) provided a detailed

outline of the prominent strategies to model change depending on the research objectives. Choosing an appropriate strategy also depends on whether the relationship between dynamic factors and outcome is of primary interest or whether it is examining change over time. The following studies have employed a multi-wave assessment design to study dynamic risk or protective factors. Although various analytical approaches were utilized across the studies, each contribute to the knowledge around the utility of assessing dynamic factors over time.

In order to accurately incorporate changes in risk into case management practices, a refined understanding of offender change is required. Simply identifying whether there were increases or decreases in risk is likely too generic. In an effort to identify whether reassessment of theoretically dynamic items predicted imminent recidivism, Lloyd (2015) examined scores from repeated assessments for a population of offenders (*n* = 3,421) supervised in the community after incarceration in New Zealand. Offenders were assessed on the DRAOR (Serin, 2007) at every meaningful contact with their supervising officer, which resulted in most offenders having an assessment completed weekly. This study sought to determine whether more proximal assessments were related to imminent recidivism (i.e., within 6 weeks of the most recent assessment; Lloyd, 2015).

A comparison of absolute change from baseline to censoring revealed that non-recidivists demonstrated approximately twice as much reduction in risk relative to recidivists, although recidivists still evidenced reductions in risk (Lloyd, 2015). Examining the average change week-to-week yielded different findings, indicating that future

recidivists maintained higher, but frequently changing scores, whereas non-recidivists tended to gradually decrease in risk (Lloyd, 2015). The predictive utility of incorporating repeated assessments was examined by adding the time-varying scores to discrete-time hazard models for each of the DRAOR subscales (i.e., stable, acute, protective) after controlling for baseline scores. Univariate models (i.e., examining each subscale independently) suggested that incorporating time-varying scores incrementally added to the prediction of recidivism. Although the magnitude of the increased predictive accuracy over baseline models was marginal (change in *c*-index (Harrell, 2015) = .00 for stable, .02 for acute, and .01 for protect), the inclusion of time-varying scores increased the variance accounted for by 4% for stable, 8% for acute, and 4% for protect. A multivariate model with stable, acute, and protect scores entered alongside static risk suggested that the dynamic model (considering time-varying scores) performed better than a baseline only model. Interestingly, the stable scores did not emerge as a significant predictor of recidivism, while controlling for static risk as well as acute and protective scores.

Lloyd (2015) also investigated whether proximal assessments were better predictors than taking the rolling average of all previous assessments. Results indicated that for stable, acute, and protective scores, the addition of the most proximal assessment added significantly to the prediction of recidivism. Analyzing the most proximal stable scores resulted in 3% more variance explained compared to the rolling average. The effects for acute and protective scores were less pronounced, although the variance explained in recidivism slightly increased (2%) when the most proximal acute

score was considered rather than the rolling average, whereas there was no increase in the effect size for protective scores. Overall, Lloyd's (2015) examination of the frequent reassessment of dynamic risk and protective factors and corresponding relationship to imminent recidivism suggested that taking the most proximal assessment was the optimal prediction strategy. It is important to note, however, that this study was interested in predicting short-term recidivism (up to 6 weeks) following a meeting between the offender and supervising officer. It may be that when the objective is to predict criminal behaviour over a longer time frame, an alternative combination strategy of reassessments is more effective.

Davies (2019) aimed to replicate and extend Lloyd's (2015) approach to examining the relationship between reassessments of dynamic risk and protective factors and imminent recidivism. Specifically, Davies' thesis aimed to determine: (a) whether reassessments were better predictors of imminent recidivism than baseline assessments, (b) whether aggregation across multiple reassessments was a better predictor of imminent recidivism compared to the most proximal score, and (c) the extent to which mean scores and change scores, when used alongside the most proximal assessment, enhanced the prediction of imminent recidivism. The sample consisted of 13,714 DRAOR assessments completed for 966 high-risk male parolees released from long prison sentences in New Zealand. All DRAOR assessments completed within 6 months of release that preceded any recidivism were examined using discrete-time hazard models. There was evidence of slight intra-individual changes between the first and last DRAOR domain scores, amounting to an average of half a point on the

stable and protective domains and nearly a full point on the acute domain. Results indicated that proximal measures of each DRAOR domain were better predictors of imminent recidivism compared to the respective baseline scores, however the effects were small and the predictive accuracy throughout most of the follow-up tended to be comparable. Importantly, models testing baseline versus proximal scores on the acute domain demonstrated a larger effect relative to models examining the stable and protect domains. A comparison of predictive accuracy for proximal scores compared to various aggregation techniques (e.g., rolling mean, mean of last 2 weeks) indicated that the proximal score tended to emerge as the optimal predictor. Lastly, an empirical test of the relationship between intra-individual change and imminent recidivism revealed that change scores on each DRAOR domain were significantly associated with recidivism after controlling for baseline scores. Results indicated that recidivism became more likely the less scores decreased over time. However, it is important to note that the inclusion of change scores into the prediction models resulted in a small improvement in predictive accuracy. Next, a test of the unique (i.e., controlling for proximal score) association between change and imminent recidivism suggested that very few change scores remained significant, none of which meaningfully improved the predictive accuracy. Taken together, the results highlighted that reassessment of dynamic risk and protective factors is important, but it appeared that simply obtaining the most current risk information is what contributed to the enhanced predictive accuracy, rather than accounting for intra-individual change that had occurred between the assessments.

An alternative approach to examining changes across repeated DRAOR assessments was employed by Hanby (2013). She examined a series of multilevel growth models to explore patterns of change over time and determined whether dynamic risk and protective information added to the prediction of recidivism above static risk. The stability of the factor structure of the DRAOR was first examined to ensure that any observed changes in scores were due to actual changes, rather than a changing factor structure. This was accomplished by examining confirmatory factor analysis models at the middle assessment (i.e., 2 assessments after the baseline assessment) and the last assessment (i.e., the most proximal assessment to study end date or recidivism date). Results suggested that the models adequately fit the data, indicating that the factor structure remained constant over time. On average, the sample had 29 (*SD* = 16.13) DRAOR assessments completed throughout the study period. Results from the multilevel growth models indicated that, on average, offenders decreased in risk (i.e., improved) throughout the follow-up (stable change = .003 per day, acute change = .01 per day, and protective change = .004 per day). Comparisons between recidivists and non-recidivists were made to examine the extent that DRAOR scores and rates of change were different across the groups. Results suggested that recidivists were scored higher on the stable domain, although the average rate of change did not differ across recidivism status. A similar trend was observed with the protective scores, whereby recidivists started lower (i.e., indicating fewer protective factors) than non-recidivists, but the average change over time was consistent across groups. Recidivists also scored higher on the acute domain and evidenced a flattened average reduction in scores over

time, relative to those who were not convicted throughout follow-up. The patterns of these findings remained consistent when considering only criminal reconvictions (excluding breaches of a supervision order).

Given that DRAOR scores evidenced change, it was of interest to determine whether the DRAOR was a useful predictor of recidivism at each time point. It was expected that the assessment closest to recidivism would be the strongest predictor, but the DRAOR scores would remain predictive across all time points. A series of Cox regression survival analysis models were constructed to examine whether later month assessments were stronger predictors of recidivism than earlier assessments. Stable risk scores were significantly related to reconvictions throughout the 12 months of assessment. Protective scores were only related to reconvictions for the first 4 months. Acute scores did not demonstrate a significant relationship with survival time, while controlling for average protective and stable risk scores. A comparison of the model effect size (R^2) suggested that average scores immediately following parole start had a stronger relationship between the covariates and survival time compared to later months. Examining the more stringent outcome of criminal recidivism yielded comparable findings. Specifically, average stable scores represented the strongest contributor, being significantly related to survival time in all but month 10. Protective factors remained a significant predictor for the first 4 months, and the acute scores did not emerge as a significant predictor across the 12 months of assessment. This method of assessing the predictive validity over time does not account for changes from one month to the next; rather, the newer information that is entered into the model is

viewed as original information, without considering the scores from the previous model. Utilizing a procedure, such as Cox regression with time dependent covariates would allow for an examination of how fluctuations in scores over time are related to the outcome. Despite the finding that later assessments were not stronger predictors of recidivism, the results from this study highlight the importance of the re-assessment of risk, as the patterns of change between recidivists and non-recidivists were discrepant. This suggests that incorporating this information could help improve the way that supervision officers respond to offenders on their caseload. For example, if an individual demonstrates an increase in their risk scores over the duration of a month, an officer should ensure that the offender has access to appropriate rehabilitative services to attempt to mitigate the increase in risk.

A similar approach was taken by Babchishin (2013) to examine whether there were changes in risk-relevant propensities (as measured by the Acute-2007; Hanson, Harris, Scott, & Helmus, 2007) for male sexual offenders (*n* = 317) on community supervision. This sample was utilized to examine whether the factor structure of the Acute-2007 remained consistent over time. Each offender had at least 3 assessments, which were completed every 2 months throughout a 6 month follow-up. Exploratory Structural Equation Modelling (ESEM; Asparouhov & Muthén, 2009) was used to ensure that observed changes in risk scores were not the result of the factor structure changing over time. Results indicated that the Acute-2007 was measuring the same construct over time (i.e., invariant thresholds and factor loadings), which provided support for examining changes in risk across repeated assessments.

Change in Acute-2007 scores was then examined through a series of multilevel models that included 575 sex offenders (6,568 assessments) supervised in the community. On average, total scores for the Acute-2007 demonstrated reductions per month throughout the follow-up. A series of univariate models were constructed to examine the extent to which covariates (e.g., age, static risk, stable risk) explained the variance in the rate of change in total Acute-2007 scores, controlling for baseline scores. Age was not significantly related to the rate of change in scores. Offenders with higher baseline (i.e., first assessment) Stable-2007 scores demonstrated significantly less change on the Acute-2007 than lower risk offenders classified by the Stable-2007. A similar trend was observed with static risk, as measured by the Static-99, whereby higher risk offenders demonstrated lower rates of change on the Acute-2007, relative to low risk offenders.

To determine whether the rates of change in the Acute-2007 scores were partially explained by the supervising officer, Babchishin (2013) examined a 3-level multilevel model. Results indicated that the rate of change did vary across supervising officers. Accounting for supervision officers when examining change in the Acute-2007 total score yielded 3.2% variance explained. To attempt to explain the variation in the rate of change across officers, a proxy for the officer's conscientiousness was created by assessing the extent to which officers provided the information that was requested of them. Conscientiousness, defined as completing the requirements of the research project (e.g., research materials, attending regular meetings), has previously been demonstrated to differentiate between officers who supervise offenders with improved

outcomes relative to offenders supervised by officers not fully participating in the research project following delivery of a structured training program based on the best practices for community supervision (Bonta et al., 2011). A similar definition of conscientiousness was utilized by Hanson and colleagues (2007) to investigate the effect of officer conscientiousness on the predictive accuracy of various sexual offender risk assessments. Results provided further support for considering the conscientiousness of officers, as the predictive accuracy of a combined STATIC-99/STABLE-2007 score increased substantially (*AUC* = .76 to *AUC* = .84) for sexual reoffence. Babchishin's (2013) results indicated that the proxy for conscientiousness did not explain the variation in changes in the Acute-2007 total scores, highlighting the need to collect alternative information pertaining to this third level (i.e., staff). Ideally, demographic information for supervising officers (e.g., years of experience, orientation towards rehabilitation) would be available to incorporate into the models to see whether these variables are related to different rates of change among offenders.

Babchishin (2013) also examined the extent to which incorporating changes in risk complemented recidivism prediction. Cox regression models that incorporated the initial assessment were utilized to examine the extent that adding repeated assessments provided unique information to the prediction of recidivism. Unique intercepts and slopes were calculated for all offenders, which were then used to estimate the Acute-2007 total scores for every 2 month interval following the start of supervision for a total of 12 months. Results suggested that the inclusion of repeated assessments incrementally added to the prediction of any recidivism and sexual recidivism across all

assessment intervals, but, when controlling for the initial assessment, reassessment information was not a significant predictor of violent recidivism. The prediction models examining the incremental utility of change information were exploratory in nature and were underpowered, which likely explains the lack of a significant finding related to violent recidivism.

Earlier evidence from Brown and colleagues (2009) demonstrated how the combination of dynamic risk and static risk information can improve the prediction of recidivism. Risk assessment data was gathered for 136 male federal offenders in Canada to assess whether change on measures of cognitive appraisals (e.g., extent that you are worried about employment), situational triggers (e.g., difficulty with marital issues), and response mechanisms (e.g., substance abuse) was occurring and the extent that accounting for such change improved the prediction of recidivism. Data were collected across three-waves, with the first occurring up to 45 days prior to release from custody and the remaining assessments occurring 1-month and 3-months after release in the community. Multivariate repeated measures analysis indicated that scores related to triggers and response mechanisms significantly improved throughout the follow-up for non-recidivists. Significant univariate level repeated-measures change was also observed for non-recidivists among 13 of the 18 hypothesized dynamic risk factors. Importantly, improvements were noted among many prominent risk factors, such as employment problems, leisure problems, global stress, perceived problems, negative affect, strong social support, positive coping ability, criminal associates, and substance abuse problems. A series of Cox regression models were then developed to evaluate the

predictive utility of the static risk and dynamic risk items. The best combination of items was retained to represent the optimal prediction model for a given cluster of items. A measure of static risk and the number of prison misconducts combined to form the best static risk prediction model (*AUC* = .81). When considering the response mechanism/trigger variables within the stepwise Cox regression, only substance abuse and impression management remained significant predictors of recidivism (*AUC* = .77). Lastly, the 18 purportedly dynamic risk factors were entered into a univariate time-dependent Cox regression. Significant predictors were forced to compete for variance, yielding a final solution that included substance abuse, employment difficulties, social support, single/unsupportive partner, and benefits of crime (*AUC* = .85). Combining the static and dynamic models resulted in an improvement in recidivism prediction (*AUC* = .89). Overall, this study provided empirical support for assessing dynamic risk factors and including them alongside static factors when predicting recidivism at the group-level. Additional support for the re-assessment of dynamic factors (within-individuals) was also obtained, suggesting that doing so can enhance the differentiation between recidivists and non-recidivists. Further, the results highlighted that dynamic factors can change rapidly, as assessments were taken just prior to release and during the initial months on supervision.

Wooditch, Tang, and Taxman (2014) explored changes in dynamic risk factors, including substance abuse, antisocial cognition, negative associates, family relations, employment and leisure activities among 251 probationers over a 12-month period. It was a primary interest to determine whether there are specific criminogenic needs that

are responsible for the greatest reduction in reoffending. Participants were assessed at 4 time points over the course of the 12 months. Conditional change regression models were used to determine whether intra-individual change on criminogenic needs during the first 6 months of assessment predicted changes in self-reported criminal behaviour. Those who demonstrated decreases in problem areas of family, finances, and alcohol use from baseline to 6 months were more likely to decline in criminal behaviour from 6 months to 12 months. These findings highlight that dynamic risk factors can change over a short period of time (less than 6 months) and that decreases on certain criminogenic need areas are related to subsequent decreases in criminal behaviour.

Horney, Osgood, and Marshall (1995) conducted interviews with 617 incarcerated male offenders to examine the relationship between changes in life circumstances (e.g., work, school, drug use) and the likelihood of future criminal behaviour. Retrospective month-to-month accounts (required a minimum of 10 months of information, *M* = 28.36 months) of criminal offences and life circumstances were provided to evaluate the extent that these social controls influenced offending behaviour in the short-term. Hierarchical linear modelling was utilized to allow for a within-individual analysis of factors that explain observed patterns of offending. Results indicated that changes in life circumstances were consistently related to short-term criminal behaviour, while controlling for probability of reoffending. Specifically, experiencing changes in the use of illegal drugs (OR = 6.1), living with a girlfriend (OR = 1.64), and school (OR = 0.48) was significantly related to any criminal behaviour. Moving from employed to unemployed was significantly related to committing a property crime

(OR = 1.28), as was heavy drinking (OR = 1.88) and illegal drug use (OR = 1.54). Although this study relied on respondents' accounts of their experiences prior to incarceration (and simply asked for dichotomous responding to each life circumstance), rather than structured assessments related to these domains, the results underscore the potential utility of capturing change information via self-report when determining risk to reoffend.

Although several studies have found that the inclusion of dynamic risk increased the overall prediction of recidivism, the results are not conclusive. An examination of static and dynamic risk factors among 133 male offenders released into the community in Texas suggested that change among dynamic factors was not related to offender outcome (Morgan, Kroner, Mills, Serna, & McDonald, 2013). Offenders were assessed a total of 7 times over 6 months in an effort to capture the fluctuations that may be apparent among dynamic factors. Results suggested that dynamic risk factors did not significantly improve predictive accuracy above static risk. Although change on the measures were not directly modelled (through difference scores or prediction models that account for the changes in scores), examining the assessment closest to outcome did not yield an improvement in outcome. There is the potential that the 6-month follow-up was too brief to have sufficiently captured offending, but these results highlight that identifying the optimal assessment schedule to enhance prediction has yet to be accomplished.

Summary. The majority of the extant literature examining dynamic risk and protective factors has examined change across two time points. Although results from

these studies provide an index of whether change is occurring and allow for assessing whether incorporating change significantly improves the prediction of recidivism, multi-wave studies are necessary to reliably measure patterns or trajectories of change over time. Although only a few multi-wave studies examining dynamic risk and protective factors have been conducted, results have largely supported that scores on these factors do demonstrate change throughout supervision. Recidivists tend to demonstrate either less improvement, or deterioration, on risk factors compared to non-recidivists. Including re-assessment data has consistently shown to improve the prediction of recidivism above and beyond models including static risk, but also over baseline models that include dynamic risk factors. Recent advancements in modelling the relationship between reassessment data and recidivism have demonstrated that the most proximal assessment tends to be the best predictor of imminent recidivism (Davies, 2019; Lloyd, 2015; Lloyd et al., in press). However, although these approaches highlight the importance of capturing more recent assessments, and change on risk and protective factors as a result, the methodology does not permit an examination of the patterns of change over time.

Current Study

The purpose of the current study is to further develop the field's understanding of offender change. By extending previous research conducted on the DRAOR, and utilizing data from a new sample, the study will examine: (a) whether items captured within the DRAOR are changing over time, (b) whether the rate of change can be explained by variables such as age, static risk, and race, and (c) whether accounting for

changes in risk enhances the overall predictive ability of the DRAOR. Analyzing trajectories of change will also enhance our understanding of the process of offender change and the relationship to future reoffending. Identifying patterns of change, and potential predictors of variations in change, will be useful for allocating resources to offenders who demonstrate characteristics associated with trajectories of change linked to heightened risk for recidivism. Lastly, understanding whether each of the DRAOR domains contribute unique information to the prediction of recidivism will be useful for informing the optimal use of the DRAOR in case management practices.

Examining change across repeated risk assessments requires a multi-pronged analytic process; one that requires building from a foundational question regarding whether the risk assessment is measuring the same construct over time, towards questions pertaining to rates of change and the relevance of incorporating change in predicting community outcomes. As a result, the following research questions and hypotheses are stated.

Research questions and hypotheses.

1. Will the DRAOR demonstrate sufficient psychometric properties and stability in factor structure over time?

   Hypothesis 1. Based on previous research (e.g., Hanby, 2013, Lloyd, 2015; Lloyd et al., in press; Yesberg & Polaschek, 2015) it is expected that the DRAOR's internal structure will be empirically supported and it will remain consistent over time. Results from other dynamic risk assessments have also found that the structure of the risk assessment remained constant over time (e.g., Babchishin,

2013; Bergeron & Miller, 2013). This will allow for an examination of change scores over the duration of the follow-up. Additionally, consistent with previous research on the DRAOR in Iowa (Chadwick, 2014; Smeth, 2013) and in New Zealand (e.g., Hanby, 2013; Lloyd, 2015; Yesberg & Polaschek, 2015) the psychometric properties (internal consistency, concurrent validity) of the DRAOR are expected to be sufficient.

2. Do offenders demonstrate changes in the stable, acute, and protect domains throughout the course of community supervision?
    a. Does static risk, age, and race predict DRAOR domain scores at the start of supervision and/or the rate of change over time?
    b. When examining a given DRAOR domain (e.g., stable), do scores on the remaining domains (e.g., acute and protect) predict initial score or rate of change over time?

Hypothesis 2. It is expected that DRAOR scores will change throughout the course of community supervision. Previous examinations of patterns of change on the DRAOR (e.g., Hanby, 2013; Polaschek & Yesberg, 2018) has found that stable and acute scores tend to decrease over time, while protect scores increase.

Hypothesis 2a. It is expected that there will be variation in both the initial domain scores at the start of supervision and the rate of change. Previous research on predictors of initial score or change has been mixed. Consistent with Hanby (2013), it is expected that individuals with higher static risk will start

community supervision with higher stable and acute scores, and lower protect scores. The examination of potential predictors of rates of change is largely exploratory. Hanby (2013) found that higher risk was associated with higher rates of change on acute scores, while Babchishin (2013) found that higher risk sex offenders demonstrated less change on a measure of dynamic risk. In both studies, age did not appear to be a consistent predictor of either initial status or change over time. Race is included as an exploratory predictor of initial status and/or rate of change.

Hypothesis 2b. Previous research on the DRAOR has not examined how scores on one domain may explain variations in initial scores or rates of change on another. Given the correlations between the domains, it is expected that offenders with higher stable scores will start supervision with higher acute and lower protect scores. The same pattern is expected with acute scores, whereas it is expected that those with higher stable and acute scores will have lower protect scores at the start of supervision. The relationship between domain scores and rates of change is exploratory, so no hypotheses are posited.

3. Does the inclusion of changes in risk and protective factors enhance the prediction of either revocations of community supervision or new convictions over baseline models?

Hypothesis 3. It is expected that results will replicate previous findings that reassessment models significantly improve the prediction of recidivism (e.g., Babchishin, 2013; Davies, 2019; Hanby, 2013; Lloyd, 2015). Incorporating

changes in risk and protective factors provides a more comprehensive and real-time estimate of the likelihood of recidivism, which is expected to translate to enhanced prediction over baseline models that do not capture change.

4. Does information about initial DRAOR scores and change over time improve the prediction of community outcomes above static risk assessments?

   Hypothesis 4. Based on previous research (e.g., Brown et al., 2009; Davies, 2019; Jones et al., 2015; Lloyd, 2015; Lloyd et al., in press) it is expected that including dynamic risk and protective factors alongside static risk will result in a significantly improved prediction model.

Method

Correctional Context

Probation and parole. This dissertation relied on an administrative dataset consisting of offenders supervised by Iowa's Community Based Corrections. In general, crimes in Iowa are classified as felonies or misdemeanors. Felonies consist of four classes (A through D), which descend in offence severity. Misdemeanors are classified in descending order as aggravated, serious, or simple. Both felonies and misdemeanors can carry uniform minimum and maximum penalties (Iowa Code, 2020). In cases that do not carry a mandatory prison or jail term, the court has the discretion to defer a judgement or sentence, or suspend a sentence and place the offender on probation. The court decides the length of probation and the associated conditions that must be followed. In felony cases that do not require a mandatory prison sentence, the court has discretion to sentence an offender to a minimum of two years and a maximum of five

years probation. In misdemeanor cases the minimum probationary period is one year, and the maximum period is two years (Iowa Code, 2020). Offenders placed on probation undergo an assessment process to determine an appropriate level of supervision to address community risk and case planning needs. In the case of an offender being sentenced to serve a period of confinement in prison, they may be eligible for parole before the expiration of their sentence, as determined by the board of parole (Iowa Code, 2020). Offenders placed on parole undergo the same assessment, case planning and treatment processes that are provided to probationers. The parole board can also grant an offender a work release, to facilitate a transition from prison to the community. Iowa Community Based Corrections also provides supervision for arrestees prior to disposition of their criminal charges, if deemed necessary, and provides special sentence supervision to sex offenders, which is either five years or lifetime, depending on the nature of the convicted offence.

During the timeframe that the study sample was drawn, the population of offenders served by Iowa's Community Based Corrections remained relatively consistent. At any given time throughout the year, there were approximately 30,000 offenders actively supervised in the community (see Table 2 for a detailed breakdown of demographic information for the offender population). Slightly more than 70% of the population tend to be supervised on probation orders, while approximately 13% are supervised on parole. During fiscal years 2014 thru 2016, the majority (approximately 70%) of community supervision orders that were closed were completed successfully (Iowa DOC, 2014; 2015; 2016).

Response to violations. The response to a potential violation of conditions associated with a community supervision order involves several decisions throughout the process. First, the supervising officer who learns of the violation discusses it with their supervisor to determine whether to proceed with filing a report of violation or whether an informal sanction is more appropriate. If it is determined necessary to proceed with the report of violation, the report goes to the necessary decision making body for processing. For those on probation, the court has several options available, it can: (1) choose to modify the conditions of probation, (2) sentence the probationer to a jail term while maintaining the probation status, (3) extend the period of probation for up to one year, and (4) revoke the probation status and require the offender to serve the jail or prison sentence imposed or any lesser sentence (Iowa Code, 2020). In the case of a violation of the terms of parole, a parole revocation hearing will take place, which will establish whether the violation occurred and whether the parole should be revoked. If the violation is established, the parole judge may continue the parole with or without any modification to the conditions of parole, or they may revoke the parole and require the offender to serve the original sentence imposed, or revoke parole and reinstate the parolee's work release status (Iowa Code, 2020).

DRAOR implementation in Iowa. The DRAOR began a pilot implementation in August 2010. A group of 36 community supervision officers representing all eight judicial districts volunteered to participate in the training. An initial training that emphasized the general literature on dynamic risk, along with the composition and scoring guidelines for the DRAOR was provided by the scale developer. An advisory group of

trained officers was established to help address any issues that arouse throughout the implementation. Following a successful pilot implementation in 2011, Iowa DOC began a system-wide roll-out of the DRAOR. This involved establishing a certified trainer who was responsible for credentialing all officers expected to score the DRAOR. An enhanced training program was established in June 2014 to address observed challenges with implementation, including scoring inconsistencies and improving the link between DRAOR scores and case management strategies. The use of the DRAOR was formally included in Iowa DOC's policy in April 2015, along with detailed quality assurance activities meant to ensure that the DRAOR is being scored appropriately and used as intended in supervision related decisions.

Procedure

Offenders sentenced to community-based corrections in Iowa are first assessed on the Iowa Risk Assessment Revised (IRR) to determine their initial level of supervision. Supervision contact standards and level of rehabilitative programming increase as the level of supervision increases. For example, offenders supervised at level 3 (i.e., low normal) are expected to report to their supervising officer monthly and are likely to receive moderate levels of correctional programming. Offenders supervised at level 5 (i.e., intensive) report to their supervising officer weekly and are expected to participate in intensive levels of programming. DRAORs are completed for all offenders supervised at level 3 or above, on probation, parole, work release, and Operating While Intoxicated (OWI) supervision. DRAORs are used to refine the initial level of supervision, if

necessary, and contribute to identifying treatment goals and case management strategies to reduce overall risk.

Policy stipulates that DRAORs are to be completed no later than 90 days from the date of sentencing. As part of ongoing case-management efforts, acute risk factors are expected to be reviewed, updated, and documented at least monthly, or more often if needed (State of Iowa, Department of Corrections – CBC-01, 201-40.1, originated May 2016, revised January 2019). Stable and protect scores are reviewed and updated at least quarterly. At each structured appointment, supervision officers are expected to rely on the results of the DRAOR to inform their approach to manage and reduce risk. Progress on each of the DRAOR items is documented, along with updated DRAOR assessments, in the DOC's information management system, known as Iowa Corrections Offender Network (ICON).

All supervision officers who supervise offenders classified as level 3, 4, or 5 are provided training on the DRAOR. Certified trainers were established in Iowa through a train-the-trainers model. Supervision officers who undergo training on the DRAOR must demonstrate proficiency before scoring the tool with the offenders they supervise. Policy establishes that ongoing proficiency is monitored through quality assurance methods. Specifically, at least two DRAOR assessments each year per supervising officer are reviewed by quality assurance specialists. One interview is audiotaped, which allows the auditor to score the DRAOR independently and assess the officer's scores for accuracy. If there is a score deviation of two on any DRAOR item (e.g., officer scores anger as a definite problem but the auditor scores anger as not a problem), the officer is

determined to be non-proficient and is provided with additional coaching and a skill development plan. Feedback is provided for all assessments that are reviewed, and ongoing training or booster sessions are made available to officers, along with updates and new information, when applicable.

All data used in this dissertation were collected through ICON, extracted by staff of Iowa DOC. The data files were then received and subsequently organized to facilitate analyses. Given that the completion of DRAORs is routine practice for Iowa DOC, recruitment efforts were not necessary and informed consent and debriefing forms were not administered. Identifying information was necessary in order to identify unique supervision periods for the same offender. However, a unique identification number was created as soon as feasible, allowing identifying information to be stripped from the datasets used throughout the analyses. Ethics approval to use the administrative data for research purposes was obtained from Carleton University (CUREB-B Clearance #104557) and Iowa DOC.

Materials

Demographic information. Demographic information is routinely collected when an individual comes into contact with Iowa DOC. The offender's age, sex, race, highest education achievement, and marital status was documented. Additional information pertaining to the charges associated with the index offence (e.g., violent vs. non-violent) and criminal history was recorded. The information management system also tracks the type of community supervision order an individual is serving (e.g., parole, probation, work release) and details the start and end dates of supervision.

Dynamic Risk Assessment for Offender Re-entry. The DRAOR (Serin, 2007) assesses stable and acute dynamic risk and protective factors. The three domains provide supervising officers with information that can be used to monitor an offender's current experiences on a consistent basis. This information is then expected to inform modifications to supervision strategies that respond to observed changes in an offender's situation. The DRAOR items are based on research in the correctional, desistance, and reintegration literature (see Hanby, 2013 for thorough discussion of item content). The DRAOR's 19 items are organized conceptually into three domains representing stable dynamic risk, acute dynamic risk, and dynamic protective factors.

All DRAOR items are scored on a 3-point scale with 0 reflecting *not a problem* and 2 representing *definite problem* for the risk items, and 0 indicating *not an asset* and 2 suggesting a *definite asset* for the protective items. As defined by Hanson and Harris (2000) the stable domain incorporates dynamic risk factors that are expected to change gradually over time, consisting largely of enduring traits and behaviour patterns. These stable dynamic factors were drawn from the work of Andrews and Bonta (2010) and Hanson and Harris (2000). Items included in this domain are: peer associations, attitudes towards authority, impulse control, problem-solving, sense of entitlement, and attachment with others.

The acute domain is comprised of seven items that reflect situations and behaviours that can change rapidly. It is thought that acute items possibly signal when criminal behaviour is imminent. Acute items were developed based on the work of Andrews and Bonta (2010), Hanson and Harris (2000), and Hanson and colleagues

(2007). Included items are: problematic substance use, anger/hostility, access to victims/opportunity to offend, negative mood, lack of employment, problematic interpersonal relationships, and unstable living situation.

The protective (i.e., protect) domain includes strength factors that are conceptualized as being independent from risk. The desistance literature informed the development of this domain, namely the work of Maruna (2001) and Sampson and Laub (2005). Items included in this domain are: being responsive to advice, having a prosocial identity, realistic expectations regarding reintegration, prosocial reward contingencies, having prosocial support, and being influenced by social supports.

Previous research on the DRAOR in Iowa suggests that the domain scores are valid predictors of technical violations for general offenders (*AUC*s for stable = .62, acute = 0.59, protect = .58; Chadwick, 2014; *AUCs* for stable = .63, acute = .65, protect = .66; Serin & Chadwick, 2019), sex offenders (*AUC*s for stable = .69, acute = .65, protect = .61; Smeth, 2013), and offenders convicted of intimate partner violence in Iowa (*AUCs* for stable = .65, acute = .65, protect = .62; Perley-Robertson, 2018). Additionally, research in New Zealand has indicated that the DRAOR domains predict new offences (*AUC* for stable = .62, acute = .67, protect = .62; Hanby, 2013) and new convictions (*AUC* for stable = .61, acute = .57, protect = .60; Yesberg & Polaschek, 2015). The DRAOR total score has also been found to predict reconviction equally for men and women offenders (Scanlan, Yesberg, Fortune, & Polaschek, 2020; Yesberg et al., 2015); although the relationship between each DRAOR domain and reconviction was different for men (*AUC*s for stable = .59, acute = .62, protect = .59) and women (*AUC*s for stable = .63,

acute = .58, protect = .64; Scanlan et al., 2020). DRAOR domain scores have also been found to predict offender outcome above and beyond static risk scores (e.g., Chadwick, 2014; Davies, 2019; Hanby, 2013; Lloyd, 2015; Scanlan et al., 2020; Smeth, 2013). Previous examination of the stability of the factor structure of the DRAOR over time has supported that it continues to measure the same constructs through repeated assessments (Davies, 2019; Hanby, 2013; Lloyd et al., in press; Yesberg & Polaschek, 2015).

As discussed previously, Lloyd (2015) and Davies (2019) found that incorporating time varying scores across assessments for New Zealand parolees improved the prediction of recidivism over baseline assessment scores. Further, Lloyd's (2015) results highlighted that the updated DRAOR scores incrementally predicted (except the stable scores) over static risk. Hanby (2013) conducted a thorough examination of the patterns of change across re-assessments for recidivists versus non-recidivists and found that generally, offenders decreased in their risk, and increased in protective scores, throughout time. Change between 2 time points has also been examined on the DRAOR domains among probationers in Iowa, which revealed that change is occurring and that including change, in addition to the total scores, significantly improved the prediction of recidivism (Serin et al., 2016). Overall, previous research on the DRAOR situates it as a valid measure for capturing changes in dynamic risk and protective factors for various correctional populations.

Iowa Risk Assessment Revised. The Iowa Risk Assessment Revised (IRR) relies on the Iowa Violence and Victimization Instrument (IVVI) as a baseline assessment of risk of

reoffending. The IVVI is a predominantly static measure of risk that is used to inform the intensity of community supervision required to safely manage the offender. The IVVI consists of 13 items that relate to the index offence, the volume and seriousness of criminal history, whether the individual has confirmed gang membership, and current age (Prell, 2013). The IRR supplements the IVVI by considering stability factors that assess the pervasiveness of alcohol or drug use, and accommodation and employment issues, in addition to incorporating whether there is a history of revocations of community supervision (Prell, 2013). Each item is scored on a victimization risk and a violence risk scale. Violence risk focuses on predicting the likelihood of a conviction for any new violent crime within 30 months following admission to community supervision. Victimization risk incorporates risk for violent crimes, but also includes property offences where there are quantifiable economic costs associated with the criminal behaviour. The results inform the classification of offenders into 5 risk levels (administrative, minimum, low normal, high normal, and intensive). Classifying risk is a two-step process that begins with examining the victimization score, and then upwardly adjusting the risk level according to the violence score, if appropriate. Given that the victimization score forms the basis of the risk level, and is related to criminal behaviour more broadly, the victimization score was used in this dissertation as the measure of static risk in any multivariate model.

In a validation study of the IVVI, which included 1,961 male probationers and parolees, offenders tended to be evenly distributed across intensive (23%), high normal (27%), low normal (19%), and minimum (30%) risk levels (Prell, 2013). As expected, rates

of any new crime or technical violation within 3 years of supervision start were highest for those classified as intensive (69%), followed by high normal (45%), low normal (35%), minimum (23%), and administrative (9%). Both violence and victimization scores were moderately related to any recidivism over a 30-month follow up (*AUC* = .63 for violence scores, *AUC* = .65 for victimization; Prell, Vitacco, & Zavodny, 2016).

Recidivism. Available data facilitated analysis of two types of recidivism: (a) revocation of community supervision, and (b) conviction for new criminal charges incurred throughout the community supervision period. Details pertaining to each outcome, along with the base rates, are provided below.

Revocation of community supervision. Supervision periods were examined to determine whether they ended with a revocation due to a breach of the conditions of supervision. Only revocations that resulted in being removed from community supervision were considered for this outcome.[1] As a result, if an offender was being supervised on parole, subsequently had that order revoked, and then was immediately released on a work release, they did not contribute to this outcome. In this case, the supervision period was treated as a continuous period of time since the offender remained at risk in the community and DRAOR assessments continued to be scored and provided. A total of 1,087 (27%) offenders had their community supervision period end with a revocation. The reason for revocation was not available. The average time to the

[1] Note that the definition of revocation of community supervision used here closely aligns with what some jurisdictions may refer to as termination of community supervision, as the offender's community supervision order has been terminated or suspended as a result of violating the conditions of their community supervision order.

end of community supervision resulting from a revocation was 480 days (*SD* = 287.93). For those that did not have their community supervision revoked (*n* = 2,913), the average follow-up was calculated as the time between supervision start and supervision end date, or the date of data collection (April 1, 2019), whichever came first. However, it was possible to be convicted of a new charge without having a revocation of supervision. In this case (*n* = 187), the end of follow-up for the revocation outcome was the date of the offence. On average, the follow-up for those without a revocation was nearly 2 years (*M* = 687.45, *SD* = 350.90).

Conviction from new charges. Official records of new charges for which an offender received a sentence (i.e., charges resulting in a conviction) were recorded within Iowa DOC's information management system, which were retrieved April 1, 2019. Any new charges that were incurred at least 30 days after the start of community supervision and no more than 100 days after the end of the supervision period were considered as recidivism.[2] The 30-day period was selected to mitigate against identifying a new charge as reoffending when in fact it is associated with the current supervision order.[3] In the case of multiple convictions throughout the supervision period, the first conviction was selected, regardless of offence severity. In the case of multiple charges occurring on the same date, the most serious charge was retained.

---

[2] The presence of new convictions was examined over the course of the supervision period. Any conviction that was associated with a date falling 100 days after the supervision period ended was retained, given the close proximity to the supervision period.

[3] This resulted in removing 6% of instances of a new conviction (324 of 5,805), keeping in mind that a given offender could have multiple convictions listed over time. Notably, 206 offenders were treated as nonrecidivists due to this requirement, despite having a new conviction within 30 days of the start of supervision.

A total of 427 (11%) offenders were convicted of a new charge during the follow-up. The average time to a new conviction was slightly less than one year (*M* = 319.19 days, *SD* = 180.69). The most serious offence type associated with the first reconviction was most commonly classified as public order (e.g., operating a vehicle while intoxicated, unlawfully at large; 37%, *n* = 158), followed by violent (26%, *n* = 109), property (17%, *n* = 74), drug (18%, *n* = 76), and other (2%, *n* = 10). Given the small number of offenders who were convicted of a violent offence, analyses were restricted to examining new convictions for any type of charge. Follow-up time was calculated for non-recidivists based on either the date their community supervision period ended or the date of data collection (April 1, 2019), whichever came first. On average, non-recidivists (*n* = 3,573) were followed for nearly 2 years (*M* = 668.57 days, *SD* = 343.66).

Participants

Sample selection. Data were provided for all active offenders in Iowa as of November 2016 with a completed DRAOR assessment on file. Updated DRAOR assessments were gathered again in April 2019 for any offender identified in the original extract. This amounted to 103,443 DRAOR assessments representing 18,938 unique offenders serving community supervision orders. DRAOR assessments were provided for individuals already under community supervision, which resulted in receiving data for offenders who began their period of community supervision ranging from 1997 to 2018. However, the vast majority (85%) of supervision orders began between 2014 to 2016. During this time, it was possible for an offender to begin distinct community supervision orders. For example, an offender could have been under a community supervision order

for one year beginning in 2014 and then later sentenced to another order that commenced in 2016. Distinct supervision periods were defined as either having a longer than six-month gap between the end of one community supervision order and the start of the next, or having two community supervision orders interrupted with a term of confinement in either jail or prison. Any change in community supervision status (e.g., parole to work release) that did not result in the offender being removed from the community was treated as a continuous period of community supervision. Given that the primary aim of this dissertation was to examine changes in DRAOR scores throughout the supervision period, it was determined that the level of analysis would be on unique supervision sequences, rather than unique offenders. Examining periods of community supervision for the 18,938 unique offenders indicated that there were 19,850 distinct community supervision sequences (18,052 had one supervision sequence, 860 had two distinct supervision periods, and 26 had three distinct supervision periods). Only sequences that had a completed DRAOR assessment that occurred after the start of supervision were retained for further consideration. As a result, 19,293 community supervision sequences were identified, representing 18,674 unique offenders.[4] Information on new convictions or revocations of community supervision was then compared to the supervision sequences to ensure that DRAOR assessments were completed before any new conviction or termination of supervision occurred. Of the valid supervision sequences with at least one completed DRAOR

[4] Some offenders did not have a DRAOR assessment completed following the start of their community supervision order, explaining the decrease in the number of unique offenders eligible for inclusion in the analyses.

assessment, 2,349 had a new conviction associated with it. Importantly, there were 1,574 supervision sequences that had a new conviction associated with it, but there were no completed DRAOR assessments prior to the new conviction. As a result, these supervision sequences were removed form the sample. This resulted in a sample of 17,719 unique supervision sequences representing 17,279 unique offenders.

This sample was then assessed against the eligibility criteria established for arriving at the analysis sample (see Figure 2). The exclusion criteria were developed to establish an analysis sample that remained statistically robust, while maintaining generalizability to the relevant offender subpopulations in Iowa. Since this was the first large-scale study of the DRAOR in a U.S. community supervision sample, it was preferable to exclude women offenders to more readily compare the results from the current study to previous research on the DRAOR in New Zealand and the general literature on dynamic risk and protective factors. It is important to note that previous research examining dynamic risk factors for men and women offenders remains mixed in terms of whether core domains are gender-neutral or if they vary as a function of gender (Manchak, Skeem, Douglas, & Siranosian, 2009). As a result, the examination of change across dynamic risk and protective factors among women offenders likely requires consideration of an additional body of gender-responsive literature as well as the potentially unique experiences of women on community supervision. Since women represented approximately 17% of the sample, rather than having them represent the minority of the sample, it was decided that a focussed analysis of change on the DRAOR

for men was more appropriate.[5] Similar logic was applied when excluding supervision sequences associated with offenders identifying as either Hispanic (5%), Indigenous (1%), Asian or Pacific Islander (1%), and those with missing race information (0.02%). Risk assessment research has increasingly focussed on examining the suitability of a given tool for use with a particular racial group (e.g., Skeem & Lowenkamp, 2016). That being said, research remains limited in terms of identifying whether dynamic risk and protective factors are equally relevant across racial subgroups. Given that the sample sizes for these groups were too small to facilitate independent meaningful examinations of change, their inclusion in the overall sample could be misleading by generalizing findings predominantly based on Black and White offenders to subsamples that may reasonably differ in terms of their experience on community supervision, or in terms of how the DRAOR performs for these groups (i.e., do the items perform equivalently across racial groups?).

Next, supervision sequences were removed from the analysis sample if they did not have 3 completed DRAOR assessments after the start of supervision. Ensuring that each sequence had 3 assessments allowed for a more robust examination of trajectories of change. Characteristics of offenders contributing supervision sequences in the remaining (n = 7,182) were generally comparable on risk relevant characteristics (e.g., age at start of supervision, type of supervision, months sentenced to prison, index offence type) as those removed (n = 6,481), although offenders with eligible supervision

[5] It is important to note that the sample size of women who met the remaining eligibility criteria was n = 808, which further supports the decision to examine the results for men only.

sequences did have slightly higher static risk scores (*M* = 8.11, *SD* = 3.93 vs. *M* = 7.74, *SD* = 4.13, *t*(13,661) = -5.19, *p* < .001, *d* = 0.09).

Among the remaining supervision sequences (*n* = 7,182), only those with the first DRAOR completed within 90 days of starting supervision were retained for analysis (*n* = 3,115 supervision sequences were removed). This requirement was established to ensure that baseline assessments were accurately capturing the offender's first experiences on community supervision. This transition to community supervision is a challenging time where offenders are typically at a heightened risk for reoffending, particularly for offenders released from custody (Brown et al., 2009). A comparison on all risk relevant characteristics mentioned above revealed that offenders representing the unique supervision sequences in the analysis sample (*n* = 4,067) were similar as those excluded (*n* = 3,115), although the analysis sample continued to have a slightly higher average static risk score (*M* = 8.57, *SD* = 3.79 vs. *M* = 7.45, *SD* = 4.04, *t*(6,756) = -11.64, *p* < .001, *d* = 0.29). Lastly, the final analysis sample required that a completed static risk assessment was associated with the supervision sequence. This resulted in the removal of 67 supervision sequences, bringing the final analysis sample to 4,000 supervision sequences with 28,023 assessments, representing 3,976 unique offenders.

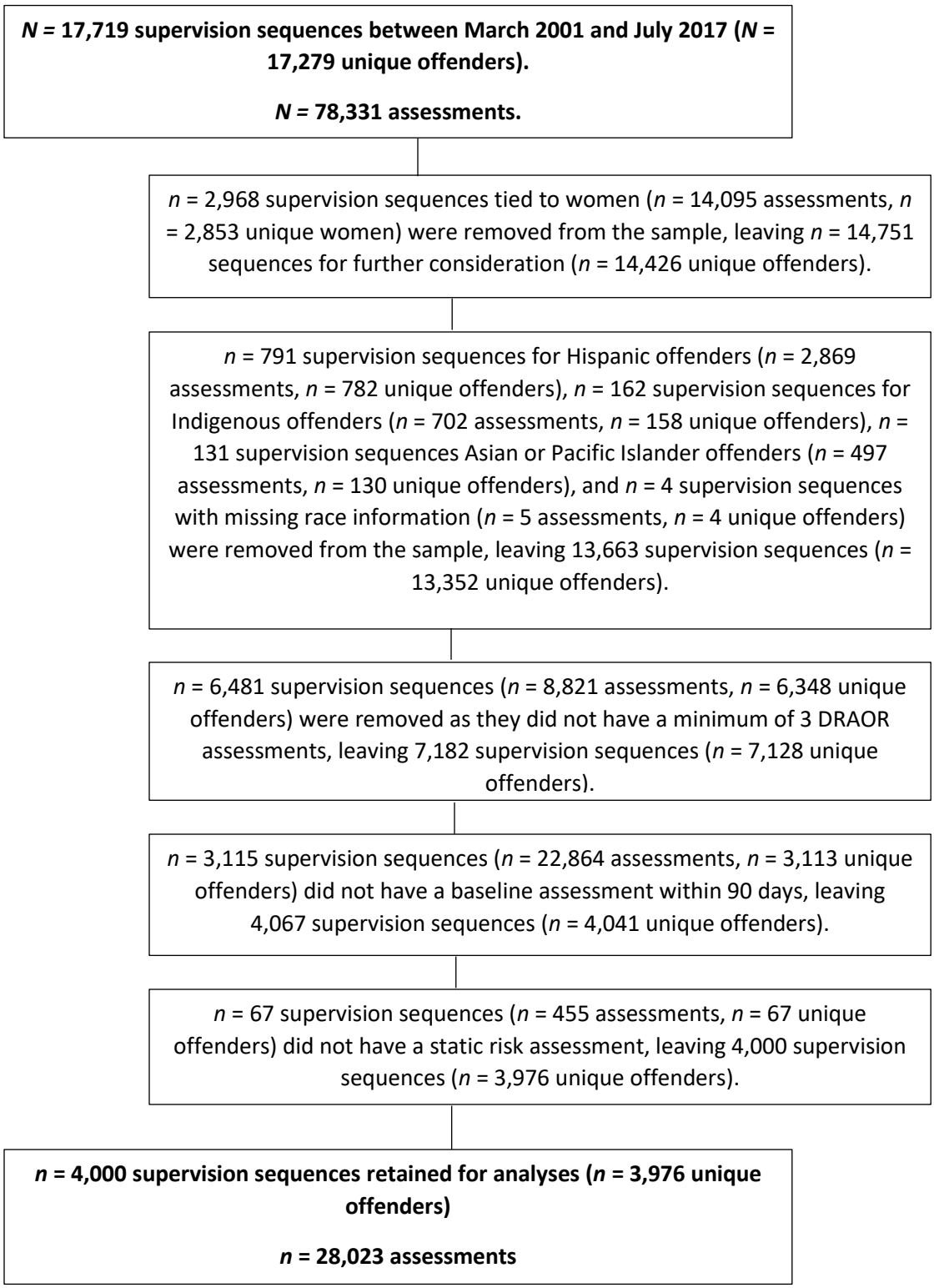

Figure 2. Flowchart describing sample selection procedure for analysis. Note that the number of unique offenders removed at each stage does not sum to the total removed.

Figure 2 continued. An offender could have had a supervision sequence retained in the analysis sample while also having a supervision sequence removed due to an exclusion criterion. In this case, they would represent a unique offender in both the analysis sample and the excluded sample.

Sample characteristics. Descriptive information for the analysis sample is presented in Table 1. The majority of supervision sequences consisted of offenders who were single (62%), a large portion had some post-secondary education (41%) or a high school diploma/GED (36%), and most supervision periods were associated with a probation order (57%), followed by parole (23%) and work release (13%). Offender age at the start of supervision ranged from 16 to 87 (*M* = 33, *SD* = 11). On average, the sample was considered high risk, requiring an intensive level of supervision (i.e., level 5), according to the victimization (*M* = 8.6, *SD* = 3.3) and violence (*M* = 8.6, *SD* = 3.8) scores. Sample participants were most commonly sentenced for a violent offence (32%, e.g., assault), followed by drug related (26%, e.g., trafficking), property (24%, e.g., theft), public order (15%, e.g., operating while intoxicated), and other (4%, e.g., conspiracy to commit offence). The majority (72%, *n* = 2,870) of the sample had a prison sentence associated with the most serious index charge. Recall that the majority of the sample were supervised on probation, so their associated prison sentence would have been suspended while serving the probation order. Among those with a prison sentence, the average sentence was 94 months (*SD* = 87 months).

Table 1

Descriptive Information for the Analysis Sample (n = 4,000 Supervision Sequences)

Variable	%	*n*
Race		
Black	27	1,076
White	73	2,924
Marital Status		
Single	62	2,480
Married/Common-Law	15	601
Divorced/Separated	16	628
Missing	7	291
Education		
Some Grade School	1	30
Some High School	18	726
High School Diploma/GED	36	1,432
Post-Secondary	41	1,630
Unknown	1	152
Missing	3	130
Supervision Status		
Probation	57	2,258
Parole	23	928
Work Release	13	537
Other[a]	7	277
Age, *M (SD)*	33.4 (11.0)	4,000
Iowa Risk Assessment Revised – Victimization, *M (SD)*	8.6 (3.3)	4,000
Iowa Risk Assessment Revised – Violence, *M (SD)*	8.6 (3.8)	4,000

[a]Other includes pretrial release with supervision (0.6%), Interstate compact parole (0.9%) or probation (1.5%), a special sentence (1.3%), Operating While Intoxicated (OWI) Continuum (2.2%), and Federal supervision (0.4%).

The sample was compared to the population of offenders supervised in Iowa from 2014 to 2016 to assess the representativeness of the sample (see Table 2). Given that the inclusion criteria for the analysis sample (e.g., Black and White men) distorted the proportions across the key statistics presented, comparisons are made between the study sample from which the analysis sample was drawn. The study sample appeared to

have a higher proportion of men (83%) than what is expected among the population of offenders on community supervision orders in Iowa (approximately 74%). Proportionally, there appeared to be fewer White offenders (71%) and more Black offenders (22%) in the study sample than what is observed in the population (76% and 16%, respectively). The distribution of offenders across the age categories approximated the offender population, although there was a slightly higher proportion of offenders falling into the 18 to 30 age group in the study sample (48% versus approximately 45%). There was a greater proportion of offenders in the study sample who committed a violent crime as their most serious index charge (30%) compared to the overall offender population (20%) and a smaller proportion who committed a public order offence (14% vs. 26%). Lastly, a greater portion of the study sample was serving a parole sentence (20%) compared to the offender population (13%), and fewer were sentenced to a probation order (62% vs. 73%). Overall, given that DRAOR assessments are completed for those considered to be higher risk in Iowa (i.e., level 3 and above), the observed differences between the study sample and the population of offenders on community supervision in Iowa are in the expected direction.

Table 2

Sample Representativeness Compared to Iowa's Community Based Corrections Population Snapshots from 2014 to 2016

	2014 (*N* = 30,666)		2015 (*N* = 29,928)		2016 (*N* = 30,033)		Study Sample (*n* = 17,719[a])		Analysis Sample (*n* = 4,000[a])	
Characteristic	n	%	n	%	n	%	n	%	n	%
Sex										
Women	7,994	26	7,684	26	7,611	25	2,968	17	-	-
Men	22,668	74	22,216	74	22,381	75	14,751	83	4,000	100
Unknown	4	0.0	28	0.1	41	0.1	0	0	-	-
Race										
Asian	322	1	353	1	331	1	146	1	-	-
Black	4,700	15	4,870	16	4,855	16	3,830	22	1,076	27
Hispanic	1,502	5	1,565	5	1,584	5	894	5	-	-
American Indian	311	1	294	1	288	1	231	1	-	-
White	23,641	77	22,727	76	22,818	76	12,614	71	2,924	73
Unknown	190	1	119	0.4	157	0.5	4	0.02	-	-
Age										
Under 18	NR		13	0.04	36	0.1	110	1	30	1
18-30	14,361	47	13,615	45	13,328	44	8,472	48	1,859	47
31-50	12,678	41	12,611	42	12,909	43	7,631	43	1,745	44
Over 50	3,627	12	3,689	12	3,760	13	1,506	9	366	9
Crime Type										
Violent	5,831	19	5,848	20	6,196	21	5,352	30	1,277	32
Property	7,242	24	7,102	24	7,108	24	4,502	25	947	24
Drug	8,434	28	8,303	28	8,127	27	4,693	26	1,022	26
Other	742	2	718	2	764	3	624	4	167	4
Public Order	8,417	27	7,957	27	7,338	24	2,548	14	587	15

Supervision Status										
Probation	22,667	74	21,326	71	22,030	73	10,980	62	2,258	57
Parole	3,973	13	3,551	12	3,911	13	3,544	20	928	23
Special Sentence Parole	629	2	685	2	770	3	465	3	51	1
Pretrial Release	1,661	5	1,292	4	1,539	5	121	1	25	0.6
Residential Facilities	1,728	6	1,822	6	1,773	6	NR	-	NR	-
Other	8	0.0	1,252	4	10	0.03	NR	-	NR	-

Note: [a] represents distinct supervision periods, not unique offenders. 2014 numbers were reported as a snapshot on July 1, 2014. 2015 numbers were reported as a snapshot on June 30, 2015, and 2016 numbers were reported as a snapshot on June 30, 2016 (source: Quarterly Quick Facts, Iowa DOC https://doc.iowa.gov/data/quick-facts). NR = Not Reported. Empty cells indicate that a comparison was not possible either due to inconsistent information available or as a result of sample constraints. Percentages may not add to 100 due to rounding.

Data Analysis

Iowa DOC provided copies of administrative datasets for all offenders with a completed DRAOR assessment during the study timeframe. SPSS 25.0 was used to organize the data into the person-level and person-period datasets. Data cleaning and examination of basic assumptions was performed in SPSS. Measurement invariance was assessed through Mplus 8.0 (Muthén & Muthén, 1998-2017), growth modelling was performed using SAS software, version 9.4, and the Survival package in R 3.5.1 (Therneau, 2015) was used for the prediction models. The analysis plan followed a sequential process beginning with data screening and descriptive results for all DRAOR assessments, followed by an examination of the DRAOR's psychometric properties and the stability of the factor structure over time, then preceding to model change trajectories on the DRAOR domains throughout the course of supervision, concluding with an analysis of the relationship between change on the DRAOR and revocation of community supervision or a new conviction. The analytic approach for each component is detailed below.

Measurement invariance. Longitudinal exploratory factor analysis (EFA) within the exploratory structural equation modelling framework (ESEM; Asparouhov & Muthén, 2009) was used to assess whether the factor structure of the DRAOR remained consistent (i.e., invariant) over time. Measurement invariance involves testing various degrees of model constraints to evaluate whether the measured constructs remain consistent over time. This approach has been applied to correctional risk assessment data previously (Babchishin, 2013; Bergeron & Miller, 2013). ESEM incorporates

elements of exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and structural equation modelling (Marsh, Morin, Parker, & Kaur, 2014). The ESEM framework allows for more flexibility in modelling the measurement structure, specifically allowing for cross-loadings across constructs, which represents an important improvement over the CFA approach (Asparouhov & Muthén, 2009).

There are different levels of measurement invariance that can be achieved, and these levels dictate the conclusions that can be drawn from the results. When testing measurement invariance on categorical data, there are two primary levels of invariance that are of interest: (a) configural, and (b) strong or scalar. Configural invariance tests a model that freely estimates factor loadings and thresholds over time (Widaman & Reise, 1997). Under this model, the same items are associated with the same constructs over time, but the factor loadings and item thresholds can vary over time. This means that the relationship between each item and the underlying latent construct can vary over time, as can the probability of receiving a particular score on an item. Obtaining configural invariance is required to sufficiently compare factor variances and covariances across time. For latent variable means to be appropriately compared across time, strong invariance must be obtained. Strong invariance requires factor loadings and thresholds to be equal across the multiple assessments (Meredith, 1993). Under the strong invariance model, residual variances of the indicators are still freely estimated. Given that the primary interest for the current study was to compare variances and means over time, strong invariance was the threshold of interest.

To evaluate measurement invariance, competing models are compared to determine whether additional constraints lead to a significant deterioration in model fit. In addition to performing likelihood ratio tests between the strong invariance model relative to the configural model, both the Root Mean Square Error of Approximation (RMSEA; threshold < .08 for good fit), Comparative Fit Index (CFI; threshold > .95 for good fit), and Tucker-Lewis Index (TLI; threshold > .95 for good fit; Hu & Bentler, 1999; Matsunaga, 2010) were examined as estimates of how well the model fit the data. If the CFI changes by less than or equal to 0.01 (Cheung & Rensvold, 2002) or the RMSEA changes by less than or equal to 0.015 when comparing models, there is sufficient support that the constrained model does not fit the data worse than the model that freely estimated factor loadings and thresholds (Chen, 2007).

Since the DRAOR items are categorical, a weighted last squares (WLSMV) estimation technique, with theta parameterization, was utilized instead of the typical maximum likelihood estimation (Asparouhov & Muthén, 2009). Importantly, likelihood ratio tests must be conducted using adjusted chi-square values and degrees of freedom when the WLSMV estimator is utilized, which is accomplished with the DIFFTEST in Mplus (Muthén & Muthén, 1998-2017). Consistent with Marsh and Hau's (1996) recommendation, correlations between the same items across measurement occasions were included in the model to reduce the potential for biasing the estimation of model parameters.

The first step of a longitudinal EFA often involves conducting an exploratory factor analysis across each time point of interest. This is performed to identify the

optimal factor structure and whether it is consistent over time. Model constraints (i.e., factor loadings and thresholds) are then imposed on the optimal factor structure to determine if measurement invariance holds. Given that previous research on the factor structure of the DRAOR in Iowa has generally supported the initial structure consisting of the stable, acute, and protect domains (Chadwick, 2014; Serin & Chadwick, 2019), analyses proceeded directly to examining the constrained models over time. Longitudinal EFA requires completed assessments for each individual in the sample across the time points that make up the analysis of measurement invariance. It was of interest to identify 3 distinct time points to test measurement variance. Since the timing of the assessments varied for the current sample, assessments were organized into various intervals to determine the approach that maximized the sample size eligible to be included in this analysis, while remaining theoretically and practically sensible. Table 3 presents the results of exploring the number of offenders who met the requirement of having at least one completed assessment in each interval. Consistent with Babchishin (2013), in the case of multiple assessments occurring during an interval, one of the assessments was randomly selected. Unsurprisingly, intervals that encompassed a greater amount of time included a greater proportion of the total sample. As a result, the assessment schedule used to test measurement invariance included an assessment occurring at some point during the first 3 months after release, from 4 to 6 months after release, and from 7 to 9 months after release.

Table 3

Sample Sizes for Possible Time-Intervals for Measurement Invariance

Assessment completed	*n*
1, 2, and 3 months after release	640
1-2, 3-4, and 5-6 months after release	1,506
1-3, 4-6, and 7-9 months after release	2,055

Note. The timing of 28,023 assessments from 4,000 supervision sequences was examined to determine the placement in these assessment intervals.

Change over time and the relationship with recidivism. A two-stage model (Yang, Guo, Olver, Polaschek, & Wong, 2017) was explored to examine whether DRAOR domain scores changed over time and whether those changes were related to recidivism. The first stage involved performing a series of multilevel models to identify and explain patterns of change. The second stage then extracted individual growth parameters from the models and assessed the relationship between change on DRAOR domains and recidivism through Cox regression survival analysis. Relevant considerations for each stage of this process are described below.

Examining change over time. Multilevel modelling (i.e., growth modelling, hierarchical level modelling) was used to analyze individual trajectories of change on DRAOR domain scores while supervised in the community. Multilevel modelling has the advantage of being able to analyze non-normal data and does not require a balanced assessment schedule, unlike other repeated measures approaches, such as repeated measures MANOVA (Singer & Willett, 2003; Tabachnick & Fidell, 2013; Tasca & Gallop, 2009). In other words, individuals do not need to be assessed at every time point to contribute to the analyses. Further, multilevel models can accommodate unequal

intervals of time between assessments (Singer & Willet, 2003). Multi-level modelling also has the advantage of being able to incorporate the fact that assessments are nested within the individual, leading to interdependence between observations (Singer & Willett, 2003). This interdependence affects the variance of the outcome and the associated estimates of standard errors, which makes conventional regression approaches largely inappropriate (Singer & Willett, 2003). Multilevel modelling addresses the issues caused by interdependence of observations by modelling the degree of relatedness between observations within an individual, and corrects the estimated standard errors, reducing the problem of Type I errors.

Although individuals do not need to contribute the same amount of assessments to the analyses, multilevel models do make the assumption that any missing assessment is missing at random. Data are said to be missing at random when the probability of missing an assessment at a time point is independent of the missing data given the observed data (Raudenbush & Bryk, 2002; Singer & Willett, 2003). Raudenbush and Bryk (2002) posit that it is reasonable to assume that the data are missing at random when the observed data capture key confounding variables that are related to both attrition and the outcome of interest. Although there is no direct test to assess whether data are missing at random, Singer and Willett (2003) recommend considering whether it is reasonable to assume that the probability of missingness is unrelated to unobserved concurrent outcomes. For the current dissertation, missing data is inherently introduced to the dataset once an offender is either revoked from supervision or receives a new convicted. As a result, recidivists are less likely to be assessed as follow-up time

increases, which is expected to coincide with an unobserved increase in the risk for reoffending. However, since key confounding variables (e.g., static risk, age, DRAOR scores) are included in the dataset and used in the analyses, it is expected that the multilevel estimation will be unbiased (Raudenbush & Bryk, 2002).

Level of analysis. In a typical repeated measures multilevel model, the lowest level of analysis focusses on the assessments, which are grouped into the higher level of individuals. It is possible to model additional levels of grouping, such as offenders grouped within supervising officers, who are grouped into offices and districts. Multilevel models allow for an examination of average change on a dependent variable (e.g., dynamic risk) across time, as well as understanding how predictors at each level explain variability in trajectories. The optimal functional form of change (e.g., linear, quadratic, cubic) can be modelled to explain variation in the rates of change over time. Additionally, level-1 predictors, or time-varying covariates, can be modelled to explain within-person variation in change over time. A notable strength of the multilevel approach is the ability to disaggregate within-person and between-person effects of a time-varying covariate (Curran & Bauer, 2011; Hoffman & Stawski, 2009; Howard, 2015). Within-person effects of a time-varying covariate can explain shifts or departures from an underlying trajectory of change. For example, an increase in protective factors over time might explain a shift to a change trajectory characterized by lower risk. Level-2 predictors are constant over time, either by nature (e.g., race) or due to a single measurement of a given variable (e.g., age at the start of supervision). Level-2 variables

can be incorporated into the multilevel model to explain variations in either the initial status or change over time.

In order to derive estimates of parameters associated with each level of the multilevel model, an estimation method is required. The maximum likelihood (ML) method was used to fit the various models of change explored in this dissertation. Utilizing ML estimation is desirable when sample sizes are large, as parameter estimates are efficiently obtained and are likely to be unbiased and be normally distributed (Singer & Willett, 2003). Importantly, models based on ML estimation can be compared when they differ in either their fixed effects or random effects, which supports the intended model building process discussed below.

Model building. In the current dissertation, each DRAOR domain (stable, acute, protect) served as an outcome variable in a sequence of multilevel models. The first, an unconditional means model, examined only the grouping variable (i.e., individuals) without predictors to allow for the variance in the dependent variable to be partitioned between and within individuals. These models serve as a baseline for comparisons with more complex models that include predictors of change in the outcome variables. Within the unconditional means model, an intraclass correlation (ICC) is calculated as the proportion of total variation in the dependent variable that is attributable to individual differences (Singer & Willet, 2003). In the context of repeated measures multilevel modelling, when ICCs approach 1, differences in the outcome variable (e.g., DRAOR score) are the result of between-group differences, suggesting that scores within an individual remain consistent over time.

Next, a time predictor was added to an unconditional growth model. Time was defined as the number of months on community supervision. The optimal functional form of time (e.g., linear, quadratic, cubic) was determined based on a descriptive analysis of individual growth curves and a comparison of model fit indices for nested models. A random effect was modelled for time, allowing trajectories of change to vary across offenders. Results from these models inform the individual's score on each of the DRAOR domains at the start of community supervision (i.e., intercepts) and the average rate of change throughout the study period (i.e., slopes). Unconditional growth models separate outcome variation into within-person variance that summarizes variation around each person's change trajectory, and between-person variability in the true initial status or true rate of change (Singer & Willett, 2003). The change in the level-1 variance from the unconditional means model can be examined to determine the amount of within-person variation that is associated with time. If within-person variance remains significant after modelling time, there is support for modelling additional time-varying covariates. In the case of significant within-individual variance, each DRAOR specific model examined scores on the remaining DRAOR domains to examine whether within-individual changes contributed to the explanation of initial status and change over time. Time-varying covariates contribute to explaining both within- and between-person variance. Within-person variance is explained by modelling a person mean centered variable that details how an individual has changed around their mean at a given measurement occasion (Howard, 2015). The intercept and slope parameters are then interpreted as the expected initial status and rate of change, when an individual is

at their average score on the time varying covariate. Between-level effects for a time-varying covariate can be modelled by including a grand-mean centered variable. The level-2 variance components quantify the amount of unpredicted variability in the true initial status and rates of change. When these variance components are significant, it suggests that there may be between-group predictors that can help explain some of the variation in either initial score at the start of supervision or change over time.

Next, a model examining between-individual differences in change was explored. The effects of static risk, race (Black and White), age, and the remaining DRAOR domain scores (e.g., acute and protect when examining change on stable scores) on an individual's initial DRAOR score and change over time were examined. The dichotomous variable, race, was not centered, but static risk, age, and DRAOR domain scores were grand mean centered so that the level-2 fitted intercepts describe the average DRAOR domain score at the start of supervision and average rate of change for a White offender of average age, static risk, and remaining DRAOR domain scores (Singer & Willett, 2003).

Consistent with the model building process recommended by Singer and Willett (2003), various conditional models were explored to identify the optimal combination of covariates and their relationship with initial status and change over time. Model fit indices (deviance, AIC, BIC) were explored to determine whether one model performed better than another. Given that ML estimation was utilized, the deviance statistic was used to assess the entire fit of the model (Singer & Willett, 2003). Model fit statistics were supplemented with an examination of several Pseudo R^2 statistics, which provide

estimates of the proportion of outcome variation that is explained by the model predictors (Singer & Willett, 2003).

Although Singer and Willett (2003) acknowledge that the field has yet to agree on appropriate summaries of explained outcome variance, they maintain that Pseudo R^2 statistics can be useful analytic tools to draw on throughout the model building process. Pseudo $R^2_{y,y}$ represents the proportion of total variation in DRAOR scores that is explained by each model's combination of predictors. The remaining statistics examine the residual variance in the outcome, to determine how the addition of predictors at each level reduces the level of residual variance. A decrease in the within-person variance (R^2_e) is examined by comparing the residual variance from the unconditional means model to the unconditional growth model. This yields a statistic that summarizes the proportion of within-person variance that is explained by the introduction of time. Similarly, change in residual variance associated with the initial status (R^2_0) and slope (R^2_1) can be examined, comparing conditional models that include level-2 predictors (e.g., static risk, age) to the unconditional growth model. Final conditional models were selected when the combination of predictors appeared to fit the data according to the suite of fit statistics and variance components, while also remaining parsimonious.

Do growth trajectories predict recidivism. Results from the trajectories of change models were used to examine the relationship between change on the DRAOR and recidivism. Individual change trajectories were constructed from the unconditional growth model by extracting the random coefficients representing each individual's variation around the fixed effect for the intercept and slope (Yang et al., 2017). Yang and

colleagues (2017) caution that a potential methodological problem of using random effects derived from the multilevel model is that they are known to contain estimation errors and shrinkage toward the grand mean. The effect of shrinkage could be large when ICCs are small, which could lead to biased estimates that are transferred to the second stage of the analyses. This in turn can cause biased estimation of the effect of change in risk on recidivism. A shrinkage factor for the random intercept can be computed to assess the potential for bias by dividing the total variance in a standard ICC calculation by the number of assessments for each individual. When the shrinkage factor approaches 1, the estimated random intercepts will be closer to the raw mean for each individual, with less shrinkage to the grand mean. When there is evidence of little shrinkage in the estimate of the random intercepts, Yang and colleagues (2017) demonstrate that it is appropriate to proceed with the two-stage model. ICCs calculated for each supervision sequence ranged from .91 to .99 for stable, .88 to .99 for acute, and .93 to .99 for protect. As a result, there was evidence of little shrinkage in the random intercept for any of the models, which provides support for continuing with the two-stage model.

The second stage of the model involves conducting a regression examining the relationship between the individual change trajectories and recidivism. Given the binary nature of the outcomes examined in this dissertation (i.e., new conviction or no new conviction, and revocation of community supervision vs. no revocation) either logistic regression or survival analysis are appropriate regression models (Tabachnick & Fidell, 2013). Cox regression survival analysis is able to incorporate variable follow-up times

and sample censoring (Brown et al., 2009; Tabachnick & Fidell, 2013), making it more suitable for the current analyses. Cox regression is a semi-parametric test that models the relation between predictors and the event (e.g., new criminal conviction), while accounting for time to the event occurrence. This approach allows multiple predictors to be measured simultaneously in order to determine their independent and unique contributions to the outcome variable (Tabachnick & Fidell, 2013). Cox regression analyses produce a hazard ratio, which represents the predicted change in the hazard (i.e. recidivism) for a unit increase in the predictor (i.e., stable, acute, protect).

Preliminary models assessed the relationship between community outcomes (new conviction and revocation) and both the initial score and change over time for each DRAOR domain, without considering other potentially relevant covariates. Static risk, race, and age were explored as relevant covariates to assess the relationship between DRAOR scores and recidivism, while accounting for their effects. Lastly, a multivariate model that incorporated individual change trajectories on stable, acute, and protect into one model was investigated to determine the overall relationship between DRAOR scores and recidivism.

Harrell's concordance index (i.e., Harrell's c or c-index; Harrell, Califf, Pryor, Lee, & Rosati, 1982) was calculated as an effect size for the survival analysis models. The c-index has the advantage of being able to account for data with variable follow-up time (Helmus & Babchishin, 2017). This index is interpreted similarly to an AUC, in that the statistic represents the probability that among two randomly selected individuals, the higher score will have been associated with the recidivist, rather than the non-recidivist

(see Harrell, 2015; Lloyd, 2015). Similar interpretations of effect size magnitude as an AUC statistic can be applied (e.g., effects of .56, .64, and .71 would be considered small, moderate, and large, respectively; Rice & Harris, 2005).

Results

Descriptive Information for Assessment Data

Descriptive information for DRAOR items and subscales is presented in Table 4. Results are based on the assessment level for each community supervision sequence included in the analysis. An examination of data entry accuracy indicated that there were no missing or out of range values for all valid assessments identified for inclusion in the sample. The presence of univariate outliers was examined among the stable, acute, and protect scores, the IRR-Victimization score, and offender's age. There were no univariate outliers identified for the DRAOR domain scores, however, there were 41 assessments (n = 7 offenders) where the standardized z-score for IRR-Victimization surpassed the 3.29 threshold ($p < .001$, two-tailed test) and 21 assessments (n = 4 offenders) where the standardized value for age surpassed the threshold. Univariate outliers were retained for all analyses, given that the large sample size made it unlikely that their inclusion affected the results. Mahalanobis distance was examined to reveal whether there were any multivariate outliers, defined as those that surpassed the critical value of 20.52 ($df = 5$, $p = < .001$). There were 57 assessments, representing 19 offenders, that were identified as a multivariate outlier. Each potential multivariate outlier was examined to determine why the combination of scores appeared atypical. In most cases (41 of 57 sequences examined), conceptually inconsistent patterns of

DRAOR scores and static risk were observed, for example an offender would be classified as low risk across stable, acute, and Victimization scores, but did not have any indication of protective factors. The remaining multivariate outliers were associated with two offenders classified as univariate outliers because of their age at the start of supervision. All models were computed with and without multivariate outliers included in the sample. Results reported in this dissertation were from analyses that included the multivariate outliers, as their removal had no meaningful impact on the results.

The distribution of DRAOR scores and the IRR-Victimization score were assessed for normality through visual inspection and an examination of the skewness and kurtosis coefficients. Given the large sample size, the skewness and kurtosis coefficients were examined, rather than performing formal significance testing (Tabachnick & Fidell, 2013). Results suggested that there were no severe departures from normality across DRAOR domains (stable skewness = 0.07, kurtosis = -0.21; acute skewness = 0.21, kurtosis = -0.26; protect skewness = 0.08, kurtosis = -0.07) and IRR-Victimization (skewness = 0.35, kurtosis = 0.23). All pairs of variables were visually inspected to assess linearity and homoscedasticity. All variables appeared to be linearly related, and there were no clear departures from homoscedasticity.

Table 4

Descriptive Information Across All Assessments (n = 28,023) Included in the Study

Variable	Range	*M*	*SD*
Stable	0-12	6.0	2.5
Peer associations	0-2	1.3	0.6
Attitudes toward authority	0-2	0.8	0.7
Impulse control	0-2	1.3	0.6
Problem solving	0-2	1.2	0.6
Sense of entitlement	0-2	0.7	0.7
Attachment with others	0-2	0.7	0.7
Acute	0-14	6.5	2.7
Substance abuse	0-2	1.3	0.8
Anger/hostility	0-2	0.8	0.7
Opportunity/access to victims	0-2	0.9	0.7
Negative mood	0-2	0.8	0.7
Interpersonal relationships	0-2	1.0	0.7
Employment	0-2	0.9	0.8
Living situation	0-2	0.8	0.7
Protect	0-12	5.6	2.6
Responsive to advice	0-2	1.0	0.6
Prosocial identity	0-2	0.9	0.6
High expectations	0-2	1.0	0.6
Cost/benefits	0-2	0.9	0.6
Social support	0-2	1.0	0.6
Social control	0-2	0.8	0.6
IRR - Victimization Score	-3-25	8.5	3.7

Psychometrics of the DRAOR

Hypothesis 1. The DRAOR will demonstrate sufficient psychometric properties, including maintaining measurement invariance and internal consistency.

As described in the methodology, measurement invariance was examined for a subsample (*n* = 2,055) of supervision sequences with at least one completed DRAOR assessment occurring at some point during the first 3 months after release, from 4 to 6 months after release, and from 7 to 9 months after release. Two models were tested

and compared to establish whether the DRAOR was measuring the same constructs over time. The strong invariance model, which imposed equality constraints on factor loadings and item thresholds across each measurement occasion, was compared to a less stringent baseline model, which freely estimated the factor loadings and item thresholds, but fixed the variance components in the model. Table 5 presents the results from both models. Each fit the data well, and importantly, the strong invariance model did not result in a significant deterioration when considering the RMSEA, CFI, and TLI. Although the difference test based on the adjusted chi-square, corrected for WLSMV estimation, indicated that the model with factor loadings and threshold constraints provided a significantly worse fit, this is likely influenced by the test's sensitivity to large sample sizes (Chen & Rensvold, 2002). Based on these findings, there is sufficient evidence to conclude that the DRAOR is measuring the same constructs in the same way over time. This provides support for comparing means on the latent variables across measurement occasions. Measurement invariance was also tested separately for White and Black offenders to ensure that the relationships between the items and corresponding constructs were consistent across each subsample. Results based on the subsamples were consistent with the overall finding that the DRAOR met the threshold for measurement invariance (see Appendix A).

Table 5

Measurement Invariance for Overall Sample (n = 2,055)

Model	# of free parameters	RMSEA [90% CI]	CFI	TLI	X^2 (df)
Baseline/Configural	360	0.036 [0.035, 0.037]	0.993	0.991	
Strong	188	0.036 [0.035, 0.037]	0.992	0.991	
Changes in fit indices		0.000	0.001	0.000	836.43** (172)

Note: A ΔCFI of less than 0.01 or an ΔRMSEA of less than 0.015 provides sufficient support that the constrained model does not fit the data worse than the model allowed to vary. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TFI = Tucker-Lewis Index.

Psychometric properties for DRAOR assessments at each of the three time points used for testing measurement invariance are presented in Table 6. Revelle's Omega Total (ω_{RT}; Revelle, 2019) was used as an estimate of the internal consistency for DRAOR domains. There have been recent calls to consider alternative indices of internal consistency (e.g., Dunn, Baguley, & Brunsden, 2014; McNeish, 2018), given the often untenable assumptions associated with Chronbach's alpha and the tendency to underestimate reliability (see McNeish, 2018 for greater discussion). Revelle's Omega Total is based on a factor analytic model and assesses reliability as a function of individual item variation that is explained by the total variance of all items (Kroner & Riordan, 2019; McNeish, 2018). Additionally, the categorical nature of the items can be accounted for by computing internal consistency based on a polychoric correlation matrix. According to Revelle's Omega Total, DRAOR stable and protect were internally consistent at each time point examined, while the point estimates for acute were smaller in magnitude. This trend is consistent with previous examinations of internal

consistency on the DRAOR domains (Chadwick, 2014; Davies, 2019; Hanby, 2013; Lloyd, 2015).

Correlations between the DRAOR domains were in the small to moderate range. The magnitude of the relationships between stable and both acute and protect remained largely consistent over the time periods examined, while the relationship between acute and protect tended to increase over time. Correlations with static risk were in the low range, with evidence of further weakening over time. Given that the static risk score remained constant over the course of the follow-up, it is unsurprising that the relationship would weaken over time. Observed relationships tended to be smaller in magnitude than what was previously found in New Zealand (Davies, 2019; Lloyd, 2015), but the pattern of relationships over time were similar.

Table 6

Psychometric Properties of DRAOR Scores Used to Test Measurement Invariance

Time point	Internal consistency (Revelle's Omega Total ω_{RT})			Inter-subscale correlations			Relationship with static risk		
	Stable	Acute	Protect	Stable & Acute	Stable & Protect	Acute & Protect	Stable & IRR	Acute & IRR	Protect & IRR
0-3 months	.83	.63	.90	.46	-.40	-.27	.18	.16	-.12
4-6 months	.84	.63	.90	.48	-.42	-.30	.16	.12	-.09
7-9 months	.86	.70	.90	.53	-.49	-.37	.13	.10	-.07

Note: All correlations significant at $p < .001$. ωRT = Revelle's omega total. Point estimates for internal consistency are based on polychoric correlation matrices. IRR = Iowa Risk Assessment Revised.

Taken together, results support Hypothesis 1, in that the psychometric properties of the DRAOR were sufficient based an examination of DRAOR scores observed during 3 distinct time periods (0-3 months, 4-6 months, 7-9 months). It is important to note that although the magnitude of the relationships both within DRAOR domains and with static risk were smaller than previously reported, they were in the expected direction. Measurement invariance results indicated that holding the factor loadings and item intercepts constant over time did not result in a significantly worse fit to the data, supporting that the original 3-factor structure of the DRAOR was measuring the same constructs in the same way over time. As a result, it was deemed appropriate to proceed with examining trajectories of change. This finding also enhances the confidence that observed change is the result of true change, plus measurement error, rather than the composition of each of the constructs changing over time.

Trajectories of Change

To be included in the analysis, each supervision sequence had to have a minimum of 3 completed DRAOR assessments. Assessments were considered valid for inclusion in the analyses if they were completed after the beginning of community supervision and before the first instance of a new conviction or end of community supervision due to a revocation. On average, there were 7 completed DRAOR assessments (*SD* = 4.8; range of 3 to 45) in each supervision sequence. Supervision sequences that ended with a new conviction had fewer completed assessments (*M* =

5.67, *SD* = 3.26) on average,[6] than compared to those sequences where the offender remained crime free in the community (*M* = 7.17, *SD* = 4.94). The number of days between the first and last assessment for each supervision sequence ranged from 3 to 1,532, with a mean of 382.84 (*SD* = 274.35). Recidivists had a shorter time between first and last assessment (*M* = 6.74 months, *SD* = 4.96), compared to nonrecidivists (*M* = 13.29 months, *SD* = 9.14).

Table 7 provides descriptive information for each assessment occasion included in the examination of change over time. On average, first assessments tended to be completed just after one month on supervision (*M* = 1.22, *SD* = 0.77). Generally, each subsequent assessment occasion consisted of assessments that were completed on average 1 to 2 months later. Descriptive results across all assessment occasions suggested that scores on stable, acute, and protect slightly fluctuated over time. During each of the first 8 assessment occasions, it appeared that, at the aggregate group-level, scores on the stable and acute decreased slightly, while protect scores increased slightly. It is important to note that only 25% (*n* = 985 of *n* = 4,000) of the supervision sequences in the sample had 9 or more completed assessments, which may explain the reversal of descriptive trends in stable, acute, and protect scores for these later assessment occasions.

[6] A similar trend was observed when considering revocations of community supervision. Offenders revoked from community supervision had an average of 6 completed DRAOR (*SD* = 3.7) assessments.

Table 7

Descriptive Statistics for DRAOR Assessment Occasions (n = 4,000 Supervision Sequences)

Assessment occasion	Supervision sequence	Stable score *M* (*SD*)	Acute score *M* (*SD*)	Protect score *M* (*SD*)	Time on supervision in months *M* (*SD*)	No. of individuals with *n* assessments
1	4000	6.15 (2.48)	6.85 (2.67)	5.50 (2.67)	1.22 (0.77)	
2	4000	6.04 (2.47)	6.62 (2.64)	5.55 (2.63)	3.59 (2.80)	
3	4000	5.90 (2.53)	6.39 (2.72)	5.71 (2.66)	5.94 (4.44)	810
4	3190	5.87 (2.50)	6.31 (2.70)	5.78 (2.60)	7.61 (5.26)	649
5	2541	5.87 (2.46)	6.28 (2.65)	5.73 (2.55)	9.12 (6.05)	517
6	2024	5.88 (2.47)	6.27 (2.66)	5.73 (2.52)	10.48 (6.67)	437
7	1587	5.92 (2.44)	6.29 (2.67)	5.72 (2.53)	11.55 (6.78)	331
8	1256	5.90 (2.44)	6.26 (2.68)	5.75 (2.54)	12.61 (6.77)	271
9	985	5.92 (2.43)	6.24 (2.68)	5.76 (2.51)	13.30 (6.56)	215
10	770	6.00 (2.40)	6.25 (2.63)	5.66 (2.44)	14.18 (6.34)	142
11	628	5.94 (2.42)	6.21 (2.57)	5.64 (2.41)	15.42 (6.54)	125
12	503	6.05 (2.50)	6.36 (2.72)	5.64 (2.48)	16.25 (6.06)	92
13	411	6.13 (2.47)	6.44 (2.69)	5.57 (2.47)	17.48 (6.28)	70
14	341	6.33 (2.46)	6.57 (2.63)	5.45 (2.45)	18.69 (6.44)	56
15	285	6.40 (2.36)	6.57 (2.62)	5.47 (2.42)	19.95 (6.84)	45
16	240	6.53 (2.45)	6.48 (2.69)	5.33 (2.43)	20.80 (6.24)	46
17	194	6.61 (2.46)	6.56 (2.64)	5.22 (2.50)	21.49 (5.74)	23
18	171	6.63 (2.42)	6.69 (2.69)	5.21 (2.48)	22.77 (5.67)	26
19	145	6.74 (2.42)	6.74 (2.69)	5.26 (2.27)	23.66 (5.82)	27
20	118	6.70 (2.47)	6.64 (2.81)	5.24 (2.18)	24.35 (6.02)	16
21	102	6.78 (2.51)	6.79 (2.77)	5.02 (2.03)	25.46 (5.99)	18
22	84	6.80 (2.40)	6.68 (2.76)	4.89 (2.02)	26.35 (5.95)	16
23	68	6.91 (2.50)	6.63 (2.72)	4.84 (2.12)	27.49 (5.46)	11

Assessment occasion	Supervision sequence	Stable score *M* (*SD*)	Acute score *M* (*SD*)	Protect score *M* (*SD*)	Time on supervision in months *M* (*SD*)	No. of individuals with *n* assessments
24	57	6.82 (2.48)	6.33 (2.75)	4.84 (2.10)	28.39 (5.56)	6
25	51	6.90 (2.51)	6.61 (2.89)	4.71 (2.21)	29.66 (5.77)	7
26	44	6.70 (2.43)	6.55 (2.70)	4.77 (2.32)	30.23 (5.88)	2
27	42	6.62 (2.51)	6.52 (2.71)	4.90 (2.41)	31.17 (6.16)	5
28	37	6.89 (2.39)	6.59 (2.51)	4.81 (2.45)	32.39 (6.69)	8
29	29	6.79 (2.27)	6.45 (2.64)	5.31 (2.29)	32.79 (7.42)	4
30	25	7.04 (2.13)	6.56 (2.42)	5.32 (2.21)	32.15 (5.74)	8
31	17	7.00 (2.32)	6.18 (2.70)	4.94 (2.08)	31.86 (5.32)	1
32	16	7.06 (2.41)	6.62 (3.01)	4.88 (2.13)	32.87 (5.25)	2
33	14	6.93 (2.59)	6.14 (2.77)	5.00 (2.15)	33.56 (5.62)	2
34	12	6.92 (2.47)	6.25 (2.26)	5.08 (2.31)	36.49 (3.19)	5
35	7	7.14 (2.41)	5.43 (3.21)	5.29 (1.80)	37.56 (2.66)	2
36	5	6.20 (1.30)	4.40 (2.41)	6.20 (0.84)	38.31 (2.82)	0
37	5	6.20 (1.30)	4.60 (2.61)	6.20 (0.84)	39.31 (3.00)	0
38	5	6.20 (1.30)	4.40 (2.41)	6.20 (0.84)	40.31 (2.92)	1
39	4	6.00 (1.41)	5.25 (1.71)	6.25 (0.96)	41.36 (3.43)	1
40	3	6.00 (1.73)	5.67 (1.16)	6.33 (1.16)	44.16 (2.41)	1
41	2	7.00 (0.00)	6.00 (1.41)	6.00 (1.41)	43.93 (1.67)	0
42	2	7.00 (0.00)	6.50 (2.12)	6.00 (1.41)	44.86 (1.65)	1
43	1	7.00 (-)	5.00 (-)	7.00 (-)	44.65 (-)	0
44	1	7.00 (-)	5.00 (-)	7.00 (-)	45.57 (-)	0
45	1	7.00 (-)	5.00 (-)	7.00 (-)	46.95 (-)	1

Multilevel models were conducted to estimate the average rate of change for offenders on the DRAOR domains, the relationship between initial scores on each domain and the respective rate of change, and whether covariates (e.g., static risk, age, race, remaining DRAOR domains) could explain variations in change on each DRAOR domain. Multilevel models explored the optimal combination of covariates for each DRAOR domain independently. Each set of analyses began with assessing the degree of variation in the dependent variable that was attributable to between-individual differences (i.e., ICC). Next, the unconditional growth model was examined, which required testing whether a linear rate of change best fit the data and whether there was variability in growth across offenders. Lastly, a final conditional model that incorporated relevant predictors of initial status and change over time was explored. The optimal combination of predictors was established first through examining each predictor independently (see Appendix B). Predictors that had a significant relationship with either initial status or change were then combined to arrive at a final model. Model fit statistics and Pseudo R^2 estimates were relied on to test whether the addition or removal of predictors meaningfully altered the overall results for a given model. Once final conditional models were established, the tenability of model assumptions was tested.

Model assumptions. Multilevel modelling invokes three assumptions that are testable: (a) the true functional form between outcome and predictors, (b) residuals at level-1 and level-2 are normally distributed, and (c) variances of the level-1 and level-2 residuals are equal at each level of every predictor (Singer & Willett, 2003). Assumptions

were assessed using the final models for each DRAOR domain with all relevant level-1 and level-2 predictors included. Findings remained consistent regardless of the DRAOR domain being examined. As a result, a summary of the findings pertaining to assumption checking across all final models is provided. A more thorough discussion of the functional form at level-1 is presented in the results for each DRAOR domain, but overall, there was support for linear change trajectories. An examination of the relationship between predicted individual growth parameters and relevant level-2 predictors also supported the linearity assumption. Normal probability plots of residuals against associated normal scores were visually inspected to determine whether the assumption of normality was tenable at both level-1 and level-2. Across the DRAOR domains, results indicated that residuals at both levels tended to be normally distributed, with minor departure evident in the normal probability plots. Cases with extreme standardized residuals (defined as the 1st or 99th percentile) were identified and removed to conduct a sensitivity analysis. The removal of observations associated with extreme residuals did not meaningfully alter the conclusions drawn from the final model for a given DRAOR domain, providing support to retain the full sample of offenders and assessments.

To evaluate the tenability of the assumption of homogeneity of variance for level-1 and level-2, residuals were plotted against predictors for each DRAOR domain. At level-1, residuals appeared to be evenly distributed across time, with a trend toward reduced variability after a considerable time on supervision has passed (i.e., after 18 months on supervision). Results for each DRAOR domain also indicated that variability in

intercept and slope parameters was approximately equal across values of each predictor (e.g., acute and protect scores when examining change on stable scores). Overall, the data appeared to meet the assumption of homogeneity of variance for level-1 and level-2 residuals across the three final models presented below. As a result, the findings presented are based on all observations available (n = 28,023 assessments for n = 4,000 supervision sequences).

Hypothesis 2. Scores on stable, acute, and protect will change throughout the course of community supervision. Further, there will be variation in initial score and rates of change, which will be partially explained by relevant covariates.

Change on stable scores. An examination of the ICC calculated from the unconditional means model suggested that approximately 77% of the variance in stable scores existed between offenders (see Table 8). The significant residual variance in this model indicated that the average offender's stable score varies over time, and the significant variance for the intercept suggests that offenders differ from each other in stable scores. Variability in rates of change was also observed when predicted stable scores, derived from an unconditional growth model, were plotted over time. A formal comparison of a model that did not permit variability in the rate of change (i.e., fixed effect for time) relative to a model that included random variability in the rate of change indicated that the model allowing for random variability in change was a better fit to the data (Δ-2LogLikelihood χ^2 (1) = 8501.8, p < .001). Next, a descriptive analysis of smoothed change trajectories for a random sample of 500 offenders suggested that

there was a linear decrease in stable scores over time (see Figure 3).[7] However, a formal test of model fit comparing a linear effect of time to a quadratic effect suggested that a quadratic effect of time provided a statistically superior fit (Δ-2LogLikelihood χ^2 (1) =26.9, $p < .001$). Given that previous research on the DRAOR has found support for a linear effect of time (Hanby, 2013; Polaschek & Yesberg, 2018), which was also supported by visual inspection of offender change patterns, the final unconditional growth model included a linear effect of time.

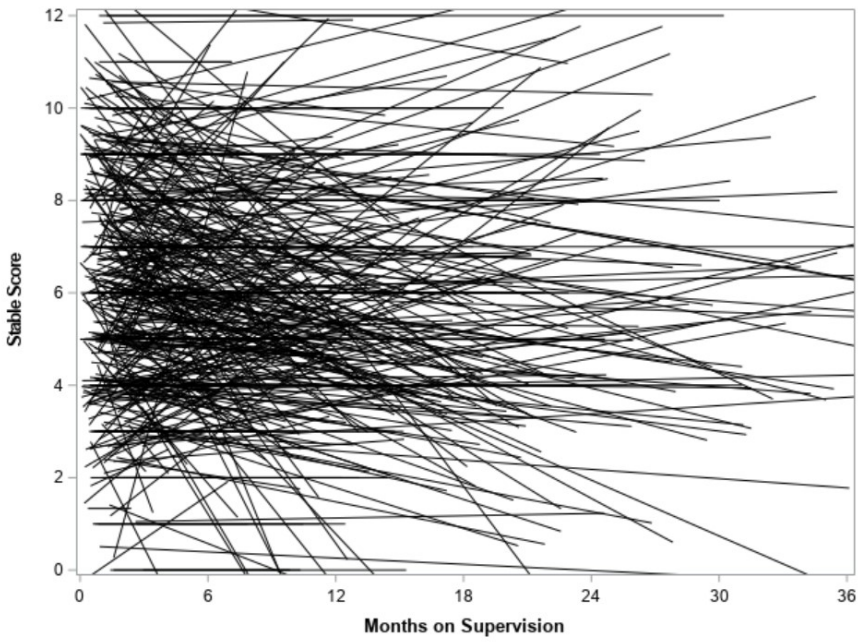

Figure 3. Individual change trajectories on stable domain during community supervision for random sample of supervision sequences (*n* = 500).

[7] A random sample of 500 supervision sequences was selected to be able to visually distinguish individual growth curves, which was challenging when the entire sample of supervision sequences (*n* = 4,000) was examined.

Results from the unconditional growth model provide an estimate of the initial stable score at the start of supervision and the rate of change for every month an offender remained on community supervision (see Table 8 and Figure 4). The fixed effect for the intercept indicated that an average offender begins their community supervision order with a stable score of approximately 6 and their score decreases by 0.06 for every month that they remain on community supervision. There was significant variation associated with the estimates of initial status and rates of change, suggesting that it was worthwhile to model predictors to explain the heterogeneity. Modelling a linear effect of time resulted in a 49% decrease in the within-individual variance associated with stable scores (1.46 to 0.74). This suggested that, although half of the within-individual variance was explained by time, there was a large portion of variance around each offenders' change trajectory that could be explained by the inclusion of time-varying predictors. Lastly, the covariance of the residuals was converted to a correlation coefficient to examine the relationship between initial status and change. There was a strong negative relationship ($r = -0.69$), indicating that higher scores at the start of supervision were associated with higher rates of change (i.e., negative scores).

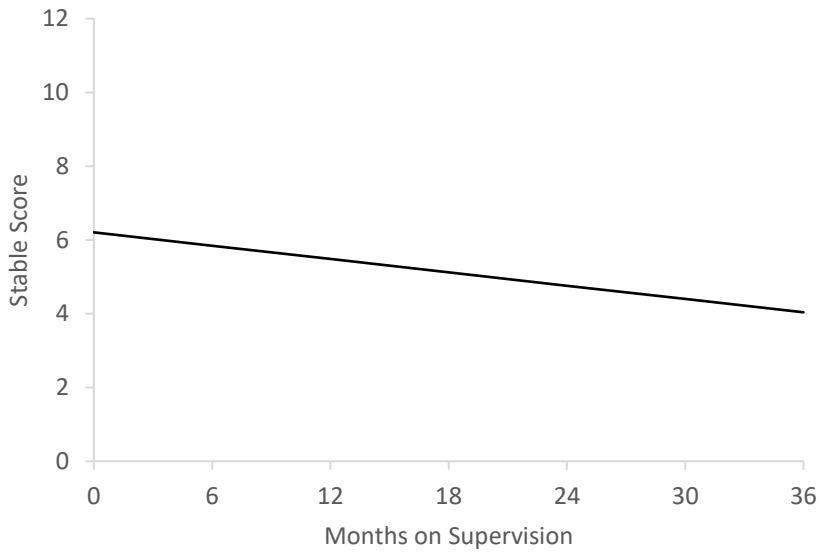

Figure 4. Unconditional growth on stable scores as a function of time on supervision.

Table 8

Results of Multilevel Models for Change on DRAOR Stable Scores (n = 4,000 Supervision Sequences)

	Model A Estimate (SE)	Model B Estimate (SE)	Model C Estimate (SE)
Fixed effects			
Initial Status			
Intercept	5.86 (0.04)**	6.21 (0.04)**	6.00** (0.03)
Protect – within			-0.34** (0.01)
Protect – between			-0.28** (0.01)
Acute – within			0.31** (0.01)
Acute – between			0.38** (0.01)
Rate of change			
Time		-0.06 (0.003)**	-0.01** (0.01)
Acute – within			0.001** (0.0004)
Acute – between			0.005** (0.001)
Variance Components			
Within-person	1.46 (0.01)**	0.74 (0.01)**	0.46** (0.005)
Initial status	4.92 (0.12)**	6.26 (0.151)**	4.22** (0.10)
Covariance		-0.18 (0.01)**	-0.12** (0.01)
Rate of change		0.03 (0.001)**	0.01** (0.001)
Pseudo R^2 and Fit statistics			
$R^2_{y,y}$		0.00	0.36
R^2_e		0.49	0.68
R^2_0			0.33
R^2_1			0.54
-2 Log Likelihood	102248	92901	78964
AIC	102254	92913	78988
BIC	102273	92951	79063

Note: **$p < .001$. Model A is an unconditional means model; Model B incorporates the number of months on supervision at the time of each assessment in the unconditional growth model; Model C is a conditional model, which incorporates within (person-mean centered) and between effects (grand mean centered) of acute and protect scores. Models were performed with SAS 9.4 using ML estimation.

An iterative process was utilized to arrive at the final conditional model. First, the relationship between individual predictors of initial status and change over time was explored. Separate models were examined for age, race, static risk, protect scores, and

acute scores. Protect and acute scores were person-mean centered to allow for an examination of both the within-individual and between-individual effect (Howard, 2015). Initial results indicated that age, race, and static risk did not meaningfully explain variation in either the initial stable score at the start of supervision or change over time (see Appendix B). Results did identify both the time-varying and between-individual effects of protect and acute as relevant predictors of variations in trajectories of change on stable scores, thus forming the final conditional model. The final model included the within and between effects of acute as a predictor of both initial status and change, but the effects of protect were only included as a predictor of initial status. Figure 5 demonstrates the within and between effect of protect scores separately, while holding all other predictors at their average value. According to this model, when an individual increased their protect score by one standard deviation, they were expected to shift to a trajectory that would have started community supervision with a stable score that was 0.34 points lower, while holding the between group effects of protect and acute, and the within effect of acute constant. Similarly, each 1-point increase in protect scores was associated with a 0.28 decrease in stable scores at the start of supervision, while holding the effects of the remaining predictors constant. For example, an offender with protect score 1 standard deviation above the average was expected to start community supervision with a stable score of approximately 5, whereas an offender 1 standard deviation below the average was expected to have a stable score of nearly 7.

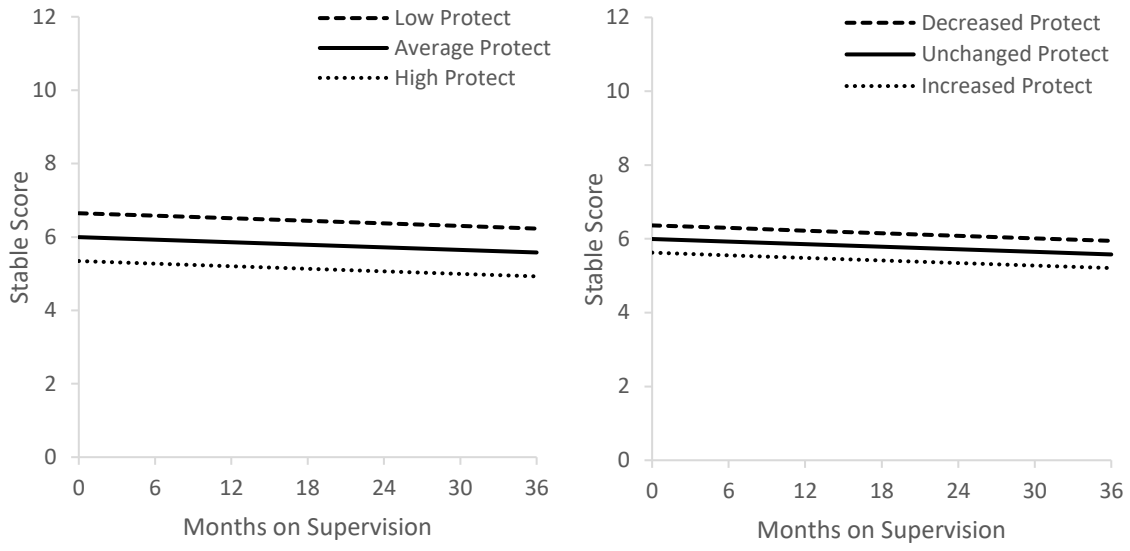

Figure 5. Prototypical growth trajectories on stable score considering between-offender effect of protect (left) and time-varying effect of protect (right).

Conversely, acute scores were positively associated with both initial stable scores and change over time. Offenders with higher acute scores tended to start community supervision with higher stable scores (0.38) and tended to increase in their stable scores over time (0.005). Further, when an offender scored higher than their personal average, they were expected to have higher initial stable scores (0.31) and slightly increase in stable scores over time (0.001). Figure 6 depicts fitted trajectories for the between-individual effect of acute scores on the initial status and change, while holding the remaining effects constant. A simple slope analysis indicated that offenders scoring 1 standard deviation below the mean acute score tended to start community supervision with a stable score of 5 and decreased by 0.02 for each month they remained on community supervision. Conversely, offenders with an acute score 1

standard deviation above the mean tended to start supervision with a stable score of nearly 7 and did not demonstrate significant change over the course of the supervision period (slope = -0.001, *p* = .85).

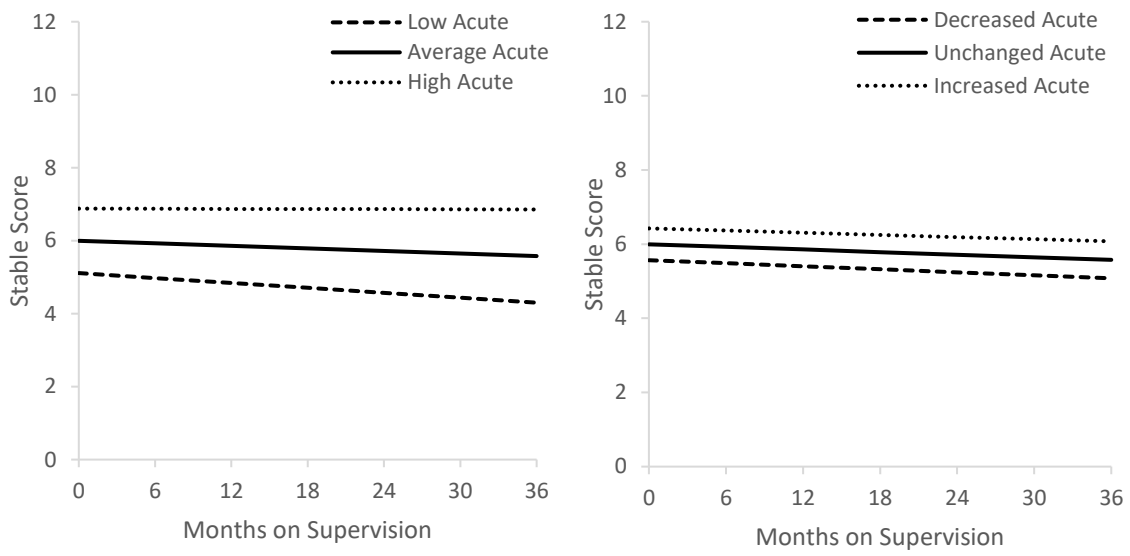

Figure 6. Prototypical growth trajectories for stable score when considering between-offender effects of acute (left) and time-varying effects of acute (right).

The within-individual effect of acute indicated that when an offender scored higher than their average score, their trajectory shifted upward leading to higher stable scores. Although results did suggest that the within-individual effect of acute scores changed over time (0.001, $p < .001$), Figure 6 demonstrates that the expected trajectories for individuals who were one standard deviation above and below their personal average were relatively parallel. As time went on, the effect of increasing individual acute scores was associated with slightly more change in stable scores than compared to the same increase in acute scores at the beginning of supervision.

An examination of the variance components associated with this model indicated that there remained unpredicted variation that might be able to be explained by modelling additional predictors. The combination of between and within effects of acute and protect explained approximately 36% of the total variation in stable scores. Including the time-varying effects of acute and protect contributed to explaining the majority (68%) of within-person variability in stable scores. There were still large portions of variance in initial status (67%) and change (46%) that remained unexplained by the combination of predictors. The partial covariance between the initial status and change also remained significant. As a result, after accounting for the effects of the predictors in the model, results indicated that there was a negative relationship (r = -0.47) between initial stable scores and change over time.

Change on acute scores. The ICC from the unconditional means model indicated that 70% of the variation in acute scores existed between offenders (see Table 9). The significant residual variance in this model indicated that the average offender's stable score varied over time, and the significant variance for the intercept suggested that offenders differed from each other in stable scores. Plotting predicted acute scores from an unconditional growth model also suggested that there was variability in the rates of change over time. A formal comparison of a model that did not permit variability in the rate of change (i.e., fixed effect for time) relative to a model that included random variability in the rate of change indicated that the model allowing for random variability in change was a better fit to the data (Δ-2LogLikelihood χ^2 (1) = 6550.3, p < .001). Next, a descriptive analysis of smoothed change trajectories for a random sample of 500

offenders indicated that acute scores decreased linearly with time (see Figure 7). Consistent with the analysis of functional form for stable scores, a likelihood ratio test between models incorporating a linear effect of time and a quadratic was significant (Δ-2LogLikelihood χ^2 (1) =73.9, $p < .001$). The support from the observed change trajectories and previous research outweighed the formal hypothesis test for the optimal functional form, leading to an unconditional growth model that incorporated a linear effect of time on acute scores.

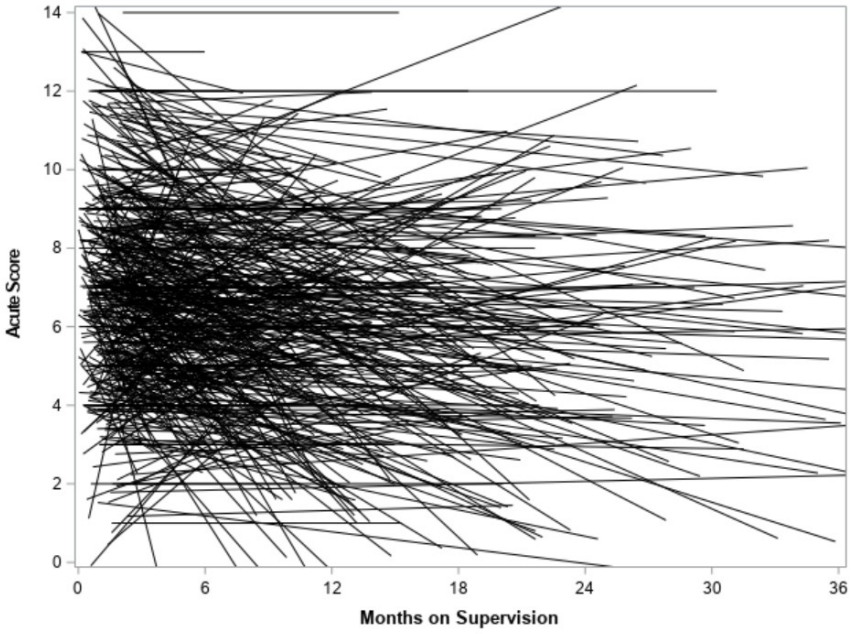

Figure 7. Individual change trajectories on acute domain during community supervision for random sample of supervision sequences (n = 500).

Table 9

Results of Multilevel Models for Change on DRAOR Acute Scores (n = 4,000 Supervision Sequences)

	Model A Estimate (SE)	Model B Estimate (SE)	Model C Estimate (SE)
Fixed effects			
Initial Status			
Intercept	6.35** (0.04)	6.86** (0.04)	6.66** (0.04)
Protect – within			-0.26** (0.01)
Protect – between			-0.13** (0.01)
Stable – within			0.58** (0.01)
Stable – between			0.49** (0.02)
Rate of change			
Time		-0.09** (0.004)	-0.04** (0.003)
Variance Components			
Within-person	2.18** (0.02)	1.24** (0.01)	0.84** (0.01)
Initial status	5.21** (0.13)	6.88** (0.17)	5.22** (0.13)
Covariance		-0.22** (0.01)	-0.17** (0.01)
Rate of change		0.04** (0.001)	0.01** (0.001)
Pseudo R^2 and Fit statistics			
$R^2_{y,y}$		0.005	0.30
R^2_e		0.43	0.61
R^2_0			0.24
R^2_1			0.49
-2 Log Likelihood	112315	104557	93233
AIC	112321	104569	93253
BIC	112340	104607	93316

Note: **p<.001. Model A is an unconditional means model; Model B incorporates the number of months on supervision at the time of each assessment in the unconditional growth model; Model C is a conditional model, which incorporates within (person-mean centered) and between effects (grand mean centered) of stable and protect scores. Models were performed with SAS 9.4 using ML estimation.

Results from the unconditional growth model provided an estimate of the initial acute score at the start of supervision and the rate of change for every month an offender remained on community supervision (see Table 9 and Figure 8). The fixed effect for the intercept indicated that the average offender began their community

supervision order with an acute score of nearly 7 and their score decreased by 0.09 for every month that they remained on community supervision. There was significant between-offender variation associated with these estimates of initial status and rates of change, suggesting that it was worthwhile to model predictors to explain the heterogeneity. Modelling a linear effect of time resulted in a 43% decrease in the within-individual variance associated with stable scores (2.18 to 1.24). This suggested that there remained a large portion of variance around each offenders' change trajectory that could be explained by the inclusion of time-varying predictors. Further, according to the pseudo-R^2 for total outcome variability, modelling an effect for linear time did not explain any variability in acute scores. Results did indicate that there was a moderate negative relationship between the true rate of change in acute scores and the initial score at the start of supervision (r = -0.44).

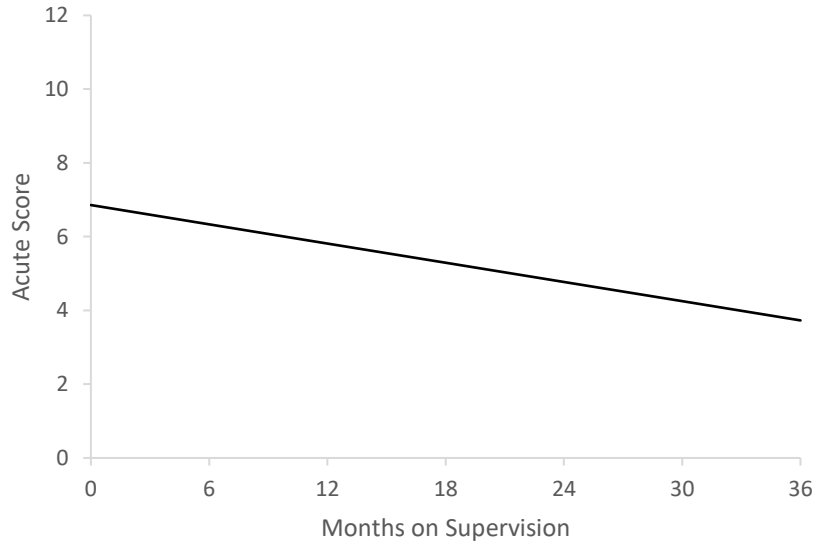

Figure 8. Unconditional growth on acute scores as a function of time on supervision.

Several predictors were explored in an attempt to explain some of the variability in acute scores at the start of supervision and change over time. Initially, separate models were examined for static risk, age, race, protect scores, and stable scores. Results from these preliminary models indicated that static risk, age, and race were not strong predictors of either initial status or change over time (see Appendix B). Stable scores and protect scores appeared to be the most relevant predictors, although neither was related with change over time. As a result, a final conditional model incorporated the time-varying and the between-individual effects of stable and protect scores on the initial acute score. Results indicated that an offender with average protect and stable scores who was also at their personal average on both constructs was expected to start community supervision with an acute score of 6.66. For every 1-point increase in protect scores, there was a corresponding 0.13 decrease in initial acute score, while holding the effects of the remaining predictors constant. This amounted to little separation between those scoring one standard deviation above and below the average on protect (see Figure 9). While holding the effects of the remaining predictors constant, an offender scoring one standard deviation below the mean of protect at the start of supervision was expected to have an acute score of 7, while an offender with a protect score 1 standard deviation above the mean was expected to have an acute score of approximately 6. There was also evidence of a within-individual effect of protect scores of a similar magnitude, such that an offender with increased protect scores one standard deviation over their average was expected to shift to a trajectory that started

at an acute score of approximately 6, whereas an offender who decreased one standard deviation from their average was expected to shift to a trajectory with an acute score starting at 7.

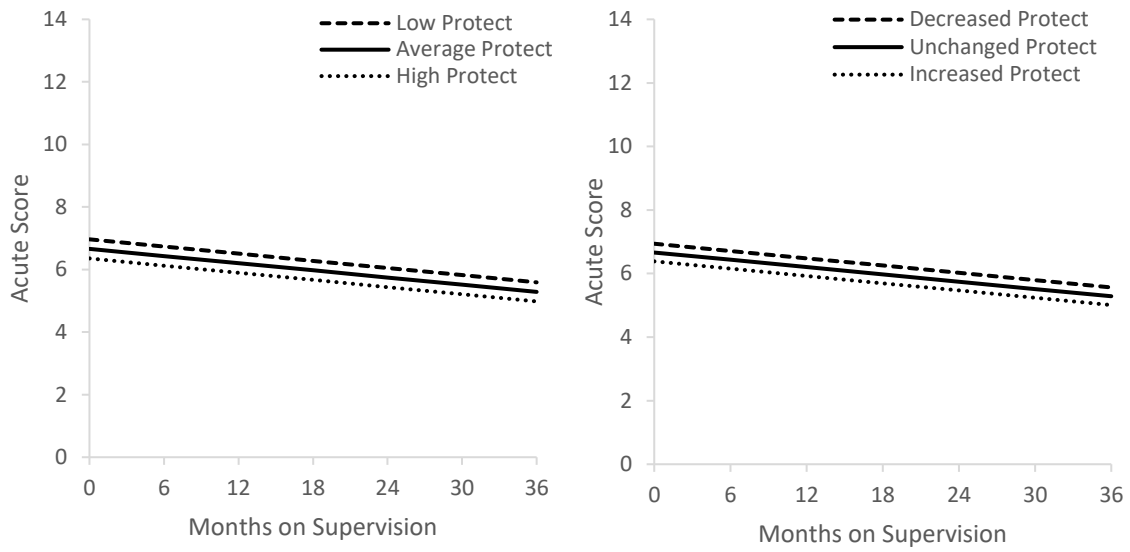

Figure 9. Prototypical growth trajectories on acute scores when considering the between-offender effects (left) and the time-varying effects (right) of protect scores.

There was also evidence of a positive relationship between stable scores and the acute score at the start of supervision. As displayed in Figure 10, For every 1-point increase in stable score, there was an expected increase of 0.58 in the initial acute score, while holding the effects of the remaining predictors constant. In other words, offenders with higher stable scores tended to start community supervision with higher acute scores. Given that there was no interaction with time, the differences in the initial acute score corresponding to different stable scores remained constant throughout community supervision. The time-varying effect for stable scores was similar to the

effect of protect scores, in that offenders who increased their stable scores one standard deviation above their average were expected to shift their trajectory so that their initial acute score was 7, whereas a decrease of one standard deviation below their personal average was associated with an initial acute score of 6.

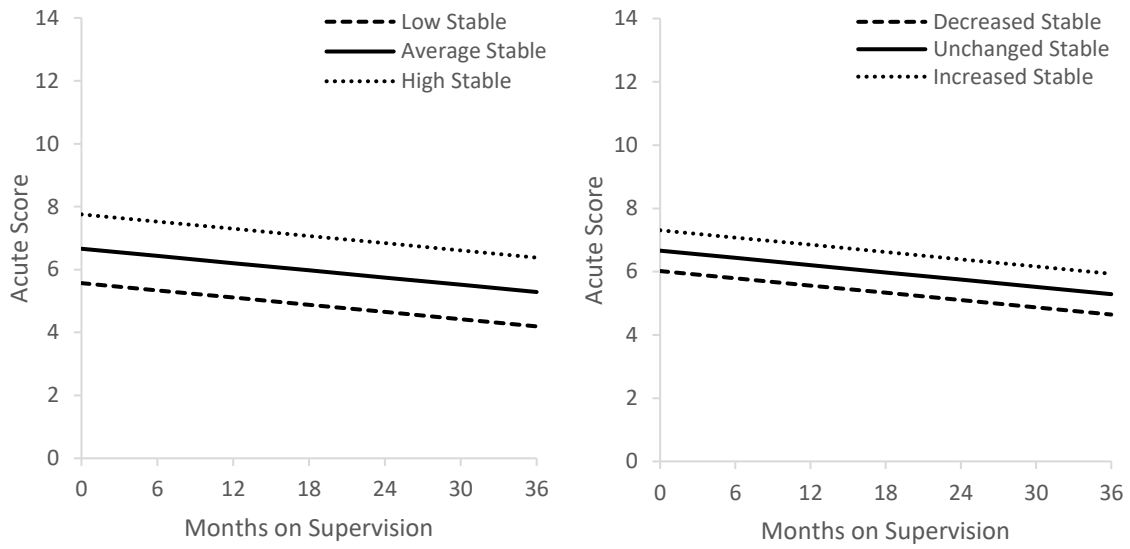

Figure 10. Prototypical change trajectories on acute when considering the between-offender effects (left) and the time-varying effects (right) of stable scores.

Given that the modelling process did not yield significant predictors of change, it was unsurprising that there remained considerable unexplained variation in acute scores. Collectively, the between and within effects of stable and protect, alongside linear time, explained approximately 30% of the total variation in acute scores. Modelling the within-individual effects of stable and protect increased the proportion of within-person variation explained (61%) compared to the unconditional growth model (43%). The model explained a small portion of the variance (24%) in initial status and

half of the variance associated with change over time (49%), indicating that further analyses that incorporate additional predictors may be worthwhile. It is noteworthy that the partial covariance between the initial acute scores and change over time remained significant. Converting this to a correlation indicated that, after controlling for the effects of the predictors in the model, there was a negative relationship (r = -0.52) between initial acute scores and change over time.

Change on protect scores. An examination of the variance components associated with the unconditional model indicated that 81% of the variation in protect scores existed between offenders. Variance associated with the initial status and within-individuals indicated offenders differed from each other in protect scores and the average offender's protect score varied over time. To examine whether rates of change on protect score vary across individuals, an unconditional growth model that fit a random effect for linear time was computed. A formal comparison between a model that fixed the effect of time across all individuals relative to a model that permitted variability in the rate of change provided support for a random effect of time ((Δ-2LogLikelihood χ^2 (1) = 8771.7, p < .001). Visual inspection of smoothed individual change trajectories for a random sample of 500 offenders indicated that there was a linear increase in protect scores over time (see Figure 11). A linear effect of time was further supported by the lack of a significant improvement in model fit when exploring a quadratic trend for time ((Δ-2LogLikelihood χ^2 (1) = 4.00, p = .045). As a result, an unconditional growth model that incorporated a random effect for linear time was fit to the data.

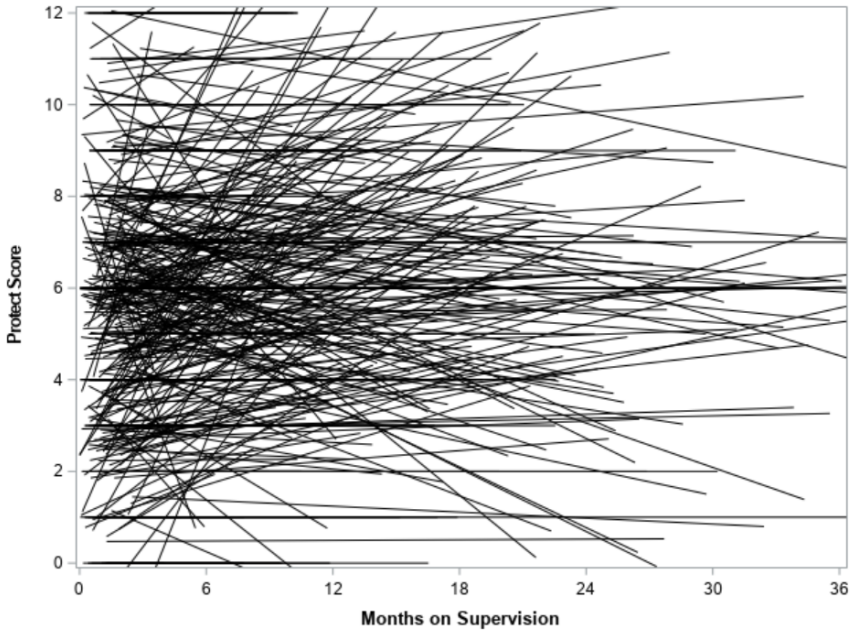

Figure 11. Individual change trajectories on protect domain during community supervision for random sample of supervision sequences (*n* = 500).

The fixed effects from the unconditional growth model indicated that the average offender began community supervision with a protect score of nearly 6 and their score increased by 0.05 every month on community supervision (see Table 10 and Figure 12). There was significant between-offender variation associated with these estimates, indicating that there was the potential to explain some of the heterogeneity by modelling predictors. The within-person variance decreased by 49% in the conditional growth model (1.33 to 0.69), indicating that half of the observed within-person variance was associated with modelling linear time. The remaining half of the within-person variance around each offender's linear change trajectory could potentially be explained by the inclusion of time-varying predictors. Further, the pseudo-R^2 for total

outcome variability indicated that simply modelling the linear effect did not explain any of the observed variation in protect scores. The covariance between the initial status and slope suggested that there was a negative relationship between the true rate of change in protect scores and the initial score at the start of supervision (*r* = -0.40).

Table 10

Results of Multilevel Models for Change on DRAOR Protect Scores (n = 4,000 Supervision Sequences)

	Model A Estimate (SE)	Model B Estimate (SE)	Model C Estimate (SE)
Fixed effects			
Initial Status			
Intercept	5.72** (0.04)	5.41** (0.04)	5.55** (0.04)
Acute – within			-0.16** (0.005)
Acute – between			-0.11** (0.02)
Stable – within			-0.37** (0.01)
Stable – between			-0.39** (0.02)
Rate of change			
Time		0.05** (0.003)	0.02** (0.003)
Acute—between			-0.006** (0.001)
Variance Components			
Within-person	1.33** (0.01)	0.69** (0.01)	0.50** (0.01)
Initial status	5.67** (0.13)	7.10** (0.17)	5.75** (0.14)
Covariance		-0.18** (0.01)	-0.14** (0.01)
Rate of change		0.03** (0.001)	0.02** (0.001)
Pseudo R^2 and Fit statistics			
$R^2_{y,y}$		0.00	0.22
R^2_e		0.49	0.63
R^2_0			0.19
R^2_1			0.40
-2 Log Likelihood	100690	91321	82163
AIC	100696	91333	82185
BIC	100715	91371	82255

Note: **p<.001. Model A is an unconditional means model; Model B incorporates the number of months on supervision at the time of each assessment in the unconditional growth model; Model C is a conditional model, which incorporates within (person-mean centered) and between effects (grand mean centered) of stable and acute scores. Models were performed with SAS 9.4 using ML estimation.

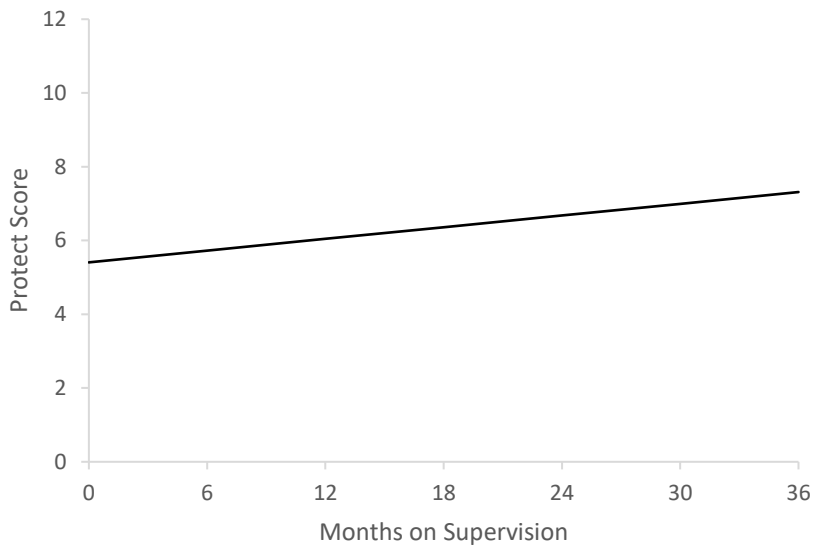

Figure 12. Unconditional growth on protect as a function of months on supervision.

Next, several variables were explored as potential predictors of the variability in either the initial status of protect or change over time. Separate models were examined for static risk, age, race, stable scores, and acute scores. Consistent with the results from the growth models examining stable and acute scores, there was no evidence that static risk, age, or race were strong predictors of either initial status or change on protect scores (see Appendix B). Stable and acute scores, however, did emerge as relevant predictors and were incorporated into a conditional model that examined the time-varying and the between-individual effects. Initially, models explored whether these predictors were related to both initial status and change over time. The between-individual effect for acute scores was the only predictor that significantly explained some of the variation in the rates of change, leading to a final conditional model that

incorporated the remaining covariates as predictors of initial status only. The fixed effects from this final model indicated that an offender with average stable and acute scores, also scoring at their personal average on both constructs, was expected to start supervision with a protect score of 5.55. The between-individual effect for acute score on initial status was negative, indicating that an offender with higher than average acute scores, relative to other offenders, was expected to start community supervision with a lower protect score. A simple slope analysis probing the significant interaction with change over time indicated that those with high acute scores maintained their level of protect score over time, while offenders with average acute or low acute scores increased their protect score over time (see Figure 13). There was also a significant time-varying effect of acute scores, such that when an offender scored above their personal average, they were expected to shift to a change trajectory that had lower protect scores. The magnitude of this effect was small, however, resulting in little differentiation in initial protect scores between an offender scoring one standard deviation higher than their average acute score (5.33) versus an offender scoring one standard deviation below their average (5.77).

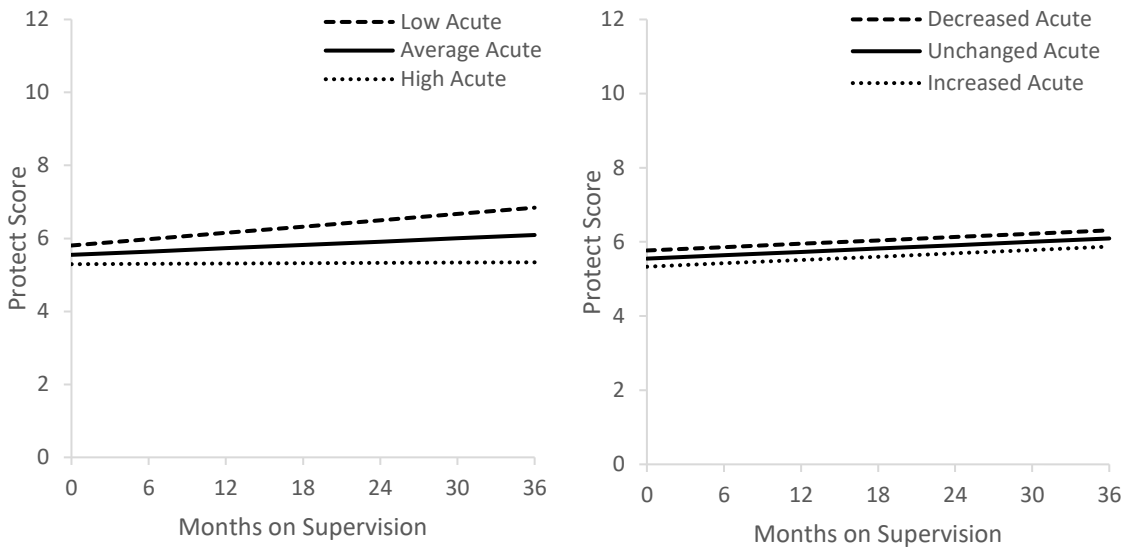

Figure 13. Prototypical change trajectories on protect considering the between-offender effects (left) and time-varying effects (right) of acute scores.

There was also evidence of significant between- and within-individual effects for stable scores on the initial status of protect. Similar to acute scores, the direction of the between-individual effect of stable scores was negative, indicating that an offender who scored higher than the average stable score for all offenders was expected to have a lower protect score at the start of supervision. For example, Figure 14 demonstrates that, while holding the effects of all additional covariates constant, an offender who scored one standard deviation above the mean stable score was expected to start supervision with a protect score of 4.7, whereas an offender scoring one standard deviation below the mean was expected to begin supervision with a protect score of 6.4. There was a similar result for the time-varying effect of stable scores, such that at

times when an offender scored higher than their average, they were expected to shift to a change trajectory that was characterized by lower protect scores. Consistent with the time-varying effect of acute scores, the magnitude of this effect was small, resulting in little differentiation in initial protect scores between an offender scoring one standard deviation higher than their average stable score (5.13) versus an offender scoring one standard deviation below their average (5.97).

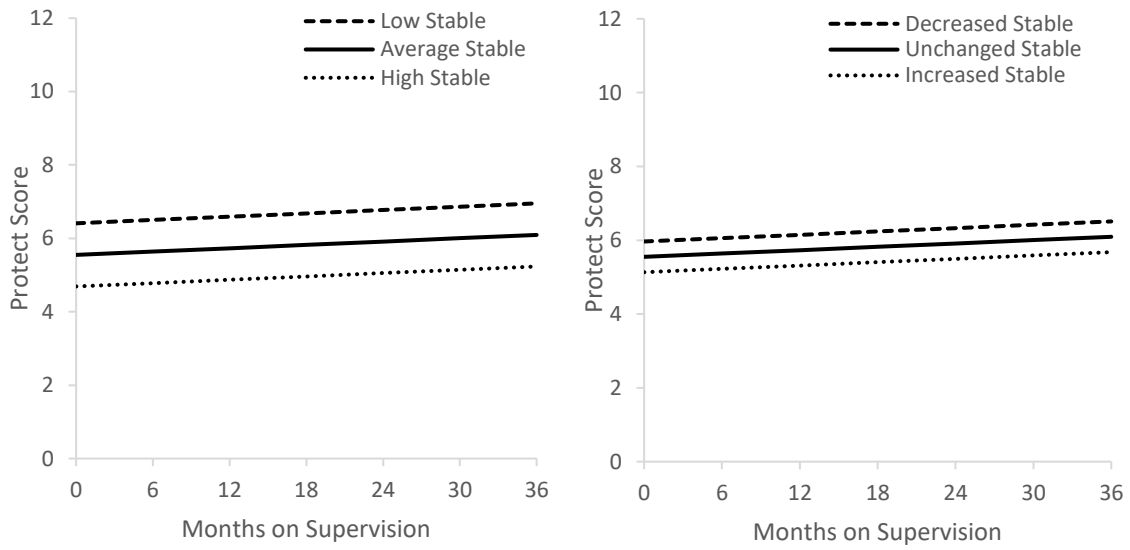

Figure 14. Prototypical change trajectories on protect considering the between-offender effects (left) and time-varying effects (right) of stable scores.

Despite the presence of significant predictors explaining variability in initial status and change, each variance component in the conditional model remained significant. The combination of predictors in the conditional model only explained approximately 22% of the total variation in protect scores. The majority of within-person variance (63%) was explained by the inclusion of linear time and the time-

varying effects of stable and acute scores. A small portion (19%) of the variation in the initial protect scores at the start of supervision was explained by the collection of predictors included in the model, whereas 40% of the variation in the change over time was explained by linear time and the between-individual effect of acute scores. Taken together, the large portions of unexplained variation for each component signify that it may be fruitful to consider modelling additional predictors to better understand change trajectories on protect scores. The partial covariance between the initial status and change also remained significant. As a result, after accounting for the effects of the predictors in the model, initial protect scores were negatively related to change over time (r = -0.45).

Summary of growth curve results. A series of growth curve models were examined to determine whether DRAOR domain scores changed throughout the course of community supervision. Further, in the presence of variability in initial status and change over time, it was of interest to determine whether several within and between predictors were relevant for explaining the variation. Overall, results supported hypothesis 2, in that there was evidence that scores on the DRAOR were changing over time. Specifically, stable and acute scores decreased throughout community supervision, on average, while protect scores increased. Interestingly, static risk, age, and race did not emerge as significant predictors of variations in either initial DRAOR scores or rates of change, which was counter to hypothesis 2a. As hypothesized, results indicated that scores on a given DRAOR domain tended to explain variations in initial

scores on another. Core findings from the growth curve models are highlighted below, separated by DRAOR domain.

Stable. Results indicated that the typical offender would be expected to start community supervision with a stable score of 6.2 and decrease linearly throughout community supervision. Scores on DRAOR acute and protect were found to explain a portion of the variation in both initial status and slope. Between group effects for acute indicated that higher scores were associated with higher stable scores at the start of supervision and less improvement (i.e., reductions) in stable scores over time, while holding the effects of the remaining predictors constant. Protect scores explained variation in the initial stable score, but there was no relationship with rates of change in stable scores over time. Time-varying effects were noted for acute on both initial status and change, such that an individual who decreased their acute score was expected to shift to a trajectory characterized by a lower initial stable score and slightly more rapid decreases throughout community supervision. There remained significant unexplained variation around each individual's growth trajectory as well as between individuals in both initial status and change over time, indicating that modelling additional predictors may be worthwhile.

Acute. The typical offender was expected to start community supervision with an acute score of nearly 7 and decrease linearly throughout the course of community supervision. There were no significant predictors of variation in change in acute scores over time. However, between group differences on stable and protect were found to explain variation in the acute score at the start of supervision. Higher stable scores were

related to higher initial acute scores, while holding the effects of the remaining predictors constant. There were also time-varying effects of stable and protect, although changes in protect score were associated with minimal changes in initial acute score. An offender who decreased their stable score was expected to shift to a trajectory that was associated with a lower initial acute score. Similar to the conditional model for stable scores, there remained significant unexplained variation around each individual's growth trajectory as well as between individuals in both initial status and change over time, indicating that modelling additional predictors may be worthwhile.

Protect. Results indicated that protect scores tended to increase linearly throughout the course of community supervision, with the average offender beginning supervision with a score of 5.4. Predictors of initial protect score and change over time were identified. Between-individual scores on acute suggested that those with lower acute scores tended to begin supervision with higher protect scores and increased their score more rapidly throughout community supervision. Importantly, offenders with high acute scores did not change on their protect score throughout community supervision, while holding the effects of the remaining covariates constant. Within-individual effects of stable and acute on the initial protect score were identified, indicating that offenders who increased in either score were expected to shift their growth trajectory to one with lower protect scores at the start of supervision. There remained significant unexplained variation around each individual's growth trajectory as well as between individuals in both initial status and change over time, indicating that modelling additional predictors may be worthwhile.

Do Growth Trajectories Predict Community Outcomes?

Offender specific growth trajectories were constructed from the results of the unconditional growth models for each of the DRAOR domains. The trajectories consisted of two parameters, representing the initial starting point and the slope coefficient for each individual (see Table 11). It is important to note that although the average rate of change per month across each of the domains was relatively small, there was considerable variation in these estimates. Table 12 presents the expected rate of change on each DRAOR domain extrapolated over a 12-month period to provide a summary of the distribution of change observed among the sample. Consistent with the average rates of change, assessing the distribution of change over a given year indicated that for the majority of the sample domain scores were not changing by more than 1 point. That being said, sizable proportions of the sample decreased by at least 2 points within a year on stable and acute (17% and 24%, respectively), while 14% of the sample increased by two or more points on the protect domain within a year on supervision.

A series of Cox regression survival analyses were conducted, first to examine the relationship between initial scores and change on each domain with both revocations from community supervision and any new conviction. Following the examination of each DRAOR domain independently, models that included relevant covariates were explored. Static risk, age, and race were first independently examined to determine whether each was significantly related to the community outcomes. Static risk and age emerged as significant predictors of both revocations of community supervision and new convictions, but race did not appear to be a relevant predictor (see Appendix C). As a

result, the effects of static risk and age were accounted for in multivariate models to determine whether DRAOR scores (initial status and change) remained related to the community outcomes. Lastly, a final model was explored that combined initial scores and change scores for each DRAOR domain, while accounting for static risk and age.

Table 11

Average Initial and Change Scores for Stable, Acute, and Protect (n = 4,000 Supervision Sequences)

	M (*SD*)	Range
Stable		
Initial score	6.21 (2.42)	-0.80 to 14.00
Change	-0.06 (0.15)	-0.95 to 0.87
Acute		
Initial score	6.88 (2.51)	-0.88 to 14.34
Change	-0.09 (0.16)	-0.95 to 0.73
Protect		
Initial score	5.41 (2.60)	-2.18 to 13.20
Change	0.05 (0.14)	-0.70 to 0.98

Note. Estimates for each individual's initial score and change over time were derived from unconditional growth multilevel models. Initial scores represent the random intercept coefficients and the change scores represent the random slope coefficients.

Table 12

Distribution of Change on Each DRAOR Domain Over a 12-Month Period

	Stable	Acute	Protect
Degree of change	% (n)	% (n)	% (n)
Decreases of 2 or more	17 (682)	24 (954)	4 (140)
Decreases greater than 1, but less than 2	17 (659)	23 (935)	5 (209)
Negligible change	56 (2,254)	43 (1,715)	63 (2,502)
Increase greater than 1, less than 2	6 (228)	5 (215)	15 (592)
Increases of 2 or more	4 (177)	5 (181)	14 (557)

Note. The magnitude of change over a 12-month period was determined by multiplying each individual's monthly change parameter by 12. Deviations in domain score of less than one after 12 months were considered negligible.

Prior to examining each of the models described above, the tenability of the proportionality of hazards assumption was evaluated. Cox Regression invokes an assumption that the relationship between survival rate and time is equivalent at all levels of each covariate included in the model (Tabachnick & Fidell, 2013). To test this assumption, an interaction term between each covariate and the natural logarithm of time was created and entered into a multivariate model. Alpha was adjusted by the number of covariates to protect against Type 1 errors. For example, the final model included 8 interaction terms, resulting in an adjusted alpha of .006 (.05/8) for testing the proportional hazards assumption for this model. Across each of the models for both revocation and new convictions, the interaction terms between covariates and time did not exceed this threshold, indicating that the proportionality of hazards assumption was satisfied for all models of interest.

Hypothesis 3. Changes on stable, acute, and protect will enhance the prediction of revocations of community supervision and new convictions over initial scores.

Hypothesis 4. While considering static risk, both initial and change scores on stable, acute, and protect will be related to revocations of community supervision and new convictions.

Stable scores. Table 13 presents results examining the incremental validity of incorporating stable change scores, in addition to the initial stable score, when predicting revocations of community supervision and new convictions. Results indicated that both initial stable score and change on stable scores were significant predictors of revocations of community supervision. However, the c-index associated with this model

was .62, indicating that collectively, initial and change scores represented a small effect. Next, it was of interest to determine whether the relationships between initial status and change with revocations were maintained when considering static risk and age. Initial stable score and change remained significant predictors of revocations of community supervision, such that increases in initial score and change (indicating increasing risk over time) were associated with an increased hazard of being revoked. The collection of predictors was associated with a c-index of .67, representing a moderate effect and an improvement over the model that examined only stable scores.

The findings did not remain consistent when examining initial stable scores and change in relation to new convictions. Change scores did not significantly predict new convictions, while accounting for initial stable score. It is noteworthy that the direction of the effect for change in stable scores was consistent, such that increases in stable scores over time were associated with an increased hazard of being convicted of a new charge. The initial stable score did remain a significant predictor of new convictions, while accounting for the effects of static risk, age, and change over time. Although the combination of initial stable score and static risk improved the prediction of new convictions, the effect size associated with this model remained small (c-index = .62).

Table 13

Cox Regression Results for Stable Scores Derived from MLM

	Predictor	*B*	*SE*	Wald	*p*	Hazard Ratio [CI]	c-index [CI]
Revocation of community supervision[a]	Model A						.62 [.60, .64]
	Initial stable score	0.14	0.01	10.8	< .001	1.15 [1.12, 1.18]	
	Change	2.91	0.24	12.0	< .001	18.33 [11.40, 29.48]	
	Model B						.67 [.65, .68]
	Initial stable score	0.11	0.01	8.41	< .001	1.12 [1.09, 1.15]	
	Change	2.74	0.24	11.40	< .001	15.53 [9.69, 24.88]	
	Static risk	0.09	0.01	11.83	< .001	1.10 [1.08, 1.12]	
	Age	-0.01	0.003	-3.19	.001	0.99 [0.99, 1.00]	
New charge resulting in a conviction[b]	Model A						.56 [.53, .59]
	Initial stable score	0.08	0.02	3.67	< .001	1.08 [1.04, 1.13]	
	Change	0.63	0.36	1.76	.08	1.88 [0.93, 3.78]	
	Model B						.62 [.59, .65]
	Initial stable score	0.05	0.02	2.16	.031	1.05 [1.00, 1.09]	
	Change	0.41	0.35	1.20	.230	1.53 [0.76, 3.07]	
	Static risk	0.08	0.01	6.49	< .001	1.09 [1.06, 1.11]	
	Age	-0.02	0.004	-4.60	< .001	0.98 [0.97, 0.99]	

Note: SE = Standard Error, CI = 95% confidence interval.

[a]1,087 of 4,000 community supervision sequences ended with a revocation.

[b]427 of 4,000 community supervision sequences had a new conviction associated with it throughout the follow-up.

Acute scores. The initial acute score and change over time was examined to determine if considering change incrementally predicted revocations from community supervision. As presented in Table 14, change scores were significant predictors of revocations, while accounting for the effects of initial acute scores. However, the c-index associated with this model was .62, indicating that collectively, initial and change scores represented a small effect. Results remained consistent when static risk and age were included as predictors of revocations. Initial acute score and change over time remained significant predictors, alongside static risk and age. The effect size associated with this model was moderate (c-index = .66), indicating that a combined model including acute initial score and change, static risk, and age improved the prediction of revocations relative to a model that only considered initial acute scores and change.

Survival analyses examining new convictions produced results that were inconsistent with the results for revocations. Change on acute scores was not a significant predictor of new convictions, while accounting for the effect of initial acute score (see Table 14). The c-index for this model was small, indicating that the combined predictive effect of initial acute score and change was slightly better than chance. The overall effect size did improve when acute scores were modelled alongside static risk and age, but the magnitude of the effect remained small (c-index = .62). Initial acute scores remained a significant predictor of new convictions in this combined model.

Table 14

Cox Regression Results for Acute Scores Derived from MLM

	Predictor	*B*	*SE*	Wald	*p*	Hazard Ratio [CI]	c-index [CI]
Revocation of community supervision[a]	Model A						.61 [.59, .63]
	Initial acute score	0.13	0.01	10.57	< .001	1.14 [1.12, 1.17]	
	Change	2.43	0.22	10.91	< .001	11.33 [7.32, 17.52]	
	Model B						.66 [.64, .68]
	Initial acute score	0.11	0.01	9.02	< .001	1.12 [1.09, 1.15]	
	Change	2.28	0.22	10.36	< .001	9.76 [6.34, 15.03]	
	Static risk	0.10	0.01	12.25	< .001	1.10 [1.09, 1.12]	
	Age	-0.01	0.003	-3.26	.001	0.99 [0.98, 1.00]	
New charge resulting in a conviction[b]	Model A						.54 [.51, .57]
	Initial acute score	0.07	0.02	3.17	.002	1.07 [1.03, 1.11]	
	Change	0.47	0.34	1.38	.17	1.59 [0.82, 3.07]	
	Model B						.62 [.59, .65]
	Initial acute score	0.05	0.02	2.23	.026	1.05 [1.01, 1.09]	
	Change	0.28	0.33	0.85	.397	1.32 [0.69, 2.53]	
	Static risk	0.08	0.01	6.62	< .001	1.09 [1.06, 1.12]	
	Age	-0.02	0.005	-4.72	< .001	0.98 [0.97, 0.99]	

Note: SE = Standard Error, CI = 95% confidence interval.

[a]1,087 of 4,000 community supervision sequences ended with a revocation.

[b]427 of 4,000 community supervision sequences had a new conviction associated with it throughout the follow-up.

Protect scores. Table 15 presents results examining the incremental validity of incorporating protect change scores, in addition to the initial stable score, when predicting revocations of community supervision. Results indicated that both initial protect score and change on protect scores were significant predictors of revocations of community supervision. However, the magnitude of the effect of initial protect scores and change was small (c-index = .58). When protect scores were entered into a model alongside static risk and age, results indicated that both initial protect score and change over time incrementally predicted revocations of community supervision. Increases in protect scores at the start of supervision and increases in protect scores over time were associated with decreases in the hazard of a revocation of community supervision. The collection of predictors was associated with a c-index of .65, representing a moderate effect and an improvement over the model that examined only protect scores.

Results from survival analyses examining new convictions were consistent with the stable and acute domains. Namely, change on protect was not a significant predictor of new convictions while accounting for initial protect scores (see Table 15). The c-index for this model was small, indicating that the combined predictive effect of initial protect score and change was slightly better than chance. A model that included protect scores alongside static risk and age improved the prediction of new convictions, but the magnitude of the effect size remained small (c-index = .62).

Table 15

Cox Regression Results for Protect Scores Derived from MLM

	Predictor	*B*	*SE*	Wald	*p*	Hazard Ratio [CI]	c-index [CI]
Revocation of community supervision[a]	Model A						.58 [.56, .60]
	Initial Score	-0.10	0.01	-8.13	< .001	0.90 [0.88, 0.92]	
	Change	-2.49	0.26	-9.57	< .001	0.08 [0.05, 0.14]	
	Model B						.65 [.63, .67]
	Initial protect score	-0.09	0.01	-6.87	< .001	0.92 [0.89, 0.94]	
	Change	-2.42	0.26	-9.34	< .001	0.09 [0.05, 0.15]	
	Static risk	0.10	0.01	12.68	< .001	1.11 [1.09, 1.12]	
	Age	-0.01	0.003	-3.82	< .001	0.99 [0.98, 0.99]	
New charge resulting in a conviction[b]	Model A						.54 [.51, .57]
	Initial Protect Score	-0.06	0.02	-2.93	.003	0.94 [0.91, 0.98]	
	Change	-0.73	0.48	-1.89	.058	0.47 [0.22, 0.99]	
	Model B						.62 [.59, .65]
	Initial protect score	-0.04	0.02	-2.13	.034	0.96 [0.92, 1.00]	
	Change	-0.61	0.38	-1.60	.109	0.54 [0.26, 1.15]	
	Static risk	0.09	0.01	6.74	< .001	1.09 [1.06, 1.12]	
	Age	-0.02	0.005	-4.69	< .001	0.98 [0.97, 0.99]	

Note: SE = Standard Error, CI = 95% confidence interval.

[a]1,087 of 4,000 community supervision sequences ended with a revocation.

[b]427 of 4,000 community supervision sequences had a new conviction associated with it throughout the follow-up.

Combining the DRAOR domains. Combined Cox regression survival analyses were explored to evaluate whether initial scores and change across each domain remained relevant predictors of revocations or new convictions. As presented in Table 16, in nearly all cases, initial scores and change remained significant predictors of revocations of community supervision, after holding the effects of static risk and age constant. Initial scores on protect were no longer significantly associated with revocations of community supervision. However, the effect size associated with this model represented a moderate effect (c-index = .67), which was comparable to the models exploring each DRAOR domain independently. As a result, finding that initial scores and change on stable and acute incrementally predicted revocations from community supervision may not be practically beneficial. In other words, the overall predictive ability of the models was not considerably improved by considering the combined information across the DRAOR domains.

Examining the combined model for new convictions identified that initial scores and change across any of the DRAOR domains were not significantly related to the outcome. Age and static risk emerged as the only significant predictors of new convictions. Consistent with the models examining each DRAOR domain independently, the overall effect size associated with this model represented a small effect (c-index = .62). As a result, combining information on either initial status or change across the DRAOR domains did not improve the prediction of new convictions throughout the supervision period.

Table 16

Cox Regression Results for Combined Model Examining Stable, Acute, and Protect

	Predictor	*B*	*SE*	Wald	*p*	Hazard Ratio [CI]	c-index [CI]
Revocation of community supervision[a]	Initial stable score	0.06	0.02	3.39	< .001	1.06 [1.03, 1.10]	.67 [.65, .69]
	Stable change	1.56	0.35	4.50	< .001	4.75 [2.41, 9.36]	
	Initial acute score	0.06	0.02	3.96	< .001	1.06 [1.03, 1.10]	
	Acute change	0.96	0.30	3.18	.001	2.62 [1.45, 4.74]	
	Initial protect score	-0.03	0.01	-1.76	.078	0.97 [0.95, 1.00]	
	Protect change	-0.84	0.33	-2.54	.011	0.43 [0.23, 0.82]	
	Static risk	0.09	0.01	11.74	< .001	1.10 [1.08, 1.12]	
	Age	-0.01	0.002	-3.00	.003	0.99 [0.99, 1.00]	
New charge resulting in a conviction[b]	Initial stable score	0.02	0.03	0.73	.467	1.02 [0.97,1.08]	.62 [.60, .64]
	Stable change	0.13	0.52	0.26	.796	1.14 [0.42, 3.15]	
	Initial acute score	0.02	0.03	0.87	.383	1.02 [0.97, 1.08]	
	Acute change	-0.08	0.46	-0.16	.870	0.93 [0.38, 2.29]	
	Initial protect score	-0.03	0.02	-1.13	.257	0.97 [0.93, 1.02]	
	Protect change	-0.57	0.49	-1.17	.242	0.56 [0.22, 1.47]	
	Static risk	0.08	0.01	6.42	< .001	1.09 [1.06, 1.11]	
	Age	-0.02	0.01	-4.62	< .001	0.98 [0.97, 0.99]	

Note: SE = Standard Error, CI = 95% confidence interval.

[a]1,087 of 4,000 community supervision sequences ended with a revocation.

[b]427 of 4,000 community supervision sequences had a new conviction associated with it throughout the follow-up.

Summary of prediction results. A series of Cox regression survival analyses were examined to determine whether change on stable, acute, and protect incrementally predicted either revocations from community supervision or convictions for new charges. Results indicated that change on stable, acute, and protect was a significant predictor of revocations of community supervision, even after controlling for the effects of initial score, static risk, and age. However, this was not the case when examining new convictions, whereby results indicated that change was not a significant predictor across all DRAOR domains. As a result, hypothesis 3 was partially supported, in that there was evidence that incorporating change on stable, acute, and protect factors improved the prediction of revocations over initial scores, but this was not the case with new convictions.

Prediction models also explored whether stable, acute, or protect scores incrementally predicted community outcomes over static risk. Hypothesis 4 was evaluated by comparing effect sizes associated with models exploring initial score and change for each domain relative to models that incorporated static risk and age. Results indicated that prediction of revocations of community supervision was enhanced when initial and change scores were incorporated into a model with static risk and age. Across the models for stable, protect, and acute, the magnitude of the effect size increased from small to moderate. A combined model explored whether initial scores and change on each DRAOR domain remained significant predictors when they were modelled together. Change in stable, acute, and protect scores remained significant predictors of revocation of community supervision, indicating that each provided unique predictive

value. Taken together, these findings provide support for hypothesis 4. However, based on the effect-size for the combined model, the predictive utility was not meaningfully enhanced over models that considered each domain independently.

When predicting new convictions, models that incorporated initial and change scores alongside static risk and age demonstrated slight improvements in the overall effect, but the magnitude was still considered small across each domain. Further, although initial scores significantly predicted new convictions when controlling for the effects of static risk, change was not a significant predictor. When initial scores and change across all domains were combined into one model, only static risk and age emerged as significant predictors of new convictions. As a result, hypothesis 4 was not supported for prediction models considering new convictions.

Discussion

Relying on the results of validated risk assessments is vital to evidence-based decision making in corrections. Advancements in the approach to risk assessment has seen an emphasis placed on measuring purported dynamic risk factors alongside offender strengths, as both are expected to be particularly useful for identifying correctional treatment targets and measuring and adapting to changes in risk over time. Despite widespread adoption of these factors in assessment tools, little research has measured change across repeated assessments in a manner that increases the likelihood that observed change is in fact true change exhibited by an offender. As a result, the field's understanding of whether risk and protective factors expected to be dynamic actually demonstrate the characteristics associated with a dynamic factor remains limited. Namely, that scores can change over time, changes are linked to the probability of recidivism, and the change occurred prior to recidivism (Kraemer et al., 1997). Using a multi-wave administrative dataset, the current dissertation employed a three-step analytic approach to establish whether: (a) the DRAOR was assessing the same constructs over time, (b) there were trajectories of within-offender change across repeated assessments, and (c) those individual trajectories of change were related to the prediction of recidivism. Prediction models also considered whether incorporating change predicted recidivism above and beyond initial DRAOR scores, static risk and age. A summary of the key findings follows, along with a discussion of the implications for practice and directions for future research. It is important to reiterate that the findings from the current study are based on a sample of White and Black male offenders

supervised on community supervision in Iowa. As a result, the findings and corresponding implications should be viewed in this context, rather than generalizing to all offenders under community supervision in Iowa and elsewhere.

Summary of Findings

Psychometric properties. The psychometric properties and the original 3-factor structure of the DRAOR (stable, acute, protect) were examined for assessments completed during three time intervals, representing 0-3 months, 4 to 6 months, and 7 to 9 months after release. Based on previous findings on the DRAOR in New Zealand (Hanby, 2013; Lloyd, 2015), it was hypothesized that the DRAOR would demonstrate sufficient psychometric properties and meet the threshold for measurement invariance. Measurement invariance was supported, in that results indicated that the DRAOR was measuring the same constructs in the same way over time. This permitted ruling out that observed change on the DRAOR was the result of the differing relationships between items and constructs throughout the reassessment period. The stable and protect domains were deemed to have adequate internal consistency over time, while the estimate for the acute domain was lower. Lower internal consistency with the acute domain is consistent with previous findings on the DRAOR, both with baseline assessments (Chadwick, 2014; Hanby, 2013) and repeated assessments (Lloyd, 2015). DRAOR domains were correlated with each other in the expected way; stable and acute were positively related, and each were negatively related to protect scores. DRAOR domain scores were weakly related to static risk, and the relationship diminished slightly over time.

Trajectories of change. Given that the psychometric properties appeared to be sufficient over time, the next step was to examine whether individual's DRAOR scores changed throughout community supervision. This was first accomplished by examining unconditional growth models in a multilevel model framework, which indicated that scores on stable and acute tended to linearly decrease throughout community supervision, while protect scores increased linearly. These findings provided support for the hypothesis that DRAOR scores change within an offender throughout the course of supervision. However, it is important to note that, on average, scores did not change markedly throughout the course of supervision. For example, the average offender began supervision with a stable score of 6, which decreased to an estimated score of 4 after 3 years in the community. Similar magnitudes of change were evident for the average offender when considering acute and protect scores. Polaschek and Yesberg (2018) found slightly higher rates of change among a sample of high-risk violent male parolees in New Zealand, particularly for stable and protect scores, whereas Hanby (2013) reported similar rates of average change on stable and protect, but found higher rates of change on the acute domain. For example, results from the current study indicated that, without considering any predictors of growth, the average offender was expected to decrease their acute score by 0.09 each month on supervision, compared to an average decrease of approximately 0.30 per month among Hanby's sample. Among the current sample, it appears that there is the potential that acute scores are not changing as rapidly as what has been theorized. The trend for relatively minimal change among the DRAOR domains could signal that there are issues with measurement,

including the frequency of assessment, a lack of proficiency among officers scoring the items, or that the DRAOR would benefit from including other relevant dynamic risk or protective factors. It is important to note that there was evidence of both between- and within-individual variability around the estimates of growth for each of the domains, which prompted an examination of the relevance of various predictors to better explain change.

Initial models indicated that static risk, age, and race were not significant predictors of DRAOR domain scores at the start of supervision or the rate of change. When examining each DRAOR domain independently, the remaining domain scores were explored as both time-varying predictors and between-individual predictors of growth. Acute scores were relevant predictors of both initial stable score and change over time, such that, while controlling for the effects of protect, offenders scoring higher on acute scores were expected to start supervision with higher stable scores that also remained unchanged throughout supervision. Conversely, offenders with lower acute scores evidenced more rapid decreases in stable scores throughout community supervision. A time-varying effect for acute scores predicting growth on stable scores was also noted, whereby an offender who decreased their acute score was expected to shift to a trajectory characterized by lower stable scores with slightly more rapid decreases throughout community supervision. No significant predictors of the rate of change on acute scores emerged, but stable and protect scores explained some variation in the acute score at the start of supervision. Stable scores were positive related to the initial acute score, while protect scores were negatively related. When

examining change on protect scores, acute scores emerged as a significant between- and within-individual predictor of initial score and change. While controlling for stable score, higher between offender acute scores were associated with a lower protect score at the start of supervision that remained unchanged over time. Within-individual effects for stable and acute scores were noted, whereby an offender who increased either score was expected to shift to a trajectory associated with lower protect scores.

Although conditional models that incorporated the remaining DRAOR domains as predictors of variation in growth trajectories resulted in improvements to model fit over baseline models, there was evidence that modelling additional between- and within-individual variables would be worthwhile. Specifically, Pseudo R^2 statistics for each of the models indicated that a minority of the total variance in the DRAOR domain scores was explained by the combination of the covariates. Although this is consistent with Hanby (2013), it underscores the importance of collecting other offender level data that may help explain variation in DRAOR scores over time (e.g., treatment referral and completion, quality and frequency of supervision contacts).

Predicting recidivism. Consistent with the two-stage model developed by Yang and colleagues (2017), individual growth trajectories were constructed for each offender based on the results from the unconditional growth models for each DRAOR domain. Cox regression survival analysis was then conducted to determine whether model estimated initial scores and change on stable, acute, and protect were significantly related to either revocations of community supervision or a new conviction. It was hypothesized that incorporating change on stable, acute, and protect would

incrementally add to the prediction of either outcome over initial scores. There was partial support for this hypothesis, as results indicated that change on stable, acute, and protect was significantly related to revocations of community supervision, but not new convictions.

The prediction of revocations of community supervision was enhanced when initial scores and change on each DRAOR domain were examined alongside static risk and age. This indicated that dynamic risk, including both initial score and change over time, incrementally added to the prediction of revocations of community supervision. The combined effect sizes for each of the stable, acute, and protect models were in the moderate range. Given the lack of a relationship between DRAOR domains and the likelihood of a conviction for a new charge, it was unsurprising that the prediction of new convictions was not enhanced when DRAOR scores were considered alongside static risk and age.

Finally, a combined model that incorporated initial scores and change across each DRAOR domain, along with static risk and age, was examined to determine whether each explained unique variance in the community outcomes. Change on stable, acute, and protect remained significant predictors of revocation of community supervision. However, the overall effect size indicated that the predictive validity of the collection of covariates (c-index = .67) was comparable to any of the models examining each DRAOR domain alongside static risk and age (c-index ranged from .65 to .67). As a result, the predictive ability was not meaningfully enhanced when considering scores on each DRAOR domain simultaneously.

Implications for Theory and Practice, Limitations, and Future Directions

Theoretical implications for dynamic risk and protective factors. The framework used to test for the presence of change on purported dynamic risk and protective factors resulted in several implications for underlying theory. Firstly, results from the examination of measurement invariance over time reinforced that the DRAOR's conceptual distinction between dynamic risk factors and protective factors continued to fit the data well, and it was consistent over time. With protective factors consistently loading onto one construct, there is support that protective factors may not simply be the inverse of risk factors. If this were the case, the items making up the protective domain would likely load onto the other risk domains instead, alongside the risk items included in the DRAOR. It is also important to highlight that the relationship between each of the DRAOR items and domains remained consistent across White and Black men offenders. This provides preliminary support that the same dynamic risk and protective items are measuring the same constructs across both subgroups of offenders. However, this does not indicate whether some of the items are more salient risk or protective factors in terms of either predicting future reoffending or for informing case management strategies. Future research should more thoroughly investigate whether individual risk and protective factors appear to be differentially relevant for the various subgroups of offenders under community supervision (e.g., women, ethnocultural offenders, Indigenous offenders).

Results from the two-stage model examining the presence of changes in DRAOR scores over time and the relationship between change and future reoffending provide

the most relevant evidence for the concepts of dynamic risk and protective factors. As previously discussed, much of the past research investigating dynamic risk and protective factors has utilized methodologies that are unable to rigorously isolate true change from potential measurement error. As a result, knowledge of whether dynamic risk and protective factors actually change over time, and possibly operate as causal dynamic factors as defined by Kraemer and colleagues (1997), remains limited. Results from the current study add to the emerging body of literature that has utilized rigorous statistical methods to first establish that the tool measures the same constructs over time, that within-individual change is occurring, and that change incrementally predicts recidivism above baseline scores. Given that current results supported these criteria, it appears that the risk and protective factors measured by the DRAOR are truly dynamic. As a result, it is likely that the items that make up the DRAOR can at least partially explain future criminal behaviour. It is possible that the relationship between change on these factors and recidivism is being influenced by unobserved underlying causal mechanisms, but the results demonstrate that the DRAOR items can at least be viewed as proxies for explanations of recidivism.

In addition to providing theoretical support for dynamic risk and protective factors, results from prediction models for the protect domain advance the conceptualization of protective factors. Results from these models indicated that both initial protect score and change were significantly related to revocations of community supervision. As a result, the current findings provide support that the relationship between the protect domain and revocations of supervision is promotive (Jones et al.,

2015). Future studies should further investigate whether protective factors may buffer risk. It should also be noted that these findings apply only to the domain score, as item-level analyses were not conducted. Additionally, results from the prediction model that incorporated initial and change scores for stable, acute, and protect found that the effect size for the overall model was not meaningfully enhanced. This calls into question whether each of the domains are contributing unique information when considered together. Future research exploring these constructs and their relative contribution in predicting recidivism is needed.

Lastly, although scores across each DRAOR domain changed over time, results from the growth curve analyses failed to support the conceptual distinction between stable and acute risk factors. Scores on the stable and acute domains changed at a comparable rate throughout the course of supervision, on average. These findings are inconsistent with previous research on the DRAOR that used a similar methodology to assess change (Hanby, 2013) and when raw change between assessments was examined (Davies, 2019; Lloyd, 2015). Previous results from both approaches indicated that acute scores fluctuated to a greater extent than scores on stable and protect. Further, Davies (2019) found that change on the acute score incrementally predicted imminent recidivism over the most proximal DRAOR assessment. It is important to highlight that some key methodological differences exist between the previous examinations of the DRAOR and the current study, most relevant being the scheduling of reassessment. In the previous approaches, assessments were completed on an almost weekly basis, whereas the schedule of assessments was irregular for the current sample, often

resulting in an average of 2 months between assessments. As a result, the lack of support for the conceptual distinction between stable and acute risk factors should be viewed cautiously until future research can examine this further. Additionally, these findings pertain to the domain level of stable and acute risk, so future research is needed to determine whether individual items are performing in a manner consistent with their conceptualization.

Implications for practice. Results from the examination of trajectories of change over the course of supervision have important implications for decisions related to resource allocation. Although, on average, offenders demonstrated decreases in their stable and acute scores and increases in their protect scores throughout community supervision, more nuanced patterns emerged when the remaining DRAOR domain scores were used to predict variation in change on a given domain. For example, an offender with high acute scores and average protect scores would be expected to start community supervision with higher stable scores and maintain that score over time, while decreases in stable are expected at all other values of acute and protect. This underscores the importance of considering the combination of DRAOR domain scores when estimating the expected direction of change among offenders on a caseload. Officers may note that two offenders have the same stable score, but if one scores above average on the acute domain, that should direct them to prioritize the offender with higher acute scores for additional intervention and monitoring activities. Although from a prediction standpoint, the results did not support that the combination of DRAOR domains meaningfully improved the prediction of recidivism, it is clear that

knowledge of each domain score provided a more accurate assessment of the typical change trajectory for a given offender.

Timing of reassessments. Although the current dissertation did not examine the timing of the assessment in relation to recidivism (e.g., examining if more proximal assessments would be better predictors than distal assessments; see Davies, 2019; Lloyd, 2015; Lloyd et al., in press), results from prediction models indicated that incorporating model estimated trajectories of change for each individual improved the prediction of revocations of community supervision above the baseline DRAOR assessment. Further, when the score at the start of supervision and change over time on each DRAOR domain was coupled with static risk and age, the prediction of revocations of community supervision was enhanced. This highlights that greater differentiation between future recidivists vs. non-recidivists can be achieved by examining repeated assessments throughout the supervision period. As a result, initial level of supervision should continue to consider both static risk and baseline DRAOR scores, but as results from repeated assessments become available, supervision requirements and referrals to treatment should consider how scores have changed over time. Given that the initial score at the start of supervision and change over time for each DRAOR domain remained significant predictors of revocations of community supervision when considered together, both provide useful information that should be relied upon to allocate treatment resources and make modifications to supervision requirements (e.g., frequency of contact, modifications of supervision conditions).

However, results indicated that the predictive accuracy was not meaningfully enhanced when models considered initial score and change over time on each DRAOR domain simultaneously, in conjunction with static risk and age. This was inconsistent with previous research on the DRAOR, as Lloyd (2015) found that incorporating baseline and change scores on stable, acute, and protect, along with static risk, significantly improved the prediction of recidivism. As discussed in greater detail below, it is possible that a variety of factors may have limited the predictive accuracy of the DRAOR, including the quality of the implementation of the DRAOR and various methodological requirements associated with the current study (e.g., the potential impact of requiring 3 completed assessments on the occurrence of revocations of supervision).

Although the results pertaining to revocations of community supervision highlight that case management decisions would be enhanced by considering each individual's trajectory of change, implications to practice would be further enhanced if real-time changes in risk were tied to imminent recidivism. This analysis would also inform the optimal reassessment schedule, which would help ensure that supervising officers are not overburdened with conducting reassessments until there is a need from a prediction standpoint. This could be accomplished by identifying the point at which the predictive accuracy of an assessment is either meaningfully diminished or superseded by a more recent assessment. Future research should explore whether alternative models of incorporating reassessment data to predict recidivism, such as cox regression with time-dependent covariates or discrete-time hazard models, provide support for the practical utility of reassessing the DRAOR throughout community

supervision. For example, Lloyd (2015) found that taking scores from the most proximal DRAOR assessment resulted in the greatest predictive accuracy of short-term recidivism, compared to averaging across all assessments, or taking an average of the most recent 3 weekly assessments. These results would translate more directly to concrete case management decisions immediately following a given assessment than compared to determining that a given offender is demonstrating a trajectory of change linked to an increased likelihood of recidivism throughout supervision.

Impact of selection criteria on predictive accuracy. Despite the promising findings in relation to predicting revocations of supervision, results did not indicate that the prediction of new convictions was enhanced when change on DRAOR scores over time was incorporated. There are several potential explanations for why this may be the case. First, the follow-up period only considered new convictions that were incurred during the supervision period. Although this amounted to an average follow-up period of nearly 2 years, it neglected to consider new convictions incurred after an offender was discharged from their sentence. In Iowa each year, approximately 35% of high-risk offenders are returned to prison with a new conviction within 3 years of the end of their sentence (Iowa Department of Corrections, 2018). In addition to only considering convictions and revocations during the community supervision period, the observed base rate for recidivism was lower than expected, given that the sample was moderate risk and higher. The validation of the Iowa Risk Assessment found that rates of any new crime within 3 years of supervision start ranged from 35% to 69% for the risk levels that were represented in the current sample (Prell, 2013). The underestimate of recidivism

observed in the current study may be a function of the more restrictive definition that an offender had to be revoked from community supervision or incur a new conviction. Future research should consider expanding the definition of recidivism to include both violations of supervision conditions and new arrests. Considering technical violations and new arrests would also more readily facilitate an examination of the association between real-time changes in dynamic risk and protective scores and current behaviour. Results from this research would help shed light on whether the predictive accuracy of capturing changes in dynamic risk and protective factors is meaningfully enhanced when recidivism is defined more broadly.

Relatedly, since offenders were required to have 3 DRAOR assessments to be included in the analyses, any offender who was either revoked from community supervision or incurred a new conviction in a short period of time would have been removed from the sample. Although it was plausible that an offender could have had 3 assessments completed within a month (e.g., if they were seen weekly and the officer completed an assessment at each visit), the third assessment was completed nearly 6 months after supervision began, on average. Iowa's policy on DRAOR reassessments states that the acute domain should be assessed at least monthly, whereas stable and protect can be assessed as infrequently as quarterly. Given that the first 6 months represent a particularly high-risk period for offenders, marked by high rates of failure (Brown et al., 2009), it is likely that the requirement for 3 assessments contributed to underestimating the rate of recidivism among the study sample. This may have attenuated the predictive accuracy of DRAOR scores by biasing the sample toward a

group of offenders who, despite demonstrating moderate to high risk, successfully remained on supervision through the most challenging period.

It is also important to note that the timing of the first assessment took place anywhere within the first 90 days of supervision. Although on average, the first assessment was completed one month into supervision, the optimal timing of the baseline assessment is to have it completed as close to the start of supervision as is feasible. It is possible that having a baseline assessment that reflects the very first experiences on supervision would provide a more accurate starting point from which to assess change. Given that some baseline assessments were completed months after beginning supervision, it is plausible that intra-individual change had already occurred, which would have attenuated the true estimate of change over time. Future research would benefit from accounting for the time between the start of supervision and the first assessment, to determine if those with an assessment closer to the start of supervision evidence greater change throughout the course of supervision, and whether the relationship between change and recidivism is stronger. Such findings would provide empirical guidance to establish shorter timeframes in policy for the completion of the baseline DRAOR assessment.

Importance of implementation fidelity. Predictive accuracy may have also been affected by the way in which officers scored the DRAOR and used the results to inform supervision related decisions (e.g., treatment referrals). Effective use of any risk tool depends on the quality of initial training and implementation efforts (Flores, Lowenkamp, Holsinger, & Latessa, 2006; Vincent, Guy, Perrault, & Gershenson, 2016;

Vincent, Paiva-Salisbury, Cook, Guy, & Perrault, 2012). Previous research has demonstrated that the type of training probation officers receive can be related to the predictive accuracy of that tool. Flores and colleagues (2006) demonstrated that the predictive accuracy of the LSI-R was related to implementation integrity, defined as having received formal training and the agency having worked with the tool for three years or more. Scores from assessments completed by officers who received formal training were significantly related to reincarceration (r (1,633) = .21, 95% CI [.16, .26]), whereas scores from assessments completed by untrained officers were not significantly related to reincarceration (r (393) = .08, 95% CI [-.02, .18]). Results also indicated that the strength of the relationship between assessment scores and reincarceration was stronger among agencies that had implemented the tool for three years or more.

Similar results were obtained in an initial validation study of the DRAOR in Iowa, whereby stable, acute, and protect scores from formally trained officers significantly predicted general recidivism (AUC = .68 [.60, .76], .65 [.56, .73], and .62 [.53, .70], respectively, while domain scores from informally trained were not related to general recidivism (AUC = .55 [.48, .63], .52 [.45, .60], and .52 [.45, .60] for stable, acute, and protect; Chadwick, 2014). Further, Vincent and colleagues (2016) demonstrated that among sites where juvenile probation officers evidenced weaker adherence to administrative policies on risk assessment, used as a proxy for implementation fidelity, overall risk level was unrelated to supervision level and the number of service referrals. Conversely, among sites with high implementation fidelity, risk assessment results were related to supervision level, service referrals, and less frequent detention placement.

The fidelity of the implementation of the DRAOR in Iowa is unknown and was unable to be assessed in the current dissertation. Although robust quality assurance standards have been established in Iowa, the level of proficiency with the DRAOR and engagement with quality assurance activities was unknown for the officers who contributed scores on DRAOR assessments. It should be emphasized that at least two DRAOR assessments are reviewed annually by quality assurance specialists to ensure that the officer is appropriately scoring the DRAOR, with sufficient justification. If the reviews identify a scoring inconsistency, defined as a two-point deviation on any given item between the auditor and the officer, officers are provided with feedback, coaching, and a skill development plan. Although these robust quality assurance activities appear to be in place, the level of engagement in coaching and skill development and the rates of compliance during the study period were unable to be considered. It is expected that officers who engage in refresher training sessions on the DRAOR or continue to demonstrate high levels of proficiency on the DRAOR would have a greater understanding of each of the DRAOR items and would be more likely to score the DRAOR diligently. As a result, DRAOR assessments completed by officers with higher levels of competency would be expected to demonstrate a stronger relationship with recidivism (Flores et al., 2006). That being said, it remains unknown whether implementation fidelity is an area that could be improved or represents an issue that may be contributing to lower levels of predictive accuracy. Future research should explore potential implementation issues such as engagement in training and continuous

learning, as well as adherence to administrative policies, to determine whether these are potentially affecting the predictive accuracy of the tool.

Relevance of interrater reliability. Inter-rater reliability has yet to be assessed on the DRAOR, so it is unclear whether officers are scoring the tool in a consistent manner. It is important to highlight that in a recent validation of a prison-based dynamic risk and protective factor assessment developed based on the DRAOR (the Structured Dynamic Assessment Case-Management-21 (SDAC-21; Serin & Wilson, 2012)), inter-rater reliability was found to be sufficient (ICCs ranging from .54 to .96) among a small sample of audiotaped interviews (n = 20) that were also scored by a trainer of the tool (Smeth, 2019). Although there are important differences between the tools (e.g., inclusion of responsivity factors in the SDAC-21) and the settings in which they are administered (institution versus community), there is considerable overlap between the items included in the stable and protect domains on both tools. As a result, finding support for inter-rater reliability for the SDAC-21 at least highlights that it is possible to achieve consistent scoring among dynamic items that are also largely present in the DRAOR.

In general, inter-rater reliability has rarely been investigated for risk assessments. In a recent systematic review of validation studies conducted on risk assessments in the U.S., only 2 studies out of 72 samples examined had reported inter-rater reliability statistics (Desmarais, Johnson, & Singh, 2016). Although the 2 studies indicated excellent agreement between independent raters on the measures of dynamic risk (Lowenkamp, Holsinger, Brusman-Lovins, & Latessa, 2004; Walters, 2011), others

have found poor levels of inter-rater reliability, particularly for domains that consist of more subjective items or represent complex psychological constructs (e.g., leisure/recreation, family/marital, attitudes/orientation; Austin, Coleman, Peyton, Johnson, 2003; Rocque & Plummer-Beale, 2016). A recent examination of inter-rater reliability among university students (*n* = 4) trained in the Level of Service/Case Management Inventory scoring 9 adult probationers largely replicated previous findings, in that most domains had good levels of rater consistency, but lower levels for domains assessing procriminal attitudes/orientation and companions (Labrecque et al., 2018). It is important to highlight that many of the previous studies focus on assessments completed for incarcerated offenders, where circumstances do not change as frequently as they might in the community (e.g., changes in employment).

Additionally, some studies reporting on inter-rater reliability of assessments lack ecological validity, in that they are based on smaller scale research projects rather than "field studies" (Edens & Boccaccini, 2017, p. 600), where assessment data are collected and used to inform decision-making. Although psychometric properties are often assumed to generalize from research studies to field studies, the reduced internal validity associated with field studies and the increased stakes associated with the results of the risk assessment combine to make this assumption unlikely. This underscores the need to prioritize research that examines inter-rater reliability among community supervision officers in a field study, despite the operational challenges that may be present (i.e., it is resource intensive and inefficient to have raters independently assess the same offender). Inconsistent scoring across raters not only poses challenges for

ensuring that treatment resources and supervision strategies are appropriately and consistently allocated to the offenders who need them most, but low inter-rater reliability can also lower the predictive accuracy of a risk assessment (Duwe & Rocque, 2017; Edens & Boccaccini, 2017). Future research should explore inter-rater reliability of the DRAOR, ideally in a field study, to determine whether enhancements to training may be necessary.

Translating risk assessment results to practice. One of the expected benefits of reassessing dynamic risk and protective factors is that officers are provided with information regarding what is working well for a given offender and what could be triaged to promote change (Serin et al., 2016). The strategies to respond to the observed risks and strengths are likely to be developed in the form of case management plans or release plans. Although some research has demonstrated that having a high-quality release plan is related to higher rates of success in the community (Dickson & Polaschek, 2014; Richards, 2016), there is often a disconnect between conducting risk assessments and incorporating the results to inform case management decisions (e.g., Bonta, Rugge, Scott, Bourgon, & Yessine, 2008; Miller & Maloney, 2013; Viglione, 2019; Viglione, Rudes, & Taxman, 2015). Further, limited attention has been devoted to whether officers are more likely to identify relevant treatment targets and develop high quality case plans following the completion of dynamic risk assessments. The current study was also limited in this regard, as the officer's behaviour following the completion of the DRAOR was unable to be examined.

One recent study of juvenile probation officers tested whether the adoption of both a dynamic risk measure and a structured case plan template improved the overall quality of case plans and corresponding delivery of relevant interventions (Viljoen, Shaffer, Muir, Cochrane, & Brodersen, 2019). Results indicated that officers using a dynamic risk and protective factor measure alongside the structured case plan template were more likely to identify key risk and protective factors and develop an array of strategies to intervene rather than simply monitor. Further, juvenile offenders supervised by officers using the structured case plan template were more likely to have their needs addressed (Viljoen et al., 2019). These results underscore the importance of translating risk assessment results to concrete supervision related decisions, but also highlight that risk assessments may need to be supplemented with a case plan template that encourages officers to directly link the results of the assessment to actions. Although the available data for the current study did not permit examining the extent that officers integrated the DRAOR results into case management decisions, future research should examine whether repeated assessments are associated with higher quality case plans that tailor interventions and actions to the salient risk and protective factors at that occasion.

Multilevel predictors of change. Although trajectory change modelling considered that assessments were nested within offenders, it was not possible to separate out potentially differential effects of district of supervision or supervising officers. Iowa DOC is separated into eight judicial districts, which could lead to regional disparity in terms of supervision practices and responses to breaches of supervision

conditions. For example, examining the 3-year recidivism rates of offenders admitted to probation in fiscal year 2014 across the districts indicated that rates ranged from as low as 6% to as high as 20% (Iowa Department of Corrections, 2019). Given the potential for variation across districts, it is important that future research accounts for any effects associated with the supervising district when examining trajectories of change throughout the supervision period.

Further, accounting for characteristics of the supervising officer may enhance the understanding of variations in changes on the DRAOR domains over time. Babchishin (2013) modelled the effects of the conscientiousness of supervising officers to explain variations in dynamic risk for sex offenders. Offenders supervised by conscientious officers tended to have smaller decreases in dynamic risk over time, relative to offenders supervised by less conscientious officers. Although officer conscientiousness may represent one important component to understanding variability in rates of change, future research should consider other factors related to effective supervision, such as the quality of the working relationship between the supervising officer and the offender (Kennealy, Skeem, Eno Louden, & Manchak, 2012) and officer use of core correctional practices (e.g., Bonta et al., 2011; Robinson et al., 2012; Smith, Schweitzer, Labrecque, & Latessa, 2012).

The available offender level data also limited the exploration of variability in the rates of change in DRAOR scores. Despite the presence of significant variability in both the initial status and change over time for each DRAOR domain, identifying relevant predictors of variation was restricted to static risk, age, race, and remaining DRAOR

domain scores. Given that static risk, age, and race did not appear to be relevant predictors of initial status and change, it appears to be worthwhile to consider other potential predictors. Polaschek and Yesberg (2018) investigated whether completers of an intensive prison-based treatment program predicted initial DRAOR scores at the start of community supervision and change over time. Treatment completers had lower stable and acute scores and higher protect scores at the start of supervision relative to a comparison group who did not complete the intensive program, but the rates of change over time did not differ between the groups. However, the effect of treatment status disappeared when considering the proportion of sentence served prior to release on parole. Those released earlier in their sentence had lower risk and higher protect scores than those released closer to the end of their prison sentence. The effects of these covariates were not considered alongside static risk, so it is unclear whether they represent uniquely relevant predictors, but future research would benefit from further investigating the role of previous treatment in predicting change in DRAOR scores. Additionally, knowledge of exposure to correctional treatment throughout the course of supervision will provide the opportunity to explore potential explanations for the observed changes in risk and contribute to further understanding the process of offender change.

Conclusion

Consistent with research on the DRAOR in New Zealand and pilot examinations in Iowa, this study provided support for the DRAOR as a measure of dynamic risk and strength factors (i.e., variables that are associated with reductions in the likelihood of

recidivism). Results from a series of rigorous statistical methods indicated that offenders' DRAOR scores of stable, acute, and protect do change throughout the course of supervision, although the average magnitude of change was small. Variation in the patterns of change across each domain were noted, which were partly explained by considering scores on the remaining DRAOR domains. Lastly, prediction models suggested that both initial scores and change across each DRAOR domain were significantly related to revocations of community supervision, even after controlling for static risk and age. However, DRAOR scores were not related to new convictions during the community supervision period and the combination of DRAOR domain scores did not improve the prediction of either outcome above domain specific analyses.

Given that community sentences in Iowa are more likely to end with a revocation rather than a new conviction, the DRAOR remains a useful tool for monitoring changes in risk over time and prioritizing case management resources (e.g., interventions) for those who demonstrate trajectories of change associated with elevated rates of community supervision failure. Further research is needed to better understand how the DRAOR is being used in practice, whether issues with fidelity of implementation may be affecting DRAOR scores and ultimately the predictive utility of the tool, and whether the reassessment policy in Iowa should be refined to include more assessments during the highest risk phase of community supervision. For the time being, the results support the use of the DRAOR in community supervision practice and provide further support for the theory on dynamic risk and protective domains.

References

Administrative Office of the U.S. Courts (AOUSC) (2011). An Overview of the federal Post Conviction Risk Assessment. Washington, D.C.: Administrative Office of the U.S. Courts. Retrieved from https://www.uscourts.gov/services-forms/probation-and-pretrial-services/supervision/post-conviction-risk-assessment

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

Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 397-438. doi: 10.1080/10705510903008204

Austin, J., Coleman, D., Peyton, J., & Johnson, K.D. (2003). *Reliability and validity study of the LSI-R risk assessment instrument*. Washington, D.C.: Institute on Crime, Justice, and Corrections at The George Washington University

Babchishin, K. M. (2013). *Sex offenders do change on risk-relevant propensities: Evidence from a longitudinal study of the ACUTE-2007* (Doctoral Dissertation). Carleton University, Ottawa, Canada.

Baird, C. (2009). A question of evidence: A critique of risk assessment models used in the justice system. Oakland, CA: National Council on Crime and Delinquency. Retrieved from http://www.nccdglobal.org.

Bergeron, C. L., & Miller, H. A. (2013). Tracking change through treatment with the Inventory of Offender Risk, Needs, and Strengths. *Psychological Assessment, 25*, 979-990. doi: 10.1037/a0033190

Blokland, A. A. J., Nagin, D., & Nieuwbeerta, P. (2005). Life span trajectories of a Dutch conviction cohort. *Criminology, 43*, 919-954. doi: 10.1111/j.1745-9125.2005.00029.x

Bonta, J. & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed). New York, New York: Routledge.

Bonta, J., Bourgon, G., Rugge, T., Scott, T. L., Yessine, A. K., Gutierrez, L., & Li, J. (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior, 38*, 1127-1148. doi: 10.1177/00-93854811420678

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

Borum, R., Bartel, P., & Forth, A. (2006). *Structured Assessment of Violence Risk in Youth (SAVRY).* Lutz, FL: Psychological Assessment Resources.

Brown, T., A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

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*, 25-45. doi: 10.1007/s10979-008-9139-7

Chadwick, N. (2014). *Validating the Dynamic Risk Assessment for Offender Re-entry (DRAOR) in a sample of US probationers and parolees* (Master's thesis). Carleton University, Ottawa, Canada.

Chadwick, N., Serin, R. C., & Lloyd, C. (2015, June). *Measuring offender change and the impact on predicting future outcome.* Symposium presented at the 3rd North American Correctional and Criminal Justice Psychology Conference, Ottawa, ON.

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. doi:10.1080/10705510701301834

Cheung G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling,9*, 233-255. doi: 10.1207/S15328007SEM0902_5

Cohen, T. H., Lowenkamp, C. T., & VanBenchoten, S. W. (2016). Does change in risk matter? Examining whether changes in offender risk characteristics influence recidivism outcomes. *Criminology & Public Policy, 15*, 263-296. doi: 10.1111/1745-9133.12190

Cohen, T. H., & VanBenschoten, S. W. (2014). Does the risk of recidivism for supervised offenders improve over time? Examining changes in the dynamic risk characteristics for offenders under federal supervision. *Federal Probation, 78*(2), 41-56. doi: 10.2139/ssrn.2463376

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583-619. doi: 10.1146/annurev.psych.093008.100356

Davies, S. T. (2019). *An investigation of how change in dynamic risk and protective factors affects the prediction of imminent criminal recidivism* (Doctoral Dissertation). Victoria University of Wellington, Wellington, New Zealand.

de Vogel, V., de Ruiter, C., Bouman, Y., & de Vries Robbé, M. (2009). SAPROF: Guidelines for the assessment of protective factors for violence risk. Utrecht, The Netherlands: Forum Educatief.

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

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

de Vries Robbé, M., de Vogel, V., Douglas, K. S., & Nijman, H. L. I. (2015). Changes in dynamic risk and protective factors for violence during inpatient forensic psychiatric treatment: Predicting reductions in postdischarge community recidivism. *Law and Human Behavior, 39*, 53-61. doi: 10.1037/lhb0000089

Desmarais, S. L., Johnson, K. L., & Singh, J. P. (2016). Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychological Services, 13*, 206-222. doi: 10.1037/ser0000075

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic and protective factors to predict inpatient aggression: Reliability and validity of START assessments. *Psychological Assessment, 24*, 685-700. doi: 10.1037/a0026668

Dickson, S. R., & Polaschek, D. L. L. (2014). Planning to avoid risk or planning for a positive life: The relationship between release plan valence and reoffending. *International Journal of Offender Therapy and Comparative Criminology, 58*, 1431-1448. doi: 10.1177/0306624X13502631

Douglas, S. K., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*(3), 347-383. doi: 10.1037/1076-8971.11.3.347

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105(3), 399–412. doi: 10.1111/bjop.12046

Duwe, G., & Rocque, M. (2017). Effects of automating recidivism risk assessment on reliability, predictive validity, and return on investment (ROI). *Criminology & Public Policy, 16*, 235-269. 10.1111/1745-9133.12270

Edens, J. F., & Boccaccini, M. T. (2017). Taking forensic mental health assessment “out of the lab” and into “the real world”: Introduction to the special issue on the field utility of forensic assessment instruments and procedures. *Psychological Assessment, 29*, 599-610. doi: 10.1037/pas0000475

Farrington, D.P. (1995). The development of offending and antisocial behaviour from childhood: Key findings from the Cambridge study in delinquent development. *Journal of Child Psychology and Psychiatry, 36*, 929-964. doi: 10.1111/j.1469_7610.1995.tb01342.x

Farrington, D. P. (2003). Developmental and life-course criminology: Key theoretical and empirical issues – The 2002 Sutherland award address. *Criminology*, *41*, 221-255. doi:10.1111/j.1745-9125.2003.tb00987.x

Farrington, D. P. (2007). Advancing knowledge about desistance. *Journal of Contemporary Criminal Justice, 23,* 125-134. doi: 10.1177/1043986206298954

Farrington, D. P., Ttofi, M. M., & Piquero, A. R. (2016). Risk, promotive, and protective factors in youth offending: Results from the Cambridge study in delinquent development. *Journal of Criminal Justice, 45*, 63-70. doi: 10.1016/j.jcrimjus.2016.02.014

Flora, D. B., & Curran, P. J. (2004). An examination of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466-491. doi: 10.1037/1082-989X.9.4.466

Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the level of service inventory-revised: The importance of implementation integrity. *Journal of Criminal Justice, 34*, 523-529. 10.1016/j.jcrimjus.2006.09.007

Fougere, A., & Daffern, M. (2011). Resilience in young offenders. *International Journal of Forensic Mental Health, 10*, 244-253. doi: 10.1080/14999013.2011.598602

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*, 575-607. doi: 10.1111/j.1745-9125.1996.tb01220.x

Giordano, P. C., Cernkovich, S. A., & Holland, D. D. (2003). Changes in friendship relations over the life course: Implications for desistance from crime. *Criminology, 41*, 293-327. doi: 10.1111/j.1745-9125.2003.tb00989.x

Giordano, P. C., Cernkovich, S. A., & Rudolph, J. L. (2002). Gender, crime, and desistance: Toward a theory of cognitive transformation. *American Journal of Sociology, 107*, 990-1064. doi: 10.1086/343191

HaggÅrd, U., Gumpert, C. H., & Grann, M. (2001). Against all odds: A qualitative follow-up of high-risk violent offenders who were not reconvicted. *Journal of Interpersonal Violence, 16,* 1048-1065. doi. 10.1177/088626001016010005

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

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

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

Canada. Retrieved from http://www.publicsafety.gc.ca/res/cor/rep/_fl/crp2007-05-en.pdf

Harrell, F. E. (2015). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis* (2nd Ed.). doi: 10.1007/978-3-319-19425-7

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA: Journal of the American Medical Association*, *247*, 2543-2546. doi:10.1001/jama.1982.03320430047030

Harris, G. T., & Rice, M. E. (2015). Progress in violence risk assessment and communication: Hypothesis versus evidence. *Behavioral Sciences and the Law, 33*, 128-145. doi: 10.1002/bsl.2157

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis. A tutorial on parallel analysis. *Organizational Research Methods, 7*, 919-204. doi: 10.1177/109442810463675

Heagerty, P. J., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, *61*, 92-105.

Healy, D. (2010). Betwixt and Between: The role of psychosocial factors in the early stages of desistance. *Journal of Research in Crime and Delinquency, 47*, 419-438. doi: 10.1177/0022427-810375574

Helmus, M. L., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*, 8-25. doi: 10.1177/0093854816678898

Hoffman, L., & Stawski, R. S. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development, 6*(2-3), 97-120. doi: 10.1080/15427600902911189

Holgado-Tello, F., Chacón–Moscoso, S., Barbero–García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quantitative methods, 44*, 153-166. doi: 10.1007/s11135-008-9190-y

Horney, J., Osgood, D. W., & Marshall, I. H. (1995). Criminal careers in the short-term: Intra-individual variability in crime and its relation to local life circumstances. *American Sociological Review, 60*, 655-673.

Howard, A. L. (2015). Leveraging time-varying covariates to test within- and between-person effects and interactions in the multilevel linear model. *Emerging Adulthood, 3*, 400-412. doi: 10.1177/2167696815592726

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi: 10.1080/10705519909540118

Hussong, A. M., Curran, P. J., Moffitt, T. E., Caspi, A., & Carrig, M. M. (2004). Substance abuse hinders desistance in young adults' antisocial behavior. *Development and Psychopathology, 16*, 1029-1046. doi:10.1017/S095457940404012X

Iowa Code 2020, Title XVI – Criminal Law and Procedure, § 900 – 916 (2020).

Iowa Department of Corrections. (2014). *Quarterly Quick facts – July 2014.* Retrieved from https://doc.iowa.gov/data/quick-facts

Iowa Department of Corrections. (2015). *Quarterly Quick facts – June 2015.* Retrieved from https://doc.iowa.gov/data/quick-facts

Iowa Department of Corrections. (2016). *Quarterly Quick facts – June 2016.* Retrieved from https://doc.iowa.gov/data/quick-facts

Iowa Department of Corrections. (2018). *FY 2018 corrections annual report.* Retrieved from https://doc.iowa.gov/document/fy-2018-corrections-annual-report

Iowa Department of Corrections. (2019). *3-year recidivism for offenders admitted to probation in Iowa* [Data file]. Retrieved from https://data.iowa.gov/Correctional-System/3-Year-Recidivism-for-Offenders-Admitted-to-Probat/e9zy-uibf

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19. doi:10.1037/0022---006X.59.1.12

Jones, N. J., Brown, S. L., Robinson, D., & Frey, D. (2015). Incorporating strengths into quantitative assessments of criminal risk for adult offenders: The Service Planning Instrument. *Criminal Justice and Behavior, 42,* 321-338. doi: 10.1177/009385-4814547041

Kaeble, D., & Cowhig, M. (2018). *Correctional populations in the United States, 2016*. Bureau of Justice Statistics, U.S. Department of Justice. Retrieved from https://www.bjs.gov/content/pub/pdf/cpus16.pdf

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, *54*, 337-343. doi: 10.1001/archpsyc.1997.01830160065009

Kroner, D. G., & Riordan, M. J. (2019). Cronbach's coefficient alpha: Comments – part I (Research Brief). *Crime Scene, 26*(1), 12-15. Retrieved from https://cpa.ca/sections/criminaljusticepsychology/criminaljusticepublications/

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. doi: 10.1037/l-hb0000022

Labrecque, R. M., Campbell, C. M., Elliott, J., King, M., Christmann, M., Page, K., McVay, J., & Roller, K. (2018). An examination of the inter-rater reliability and rater accuracy of the level of service/case management inventory. *Corrections: Policy, Practice and Research, 3*(2), 105-118. doi: 10.1080/23774657.2017.1323253

Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation, 53*, 116-128. doi: 10.1080/10509674.2013.868389

LaBrish, C. S. (2011). Advantages of using Polychoric correlations for item-level exploratory factor analysis. *ProQuest Dissertations and Theses.* UMI: MR80604

Laub, J. H., & Sampson, R. J. (2001). Understanding desistance from crime. *Crime and Justice, 28*, 1-69.

LeBel, T.P., Burnett, R., Maruna, S., & Bushway, S. (2008). The “Chicken and Egg” of Subjective and Social Factors in Desistance From Crime. *European Journal of Criminology, 5,* 131–159. doi:10.1177/1477370807087640

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*, 150-164. doi: 10.1177/1073191112441242

Lloyd, C. D. (2015). *Can a dynamic risk instrument make short-term predictions in “real time”? Developing a framework for testing proximal assessment of offender recidivism risk during re-entry* (Doctoral Dissertation). Carleton University, Ottawa, Canada.

Lloyd, C. D., Hanson, R. K., Richards, D. K., & Serin, R. C. (in press). Reassessment improves prediction of criminal recidivism: A prospective study of 3,421 individuals in New Zealand. *Psychological Assessment*.

Lloyd, C. D., & Serin, R. C. (2012). Agency and outcome expectancies for crime desistance: Measuring offenders’ personal beliefs about change. *Psychology, Crime, and Law, 18*, 543-565. doi: 10.1080/1068316X.2010.511221

Lodewijks, H. P. B., de Ruiter, C., & Doreleijers, T. A. H. (2010). The impact of protective factors in desistance from violent reoffending: A study in three samples of adolescent offenders. *Journal of Interpersonal Violence, 25*, 568-587. doi: 10.1177/0886260-509334403

Lösel, F. & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventative Medicine, 43*(2S1), S8-S23. doi: 10.1016/j.amepre.2012.04.029

Lowenkamp, C. T., Holsinger, A. M., Brusman-Lovins, L., & Latessa, E. J. (2004). Assessing the inter-rater agreement of the level of service inventory revised (LSI-R). *Federal Probation, 68*(3), 34-38.

Manchak, S., Skeem, J., Douglas, K., & Siranosian, M. (2009). Does gender moderate the predictive utility of the Revised Level of Service Inventory (LSI-R) for serious violent offenders? *Criminal Justice and Behavior, 35,* 425-442. doi:10.1177/0093854809333058

Marsh, H. W., & Hau, K-T. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education, 64*, 364-390. doi: 10.1080/00220973.1996.10806604

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85-110. doi: 10.1146/annurev-clinpsy-032813-153700

Martin, K., & Stermac, L. (2010). Measuring hope: Is hope related to criminal behaviour in offenders? *International Journal of Offender Therapy and Comparative Criminology, 54,* 693-705. doi: 10.1177/0306624X09336131

Maruna, S. (2001). *Making good: How ex---convicts reform and rebuild their lives*. Washington, DC: American Psychological Association. doi:10.1037/10430000

Matsunaga, M. (2010). How to factor-analyze your data right: Do's don'ts and how to's. *International Journal of Psychological Research, 3*, 97-110. doi: 10.21500/20112084.854

Maume, M. O., Ousey, G. C., & Beaver, K. (2005). Cutting the grass: A re-examination of the link between marital attachment, delinquent peers and desistance from marijuana use. *Journal of Quantitative Criminology, 21,* 27-53. doi: 10.1007/s10940-004-1786-3

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*, 412-433. doi: 10.1037/met0000144

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543. doi:10.1007/BF02294825

Miller, H. A. (2006). A dynamic assessment of offender risk, needs, and strengths in a sample of general offenders. *Behavioural Sciences and the Law, 24*, 767-782. doi: 10.1002/bsl.728

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior, 40*, 716-736. doi: 10.1177/0093854812468883

Mills, J. F., & Kroner, D. G. (1999). Measures of Criminal Attitudes and Associates (MCAA). Unpublished instrument and user guide.

Morgan, R. D., Kroner, D. G., Mills, J. F., Serna, C., & McDonald, B. (2013). Dynamic risk assessment: A validation study. *Journal of Criminal Justice, 41*, 115-124. doi: 10.101-6/j.jcrimjus.2012.11004

Moulden, H. M., & Marshall, W. L. (2005). Hope in the treatment of sexual offenders: The potential application of hope theory. *Psychology, Crime & Law, 11*, 329-342. doi: 10.1080/10683160512331316361

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén

Olver, M. E., Beggs Cristofferson, 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 and the Law, 33*, 92-110. doi: 10.1002/bsl.2159

Orbis Partners. (2003). Service Planning Instrument (SPIn). Ottawa, Ontario, Canada: Author.

Paternoster, R., Bachman, R., Kerrison, E., O'Connell, D., & Smith, L. (2016). Desistance from crime and identity: An empirical test with survival time. *Criminal Justice and Behavior, 43,* 1204-1224. doi: 10.1177/0093854816651905

Perley-Robertson, B. (2018). *Predictive validity of the Dynamic Risk Assessment for Offender Re-Entry among intimate partner violence offenders* (Master's thesis). Carleton University, Ottawa, Canada.

Piquero, A. R., Blumstein, A., Brame, R., Haapanen, R., Mulvey, E. P., & Nagin, D. S. (2001). Assessing the impact of exposure time and incapacitation on longitudinal trajectories of criminal offending. *Journal of Adolescent Research, 16*, 54-74. doi:10.1177/0743558401161005

Polaschek, D. L. L. (2016). Desistance and dynamic risk factors belong together. *Psychology, Crime & Law, 22*, 171-189. doi: 10.1080/1068316X.2015.1114114

Polaschek, D. L. L. (2017). Protective factors, correctional treatment and desistance. *Aggression and Violent Behavior, 32*, 64-70. doi: 10.1016/j.avb.2016.12.005

Polaschek, D. L. L., & Yesberg, J. A. (2018). High-risk violent prisoners' patterns of change on parole on the DRAOR's dynamic risk and protective factors. *Criminal Justice and Behavior*, *45*, 340-363. doi: 10.1177/0093854817739928

Prell, L. (2013). *Iowa Risk Assessment Revised: Predicting violence and victimization among male and female probationers and parolees*. Unpublished Manuscript. Iowa Department of Corrections.

Prell, L., Vitacco, M. J., & Zavodny, D. (2016). Predicting violence and recidivism in a large sample of males on probation or parole. *International Journal of Law and Psychiatry,* 107-113. doi: 10.1016/j.ijlp.2016.08.003

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raynor, P. (2007). Risk and need assessment in British probation: The contribution of the LSI-R. *Psychology, Crime, & Law, 13*, 125-138. doi: 10.1080/10683160500337592

Revelle, W. (2019). Psych Package: Procedures for Psychological, Psychometric, and Personality Research. Version 1.8.12. https://cran.r-project.org/web/packages/psych/index.html

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r. Law and Human Behavior, 29*, 615-620. doi: 10.1007/s10979-005-6832-7

Richards, C. M. (2016). *You can get prepared, but it's still scary walking out the door: Exploring the role of dynamic re-entry factors in release planning for high-risk offenders* (Master's thesis, Victoria University of Wellington, Wellington, New Zealand). Retrieved from http://hdl.handle.net/10063/6809

Robinson, C. R., Lowenkamp, C. T., Holsinger, A. M., VanBenschoten, S., Alexander, M., & Oleson, J. C. (2012). A random study of Staff Trained at Reducing Re-arrest (STARR): Using core correctional practices in probation interactions. *Journal of Crime & Justice*, *35*, 167-188. doi:10.1080/0735648X.2012.674823

Rocque, M., & Plummer-Beale, J. (2016). In the eye of the beholder? An examination of the inter-rater reliability of the LSI-R and YLS/CMI in a correctional agency. *Journal of Criminal Justice, 42*, 568-578. doi: 10.1016/j.jcrimjus.2014.09.011

Sampson, R. J., & Laub, J. H. (2003). Life-course desisters? Trajectories of crime among delinquent boys followed to age 70. *Criminology, 41*, 555-592. doi: 10.1111/j.1745-9125.2003.tb00997.x

Sampson, R. J., & Laub, J. H. (2005). A life-course view of the development of crime. *The Annals of the American Academy of Political and Social Science*, *602*(1), 12-45. doi: 10.1177/0002716205280075

Scanlan, J. M., Yesberg, J. A., Fortune, C-A., Polaschek, D. L. L. (2020). Predicting women's recidivism using the dynamic risk assessment for offender re-entry:

Preliminary evidence of predictive validity with community-sentenced women using a “gender-neutral” risk measure. *Criminal Justice and Behavior, 47*, 251-270. doi: 10.1177/0093854819896387

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

Serin, R. (2007). *The Dynamic Risk Assessment Scale for Offender Re-Entry (DRAOR)*. Unpublished scale. Carleton University, Ottawa, Ontario.

Serin, R. C., & Chadwick, N. (2019). *Assessing dynamic risk and protective factors among male probationers and parolees in Iowa: The utility of the dynamic risk assessment for offender re-entry.* Unpublished manuscript, Department of Psychology, Carleton University, Ottawa, Canada.

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

Serin, R. C., & Lloyd, C. D. (2009). Examining the process of offender change: the transition to crime desistance. *Psychology, Crime & Law, 15*, 347-364. doi: 10.180/1068316080-2261078

Serin, R. C., Lloyd, C., & Chadwick, N. (2016). Integrating dynamic risk assessment into community supervision practice. In D. Polaschek (Ed). *The Wiley International Handbook of Correctional Psychology.*

Serin, R. C., Lloyd, C., & Hanby, L. J. (2010). Enhancing offender re-entry: An integrated model for understanding offender re-entry. *European Journal of Probation, 2*(2), 53-75.

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*, 32-53. doi: 10.1016/j.avb.20-12.09.002

Serin, R. C., & Wilson, N. J. (2012). The structured dynamic assessment case-management 21-item (SDAC-21). Unpublished user manual.

Shields, I. W., & Whitehall, G. C. (1994). Neutralization and delinquency among teenagers. *Criminal Justice and Behavior, 21*, 223-235. doi: 10.1177/0093854894021002003

Simourd, D. J. (1997). The Criminal Sentiments Scale-Modified and Pride in Delinquency Scale: Psychometric properties and construct validity of two measures of criminal attitudes. *Criminal Justice and Behavior, 24*, 52-70. doi: 10.1177/0093854897024001004

Simourd, D. J. (2004). Use of dynamic/risk need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior, 31*, 306-323. doi: 10.1177/00-93854803262507

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modelling change and event occurrence.* New York: Oxford University Press.

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology, 54*, 680-712. doi: 10.1111/1745-9125.12123

Skeem, J. L. & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*, 38-42. doi: 10.1177/0963721410397271

Skeem, J., & Mulvey, E. (2002). Monitoring the violence potential of mentally disordered offenders being treated in the community. In A. Buchanan (Ed.), *Care of the mentally disordered offender in the community,* (pp. 111–142). New York, NY: Oxford Press.

Smeth, A. (2013). *Evaluating risk assessments among sex offenders: a comparative analysis of static and dynamic factors* (Master's thesis). Carleton University, Ottawa, Canada.

Smeth, A. (2019). *Validating the structured dynamic assessment case-management 21-item (SDAC-21) in a sample of incarcerated offenders* (Doctoral Dissertation). Carleton University, Ottawa, Canada.

Smith, P., Schweitzer, M., Labrecque, R. M., & Latessa, E. J. (2012). Improving probation officers' supervision skills: An evaluation of the EPICS model. *Journal of Crime & Justice*, *35*, 189-199. doi:10.1080/0735648X.2012.674826

Statistics Canada (2019). *Adult and youth correctional statistics in Canada, 2017/2018* (Catalogue No 85-002-X). Retrieved from https://www150.statcan.gc.ca/n1/pub/85-002-x/2019001/article/00010-eng.htm

Stouthamer-Loeber, M., Wei, E., Loeber, R., & Masten, A. S. (2004). Desistance from persistent serious delinquency in the transition to adulthood. *Development and Psychopathology, 16*, 897-918. doi: 10.1017/S0954579404040064

Tabachnick, B. G., and Fidell, L. S. (2013). *Using multivariate statistics* (6th Ed.). Boston, MA: Pearson Education, Inc.

Tasca, G. A., & Gallop, R. (2009). Multileve modeling of longitudinal data for psychotherapy researchers: I. The basics. *Psychotherapy Research, 19,* 429-437. doi: 10.1080/105033-00802641444

Therneau, T. (2015). *A Package for Survival Analysis in R*. version 2.38, https://CRAN.R-project.org/package=survival.

Troquete, N. A. C., van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W., Schoevers, R. A., & Wiersma, D. (2015). Predictive validity of the short-term assessment of risk and treatability for violent behavior in outpatient forensic psychiatric patients. *Psychological Assessment, 27*, 377-391. doi: 10.1037/a0038270

Uggen, C. (2000). Work as a turning point in the life course of criminals: A duration model of age, employment, and recidivism. *American Sociological Review, 65*(4), 529–546.

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

van Mastrigt, S. B., & Farrington, D. P. (2009). Co-offending, age, gender and crime type: Implications for criminal justice policy. *British Journal of Criminology*, *49*, 552-573. doi: 10.1093/bjc/azp021

Viglione, J. (2019). The risk-need-responsivity model: How do probation officers implement the principles of effective intervention? *Criminal Justice and Behavior, 46*, 655-673. doi: 10.1177/0093854818807505

Viglione, J., Rudes, D. S. and Taxman, F. S. (2015). Misalignment in supervision: Implementing risk/needs assessment instruments in probation. *Criminal Justice and Behavior, 42*, 263-285. doi: 10.1177/0093854814548447.

Viljoen, J. L., Shaffer, C. S., Muir, N. M., Cochrane, D. M., & Brodersen, E. M. (2019). Improving case plans and interventions for adolescents on probation. *Criminal Justice and Behavior, 46*, 42-62. doi: 10.1177/0093854818799379

Vincent, G. M., Guy, L. S., Perrault, R. T., & Gershenson, B. (2016). Risk assessment matters, but only when implemented well: A multisite study in juvenile probation. *Law and Human Behavior, 40*, 683-696. doi:10.1037/lhb0000214

Vincent, G. M., Paiva-Salisbury, M. L., Cook, N. E., Guy, L. S., & Perrault, R. T. (2012). Impact of risk/needs assessment on juvenile probation officers' decision making: Importance of implementation. *Psychology, Public Policy, and Law, 18*, 549-576. doi: 10.1037/a0027186

Vose, B., Lowenkamp, C. T., Smith, P., & Cullen, F. T. (2009). Gender and the predictive validity of the LSI-R: A study of parolees and probationers. *Journal of Contemporary Criminal Justice, 25*(4), 459-471. doi: 10.1177/1043986209344797

Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior, 40*, 1383-1396. doi: 10.1177/009-3854813508916

Walters, G. D. (1995). The Psychological Inventory of Criminal Thinking Styles: Part I. Reliability and preliminary validity. *Criminal Justice and Behavior, 22,* 307–325. 10.1177/0093854895022003008

Walters, G. D. (2011). Predicting recidivism with the psychological inventory of criminal thinking styles and level of service inventory-revised: Screening version. *Law and Human Behavior, 35*, 211-220. doi: 10.1007/s10979-010-9231-7

Walters, G. D. (2012). Criminal thinking and recidivism: Meta-analytic evidence on the predictive and incremental validity of the Pyschological Inventory of Criminal Thinking Styles (PICTS). *Aggression and Violent Behavior, 17*, 272-278. doi: 10.1016/j.avb.20-12.02.010

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*, 411-419. doi: 10.1037/lhb0000185

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*, 652-659. doi: 10.1037/pas0000210

Ward, T. (2017). Prediction and agency: The role of protective factors in correctional rehabilitation and desistance. *Aggression and Violent Behavior, 32*, 19-27. doi: 10.1016/j.avb.2016.11.012

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20. Assessing the risk of violence. Version 2.* Burnaby, Canada: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.

Webster, C., Martin, M., Brink, J., Nicholls, T., & Middleton, C. (2004). *The Short Term Assessment of Risk and Treatability (START).* British Columbia: Forensic Psychiatric Services Commission.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.

Wooditch, A., Tang, L. L., & Taxman, F. S. (2014). Which criminogenic need changes are most important in promoting desistance from crime and substance use? *Criminal Justice and Behavior, 41,* 276-298. doi: 10.1177/0093854813503543

Xu, R., & O'Quigley, J. (1999). A R^2 type measure of dependence for proportional hazards models. *Journal of Nonparametric Statistics*, *12*, 83-107. doi:10.1080/10485259908832799

Yang, M., Guo, B., Olver, M. E., Polaschek, D. L. L., & Wong, S. C. P. (2017). Assessing associations between changes in risk and subsequent reoffending. *Criminal Justice and Behavior, 44*, 59-84. doi: 10.1177/0093854816678648

Yang, S., & Mulvey, E. P. (2012). Violence risk: Re-defining variables from the first-person perspective. *Aggression and Violent Behavior, 17*, 198-207. doi: 10.1016/j.avb.20-12.02.001

Yesberg, J. A., & Polaschek, D. L. L. (2015). Assessing dynamic risk and protective factors in the community: examining the validity of the Dynamic Risk Assessment for Offender Re-entry. *Psychology, Crime & Law, 21*, 80-99. doi: 10.1080/1068316X.2014.935775

Yesberg, J. A., Scanlan, J., Hanby, L. J., Serin, R. C., & Polaschek, D. L. L. (2015). Predicting women's recidivism: Validating a dynamic community-based 'gender-neutral' tool. *Probation Journal, 62*, 33-48. doi: 10.1177/0264550514562851 0264550514562851

Zamble, E., & Quinsey, V. L. (1997). *The criminal recidivism process.* Cambridge, UK: Cambridge University Press.

Appendix A: Measurement Invariance

Table A1

Measurement Invariance for Supervision Sequences Associated with White Offenders (n = 1,497)

Model	# of free parameters	RMSEA [90% CI]	CFI	TLI	X^2 (df)
Baseline/Configural	360	0.037 [0.036, 0.039]	0.991	0.989	
Strong	188	0.038 [0.036, 0.039]	0.990	0.989	
Changes in fit indices		0.001	0.001	0.000	673.80** (172)

Note: A ΔCFI of less than 0.01 or an ΔRMSEA of less than 0.015 provides sufficient support that the constrained model does not fit the data worse than the model allowed to vary. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TFI = Tucker-Lewis Index.

Table A2

Measurement Invariance for Supervision Sequences Associated with Black Offenders (n = 558)

Model	# of free parameters	RMSEA [90% CI]	CFI	TLI	X^2 (df)
Baseline/Configural	360	0.029 [0.027, 0.032]	0.995	0.994	
Strong	188	0.030 [0.028, 0.033]	0.994	0.994	
Changes in fit indices		0.001	0.001	0.000	414.33** (172)

Note: A ΔCFI of less than 0.01 or an ΔRMSEA of less than 0.015 provides sufficient support that the constrained model does not fit the data worse than the model allowed to vary. RMSEA = Root Mean Square Error of Approximation, CFI = Comparative Fit Index, TFI = Tucker-Lewis Index.

Appendix B: Multilevel Model Building

Table B1

Results of a series of multilevel models exploring combinations of potentially relevant predictors to build toward a final model for Stable scores

	Model A Estimate (SE)	Model B Estimate (SE)	Model C Estimate (SE)	Model D Estimate (SE)	Model E Estimate (SE)	Model F Estimate (SE)	Model G Estimate (SE)
Fixed effects							
Intercept	5.86 (0.04)**	6.21 (0.04)**	6.08 (0.04)**	6.21 (0.04)**	6.51 (0.08)**	6.20 (0.04)**	6.03 (0.04)**
Age				-0.01 (0.004)**			
Race					-0.42 (0.09)**		
Static Risk						0.12 (0.01)**	
Protect(pc)			-0.51 (0.01)**				
Protect(gc)			-0.38 (0.02)**				
Acute(pc)							0.42 (0.01)**
Acute(gc)							0.46 (0.01)**
Rate of change							
Time (months)		-0.06 (0.003)**	-0.03 (0.002)**	-0.06 (0.003)**	-0.05 (0.01)**	-0.06 (0.003)**	-0.02 (0.003)**
Time*Age				-0.001 (0.0003)*			
Time*Race					-0.01 (0.01)		
Time*Static Risk						-0.001 (0.001)	
Time*protect(pc)			-0.002 (0.001)*				
Time*protect(gc)			-0.005 (0.001)**				
Time*acute(pc)							0.002 (0.0004)**
Time*acute(gc)							0.01 (0.001)**
Variance components							
Initial status	4.92 (0.12)**	6.26 (0.151)**	5.14 (0.12)**	6.24 (0.15)**	6.22 (0.15)**	6.05 (0.15)**	4.74 (0.11)**
Covariance		-0.18 (0.01)**	-0.15 (0.01)**	-0.18 (0.01)**	-0.18 (0.01)**	-0.18 (0.01)**	-0.14 (0.01)**
Rate of change		0.03 (0.001)**	0.02 (0.001)**	0.03 (0.001)**	0.03 (0.001)**	0.03 (0.001)**	0.02 (0.001)**
Within-person	1.46 (0.01)**	0.74 (0.01)**	0.56 (0.01)**	0.74 (0.01)**	0.74 (0.01)**	0.74 (0.01)**	0.53 (0.01)**
Pseudo R^2 and fit statistics							

$R^2_{y,y}$		0.00	0.18	0.002	0.002	0.01	0.28
R^2_e		0.49	0.61	0.49	0.49	0.49	0.64
R^2_0			0.18	0.003	0.006	0.03	0.24
R^2_1			0.37	0.000	0.000	0.000	0.45
-2 Log Likelihood	102248	92901	84879	92872	92865	92764	82726
AIC	102254	92913	84899	92888	92881	92780	82746
BIC	102273	92951	84962	92938	92931	92830	82809

Note: *p < .01, **p < .001. Model A = Unconditional means model; Model B = Unconditional growth model; Model C = Conditional model of within and between effects of Protect scores on initial status and change; Model D = Conditional model of age on initial status and change; Model E = Conditional model of race on initial status and change; Model F = Conditional model of static risk on initial status and change; Model G = Conditional model of within and between effects of Acute scores on initial status and change. Empty cells indicate that the parameter was not estimated for the given variable in that model. (pc) indicates that the variable was person-mean centered, (gc) indicates that the variable was grand-mean centered. Models were performed with SAS 9.4 using ML estimation.

Table B2

Results of a series of multilevel models exploring combinations of potentially relevant predictors to build toward a final model for Acute scores

	Model A Estimate (SE)	Model B Estimate (SE)	Model C Estimate (SE)	Model D Estimate (SE)	Model E Estimate (SE)	Model F Estimate (SE)	Model G Estimate (SE)	Model H Estimate (SE)
Fixed effects								
Intercept	6.35** (0.04)	6.86** (0.04)	6.71** (0.04)	6.86** (0.04)	6.69** (0.08)	6.85** (0.04)	6.69** (0.04)	6.66** (0.04)
Age				0.003 (0.004)				
Race					0.23 (0.10)			
Static Risk						0.11** (0.01)		
Protect(pc)			-0.55** (0.01)					-0.26** (0.01)
Protect(gc)			-0.30** (0.02)					-0.12** (0.02)
Stable(pc)							0.68** (0.01)	0.56** (0.01)
Stable(gc)							0.53** (0.02)	0.48** (0.02)
Rate of change								
Time (months)		-0.09** (0.004)	-0.06** (0.003)	-0.09** (0.004)	-0.07** (0.007)	-0.09** (0.003)	-0.04** (0.003)	-0.04** (0.002)
Time*Age				-0.002** (0.0003)				
Time*Race					-0.02* (0.09)			
Time*Static Risk						-0.002 (0.001)		
Time*protect (pc)			-0.002 (0.001)					-0.0001 (0.001)
Time*protect (gc)			-0.01** (0.001)					-0.002 (0.001)
Time*stable (pc)							0.002* (0.001)	0.001 (0.001)
Time*stable (gc)							0.004** (0.001)	0.003 (0.001)
Variance components								
Initial status	5.21** (0.13)	6.88** (0.17)	6.33** (0.16)	6.88** (0.17)	6.87** (0.17)	6.72** (0.17)	5.27** (0.13)	5.21** (0.13)
Covariance		-0.22** (0.01)	-0.20** (0.01)	-0.22** (0.01)	-0.22** (0.01)	-0.22** (0.01)	-0.17** (0.01)	-0.17** (0.01)

Rate of change		0.04** (0.001)	0.03** (0.001)	0.04** (0.001)	0.04** (0.001)	0.04** (0.001)	0.02** (0.001)	0.02** (0.001)
Within-person	2.18** (0.02)	1.24** (0.01)	1.02** (0.01)	1.24** (0.01)	1.24** (0.01)	1.24** (0.01)	0.88** (0.01)	0.84** (0.01)
Pseudo R^2 and fit statistics								
$R^2_{y,y}$		0.005	0.13	0.01	0.004	0.01	0.29	0.30
R^2_e		0.43	0.53	0.43	0.43	0.43	0.60	0.61
R^2_0			0.08	0.0005	0.002	0.02	0.23	0.24
R^2_1			0.28	0.009	0.003	0.003	0.47	0.49
-2 Log Likelihood	112315	104557	99173	104530	104548	104467	94318	93221
AIC	112321	104569	99193	104546	104564	104483	94338	93249
BIC	112340	104607	99256	104597	104615	104533	94401	93337

Note: *p < .01, **p < .001. Model A = Unconditional means model; Model B = Unconditional growth model; Model C = Conditional model of within and between effects of Protect scores on initial status and change; Model D = Conditional model of age on initial status and change; Model E = Conditional model of race on initial status and change; Model F = Conditional model of static risk on initial status and change; Model G = Conditional model of within and between effects of Stable scores on initial status and change; Model H = Conditional model of within and between effects of Stable and Protect scores on initial status and change. Empty cells indicate that the parameter was not estimated for the given variable in that model. (pc) indicates that the variable was person-mean centered, (gc) indicates that the variable was grand-mean centered. Models were performed with SAS 9.4 using ML estimation.

Table B3

Results of a Series of Multilevel Models Exploring Combinations of Potentially Relevant Predictors to Build Toward a Final Model for Protect Scores

	Model A Estimate (SE)	Model B Estimate (SE)	Model C Estimate (SE)	Model D Estimate (SE)	Model E Estimate (SE)	Model F Estimate (SE)	Model G Estimate (SE)	Model H Estimate (SE)
Fixed effects								
Intercept	5.72** (0.04)	5.41** (0.04)	5.54** (0.04)	5.41** (0.04)	5.12** (0.08)	5.41** (0.04)	5.51** (0.04)	5.55** (0.04)
Age				-0.002 (0.004)				
Race					0.39** (0.10)			
Static Risk						-0.07** (0.01)		
Acute (pc)			-0.30** (0.01)					-0.15** (0.01)
Acute (gc)			-0.29** (0.02)					-0.11** (0.02)
Stable (pc)							-0.46** (0.01)	-0.36** (0.01)
Stable (gc)							-0.44** (0.02)	-0.39** (0.02)
Rate of change								
Time (months)		0.05** (0.003)	0.02** (0.003)	0.05** (0.003)	0.04** (0.04)	0.05** (0.003)	0.02** (0.003)	0.02** (0.003)
Time*Age				0.001* (0.0003)				
Time*Race					0.01 (0.01)			
Time*Static Risk						0.001 (0.001)		
Time*acute (pc)			-0.003** (0.0005)					-0.001 (0.001)
Time*acute (gc)			-0.01** (0.001)					-0.01** (0.001)
Time*stable (pc)							-0.002** (0.001)	-0.001 (0.001)
Time*stable (gc)							-0.004** (0.001)	-0.0003 (0.001)
Variance components								

Initial status	5.67** (0.13)	7.10** (0.17)	6.46** (0.15)	7.11** (0.17)	7.07** (0.17)	7.03** (0.17)	5.84** (0.14)	5.76** (0.14)
Covariance		-0.18** (0.01)	-0.16** (0.01)	-0.18** (0.01)	-0.18** (0.01)	-0.18** (0.01)	-0.14** (0.01)	-0.14** (0.01)
Rate of change		0.03** (0.001)	0.02** (0.001)	0.03** (0.001)	0.03** (0.001)	0.03** (0.001)	0.02** (0.001)	0.02** (0.001)
Within-person	1.33** (0.01)	0.69** (0.01)	0.57** (0.01)	0.69** (0.01)	0.69** (0.01)	0.69** (0.01)	0.52** (0.01)	0.50** (0.01)
Pseudo R^2 and fit statistics								
$R^2_{y,y}$		0.00	0.12	0.06	0.00	0.00	0.20	0.22
R^2_e		0.49	0.57	0.49	0.49	0.49	0.61	0.63
R^2_0			0.09	0.00	0.00	0.01	0.18	0.19
R^2_1			0.27	0.00	0.00	0.00	0.37	0.39
-2 Log Likelihood	100690	91321	85903	91310	91293	91276	83266	82147
AIC	100696	91333	85924	91326	91309	91292	83286	82175
BIC	100715	91371	85987	91376	91359	91342	83349	82263

Note: *p < .01, **p < .001. Model A = Unconditional means model; Model B = Unconditional growth model; Model C = Conditional model of within and between effects of Acute scores on initial status and change; Model D = Conditional model of age on initial status and change; Model E = Conditional model of race on initial status and change; Model F = Conditional model of static risk on initial status and change; Model G = Conditional model of within and between effects of Stable scores on initial status and change; Model H = Conditional model of within and between effects of Stable and Acute scores on initial status and change. Empty cells indicate that the parameter was not estimated for the given variable in that model. (pc) indicates that the variable was person-mean centered, (gc) indicates that the variable was grand-mean centered. Models were performed with SAS 9.4 using ML estimation.

Appendix C: Univariate Cox Regression Survival Analyses

Table C1

Univariate Cox Regression Survival Analysis Examining Relationship with Revocation of Community Supervision and New Convictions

	Predictor	*B*	*SE*	Wald	*p*	Hazard Ratio [CI]	c-index [CI]
Revocation of community supervision[a]	Static risk	0.10	0.01	13.13	< .001	1.11 [1.09, 1.13]	.62 [.60, .64]
	Age	-0.01	0.002	-4.75	< .001	0.99 [0.98, 0.99]	.53 [.51, .55]
	Race	-0.11	0.07	-1.79	.07	0.89 [0.78, 1.01]	.51 [.50, .52]
	Initial stable score	0.08	0.01	6.52	< .001	1.08 [1.06, 1.11]	.57 [.55, .59]
	Stable change	1.90	0.22	8.61	< .001	6.69 [4.34, 10.31]	.53 [.51, .55]
	Initial acute score	0.08	0.01	6.63	< .001	1.08 [1.06, 1.11]	.57 [.55, .59]
	Acute change	1.52	0.20	7.44	< .001	4.57 [3.06, 6.82]	.53 [.51, .55]
	Initial protect score	-0.05	0.01	-4.68	< .001	0.95 [0.93, 0.97]	.55 [.53, .56]
	Protect change	-1.64	0.23	-7.01	< .001	0.19 [0.12, 0.31]	.50 [.48, .52]
New charge resulting in a conviction[b]	Static risk	0.09	0.01	7.02	< .001	1.09 [1.07, 1.12]	.60 [.58, .63]
	Age	-0.02	0.005	-4.93	< .001	0.98 [0.97, 0.99]	.57 [.54, .59]
	Race	-0.19	0.10	-1.83	.07	0.83 [0.67, 1.01]	.52 [.50, .54]
	Initial stable score	0.06	0.02	3.23	.001	1.07 [1.03, 1.11]	.55 [.52, .58]
	Stable change	0.13	0.33	0.39	.70	1.14 [0.60, 2.16]	.48 [.45, .51]
	Initial acute score	0.05	0.02	2.85	.004	1.06 [1.02, 1.10]	.54 [.51, .57]
	Acute change	0.03	0.31	0.11	.91	1.04 [0.57, 1.89]	.48 [.45, .51]
	Initial protect score	-0.04	0.02	-2.37	.018	0.96 [0.92, 0.99]	.54 [.51, .57]
	Protect change	-0.29	0.35	-0.84	.40	0.75 [0.37, 1.48]	.46 [.43, .49]

Note: SE = Standard Error, CI = 95% confidence interval.

[a]1,087 of 4,000 community supervision sequences ended with a revocation.

[b]427 of 4,000 community supervision sequences had a new conviction associated with it throughout the follow-up.